

الجمهورية الجزائرية الديمقراطية
الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة أبي بكر بلقايد - تلمسان -

Université Aboubakr Belkaïd – Tlemcen –
Faculté de TECHNOLOGIE



MEMOIRE

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

En : Génie Biomédical

Spécialité : Imagerie médicale

Par : Mokeddem Amina & Lammari Khouloud

Sujet

Détection et classification de l'appendicite à l'aide de techniques d'apprentissage automatique

Soutenu publiquement, le 17 / 06 / 2025 , devant le jury composé de :

M MESSADI Mahammed
Mme MEZIANI Fadia
Mme BELAIDI Asma
Mme YOUBI Fatiha

Professeur
MCA
MCB
Docteur

Université de Tlemcen
Université de Tlemcen
Université de Tlemcen
Université de Tlemcen

Président
Examinatrice
Encadrante
Co-Encadrante

Dédicace

Je dédie ce mémoire à mes parents bien-aimés, En témoignage de mon profond respect et de ma gratitude infinie. Votre soutien constant, vos sacrifices silencieux, vos prières sincères et votre amour inconditionnel ont été le socle sur lequel j'ai pu m'appuyer tout au long de ce parcours.

À mes frères: Abdellah, Younes, Amine, Abdrahime, Ilyes Vos sourires, votre spontanéité ont été pour moi une source de réconfort et de motivation. Votre présence joyeuse a allégé les moments de fatigue et donné du sens à mes efforts. Je formule pour vous les vœux les plus sincères de réussite et d'épanouissement, sous la bienveillance d'Allah. Puisse votre chemin être éclairé par la sagesse, la foi et la réussite.

À mes amies et collègues : Bouchra, Khouloud, Tesnim, Aya, Chiraz, Aya, Chourouq, Votre soutien tout au long de ce parcours universitaire a été d'un grand réconfort. Merci pour votre solidarité, vos conseils et les moments partagés, qui ont enrichi cette expérience tant sur le plan humain qu'intellectuel.

Ce mémoire constitue l'aboutissement d'un cheminement exigeant, rendu possible grâce à l'appui indéfectible de mes proches, et s'inscrit dans une dynamique de reconnaissance, de foi et de gratitude.

Amina

Dédicace

Louange à Allah, le Très-Haut, le Très-Miséricordieux, qui m'a accordé la force, la patience et la persévérance pour franchir chaque étape de ce parcours.

Je dédie ce modeste travail : À mes parents bien-aimés,

Il n'existe ni mots ni pages suffisantes pour traduire la profondeur de ce que je ressens pour vous.

Merci pour votre amour sans conditions, vos sacrifices silencieux et vos invocations constantes. Vous êtes mon refuge, ma force et ma fierté.

À mes chères sœurs et à mon cher frère,

Merci pour votre présence, vos encouragements et votre confiance en moi. Vous êtes les piliers de mon équilibre.

À leurs enfants,

Vos sourires et votre innocence apportent de la lumière dans ma vie. Puissiez-vous grandir sous la protection divine et réaliser vos rêves.

À mes fidèles amis : Amina, Tesnim, Chiraz, Aya, Chourouq...

Merci pour les souvenirs, Les années d'études ont été remplies de moments inoubliables grâce à vous. Vous avez été mes compagnons de route, merci pour votre patience et votre présence à mes côtés dans les moments d'effort comme de doute. Vous avez enrichi ce voyage.

À chacun de vous, je dis avec reconnaissance :

Ce mémoire est bien plus qu'un document académique, il est le fruit d'un chemin parcouru avec vous, dans la foi, l'amour et la gratitude.

Khouloud

Remerciements

Nous tenons tout d'abord à exprimer notre profonde reconnaissance à Mme Belaidi Asma, notre encadrante, pour sa disponibilité constante, son accompagnement rigoureux et ses conseils avisés tout au long de ce travail. Sa bienveillance, sa pédagogie et son exigence scientifique ont été déterminants dans l'aboutissement de ce mémoire.

Nos remerciements s'adressent également à Mme Youbi Fatiha, notre co-encadrante, pour son soutien technique, ses remarques pertinentes et son engagement sans faille. Sa contribution a largement enrichi notre réflexion et renforcé la qualité scientifique de notre projet.

Nous remercions tout particulièrement Monsieur le Président du jury et Madame l'Examinatrice, d'avoir accepté de présider et d'examiner ce travail. Leur disponibilité, leur attention et leurs retours seront pour nous une source précieuse d'apprentissage. Enfin, nous remercions toutes les personnes qui nous ont soutenus de près ou de loin tout au long de cette aventure académique.

Résumé

Le diagnostic de l'appendicite aiguë chez l'enfant représente un défi clinique important en pédiatrie, souvent complexifié par la variabilité des symptômes. Pour relever ce défi, ce projet de fin d'études a exploré le potentiel de l'Intelligence Artificielle (IA) afin d'améliorer la précision diagnostique. Nous avons mené une étude comparative approfondie entre les méthodes de Machine Learning (ML) classiques et le Deep Learning (DL) pour la classification binaire et multiclasse de l'appendicite. En utilisant une base de données numérique, nous avons développé et évalué un système d'aide au diagnostic pédiatrique. Les résultats expérimentaux ont clairement démontré la supériorité de l'algorithme XGBoost, qui a obtenu les meilleurs scores avec un taux de classification de 97% et un F1-score de 97%. Il est important de noter que XGBoost a globalement surpassé les réseaux de neurones convolutionnels (CNN) dans ce contexte spécifique. Cette performance s'explique principalement par la nature tabulaire de notre base de données numérique. Alors que les CNN excellent dans l'analyse d'images et de textes, tirant parti de structures spatiales ou séquentielles complexes.

Mots clés

Appendicite pédiatrique, Intelligence Artificielle (IA), Machine Learning (ML), Deep Learning (DL), Classification binaire, Classification multiclasse, Prédiction de l'appendicite

Abstract

The diagnosis of acute appendicitis in children represents a major clinical challenge in pediatrics, often complicated by the variability of symptoms. To address this challenge, this study explored the potential of Artificial Intelligence (AI) to improve diagnostic accuracy. We conducted an in-depth comparative study between classical Machine Learning (ML) methods and Deep Learning (DL) for binary and multiclass classification of appendicitis. Using a digital database, we developed and evaluated a pediatric diagnostic support system. Experimental results clearly demonstrated the superiority of the XGBoost algorithm, which achieved the best scores with a classification rate of 97% and an F1-score of 97%. Importantly, XGBoost outperformed convolutional neural networks (CNNs) overall in this specific context. This performance is mainly due to the tabular nature of our digital database. Whereas CNNs excel in image and text analysis, taking advantage of complex spatial or sequential structures.

Keywords

Pediatric appendicitis, Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Binary classification, Multiclass classification, Appendicitis prediction

المخلص

يُعد تشخيص التهاب الزائدة الدودية الحاد عند الأطفال تحدياً سريرياً كبيراً في طب الأطفال، وغالباً ما تزداد صعوبته بسبب تنوع الأعراض. ولمواجهة هذا التحدي، استكشف هذا المشروع البحثي إمكانيات الذكاء الاصطناعي.

يهدف تحسين دقة التشخيص، قمنا بإجراء دراسة مقارنة معمّقة بين أساليب التعلم الآلي التقليدية وأساليب التعلم العميق لتصنيف الزائدة الدودية تصنيفاً ثنائياً ومتعدد الفئات. وباستخدام قاعدة بيانات رقمية، قمنا بتطوير وتقييم نظام دعم للتشخيص في طب الأطفال.

أظهرت النتائج التجريبية بوضوح تفوق خوارزمية XGBoost التي حققت أفضل النتائج بنسبة تصنيف بلغت درجة 97%. و F1-score بلغت درجة 97% من المهم الإشارة إلى أن خوارزمية XGBoost قد تفوقت بشكل عام على الشبكات العصبية الالتفافية CNN. في هذا السياق الخاص يرجع هذا الأداء بشكل رئيسي إلى الطابع الجدولي لقاعدة البيانات الرقمية الخاصة بنا. في حين أن نماذج CNN ممتازة في تحليل الصور والنصوص.

الكلمات المفتاحية: التهاب الزائدة الدودية عند الأطفال، الذكاء الاصطناعي، أساليب التعلم الآلي التقليدية، أساليب التعلم العميق، تصنيف ثنائي، تصنيف متعدد الفئات، تنبؤ بالتهاب الزائدة الدودية.

Table des matières

Résumé	i
Abstract	iv
contenu	iv
Liste des figures	vi
Liste de tableaux	vii
Liste d'abréviations	viii
Glossaire	viii
Introduction générale	1
1 Aspect médicaux de l'appendicite	3
1 Introduction	3
2 Définition et Physiopathologie de la maladie	3
2.1 Définition de l'appendicite	3
2.2 Statistiques épidémiologiques récentes de l'appendicite	4
2.3 Anatomie	4
2.4 Différences anatomiques de l'appendice entre l'adulte et l'enfant	5
2.5 Physiopathologie de l'appendicite	6
3 Diagnostic et bilan initiaux	6
3.1 Anamnèse	6
3.2 Examen et critères clinique	7
3.3 Examens et critères biologiques	7
3.4 Examens et critères d'imagerie	7
4 Classification et Facteurs pronostiques	9
4.1 Classification de l'appendicite	9
4.2 Étude des facteurs de diagnostic / Causes de l'appendicite	11
4.3 Défis actuels dans le diagnostic	11
5 Intégration des nouvelles technologies	12
6 Traitement	13
6.1 Traitement chirurgical	13
6.2 Traitement non chirurgical	13
6.3 Recommandations actuelles	13

7	Conclusion	13
2	Introduction aux méthodes d'apprentissage automatique	14
1	Introduction	14
2	Intelligence artificielle (IA)	14
3	L'apprentissage automatique (Machine Learning)	15
4	Approches d'apprentissage automatique	16
4.1	L'apprentissage supervisé	16
4.2	L'apprentissage non supervisé	17
4.3	L'apprentissage Semi-supervisée	18
4.4	L'apprentissage par renforcement (RL)	18
5	Algorithmes d'apprentissage automatique	19
5.1	Le Naïve Bayes	19
5.2	Régressions linéaires	20
5.3	Régressions Logistiques	20
5.4	k-plus proches voisins (k-NN)	21
5.5	Arbres de décision	22
5.6	Support Vector Machines (SVM)	23
5.7	Les méthodes d'ensemble en Machine Learning	24
5.8	Réseaux neuronaux	25
6	L'apprentissage profond « Deep Learning »	26
6.1	Réseaux de neurones profonds	27
6.2	Réseaux de neurones convolutifs (CNN)	28
6.3	Réseaux de neurones récurrents (RNN)	30
7	Applications du Machine Learning et du Deep Learning en médecine	31
8	Travaux connexes : détection de L'appendicite	32
9	Conclusion	33
3	Détection et Classification d'appendicite : Résultats et discussion	35
1	Contexte de l'étude	35
2	Environnement de développement	35
3	Implémentation des méthodes	37
3.1	Description de la base de données	38
3.2	Prétraitement de la base de données	38
3.3	Classification de L'appendicite : approches proposées	39
4	Résultats et discussion	44
4.1	Critères d'évaluation	44
4.2	Contribution 01 : Classification binaire	45
4.3	Contribution 02 : Classification multi classe	47
4.4	Discussion des résultats	51
5	Analyse comparative avec les travaux existant	51
6	Interface interactive de classification	52
6.1	Architecture de L'interface	53
6.2	Fonctionnalités de L'interface	53
6.3	Avantages pédagogiques et pratiques	55
7	Conclusion	55
	Conclusion générale	56

Bibliographie

57

Table des figures

1	Localisation de l'appendice [5]	5
2	Vascularisation appendiculaire [5]	5
3	Emplacement de l'appendice [7]	6
4	La relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond [34]	15
5	Types d'apprentissage automatique [37]	16
6	Exemple d'apprentissage supervisé [38]	16
7	Exemple d'apprentissage non supervisé [40]	17
8	Exemple d'apprentissage semi supervisé [44]	18
9	Exemple d'apprentissage par renforcement [46]	19
10	une régression logistique binaire[54]	21
11	Exemple de modèle KNN [58]	22
12	Exemple des Arbres de décision [60]	23
13	Diagram representing the SVM classification of binary class data (Positive, Negative) by plotting the optimal hyperplane [63].	24
14	Exemple de forêt aléatoire [67].	25
15	Exemple de réseau de neurones avec une couche d'entrée, une couche cachée et une couche de sortie [75].	26
16	La différence entre DL et ML[79]	27
17	Fonctionnement de réseau neuronal [79]	27
18	Les couche de CNN [82]	28
19	Une couche convolutionnelle en action [84]	28
20	Exemple de fonctionnement de la couche pooling [86]	29
21	Exemple des Couches entièrement connectées [87]	29
22	Fonctionnement des RNN [92]	30
23	Le processus de l'approche proposée	37
24	Répartition des classes dans Diagnostic	38
25	Répartition des classes dans Management	39
26	Architecture du CNN pour la classification binaire	42
27	Architecture du RNN pour la classification binaire	42
28	Architecture du CNN pour la classification multi classe	43

29	Architecture du RNN pour la classification multi classe	44
30	Analyse comparative des performances des modèles ML et DL en classification binaire	46
31	Répartition de matrice de confusion du XGBoost en classe Binaire	47
32	Analyse comparative des performances des modèles ML et DL en classification multi classe	48
33	Répartition de matrice de confusion du XGBoost en multi classe	49
34	La courbe de ROC du modèle XGBoost en classification binaire	50
35	La moyenne de courbe de ROC du modèle XGBoost en classification multiclasse	50
36	Interface interactive de classification avec XGBoost	53
37	Affichage des premières lignes de la base de données.	54
38	Rapport de classification de notre modèle XGBoost.	54
39	Sélection d'un échantillon de test.	54
40	Affichage détaillé des caractéristiques de l'échantillon, de la prédiction de l'échantillon et sa valeur réelle du diagnostic	55

Liste des tableaux

1	Score appendicite pédiatrique (PAS)	9
2	Encodage des variables catégorielles	39
3	Paramètres de création des modèles classiques	41
4	Paramètres des modèles classiques pour la classification multi classe	43
5	Résultats expérimentaux des modèles de ML et de DL pour la classification binaire	46
6	Résultats expérimentaux des modèles de ML et de DL pour la classification multi classe	48
7	comparaison des résultats avec les résultats des travaux utilisés la BDD Regensburg	52

Liste des abréviations

GBD : Global Burden of Disease Study
NFS : Numération formule sanguine
CRP : Protéine C-réactive
PCT : Procalcitonine
US : Échographie abdominale
CT : Scanner abdomino-pelvien
IRM : Imagerie par résonance magnétique
PAS : Score appendicite pédiatrique
IA : Intelligence artificielle
ML : Machine Learning
RL : L'apprentissage par renforcement
k-NN : k-plus proches voisins
SVM : Support Vector Machines
RBF : fonction de base radiale
XGBoost : Extreme Gradient Boosting
DL : Deep learning
DNN : Réseau de Neurones Profond
CNN : Réseaux de neurones convolutifs
FC : Couches fully connected
RNN : Les réseaux de neurones récurrents
LSTM : Long Short-Term Memory
GRU : Gated Recurrent Unit

Introduction générale

L'appendicite aiguë représente l'une des urgences chirurgicales les plus fréquentes à l'échelle mondiale, avec une incidence particulièrement notable en pédiatrie. Malgré les avancées significatives en imagerie médicale et en examens biologiques, son diagnostic précoce demeure un défi complexe pour les cliniciens. Cette complexité est accentuée par la variabilité des présentations cliniques chez l'enfant, rendant les signes souvent atypiques ou trompeurs. Cette incertitude diagnostique peut avoir des conséquences cliniques importantes, allant de la réalisation d'interventions chirurgicales injustifiées (entraînant des risques anesthésiques et post-opératoires inutiles) à des retards de prise en charge, augmentant ainsi le risque de complications graves telles que la perforation appendiculaire et la péritonite.

Face à cette problématique diagnostique cruciale et aux enjeux sanitaires qui en découlent, une question centrale se pose : Comment l'intégration de l'Intelligence Artificielle (IA), et plus spécifiquement des approches de Machine Learning (ML) et de Deep Learning (DL), peut-elle améliorer significativement la précision et la rapidité du diagnostic de l'appendicite aiguë chez l'enfant ?

Dans ce contexte, l'émergence de l'Intelligence Artificielle offre des perspectives prometteuses pour transformer les pratiques diagnostiques. Les techniques de Machine Learning et de Deep Learning se distinguent par leur capacité à analyser automatiquement de vastes quantités de données médicales hétérogènes (symptômes, données de laboratoire, résultats d'imagerie), en extrayant des schémas complexes et des corrélations difficilement perceptibles par l'analyse humaine seule. Leur intégration progressive dans les systèmes de santé vise à améliorer la précision diagnostique, à optimiser les décisions cliniques et, in fine, à réduire les erreurs médicales.

Dans cette optique d'innovation et d'interdisciplinarité, l'objectif principal de ce projet de fin d'études est de développer et d'évaluer un système d'aide au diagnostic de l'appendicite pédiatrique basé sur des algorithmes de Machine Learning et Deep Learning. Plus spécifiquement, ce travail vise à :

- Comprendre et synthétiser en profondeur les aspects médicaux, épidémiologiques et les défis diagnostiques de l'appendicite aiguë pédiatrique.
- Explorer et comparer les performances de différentes techniques avancées de Ma-

chine Learning et Deep Learning pour la classification binaire et multiclasse de l'appendicite.

- Démontrer l'efficacité et la robustesse des modèles développés en les validant sur une base de données clinique réelle et pertinente ("Regensburg Pediatric Appendicitis Dataset").

Pour atteindre ces objectifs, ce projet se structure de manière logique en trois chapitres complémentaires :

Le premier chapitre, intitulé « Aspects médicaux de l'appendicite », établit les fondements cliniques et épidémiologiques de l'appendicite aiguë, avec un focus particulier sur la population pédiatrique. Il met en lumière la complexité du diagnostic due à la diversité des présentations, soulignant les difficultés spécifiques chez l'enfant où les signes peuvent être particulièrement trompeurs. Ce chapitre détaillé couvre l'anatomie, la physiopathologie, les méthodes d'imagerie, les marqueurs biologiques et les scores diagnostiques couramment utilisés (Alvarado, PAS), fournissant ainsi un socle solide pour appréhender les enjeux cliniques auxquels l'IA devra répondre et identifier les données pertinentes pour son application.

Le second chapitre, « Méthodes d'apprentissage automatique et profond », constitue l'axe technique de ce mémoire. Il présente les concepts fondamentaux des algorithmes d'Intelligence Artificielle exploités dans ce projet. L'originalité de cette partie réside dans la comparaison critique des différentes approches (Machine Learning classique vs. Deep Learning), en détaillant leurs principes, leurs architectures et leur pertinence spécifique dans le domaine médical. En expliquant comment ces modèles peuvent apprendre à partir de données cliniques complexes, ce chapitre assure une transition fluide entre le besoin médical identifié et la solution technologique proposée.

Enfin, le troisième chapitre, « Détection et classification de l'appendicite – Résultats et discussion », décrit la phase de mise en œuvre concrète et d'évaluation du projet. Il détaille l'expérimentation menée sur la base de données "Regensburg Pediatric Appendicitis Dataset", en parcourant les étapes clés : le prétraitement des données, l'encodage, l'entraînement et l'évaluation des modèles, ainsi que la classification binaire et multiclassées. Ce chapitre propose également une analyse comparative et critique des performances obtenues par les différents modèles, tout en discutant des limites rencontrées et des perspectives futures.

Aspect médicaux de l'appendicite

1 Introduction

L'appendicite aiguë demeure l'une des urgences chirurgicales les plus fréquentes en pédiatrie et en médecine générale. Bien qu'elle soit décrite depuis le XIXe siècle, sa présentation clinique, sa physiopathologie et sa prise en charge continuent d'alimenter de nombreux débats en raison de sa variabilité interindividuelle et de la diversité de ses formes cliniques. Ce premier chapitre se propose de dresser un panorama global des aspects médicaux de l'appendicite, en abordant successivement sa définition, son anatomie, sa physiopathologie, les modalités diagnostiques, les classifications cliniques et évolutives, ainsi que les facteurs pronostiques associés. Une attention particulière est portée aux spécificités pédiatriques, où les signes sont parfois trompeurs et les complications plus fréquentes. La compréhension fine de ces éléments conditionne la rapidité du diagnostic, la pertinence des examens complémentaires et l'adéquation du traitement.

2 Définition et Physiopathologie de la maladie

2.1 Définition de l'appendicite

L'appendicite est une inflammation de l'appendice, un organe en forme de tube d'environ 10 centimètres de longueur, qui se rattache au cæcum, la section initiale du gros intestin. Bien que son rôle exact ne soit pas totalement déterminé, on le considère fréquemment comme un organe lymphoïde secondaire qui pourrait intervenir dans les mécanismes de défense de l'organisme [1]. Elle se caractérise généralement par une douleur abdominale, fréquemment située dans la partie inférieure droite de l'abdomen, accompagnée de symptômes tels que la fièvre, les nausées et les vomissements.

2.2 Statistiques épidémiologiques récentes de l'appendicite

D'après les données actuelles du Global Burden of Disease Study (GBD) 2021, près de 17 millions de nouveaux cas d'appendicite ont été recensés à travers le monde, ce qui correspond à un taux standardisé d'incidence par âge de 214 cas pour 100 000 personnes [2].

L'appendicite constitue une proportion significative des cas chez les enfants, avec près de 2,19 millions de nouveaux cas signalés en 2021, représentant environ 12,93 % du total mondial. L'incidence chez les enfants atteint 109 pour 100 000 d'entre eux. Cette condition touche un peu plus les filles (1 147 568 cas; 118 pour 100 000) que les garçons (1 045 452 cas; 101 pour 100 000) [3]. Entre 1990 et 2021, une légère augmentation annuelle moyenne de 0,3 % de l'incidence de l'appendicite chez l'enfant a été observée, avec un pic de fréquence dans la tranche d'âge 10-14 ans[3].

En termes de géographie, on a noté les taux d'incidence les plus importants dans des pays tels que le Bangladesh (1 349,8 pour 100 000), le Bhoutan (1 174,4 pour 100 000) et le Pérou (879,7 pour 100 000). En revanche, des pays tels que l'Éthiopie, le Kenya et l'Indonésie affichent les plus bas taux, à savoir 53,8, 65,4 et 92,5 pour 100 000 personnes respectivement [4].

Concernant l'Algérie, les données précises sur l'incidence nationale de l'appendicite restent limitées. Toutefois, en se basant sur les tendances régionales de l'Afrique du Nord, on peut supposer que le pays présente des taux similaires à ceux observés dans la région. Après avoir établi ce cadre conceptuel, il est essentiel de se pencher sur l'anatomie de l'appendice.

2.3 Anatomie

Dans environ 65 % des cas, l'appendice se trouve en position oblique, souvent dans la région pelvienne. De plus, chez 25 % des personnes, il peut être positionné de manière rétro-caecale, c'est-à-dire placé dans la cavité située derrière le caecum. En conclusion, 5 % des patients présentent un appendice vermiforme orienté vers le côté et vers le haut, situé dans la gouttière para colique droite, entre le côlon ascendant et la paroi antérieure latérale de l'abdomen. L'appendice reçoit son approvisionnement en sang grâce à l'artère terminale de la branche iléo-colique, qui est dérivée de l'artère mésentérique supérieure. Pour ce qui est du drainage lymphatique, il emprunte le parcours du côlon droit, et on trouve souvent un ou deux ganglions lymphatiques au sein du méso-appendice [5].

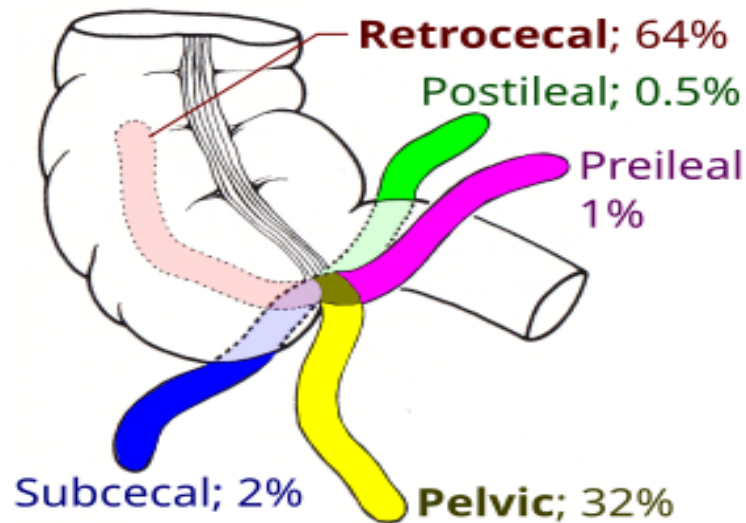


Figure 1 – Localisation de l'appendice [5]

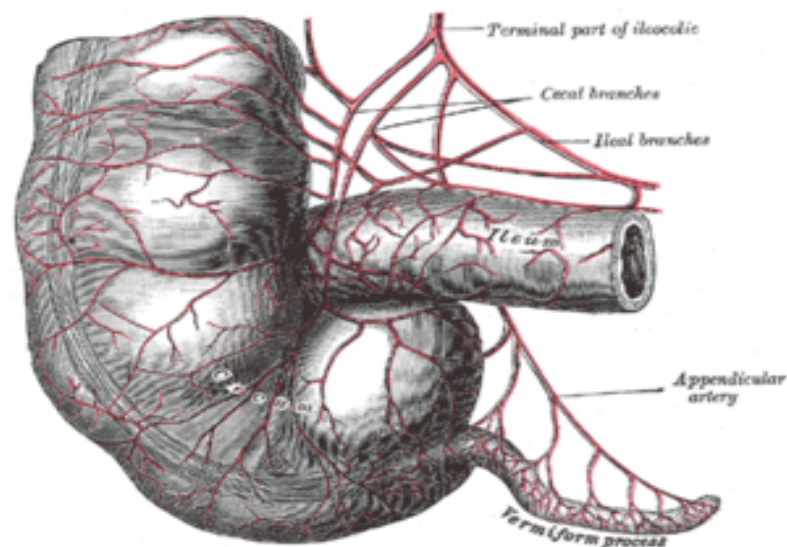


Figure 2 – Vascularisation appendiculaire [5]

2.4 Différences anatomiques de l'appendice entre l'adulte et l'enfant

L'appendice vermiforme montre des particularités anatomiques qui varient en fonction de l'âge. Chez l'adulte, sa longueur varie habituellement de 6 à 12 cm et son diamètre se situe entre 4 et 8 mm. On le trouve généralement à la rencontre de l'iléon terminal et du caecum [6].

En revanche, chez l'enfant, l'appendice est proportionnellement plus court, plus étroit, et sa localisation peut être plus difficile à déterminer en raison de la petite taille de l'abdomen et de la mobilité des structures abdominales.

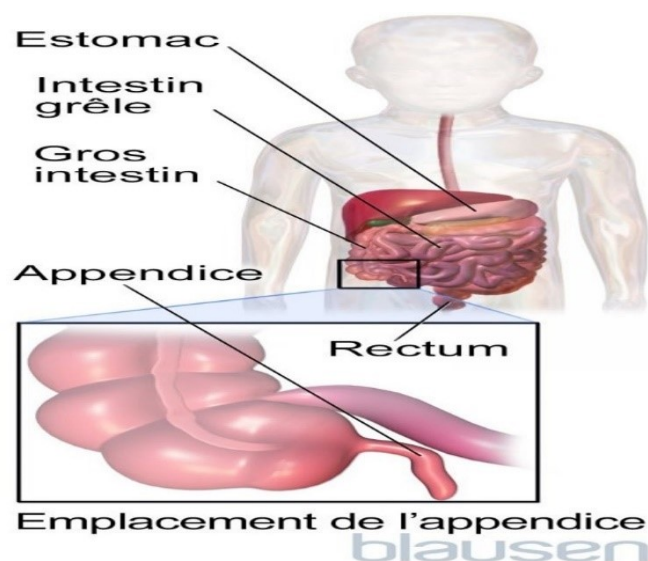


Figure 3 – Emplacement de l'appendice [7]

Une fois l'anatomie clarifiée, nous aborderons la physiopathologie de l'appendicite.

2.5 Physiopathologie de l'appendicite

La physiopathologie de l'appendicite est principalement due à l'obstruction de l'orifice appendiculaire. Cette obstruction entraîne une inflammation et une ischémie localisée, pouvant conduire à une perforation avec formation d'un abcès ou à une perforation franche, provoquant ainsi une péritonite. Les causes de cette obstruction incluent l'hyperplasie lymphoïde, des infections parasitaires, des fécalithes, ainsi que des tumeurs bénignes ou malignes. Lorsqu'une obstruction se produit, elle engendre une augmentation de la pression à l'intérieur de l'appendice, entraînant l'occlusion des petits vaisseaux sanguins et une stase lymphatique. L'appendice s'accumule alors de mucus et se distend, et avec l'aggravation de la compromission lymphatique et vasculaire, la paroi appendiculaire devient ischémique et nécrotique. Une prolifération bactérienne survient dans l'appendice obstrué, initialement dominée par des organismes aérobies, suivie par une croissance mixte d'anaérobies et d'aérobies. En cas d'inflammation et de nécrose sévères, l'appendice peut se perforer, entraînant soit un abcès localisé, soit une péritonite généralisée [5].

Pour compléter notre compréhension, nous procéderons à une analyse rigoureuse des symptômes, des méthodes de diagnostic, et des critères cliniques.

3 Diagnostic et bilan initiaux

3.1 Anamnèse

Lors de l'anamnèse, on observe généralement une douleur abdominale qui débute au niveau péri-ombilical avant de migrer vers la fosse iliaque droite.

L'intensité de la douleur, similaire à une entorse ou un spasme, s'amplifie graduellement au cours des premières 24 heures. Cette douleur est généralement intense et persistante, mais demeure localisée dans le bas-ventre à droite. Elle contraint l'individu à demeurer en position fixe, les jambes repliées. La douleur s'aggrave lors de la toux et des mouvements

de la cuisse droite.

Si l'appendice se trouve en dehors de sa position habituelle (appendice ectopique), la douleur ne sera pas localisée dans la fosse iliaque droite. Selon la position de l'appendice, elle pourrait être plus élevée, plus à gauche ou plus basse dans la région pelvienne.

3.2 Examen et critères clinique

Au cours de l'examen clinique, la palpation abdominale révèle une douleur localisée dans la fosse iliaque droite. Le signe de McBurney est également souvent présent, qui sont douleur causée par une pression sur le point de McBurney dans la fosse iliaque droite [8]. Par ailleurs, une défense musculaire localisée et un signe de Rovsing, Douleur provoquée dans la fosse iliaque droite par la pression exercée au niveau de la fosse iliaque gauche en un point symétrique de celui de McBurney, peuvent être notés [8]. Les indications cliniques comprennent une douleur abdominale déplaçant, commençant près du nombril avant de se fixer dans la fosse iliaque droite. On observe aussi des cas d'anorexie, de nausées et/ou de vomissements, ainsi que de la fièvre généralement au-delà de 38 °C [5].

3.3 Examens et critères biologiques

L'examen biologique est un outil complémentaire au diagnostic clinique et radiologique de l'appendicite aiguë. Il permet d'évaluer la réponse inflammatoire de l'organisme, bien qu'il ne soit pas spécifique. Les examens les plus fréquemment demandés sont :

Numération formule sanguine (NFS) : elle recherche une leucocytose ($> 10\,000/\text{mm}^3$) avec neutrophilie ($> 75\%$), qui reflète une réponse inflammatoire bactérienne [9].

Protéine C-réactive (CRP) : la CRP est un marqueur sensible de l'inflammation aiguë. Un taux supérieur à 10 mg/L est fréquent en cas d'appendicite. Des valeurs > 50 mg/L peuvent suggérer une forme compliquée [10].

Procalcitonine (PCT) : bien que moins sensible que la CRP, un taux $> 0,5$ ng/mL est fortement évocateur d'une appendicite compliquée, comme un abcès ou une péritonite [11].

Bilirubine totale : Une hyperbilirubinémie ($> 20\ \mu\text{mol/L}$) peut être observée en cas de perforation appendiculaire, possiblement liée à une translocation bactérienne [12].

3.4 Examens et critères d'imagerie

L'imagerie joue un rôle fondamental dans la confirmation du diagnostic d'appendicite, en particulier lorsque la présentation clinique est atypique.

3.4.1 Échographie abdominale (US)

L'échographie abdominale est moins sensible (85 %) et spécifique (90 %) que la tomodensitométrie, mais peut être utile pour éviter les rayonnements ionisants chez les enfants et les femmes enceintes[2].

En cas de l'appendicite on va voir un appendice dont le diamètre excède 7 mm une paroi de l'appendice épaissie, La présence d'un cône d'ombre associé à l'appendice (indiquant la présence d'une appendicite), et une lésion en forme de cible du liquide autour de l'appendice et aussi une lumière non compressible (phlegmon ou abcès) [5].

3.4.2 Scanner abdomino-pelvien (CT)

L'imagerie par tomodensitométrie abdominale avec administration d'un agent de contraste présente une exactitude supérieure à 95% dans le diagnostic de l'appendicite et gagne en popularité grâce à sa possibilité d'interprétation à distance. La Tomodensitométrie est efficace pour examiner les abcès, la péritonite (présence de liquide libre), les ganglions lymphatiques et le lien entre l'appendice et les autres organes. Dans les situations d'obésité abdominale qui rendent l'échographie difficile, la tomodensitométrie (TDM) est également une option privilégiée. En outre, chez les seniors qui présentent un risque accru de néoplasie, cette technique d'imagerie enrichit le diagnostic [5]. Pour le diagnostic par scanner, les paramètres de la tomodensitométrie concernant l'appendicite sont les suivants :

- une hypertrophie de l'appendice (diamètre dépassant 6 mm)
- une augmentation de l'épaisseur de la paroi appendiculaire (dépassant 2 mm)
- une infiltration graisseuse autour de l'appendice et une accentuation de la paroi appendiculaire. On observe la présence d'une appendicite chez environ 25% des patients.

Il est rare d'observer de l'air ou un contraste dans la lumière en cas d'appendicite, compte tenu de l'expansion luminale et d'une obstruction potentielle dans la majorité des situations d'appendicite. L'absence de visualisation de l'appendice n'élimine pas la possibilité d'une inflammation de l'appendice [5].

3.4.3 Imagerie par résonance magnétique (IRM) :

L'IRM est une alternative précieuse, notamment chez la femme enceinte, car elle n'utilise pas de rayons ionisants. Il a une sensibilité d'environ 96 à 100 % et une spécificité d'environ 96 à 100 % [13].

3.4.4 Scores diagnostiques :

Certains scores cliniques sont utilisés pour orienter le diagnostic d'appendicite chez l'enfant, bien qu'ils aient des limites spécifiques à l'âge pédiatrique. Certains scores cliniques sont utilisés pour orienter le diagnostic d'appendicite chez l'enfant, bien qu'ils aient des limites spécifiques à l'âge pédiatrique.

a) Score d'Alvarado (sur une échelle de 10 points) : Ce score, bien que développé pour les adultes, est souvent utilisé chez les enfants >5 ans. Il prend en compte :

- Déplacement de la douleur (1 point)
- Anorexie (1 point)
- Nausées/vomissements (1 point)
- Douleur dans la fosse iliaque droite (2 points)
- Défense dans la fosse iliaque droite (1 point)
- Température supérieure à 37,3 °C (1 point)

- Leucocytose supérieure à 10 000/mm³ (2 points)
- 1 point pour une déviation à gauche de la formule leucocytaire [14].

Interprétation :

- Un score de 7 ou plus indique une forte possibilité d'appendicite.
- Un score entre 5 et 6 nécessite une observation et une réévaluation.
- Un score de 4 ou moins indique une faible possibilité d'appendicite [14].

b) Score appendicite pédiatrique (PAS) : Mieux adapté aux enfants. Comprend :

Table 1 – Score appendicite pédiatrique (PAS)

Critère	Points
Neutrophilie >7500	1
Anorexie	1
Nausées/Vomissements	1
Toux provoquant la douleur	2
Douleur FID	2
Défense à la palpation	2
Leucocytose >10 000	1
Neutrophilie >7500	1

Interprétation :

Score 6 : suspicion forte d'appendicite [15].

4 Classification et Facteurs pronostiques

4.1 Classification de l'appendicite

L'appendicite aiguë peut être classée selon deux approches complémentaires : d'une part, son évolution anatomopathologique, et d'autre part, ses manifestations cliniques. Ces classifications permettent d'adapter la prise en charge diagnostique et thérapeutique

4.1.1 Classification selon le stade évolutif

L'appendicite aiguë peut être classée en plusieurs stades selon l'évolution anatomopathologique de l'inflammation de l'appendice. Cette classification permet de mieux comprendre la progression de la maladie et d'adapter la stratégie thérapeutique, en particulier chez l'enfant [16] [17].

a. Appendicite catarrhale (ou congestive) : Il s'agit du stade initial. L'inflammation est limitée à la muqueuse, sans suppuration ni nécrose. L'appendice est simplement congestionné, avec une hyperhémie visible à l'inspection macroscopique. Ce stade est parfois réversible [16].

b. Appendicite suppurée (ou purulente) : L'inflammation devient plus intense, atteignant toutes les couches de la paroi appendiculaire. Il y a accumulation de pus dans la lumière et/ou sur la séreuse de l'appendice. L'appendice apparaît tuméfié, avec des signes cliniques plus francs (fièvre, défense) [17].

c. Appendicite gangréneuse : L'appendicite gangréneuse se caractérise par une réduction ou une éradication totale de la douleur intense antérieurement ressentie sur le flanc droit de l'abdomen, résultant d'une atteinte du système nerveux de l'appendice. Le patient souffre de nausées et de vomissements (occasionnellement fréquents), ce qui ne procure pas de répit. L'état général du patient est critique en raison d'une lourde intoxication, qui influence aussi le rythme du pouls (100-120 pulsations par minute). L'ischémie provoque une nécrose de la paroi appendiculaire. L'appendice devient gris ou noirâtre. Ce stade est considéré comme une forme compliquée avec un risque élevé de perforation [16] [18].

d. Appendicite perforée : Si l'appendicite n'est pas dépistée et traitée, l'appendice peut se perforer (se percer) et former une poche infectée (abcès) à l'extérieur de l'intestin ou répandre le contenu intestinal dans la cavité abdominale, ce qui provoque une grave infection (péritonite). Plus l'enfant est jeune, plus le risque de rupture de l'appendice augmente [7] [16] [17].

4.1.2 Classification selon la présentation clinique

Cliniquement, l'appendicite se manifeste sous différentes formes, influencées par des facteurs individuels tels que l'âge, le sexe, et la position anatomique de l'appendice. On distingue principalement :

Appendicite aiguë typique : elle débute par une douleur abdominale péri-ombilicale, migrante vers la fosse iliaque droite. Ce tableau s'accompagne souvent de fièvre modérée, nausées, vomissements et anorexie. C'est la forme la plus fréquente [16].

Appendicite atypique : cette forme présente des symptômes moins spécifiques. Les douleurs peuvent être diffuses ou pelviennes, parfois associées à des signes digestifs ou urinaires. Elle est plus fréquente chez les enfants, les personnes âgées, et les femmes enceintes [17].

Appendicite compliquée : cette catégorie regroupe les formes associées à des complications, telles que : la perforation appendiculaire ; la péritonite ; l'abcès appendiculaire, qui correspond à une collection purulente localisée ; le plastron appendiculaire, masse inflammatoire englobant l'appendice, l'épiploon et les anses intestinales adjacentes [19].

Appendicite chronique : Moins courante, cela se produit quand l'inflammation de l'appendice n'est pas identifiée et que les symptômes persistent pendant jusqu'à trois semaines. Les symptômes peuvent survenir et s'estomper. Habituellement, l'appendicite chronique est identifiée lorsque la douleur s'intensifie et que le patient se manifeste comme s'il avait une appendicite aiguë [19].

4.2 Étude des facteurs de diagnostic / Causes de l'appendicite

Les facteurs pronostiques de l'appendicite chez les enfants sont essentiels pour évaluer la gravité de la maladie et orienter les décisions thérapeutiques.

4.2.1 Facteurs liés au patient :

L'Âge : Les enfants plus jeunes, en particulier ceux de moins de 5 ans, présentent un risque accru de complications en raison d'un diagnostic souvent retardé [20] [21].

Accès aux soins : Les enfants vivant dans des zones rurales ou ayant un accès limité aux soins médicaux sont plus susceptibles de présenter des complications [21].

Les comorbidités : Les comorbidités associées à l'appendicite comme Les infections gastro-intestinales, l'obésité et les troubles métaboliques comme le diabète qui affectent la réponse inflammatoire peuvent influencer le diagnostic, le traitement et le pronostic [7].

Le délai de consultation : Un délai prolongé entre l'apparition des symptômes et la consultation médicale est associé à un risque plus élevé de perforation et de péritonite [20].

4.2.2 Obstruction de la lumière appendiculaire :

Ce mécanisme représente le principal facteur physiopathologique de l'appendicite. Les causes possibles de cette obstruction incluent : Un fécalome (chez l'adulte), Une hyperplasie lymphoïde (chez l'enfant), Un corps étranger, Une infection parasitaire, Ou une tumeur, bien que rares [22].

4.2.3 Facteurs favorisants :

Plusieurs facteurs peuvent favoriser l'apparition de l'appendicite, notamment un régime alimentaire pauvre en fibres, la déshydratation, ainsi qu'une prédisposition génétique, comme le montre l'agrégation familiale observée.

4.2.4 Processus inflammatoire :

Après la mise en évidence de l'obstruction, une multiplication bactérienne se produit dans la lumière appendiculaire bloquée, conduisant à une dilatation de l'appendice et à une hausse de la pression à l'intérieur du lumen. Simultanément, une altération de la vascularisation de la paroi appendiculaire est observée [23].

4.2.5 Facteurs microbiologiques :

On note l'existence du phénomène de flore polymicrobienne impliquée (aérobies et anaérobies), et de germes régulièrement isolés : *Escherichia coli*, *Bactéroïdes fragilise*, *Peptostreptococcus* [24].

4.3 Défis actuels dans le diagnostic

Limites des examens biologiques : Le manque de spécificité des marqueurs inflammatoires (leucocytose, CRP et l'absence de test biologique pathognomonique).

Défis de l'imagerie : L'appendicite chez les jeunes enfants reste difficile à diagnostiquer malgré les progrès de l'imagerie. Les présentations cliniques atypiques et les complications fréquentes rendent le diagnostic précoce [25], L'échographie est l'examen de première intention, mais sa sensibilité peut être limitée en cas d'appendicite perforée ou si l'appendice n'est pas visualisé. Dans ces situations, l'expertise de l'opérateur joue un rôle déterminant [26]. Le diagnostic échographique de l'appendicite perforée chez l'enfant demeure complexe ; l'association de plusieurs signes sonographiques permet d'en améliorer la spécificité [25]. La tomodensitométrie (CT), bien que plus précise, soulève des préoccupations liées à l'exposition aux radiations ionisantes, ce qui en limite l'utilisation systématique chez les enfants [27]. L'imagerie par résonance magnétique (IRM) constitue une alternative prometteuse, notamment dans les centres spécialisés, car elle combine une excellente sensibilité avec l'absence d'irradiation [28]. Ainsi, le choix de la modalité d'imagerie dépend de l'âge de l'enfant, de la présentation clinique et de l'expertise disponible, et le coût ainsi que l'accessibilité des tests d'imagerie sophistiqués représentent également un enjeu.

Diagnostic différentiel complexe : L'appendicite partage ses symptômes avec de nombreuses autres pathologies abdominales, ce qui rend son diagnostic difficile, surtout chez les enfants, les femmes et les personnes âgées. Des affections telles que les infections urinaires, les gastroentérites, la maladie inflammatoire pelvienne ou encore les douleurs abdominales fonctionnelles peuvent imiter l'appendicite [27].

Équilibre entre sur-diagnostic et sous-diagnostic : Un excès de prudence peut mener à des interventions chirurgicales inutiles (appendicectomies négatives), alors qu'un manque de vigilance peut retarder le traitement d'une vraie appendicite, avec un risque accru de perforation. Ce dilemme diagnostique reste un enjeu constant dans la pratique clinique [27].

Émergence de l'antibiothérapie comme alternative : L'utilisation des antibiotiques seuls pour traiter certaines appendicites non compliquées remet en question l'appendicectomie systématique. Ce changement de paradigme nécessite un diagnostic très précis et une bonne sélection des patients [24].

5 Intégration des nouvelles technologies

L'évolution des technologies numériques a ouvert la voie à l'utilisation de l'intelligence artificielle (IA) dans le domaine médical, notamment pour l'aide au diagnostic. Des algorithmes basés sur l'apprentissage automatique (Machine Learning) sont actuellement en cours de développement pour analyser les données cliniques, biologiques et radiologiques afin de prédire la probabilité d'une appendicite ou orienter vers un traitement optimal. Toutefois, l'utilisation de ces outils en pratique clinique exige une validation rigoureuse de leur performance, leur sécurité, et leur intégration dans les processus décisionnels existants. Sans cela, leur application pourrait mener à des biais ou à des diagnostics inappropriés [29].

6 Traitement

Le traitement de l'appendicite dépend de la gravité de l'inflammation (non compliquée ou compliquée) et des caractéristiques du patient. Traditionnellement, l'appendicectomie (chirurgie d'ablation de l'appendice) est considérée comme le traitement standard. Cependant, des approches non chirurgicales, notamment l'antibiothérapie seule, ont émergé comme alternatives viables dans certains cas. Traitement chirurgical : l'appendicectomie. Le traitement de référence de l'appendicite aiguë, qu'elle soit simple ou compliquée, est l'appendicectomie, c'est-à-dire l'ablation chirurgicale de l'appendice vermiforme.

6.1 Traitement chirurgical

L'appendicectomie, le traitement de référence de l'appendicite aiguë, c'est-à-dire l'ablation chirurgicale de l'appendice vermiforme, notamment par voie laparoscopique, constitue le traitement standard chez l'enfant. Elle permet une récupération rapide, une réduction des complications et un faible taux de récurrence. Une étude multicentrique récente publiée dans *The Lancet* (2025) portant sur 936 enfants a démontré que l'appendicectomie est plus efficace que le traitement antibiotique seul. En effet, le taux d'échec était de 7% dans le groupe chirurgie contre 34% dans le groupe antibiothérapie [30].

6.2 Traitement non chirurgical

Le recours à l'antibiothérapie seule est une alternative envisagée dans les cas d'appendicite non compliquée, bien sélectionnés. Cette approche permet d'éviter la chirurgie et l'anesthésie, mais comporte un risque de récurrence estimé entre 16% et 25% selon les méta-analyses disponibles [31]. Elle nécessite donc un suivi médical rigoureux et une collaboration étroite avec les familles.

6.3 Recommandations actuelles

La chirurgie reste le traitement de première intention dans la plupart des cas, mais l'antibiothérapie seule peut être proposée dans des contextes particuliers. La décision thérapeutique doit être individualisée, en tenant compte de l'âge de l'enfant, de la sévérité de la maladie, des souhaits parentaux, et des ressources disponibles.

7 Conclusion

L'appendicite demeure une pathologie abdominale fréquente, à la croisée d'enjeux diagnostics, thérapeutiques et pronostiques. Ce chapitre a permis de dégager les fondements anatomiques et physiopathologiques de l'inflammation appendiculaire, en insistant sur la diversité des présentations cliniques et l'apport essentiel des examens biologiques et radiologiques dans le diagnostic. Les classifications évolutives, les scores cliniques et les nouvelles stratégies thérapeutiques, notamment l'antibiothérapie dans les formes simples, témoignent d'une prise en charge de plus en plus individualisée. Toutefois, les limites des outils actuels et la complexité du diagnostic différentiel, en particulier chez l'enfant, rappellent l'importance d'une évaluation globale et contextualisée du patient. Ce socle de connaissances constitue une base essentielle à l'approche pédiatrique de l'appendicite, objet des chapitres suivants

Introduction aux méthodes d'apprentissage automatique

1 Introduction

La classification de l'appendicite est d'une grande importance dans le secteur de la santé, surtout compte tenu de la difficulté du diagnostic et de l'exigence d'un traitement rapide et exact. Cette affection abdominale courante peut manifester divers symptômes, ce qui peut compliquer son diagnostic, même pour des praticiens chevronnés. Grâce aux avancées récentes en matière d'intelligence artificielle et de technologies de l'information, des méthodes novatrices ont vu le jour pour aider les professionnels de la santé à identifier et classer l'appendicite.

Dans ces méthodes, le recours à des algorithmes d'apprentissage automatique facilite l'extraction automatique de caractéristiques pertinentes à partir de données médicales comme les indicateurs cliniques (résultats biologiques, symptômes, etc.). Dans les années récentes, l'apprentissage profond, en particulier, a démontré son efficacité dans l'étude des données médicales, facilitant ainsi une classification automatisée et plus fiable de l'appendicite.

Ce chapitre explore les fondements des méthodes d'apprentissage automatique, en mettant l'accent sur ces types, en particulier l'apprentissage supervisé par la suite on détaille les architectures algorithmiques de cette approche telles que, les réseaux de neurone, la régression logistique, et les méthodes d'ensemble. Dans la dernière partie de ce chapitre, nous intéressons plus particulièrement à l'apprentissage en profondeur : sa définition, ces types tout en illustrant leur application concrète dans le domaine du diagnostic médical, notamment pour l'appendicite. À travers l'analyse de cas d'études récents, ce chapitre met en lumière le rôle croissant de l'IA comme outil d'aide à la décision en médecine.

2 Intelligence artificielle (IA)

Dans le domaine médical, l'intelligence artificielle (IA) fait référence à l'application des algorithmes d'apprentissage automatique et des approches basées sur les données médicales pour résoudre des problèmes biologiques et médicaux complexes [32]. L'IA est une technologie qui permet aux ordinateurs et aux machines de simuler l'apprentissage,

La compréhension, la résolution de problèmes, la prise de décision, la créativité et l'autonomie de l'être humain.

Les applications et les appareils équipés d'une IA peuvent voir et identifier les objets. Ils sont capables de comprendre le langage humain et d'y répondre. Ils peuvent tirer des enseignements de nouvelles informations et de nouvelles expériences. Ils peuvent formuler des recommandations détaillées aux utilisateurs et aux experts. Ils peuvent agir indépendamment, remplaçant ainsi l'intelligence ou l'intervention humaine [33].

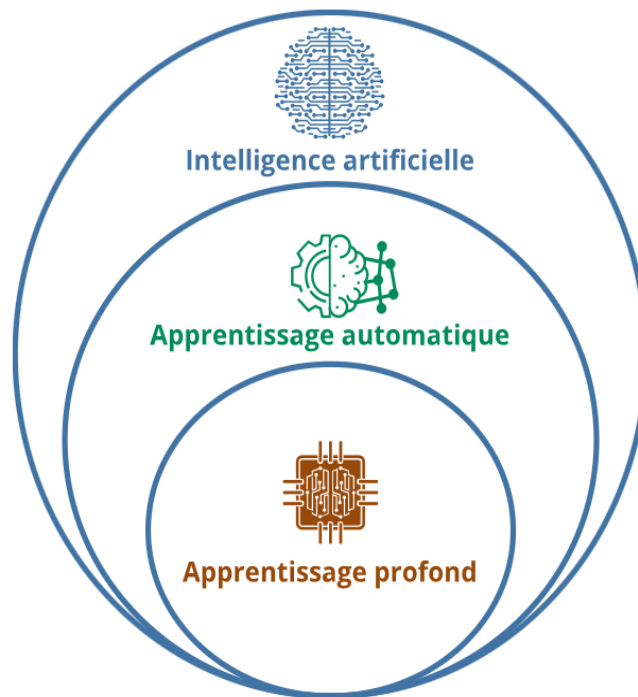


Figure 4 – La relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond [34]

3 L'apprentissage automatique (Machine Learning)

L'apprentissage automatique ou Le Machine Learning (ML) est un domaine scientifique, et plus particulièrement une sous-catégorie de l'intelligence artificielle. Elle consiste à laisser des algorithmes découvrir des « patterns », à savoir des motifs récurrents, dans les ensembles de données.

Ces données peuvent être des chiffres, des mots, des images, des statistiques. Tout ce qui peut être stocké numériquement peut servir de données pour le l'apprentissage automatique. En décelant les patterns dans ces données, les algorithmes apprennent et améliorent leurs performances dans l'exécution d'une tâche spécifique [35].

Ces algorithmes destinés à construire des applications intelligentes apprennent de l'expérience pour effectuer des prédictions intelligentes et s'améliorer au cours du temps sur la base d'un ensemble de données. Cet ensemble de données définit également le type d'algorithme que nous choisissons pour le modèle. La construction d'une application ML repose sur 4 étapes qui sont la sélection, la préparation de l'ensemble de données, le choix d'un algorithme à exécuter sur ces données, l'entraînement de l'algorithme pour

créer le modèle et enfin l'utilisation et l'amélioration du modèle [36].

4 Approches d'apprentissage automatique

Il y a une grande variété de systèmes d'apprentissage automatique peuvent être divisées en plusieurs catégories (Figure 4), notamment :

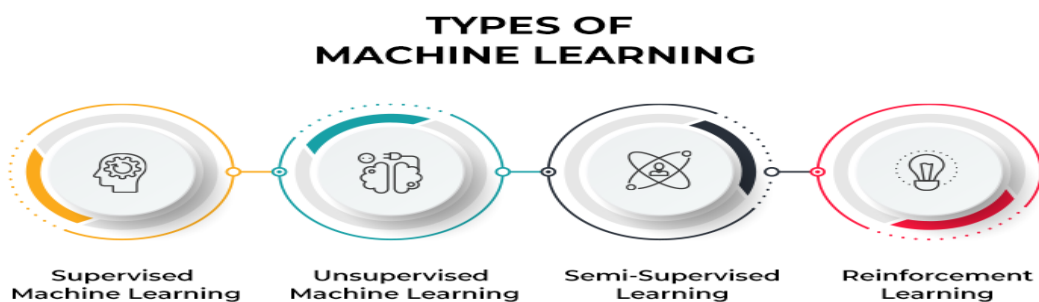


Figure 5 – Types d'apprentissage automatique [37]

4.1 L'apprentissage supervisé

L'apprentissage supervisé implique l'emploi d'un jeu de données labellisées pour entraîner un algorithme d'apprentissage automatique. Cette supervision repose sur le fait que les étiquettes associées aux données permettent au modèle de faire des prédictions précises. En d'autres termes, les données de formation comportent déjà les réponses appropriées aux résultats attendus [35]. Voir figure 5

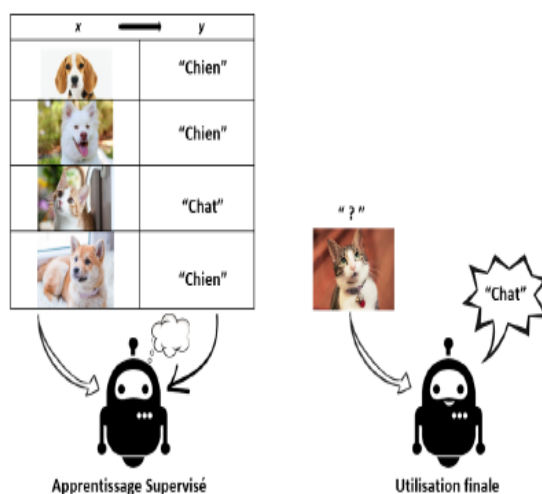


Figure 6 – Exemple d'apprentissage supervisé [38]

Il existe deux catégories principales d'algorithmes qui sont : la régression et la classification :

Les algorithmes de régression : sont employés pour évaluer ou anticiper des valeurs continues, autrement dit des montants numériques. Ces algorithmes visent à établir un lien mathématique entre les variables d'entrée et de sortie pour être en mesure de prévoir une valeur numérique pour des données inédites. On les utilise couramment dans des secteurs tels que l'économie, la finance, la météorologie et d'autres domaines où il est crucial de prévoir des valeurs continues [38].

Les algorithmes de classification : sont élaborés pour anticiper la classe ou la catégorie d'une nouvelle donnée. Ils examinent les propriétés ou les caractéristiques des données d'entrée et les lient à des catégories préétablies. Le but est de créer un modèle capable de généraliser cette association pour classer de nouvelles données. Les algorithmes de classification trouvent une utilisation répandue dans des secteurs tels que la reconnaissance d'images, le filtrage de courriers indésirables, la prévision de maladies et diverses autres applications où une catégorisation précise des données est primordiale [38].

4.2 L'apprentissage non supervisé

L'apprentissage non supervisé est basé sur l'utilisation de données non étiquetées pour former un modèle. En d'autres termes, l'algorithme apprend par lui-même à découvrir des informations à partir d'un ensemble de données. Il est surtout utilisé pour le clustering et l'association [39].

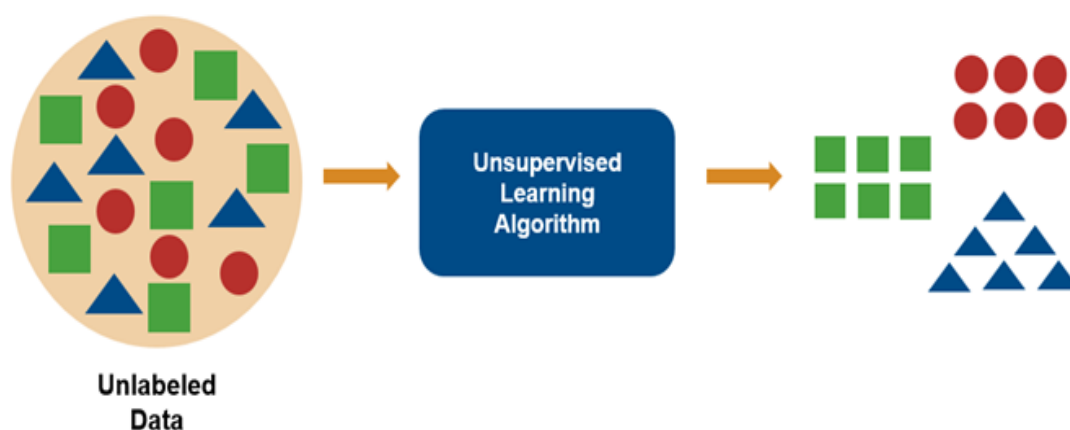


Figure 7 – Exemple d'apprentissage non supervisé [40]

Plusieurs algorithmes sont utilisés pour regrouper les données en clusters :

K-means : est un processus largement utilisé pour la classification des données non supervisé, dans lequel les points de données sont répartis en K groupes en fonction de leur distance par rapport au centre (ou centre de gravité) de chaque groupe [41]. Les méthodes de regroupement hiérarchique créent un dendrogramme hiérarchique en forme d'arbre, avec deux approches principales : agglomérative (fusion de petits clusters en plus grands) et divisivé (division de grands clusters en plus petits) [42].

4.3 L'apprentissage Semi-supervisée

L'apprentissage semi-supervisé est une méthode qui fusionne les bénéfices de l'apprentissage supervisé et non supervisé. Il se fonde sur un ensemble de données principalement non étiquetées complété par une quantité restreinte d'éléments labellisés. Le but de cette technique est d'exploiter les données présentes dans les éléments non étiquetés afin d'optimiser l'efficacité d'un modèle [43].

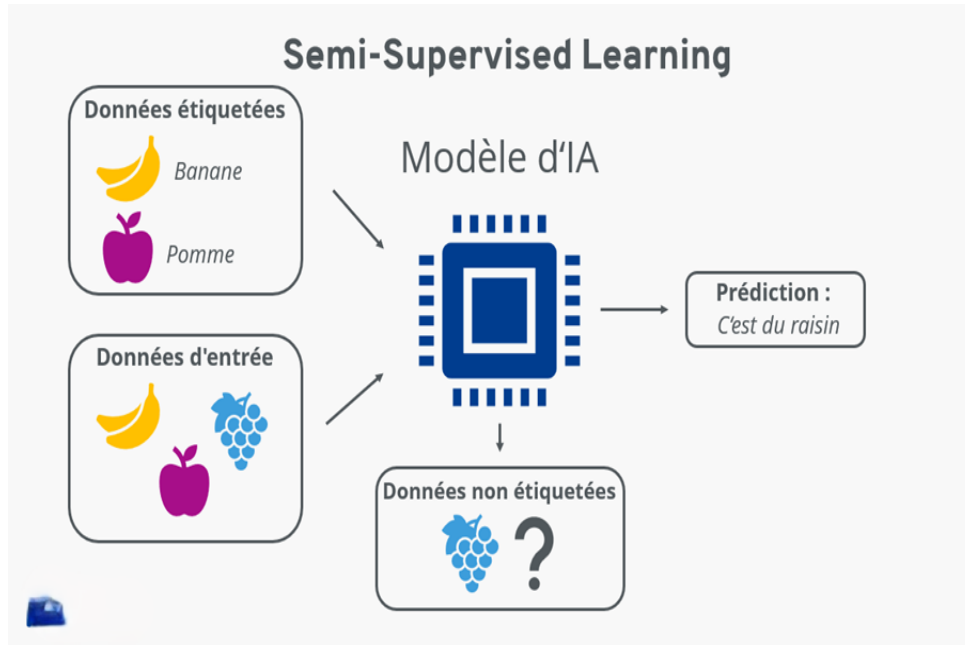


Figure 8 – Exemple d'apprentissage semi supervisé [44]

4.4 L'apprentissage par renforcement (RL)

L'apprentissage par renforcement est une méthode qui forme les programmes à faire des choix dans le but d'atteindre les meilleurs résultats. Elle reproduit la méthode d'apprentissage par essais et erreurs utilisée par les humains pour parvenir à leurs buts. Les actions qui aident à atteindre votre but sont renforcées, tandis que celles qui entravent ce dernier sont écartées [45].

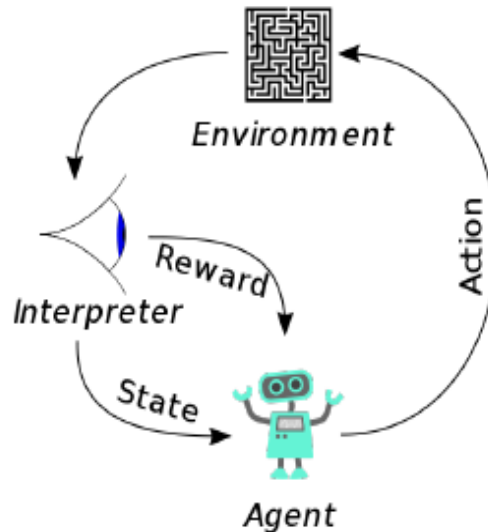


Figure 9 – Exemple d'apprentissage par renforcement [46]

5 Algorithmes d'apprentissage automatique

Dans le cadre de notre projet de fin d'études, Nous nous sommes appuyés sur l'apprentissage supervisé afin de développer un système de détection et de classification de l'appendicite.

L'approche supervisée s'est révélée pertinente pour notre problématique, car elle permet d'optimiser la précision du diagnostic en s'appuyant sur des algorithmes éprouvés de classification, voici quelques-uns des algorithmes d'apprentissage supervisé :

5.1 Le Naïve Bayes

est une méthode de classification probabiliste fondée sur le théorème de Bayes, avec une hypothèse forte (naïve) d'indépendance entre les caractéristiques. Cet algorithme est largement utilisé en apprentissage supervisé, en particulier pour les tâches de classification de texte, comme la détection de spam ou l'analyse de sentiments. Sa simplicité, sa rapidité d'apprentissage, et son efficacité sur de grands ensembles de données en font un choix populaire, même lorsque l'hypothèse d'indépendance est souvent irréaliste [47][48].

Son fonctionnement repose sur le théorème de Bayes, qui permet de calculer la probabilité qu'un exemple appartienne à une classe donnée à l'aide de la formule

$$P(A|B) = \frac{(P(B|A)P(A))}{(P(B))} \quad (2.1)$$

où $P(A|B)$ est la probabilité de la classe A étant donné les caractéristiques B, $P(B|A)$ est la probabilité des caractéristiques B étant donné la classe A, et $P(A)$ et $P(B)$ sont les probabilités a priori des classes et des caractéristiques respectivement B. L'hypothèse clé de Naïve Bayes est que toutes les caractéristiques sont indépendantes les unes des

autres conditionnellement à la classe, simplifiant ainsi les calculs en décomposant la probabilité conditionnelle en un produit de probabilités individuelles.

Il existe plusieurs types de l'algorithme Naïve Bayes, tels que :

- **Le Naïve Bayes Gaussien** : utilisé pour des attributs continus suivant une distribution normale.
- **Le Naïve Bayes Multinomial** : adapté pour des données discrètes comme les fréquences de mots.

5.2 Régressions Linéaires

La régression linéaire est une méthode statistique permettant de modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes en ajustant une équation linéaire aux données observées [49].

Elle repose sur un modèle de régression, généralement représenté par l'équation (2.2)

$$y = mx + by = mx + b \quad (2.2)$$

, où y désigne la variable dépendante, x la variable indépendante, m la pente de la droite, et b l'ordonnée à l'origine [50].

Le but est de trouver la ligne qui minimise la distance entre les points de données et la droite de régression. Pour cela, on utilise une fonction de perte, souvent la somme des carrés des erreurs (SSE), qui mesure l'écart entre les valeurs réelles et celles estimées par le modèle. L'optimisation de cette fonction est généralement effectuée à l'aide de la méthode du gradient, une technique qui ajuste progressivement les paramètres du modèle (pente et intercept) en fonction de la pente de la fonction de perte, jusqu'à trouver les valeurs optimales [51].

Les coefficients ainsi obtenus permettent d'interpréter l'influence des variables indépendantes sur la variable dépendante ; par exemple, dans un modèle simple, la pente indique de combien varie la variable dépendante lorsque la variable indépendante augmente d'une unité [52].

5.3 Régressions Logistiques

La régression logistique est une méthode statistique utilisée pour modéliser la probabilité qu'une observation appartienne à l'une des deux catégories d'une variable dépendante binaire. Contrairement à la régression linéaire, qui prédit une valeur continue, la régression logistique estime la probabilité de survenue d'un événement en utilisant une fonction logistique (ou sigmoïde) pour transformer l'entrée en une valeur comprise entre 0 et 1 [52][53]. Se présente désormais comme suit :

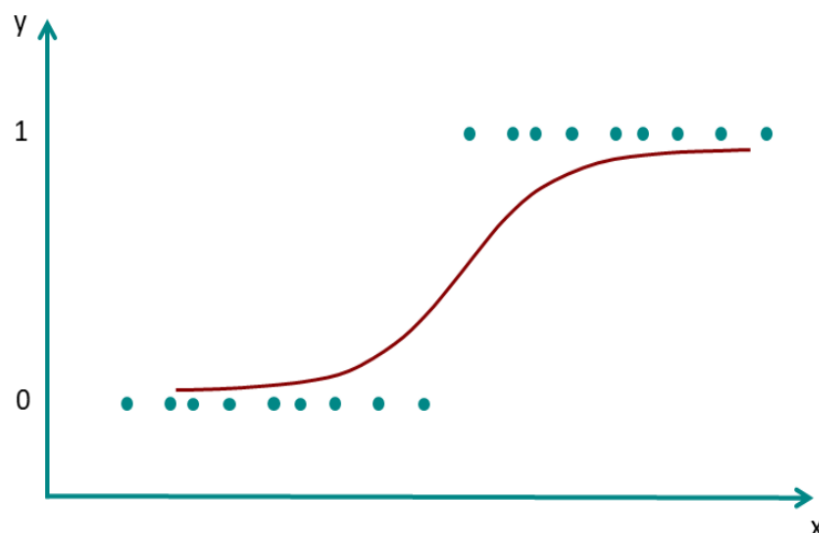


Figure 10 – une régression logistique binaire[54]

Le modèle de base repose sur l'équation : (2.3)

$$P(y = 1 | x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (2.3)$$

Où $P(y=1|x)$ représente la probabilité que l'événement se produise 0 est l'ordonnée à l'origine, et 1, 2,,n sont les coefficients des variables indépendantes x_1, x_2, \dots, x_n [55] [56]. Le modèle utilise la fonction de perte log-vraisemblance, qui mesure l'écart entre les valeurs réelles et les probabilités estimées, et cherche à la minimiser pour obtenir les meilleurs paramètres possibles [55].

Pour ce faire, des techniques d'optimisation comme la descente de gradient sont utilisées afin d'ajuster les coefficients du modèle et améliorer ses performances [56]. L'interprétation des coefficients permet de comprendre l'effet des variables indépendantes sur la probabilité de l'événement. Par exemple un coefficient positif signifie qu'une augmentation de la variable concernée augmente la probabilité que l'événement ait lieu [35].

5.4 k-plus proches voisins (k-NN)

L'algorithme des k plus proches voisins (k-NN) effectue une classification d'une observation en se basant sur les classes de ses k voisins les plus proches dans l'espace des attributs. La classe prédominante parmi ces k voisins est assignée à l'observation [57]. L'algorithme de construction du KNN se déroule généralement comme suit :[58]

- Sélectionnez le nombre K de voisins
- Calcule la distance entre le point à prédire et tous les points du jeu d'entraînement à l'aide d'une métrique telle que la distance euclidienne ou la distance de Manhattan [39][42].

$$\text{Manhattan} = \sum_{i=1}^n |x_i - y_i| \quad \text{Euclidienne} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.4)$$

- Prenez les K voisins les plus proches selon la distance calculée.
- Parmi ces K voisins, comptez le nombre de points appartenant à chaque catégorie.
- Attribuez le nouveau point à la catégorie la plus présente parmi ces K voisins.

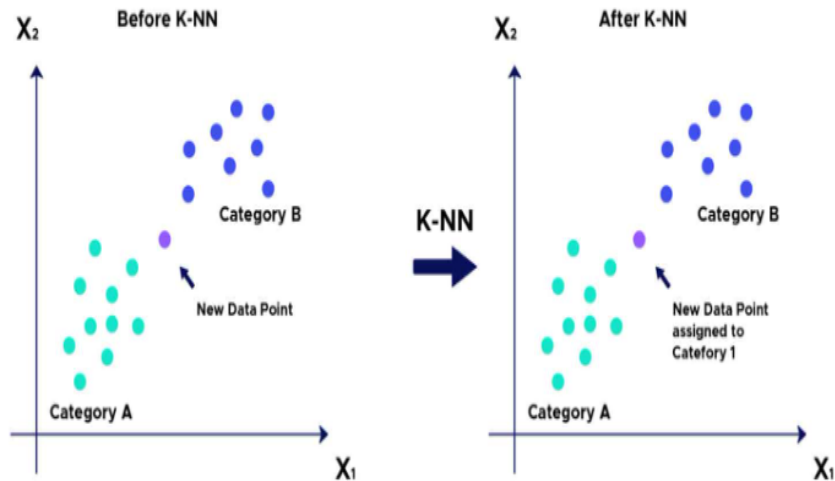


Figure 11 – Exemple de modèle KNN [58]

5.5 Arbres de décision

Les arbres de décision sont des modèles d'apprentissage supervisé non paramétriques, utilisés pour des tâches de classification et de régression. Ils reposent sur une structure hiérarchique formée de nœuds (racine, internes et terminaux) et de branches, organisant un partitionnement récursif de l'espace des données. À chaque étape, une stratégie gloutonne sélectionne la variable et le seuil optimaux afin de maximiser un critère d'homogénéité (comme l'indice de Gini ou l'entropie) [59].

Le modèle est construit de manière itérative jusqu'à ce qu'un critère d'arrêt soit atteint, mais une croissance excessive peut induire un surapprentissage. Pour y remédier, on applique une taille minimale de nœuds, une profondeur maximale, ou un élagage post-hoc, afin de contrôler la complexité et améliorer la généralisation. Cette simplification s'inscrit dans la logique du Rasoir d'Occam, qui privilégie les modèles les plus simples capables d'expliquer les données [59].

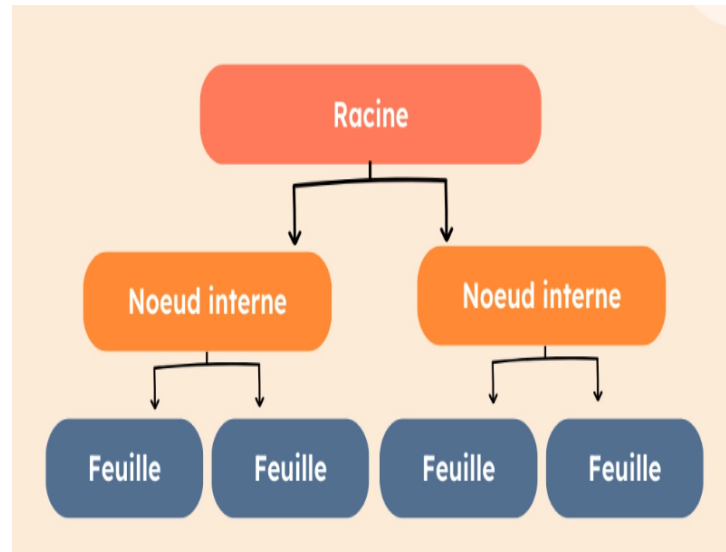


Figure 12 – Exemple des Arbres de décision [60]

5.6 Support Vector Machines (SVM)

Les SVM sont des algorithmes d'apprentissage supervisé utilisés pour la classification et la régression, dont l'objectif est de trouver une frontière de séparation optimisant la marge entre les différentes classes [61].

Les SVM reposent sur plusieurs principes fondamentaux pour fonctionner efficacement. Tout d'abord, elles cherchent à déterminer un hyperplan de séparation qui maximise la distance entre les points de données les plus proches appartenant à des classes différentes, une distance appelée marge maximale. Les points situés au plus près de cet hyperplan sont appelés vecteurs de support ; ils jouent un rôle clé, car ils déterminent l'emplacement de l'hyperplan et influencent directement la qualité de la séparation. Lorsque les données ne sont pas linéairement séparables, les SVM utilisent des fonctions de noyau pour transformer l'espace de données initial en un espace de dimension supérieure, dans lequel une séparation linéaire devient possible. Les noyaux les plus couramment utilisés sont les noyaux linéaires, polynomiaux, à fonction de base radiale (RBF) et sigmoïdes [32][61][62]. Figure 13

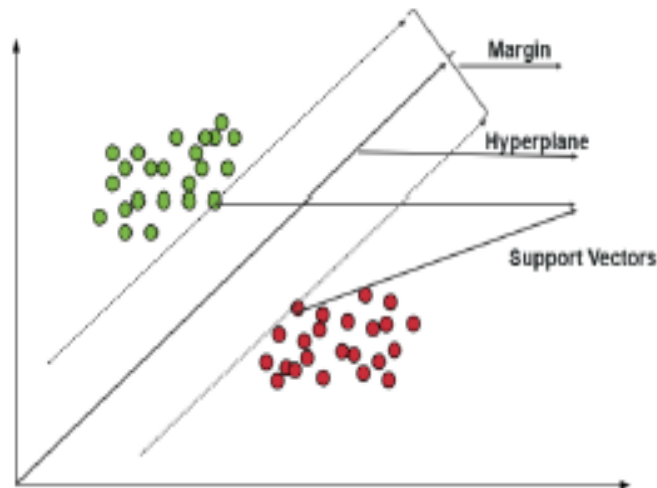


Figure 13 – Diagram representing the SVM classification of binary class data (Positive, Negative) by plotting the optimal hyperplane [63].

Enfin, pour identifier l'hyperplan optimal, les SVM s'appuient sur des techniques d'optimisation, notamment la résolution de problèmes d'optimisation quadratique, afin de maximiser la marge tout en minimisant les erreurs de classification [64].

5.7 Les méthodes d'ensemble en Machine Learning

Les méthodes d'ensemble représentent des stratégies d'apprentissage automatique qui visent à fusionner divers modèles d'apprentissage afin de renforcer la précision, la solidité et la capacité de généralisation. Elles font appel à la diversité parmi les modèles de base pour pallier aux erreurs individuelles [65].

5.7.1 Forêts Aléatoires

Les forêts aléatoires (Random Forest) sont une méthode d'ensemble en apprentissage automatique qui repose sur la combinaison de plusieurs arbres de décision afin d'améliorer la précision des prédictions et de limiter le risque de surapprentissage [66].

Le principe consiste à générer un grand nombre d'arbres de décision, chacun étant construit à partir d'un sous-ensemble aléatoire du jeu de données d'apprentissage, sélectionné selon le principe du bagging (bootstrap aggregating), c'est-à-dire par tirage avec remise. Ainsi, certaines observations peuvent être sélectionnées plusieurs fois, tandis que d'autres peuvent être ignorées. À chaque nœud de chaque arbre, un sous-ensemble aléatoire de caractéristiques est également sélectionné pour déterminer la meilleure division, ce qui introduit de la diversité entre les arbres et réduit la corrélation entre eux. Lors de la prédiction, chaque arbre de la forêt fournit une réponse : pour une tâche de classification, la classe finale est celle qui recueille le plus de votes, tandis que pour une tâche de régression, on calcule la moyenne des prédictions de tous les arbres [68][69][70].

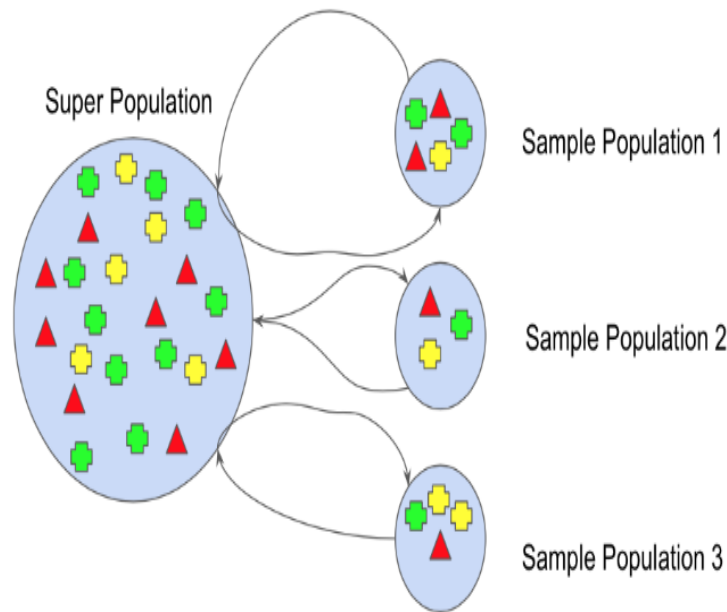


Figure 14 – Exemple de forêt aléatoire [67].

5.7.2 Gradient Boosting

suit une logique similaire mais se concentre sur la minimisation d'une fonction de perte. Il commence par entraîner un modèle initial, puis calcule les résidus (l'écart entre les prédictions et les valeurs réelles). À chaque itération, un nouveau modèle faible est entraîné pour prédire ces résidus, et le modèle global est ajusté en intégrant progressivement les nouvelles corrections. Cette méthode est particulièrement efficace lorsqu'elle est couplée avec des arbres de décision comme modèles de base [71][72].

5.7.3 XGBoost (Extreme Gradient Boosting)

est une implémentation optimisée du gradient boosting. Il améliore la précision et la vitesse d'entraînement grâce à diverses optimisations, comme la régularisation, le parallélisme, et une meilleure gestion des données manquantes. Le processus débute par l'entraînement d'un modèle de base, suivi du calcul des gradients de la fonction de perte. À chaque cycle, un nouveau modèle est entraîné pour prédire ces gradients, et le modèle principal est mis à jour en utilisant une approximation de second ordre, ce qui améliore la convergence et la performance [73].

5.8 Réseaux neuronaux

Les réseaux neuronaux s'inspirent du fonctionnement des neurones biologiques présents dans le cerveau humain. Ils sont constitués de couches de neurones artificiels interconnectés, qui permettent de modéliser des relations complexes entre les données d'entrée et les résultats de sortie à travers des connexions pondérées [74]. Leur architecture est généralement organisée en plusieurs strates :

- Une couche d'entrée qui reçoit les données brutes.
- Des couches cachées qui réalisent des transformations non linéaires complexes.
- Une couche de sortie qui fournit la prédiction finale.

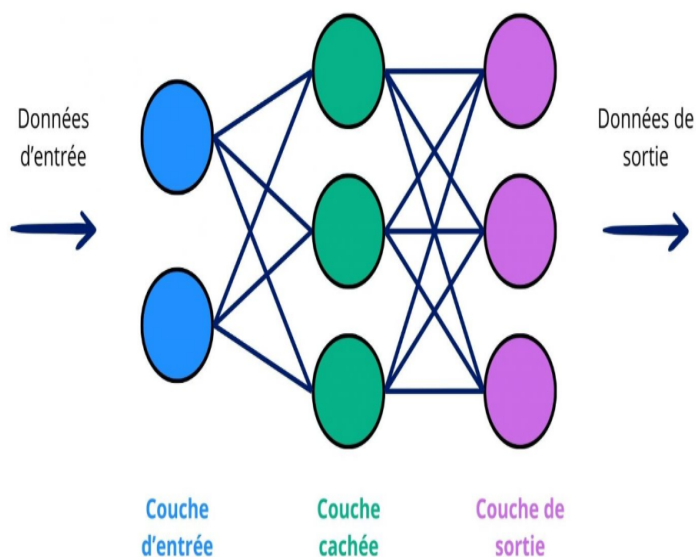


Figure 15 – Exemple de réseau de neurones avec une couche d'entrée, une couche cachée et une couche de sortie [75].

Chaque neurone utilise une fonction d'activation, telle que la sigmoïde ou ReLU (Rectified Linear Unit), pour introduire de la non-linéarité dans le modèle, ce qui permet de traiter des problèmes plus complexes. Le processus de calcul débute par une propagation en avant, où les données circulent à travers les couches du réseau, chaque neurone effectuant des opérations mathématiques et transmettant ses résultats aux neurones suivants. L'apprentissage du réseau repose ensuite sur la rétropropagation, une méthode qui ajuste les poids des connexions en minimisant l'erreur de prédiction à l'aide de la descente de gradient. Pour optimiser les performances du réseau, des algorithmes tels qu'Adam ou la descente de gradient stochastique (SGD) sont fréquemment utilisés [76][77].

6 L'apprentissage profond « Deep Learning »

L'apprentissage profond, également connu sous le nom de Deep Learning, est une branche spécifique de l'apprentissage automatique (Machine Learning), qui lui-même est un segment de l'intelligence artificielle. On entend par intelligence artificielle l'acte de faire exécuter par des machines des tâches que les humains considèrent comme complexes, telles que : permettre à une machine de raisonner, enseigner à une machine l'organisation de tâches ou encore apprendre à une machine comment structurer des connaissances. Il englobe les techniques les plus efficaces et performantes mises en œuvre au sein de la communauté du Machine Learning [78].

L'illustration ci-dessous montre la relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond

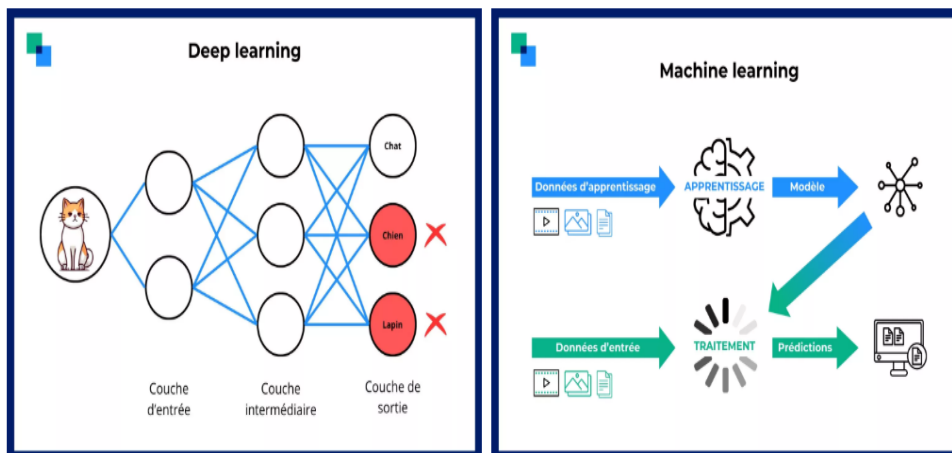


Figure 16 – La différence entre DL et ML[79]

L'apprentissage en profondeur est une technique d'intelligence artificielle qui s'appuie sur des réseaux de neurones profonds pour se former et optimiser ses performances de façon autonome.

6.1 Réseaux de neurones profonds

Un Réseau de Neurones Profond, également connu sous l'abréviation DNN pour Deep Neural Network, se caractérise par un aspect spécifique : il comporte un minimum de deux couches. Cela lui donne la capacité de gérer les données de façon complexe, en utilisant des modèles mathématiques sophistiqués [80].

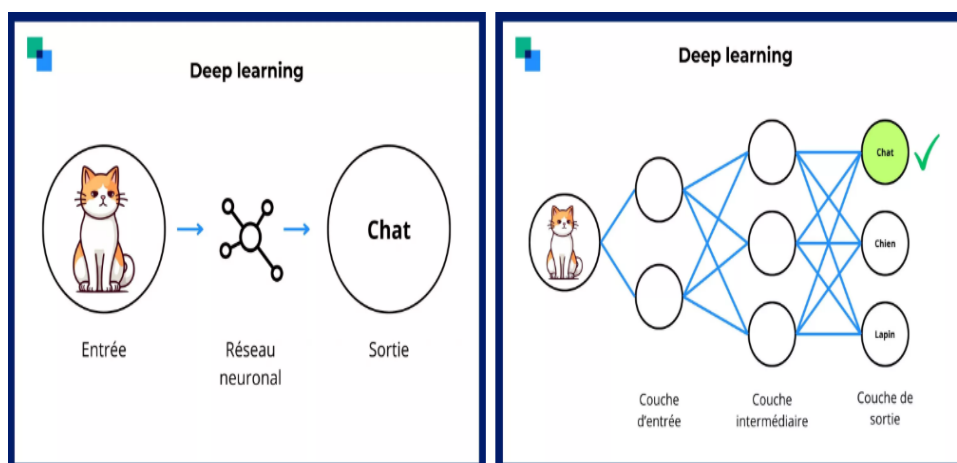


Figure 17 – Fonctionnement de réseau neuronal [79]

6.2 Réseaux de neurones convolutifs (CNN)

Les CNN, ou réseaux de neurones convolutifs, constituent une catégorie précise de réseaux de neurones qui trouvent leur utilité dans le traitement des données visuelles, telles que les images médicales. Ils sont particulièrement aptes à identifier des schémas complexes en raison de leur capacité à assimiler les caractéristiques directement à partir des données non traitées [81].

Les CNN comportent plusieurs niveaux distincts comme la montre la figure 18. Chaque couche joue un rôle spécifique dans l'acquisition des caractéristiques et la catégorisation des données :

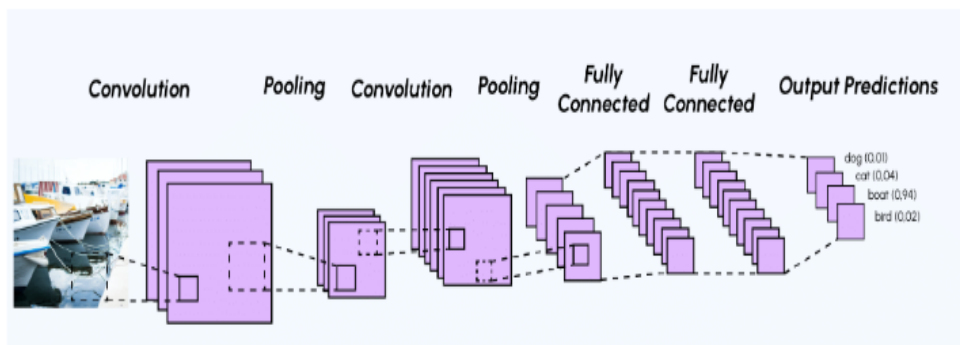


Figure 18 – Les couche de CNN [82]

1. Couche Convolutionnelle : Cette couche consiste à appliquer des filtres (également appelés noyaux) sur l'entrée (qu'il s'agisse d'une image ou d'une carte de caractéristiques) afin d'identifier des motifs locaux tels que des bords, des textures ou des formes. Chaque filtre génère une carte de caractéristiques (feature map) [83].

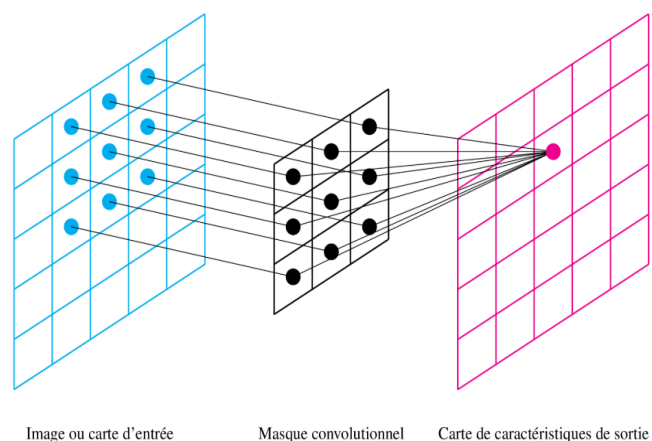


Figure 19 – Une couche convolutive en action [84]

2. Couche de Pooling (Pooling Layer) : Ces couches diminuent la dimension des données tout en préservant les informations clés. Le max pooling et l'average pooling, qui retiennent respectivement la valeur maximale ou moyenne d'un groupe de valeurs

[85]. Son but est de réduire progressivement la taille spatiale de la représentation afin de diminuer le nombre de paramètres et de calculs dans le réseau et, par conséquent, de contrôler le sur apprentissage (overfitting).

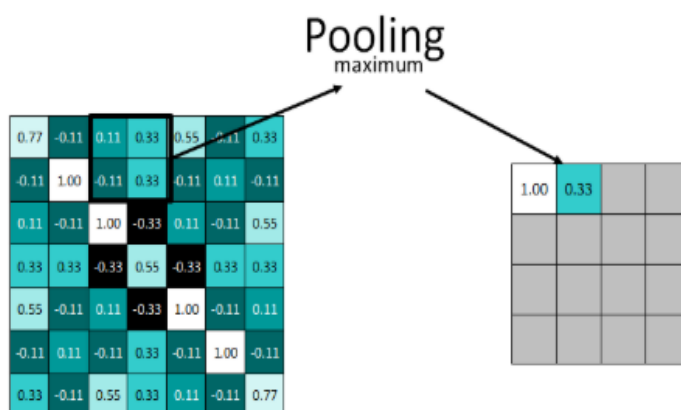


Figure 20 – Exemple de fonctionnement de la couche pooling [86]

3. Couche entièrement connectée (Fully Connected Layer / Dense Layer) : Ces couches se trouvent généralement à la fin du réseau et relie chaque neurone de la couche précédente à tous les neurones de la couche suivante (Figure 21). Elles utilisent les attributs extraits par les couches antérieures pour réaliser la classification définitive [87].

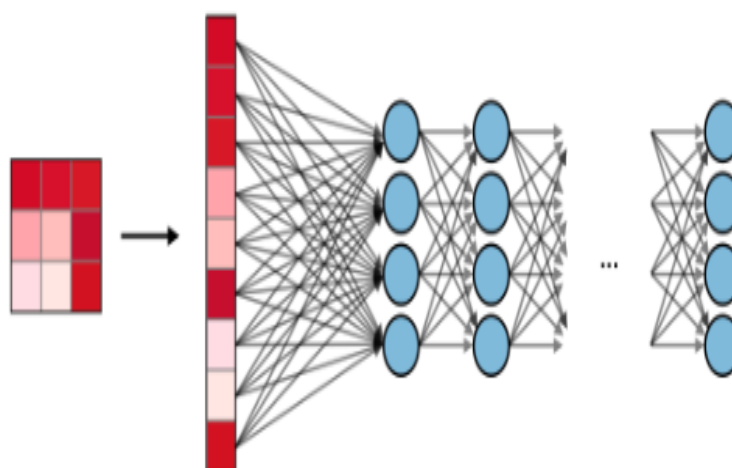


Figure 21 – Exemple des Couches entiereement connectées [87]

4. Couche de sortie (Output Layer) Elle Produit la sortie finale du modèle. Le choix de la fonction d'activation se fait selon la tâche :

Softmax : Cette fonction est fréquemment utilisée en sortie des réseaux de neurones pour convertir les scores bruts en probabilités. Elle classe les données en assignant une probabilité à chaque catégorie, facilitant ainsi la prise de décision [88].

Sigmoid : Utilisée pour les tâches de classification binaire. Elle transforme les valeurs de sortie en probabilités comprises entre 0 et 1, facilitant la distinction entre deux classes.

Rien ou linéaire pour les tâches de régression Linéaire : Utilisée pour les tâches de régression. La fonction linéaire ne modifie pas les valeurs de sortie, permettant de prédire des valeurs continues [89].

6.3 Réseaux de neurones récurrents (RNN)

Les réseaux de neurones récurrents (RNN) : sont conçus pour traiter des données séquentielles en conservant en mémoire les états précédents afin de mieux contextualiser les décisions actuelles. Contrairement aux réseaux de neurones classiques, les RNN intègrent des connexions récurrentes qui leur permettent de prendre en compte le contexte temporel antérieur dans le traitement des séquences [90][91].

Leur architecture repose sur des neurones récurrents qui reçoivent des signaux non seulement de la couche précédente, mais aussi d'eux-mêmes lors de l'étape antérieure, ce qui leur permet de mémoriser les états passés [90][91]. Le fonctionnement du RNN implique une phase de propagation des données vers l'avant suivie d'une rétropropagation de l'erreur pour ajuster les poids synaptiques [85]. Les fonctions d'activation couramment utilisées dans les RNN sont tanh ou ReLU introduisant ainsi la non-linéarité nécessaire au modèle [91].

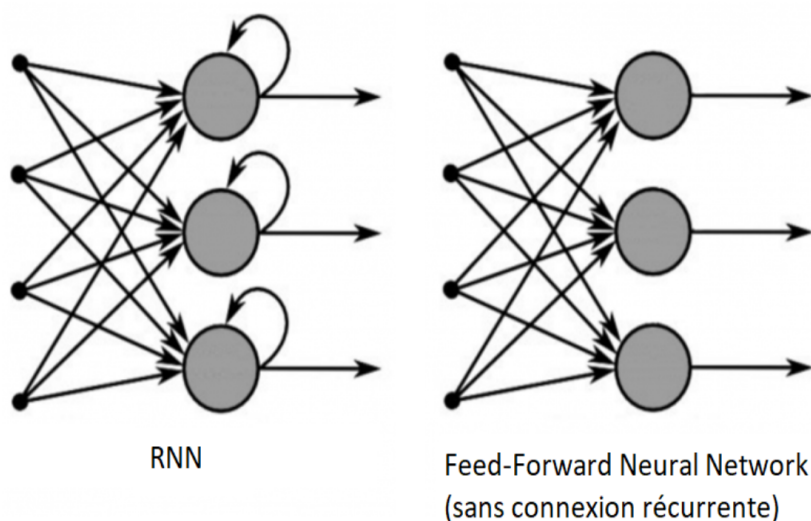


Figure 22 – Fonctionnement des RNN [92]

Il existe plusieurs variantes de RNN :

- les RNN simples qui reposent sur des connexions récurrentes basiques.
- les LSTM (Long Short-Term Memory) qui utilisent des cellules de mémoire pour pallier le problème de disparition du gradient.
- les GRU (Gated Recurrent Unit) qui proposent une structure plus simple tout en conservant une efficacité similaire à celle des LSTM [91][85].

7 Applications du Machine Learning et du Deep Learning en médecine

Au cours des dernières années, la multiplication des données médicales numériques a profondément transformé les pratiques en santé, favorisant l'intégration croissante de l'intelligence artificielle. Le Machine Learning, basé sur des données structurées, et Le Deep Learning, capable de traiter des données brutes complexes, sont désormais utilisés dans divers domaines médicaux pour améliorer le diagnostic, le traitement et la prévention.

7.0.1 Applications du Machine Learning en médecine

Diagnostic assisté du cancer : Les algorithmes de ML sont capables d'analyser des images médicales telles que les radiographies, IRM ou scanners, afin de détecter des anomalies suggérant la présence de lésions cancéreuses. Ces modèles permettent de repérer des motifs spécifiques liés aux tumeurs, souvent imperceptibles à l'œil humain [32][93].

Détection des maladies cardiovasculaires : En exploitant les données issues de l'imagerie cardiaque, ces méthodes permettent d'identifier précocement des signes de pathologies coronariennes ou de troubles du rythme, contribuant à une prise en charge rapide et ciblée [39].

Prédiction du diabète : Les modèles d'apprentissage supervisé peuvent analyser un ensemble de données cliniques (antécédents familiaux, résultats biologiques, habitudes de vie) afin d'évaluer le risque individuel de développer un diabète de type 2 [39].

Médecine personnalisée : Le Machine Learning permet également de proposer des traitements adaptés aux profils spécifiques des patients, en tenant compte de facteurs biologiques, cliniques et pharmacologiques, optimisant ainsi les résultats thérapeutiques [39].

7.0.2 Domaines d'application du Deep Learning

Analyse génomique : L'application du Deep Learning à la génomique permet d'identifier des mutations, des prédispositions génétiques ou des maladies héréditaires à partir de séquences complètes du génome. Des architectures comme les réseaux de neurones récurrents (RNN) ou les Transformers sont utilisées pour modéliser ces données séquentielles de manière efficace [94].

Détection automatisée de l'appendicite : Les réseaux de neurones convolutifs (CNN) sont utilisés pour analyser des images de tomodensitométrie abdominale, permettant d'identifier des signes caractéristiques de l'appendicite, tels que l'épaississement pariétal de l'appendice, la dilatation du diamètre ou la présence de liquide inflammatoire [95].

Assistance chirurgicale par IA : Les systèmes d'IA basés sur la vision par ordinateur assistent les chirurgiens en temps réel lors des interventions. Ces modèles sont capables de reconnaître et de délimiter les tissus inflammatoires ou pathologiques, améliorant la précision opératoire, notamment lors des appendicectomies [96].

Applications en ophtalmologie : Les CNN sont également largement utilisés dans le dépistage des maladies rétinienne à partir d'imagerie du fond d'œil. Ils permettent notamment de détecter avec fiabilité la rétinopathie diabétique et la dégénérescence maculaire liée à l'âge, deux pathologies majeures pouvant entraîner la cécité [97].

8 Travaux connexes : détection de l'appendicite

Le diagnostic de l'appendicite reste un défi clinique important, notamment pour distinguer les formes simples des formes compliquées. Ces dernières années, les méthodes d'intelligence artificielle ont suscité un intérêt croissant pour améliorer la précision du diagnostic à partir de données cliniques ou d'imagerie médicale. Plusieurs travaux récents ont exploré différentes approches du ML classique aux réseaux de neurones profonds afin d'automatiser et fiabiliser la détection de l'appendicite. Nous présentons ci-dessous une synthèse des contributions les plus marquantes :

Patia et al. (2023) ont développé des modèles de machine learning pour diagnostiquer et prédire la prise en charge de l'appendicite chez les enfants, à partir de données cliniques. Ils ont évalué 16 modèles (dont Random Forest, Gradient Boosting, SVM), et les meilleurs ont obtenu une précision de 91,3 % pour le diagnostic, 90,8 % pour la prédiction de la prise en charge (chirurgie vs non chirurgie), et 85,4 % pour la prédiction de la gravité. Les modèles ADABOOST, GB et RF se sont révélés les plus efficaces [107].

Pedro Roig Aparicio, al. (2021) ont proposé l'utilisation de modèles linéaires entiers superparsimonieux de risque (Risk SLIM) pour générer des scores de risque cliniquement interprétables afin de prédire le diagnostic, la prise en charge et les complications chez les enfants suspectés d'appendicite. Les résultats montrent que Risk SLIM identifie des variables médicalement significatives, avec des performances comparables, voire supérieures, à celles des forêts aléatoires, tout en offrant une meilleure lisibilité pour une intégration en pratique clinique. Ce travail démontre l'intérêt des modèles explicables pour renforcer l'aide à la décision médicale dans le cas de l'appendicite pédiatrique [108].

Marcinkevics et al. (2021) ont testé 13 modèles de machine learning sur des données cliniques pour diagnostiquer l'appendicite. Ils ont notamment utilisé un ensemble de données comprenant les variables cliniques classiques, et ont comparé les performances des modèles. Leurs résultats montrent que l'intelligence artificielle peut égaler ou surpasser les médecins dans certaines situations, en réduisant les erreurs de diagnostic [109].

Sotirios C Messinis et al. (2024) ont proposé une méthode d'apprentissage fédéré appelée DPS-GAT, qui combine des Graph Attention Networks (GATs) avec des mécanismes de confidentialité différentielle pour la sélection des clients et l'allocation des ressources dans des environnements médicaux décentralisés. Cette méthode vise à garantir la préservation de la vie privée tout en maintenant une performance stable du modèle dans des conditions de communication restreintes. DPS-GAT a été évalué sur trois ensembles de données, notamment le jeu de données pédiatrique de l'appendicite de Regensburg. Les résultats montrent que DPS-GAT atteint une précision moyenne supérieure à 0,7 pour

toutes les valeurs du budget de confidentialité différentiel (ϵ), avec une précision maximale d'environ 0,82 dans les scénarios d'hétérogénéité de confidentialité entre les clients [110].

Hariri et al. (2024) ont proposé un modèle appelé LesionScanNet, basé sur une architecture de deep learning personnalisée, pour la détection automatique de l'appendice dans les images CT. Leur approche a obtenu une précision de 99,25 %, un rappel de 99,52 % et un F1-score de 99,38 %, démontrant l'efficacité du modèle à identifier la présence de l'appendicite à partir d'images radiologiques [112].

Lee et al. (2023) ont proposé une technique innovante basée sur la superposition d'images en RGB pour classifier des images CT en normal, appendicite simple ou compliquée. Leurs modèles de deep learning ont atteint une précision de 94,3 %, avec une sensibilité de 91,6 % pour la détection de l'appendicite compliquée, soulignant l'intérêt d'approches de prétraitement avancées pour enrichir l'apprentissage [113].

Betül Tiryaki, al. (2024) ont utilisé l'architecture U-Net pour segmenter automatiquement l'appendice sur des images de tomodensitométrie abdominale. Leur modèle atteint un Dice Similarity Coefficient de 86,58 %, prouvant que le deep learning peut être utilisé pour identifier et délimiter avec précision l'appendice, étape cruciale dans le diagnostic automatisé [114].

Dandil et al. (2024) ont utilisé Mask R-CNN pour détecter l'appendice et prédire sa condition (normal ou inflammé). En intégrant plusieurs couches de traitement et une annotation médicale experte, leur modèle a atteint une précision de 99 %, confirmant l'intérêt des architectures de type "detection + segmentation" pour l'analyse des images médicales complexes [115].

Zhao et al. (2024) ont combiné les données radiomiques extraites des images CT avec des données cliniques pour différencier l'appendicite compliquée de la non compliquée. Leur modèle de fusion multimodale a obtenu une AUC de 0,799, contre 0,719 pour les radiologues, montrant que les modèles hybrides surpassent les approches humaines dans certains cas [116].

Liang et al. (2024) ont adopté une méthode similaire basée sur les caractéristiques radiomiques, en extrayant des paramètres d'intensité, de texture et de forme à partir des images CT. Leurs modèles ont permis une distinction efficace entre appendicite simple et compliquée, confirmant l'intérêt des descripteurs quantitatifs dans l'analyse de l'imagerie médicale [117].

9 Conclusion

Dans ce chapitre, nous avons exploré les fondements techniques et les mécanismes des méthodes d'apprentissage automatique et d'apprentissage profond, en mettant l'accent sur leur intégration croissante dans le domaine médical. L'intégration de ces approches dans le diagnostic médical constitue une avancée significative dans la prise en charge des pathologies complexes. Grâce à la capacité des algorithmes à traiter et à analyser de grands volumes de données hétérogènes (cliniques, biologiques, radiologiques) il devient possible d'atteindre une précision diagnostique parfois supérieure à celle obtenue par l'analyse humaine seule.

Les réseaux de neurones profonds, et notamment les CNN, se sont révélés particulièrement performants dans l'analyse d'images médicales, tandis que d'autres algorithmes, tels que les forêts aléatoires ou les modèles d'ensemble, ont montré leur efficacité sur des données

tabulaires. Toutefois, malgré ces avancées, des défis subsistent : qualité des données, risque de biais, et manque d'interprétabilité. Ces limites appellent à une collaboration étroite entre médecins, chercheurs et ingénieurs pour assurer une implémentation fiable, éthique et efficace de l'IA dans les pratiques cliniques.

Détection et Classification d'appendicite : Résultats et discussion

1 Contexte de l'étude

L'appendicite aiguë chez l'enfant représente l'une des urgences abdominales les plus fréquentes, mais son diagnostic reste complexe en raison de la variabilité des symptômes cliniques et de la similarité avec d'autres pathologies abdominales. Une prise en charge rapide et précise est pourtant essentielle pour éviter les complications, telles que la perforation ou l'infection généralisée. Dans ce contexte, l'intégration de l'intelligence artificielle, et plus particulièrement des méthodes d'apprentissage automatique (machine learning) et d'apprentissage profond (deep learning), ouvre de nouvelles perspectives pour améliorer l'aide au diagnostic.

Ce chapitre présente une approche expérimentale axée sur l'utilisation d'algorithmes de Machine Learning (ML) et de Deep Learning (DL) pour la détection et la classification de l'appendicite pédiatrique à partir de données cliniques, biologiques et échographiques. En s'appuyant sur la base de données «Regensburg Pediatric Appendicitis»[116], collectée auprès d'un hôpital spécialisé en Allemagne, nous avons évalué la performance de plusieurs modèles prédictifs appliqués à des tâches de classification binaire (présence ou absence d'appendicite) et multi classes (types de prise en charge). Ce chapitre détaille l'environnement de développement, les étapes d'implémentation, les résultats obtenus, et propose une discussion critique en les comparant aux travaux récents dans la littérature.

2 Environnement de développement

Afin de mettre en œuvre les modèles d'apprentissage automatique et d'apprentissage profond pour la détection de l'appendicite, nous avons choisi le langage de programmation Python comme environnement de développement principal. Ce langage, largement utilisé en science des données, nous a permis d'exploiter une variété d'outils et de bibliothèques performants qui ont grandement facilité l'implémentation, le traitement des données et l'entraînement des modèles.

Il comprend plusieurs bibliothèques standards, dont [98] :

- **TensorFlow** : une bibliothèque open-source, en 2015. Elle est conçue pour le calcul intensif et excelle dans la conception de modèles d'apprentissage automatique et d'apprentissage profond. TensorFlow s'appuie sur une structure reposant sur des graphes de flux de données, dans laquelle les opérations mathématiques sont symbolisées par des nœuds et les matrices de données sont connues sous le nom de tenseurs [99].
- **NumPy (Numerical Python)** : est une bibliothèque open-source offrant des structures de données robustes, notamment les ndarrays (tableaux multidimensionnels), ainsi que des fonctions mathématiques pour réaliser efficacement des opérations sur ces tableaux. Elle est principalement employée pour des calculs scientifiques, le traitement de données numériques et constitue la fondation de nombreuses autres bibliothèques scientifiques Python [100].
- **Pandas** : est une bibliothèque Python open-source destinée à la manipulation et analyse de données. Elle propose deux structures principales :
 - DataFrame : tableau de données à deux dimensions (lignes et colonnes, comme un tableau Excel).
 - Series : tableau unidimensionnel. Pandas facilite des opérations comme la lecture de fichiers (CSV, Excel), le nettoyage des données, la transformation, et l'analyse statistique [101].
- **Keras** : est une bibliothèque Python libre qui facilite la construction et l'entraînement de modèles d'apprentissage profond (deep learning) de façon aisée et rapide. Il s'agit d'une interface utilisateur avancée qui utilise des moteurs de calcul tels que TensorFlow, Theano ou CNTK pour réaliser les opérations intensives. Keras rend la construction de réseaux de neurones aisée grâce à une syntaxe intuitive, convenant aussi bien aux novices qu'aux chercheurs expérimentés [102].

Outils d'implémentations : Dans le cadre de ce projet, nous avons utilisé les outils suivants :

Google Colab : est une plateforme en ligne gratuite proposée par Google qui permet d'écrire, d'exécuter et de partager du code Python dans un environnement cloud. Il est principalement utilisé pour la machine Learning, l'apprentissage profond (deep learning), l'analyse de données et le prototypage rapide.

Colab est basé sur la technologie des Jupyter Notebooks, permettant de mélanger code, texte, images, formules mathématiques (LaTeX) et graphiques dans un seul document interactif [103].

Google Drive : est un service de stockage et de synchronisation de fichiers en ligne (cloud) développé par Google, lancé en 2012. Il permet aux utilisateurs de sauvegarder, partager et accéder à leurs fichiers (documents, images, vidéos, etc.) depuis n'importe quel appareil connecté à Internet.

Google Drive offre aussi une intégration native avec d'autres services Google comme Google Docs, Google Sheets, Google Slides, et Google Colab [104].

Kaggle : est une plateforme en ligne appartenant à Google qui offre des compétitions de data science, des datasets, ainsi que des notebooks interactifs pour pratiquer et développer des compétences en data science et machine learning. Fondée en 2010, Kaggle est aujourd'hui un des sites les plus populaires pour la compétition de modèles prédictifs et la collaboration entre data scientists [105].

3 Implémentation des méthodes

Cette étude a pour objectif d'analyser et de comparer les modèles les plus performants d'apprentissage automatique et d'apprentissage profond pour la détection de l'appendicite chez l'enfant. Dans ce cadre, nous menons une étude comparative entre des algorithmes classiques d'apprentissage automatique et des approches avancées d'apprentissage profond.

La figure suivante (Figure 40) illustre le processus de traitement mis en œuvre pour la classification de la base de données utilisée dans ce travail. Ce processus se décompose en deux étapes principales : le prétraitement des données et la classification.

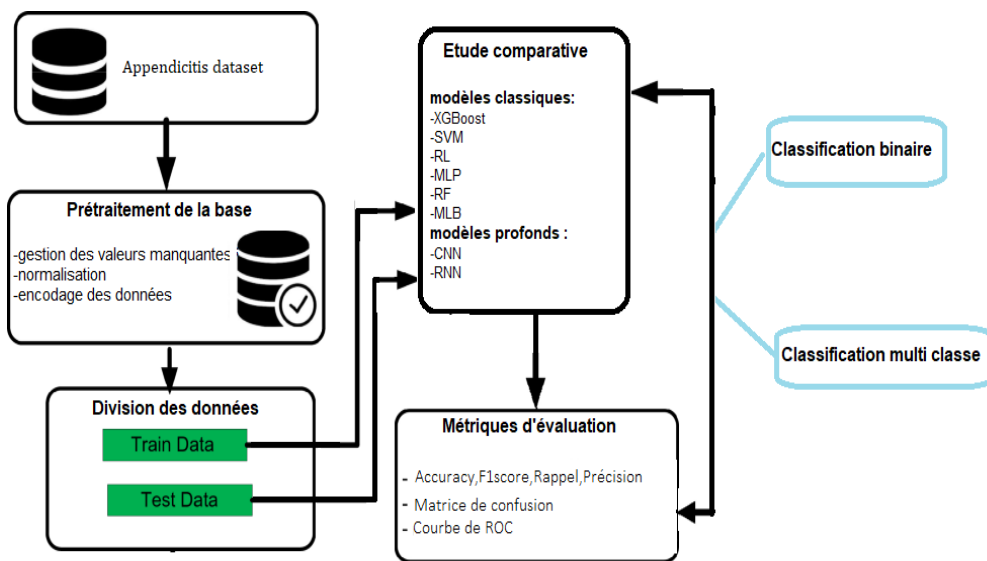


Figure 23 – Le processus de l'approche proposée

Dans un premier temps, les données brutes sont soumises à une phase de prétraitement. Cette étape vise à améliorer la qualité des données et à garantir leur compatibilité avec les algorithmes d'apprentissage. La seconde étape consiste en l'application d'algorithmes de classification supervisée, selon une approche binaire ou multiclasse. Deux familles de méthodes ont été explorées dans ce travail :

- Les techniques de Machine Learning (ML) telles que SVM, Random Forest ou XG-Boostes.
- Les méthodes de Deep Learning (DL), reposant sur des architectures de réseaux de neurones (CNN, RNN).

3.1 Description de la base de données

Dans le cadre de ce projet, nous avons utilisé un jeu de données médicales issu de la plateforme UCI [106], intitulé « Regensburg Pediatric Appendicitis ». Cette base de données est spécifiquement conçue pour la classification de l'appendicite pédiatrique. Elle provient d'une étude rétrospective menée entre 2016 et 2021 sur une cohorte de patients pédiatriques admis pour douleurs abdominales à l'hôpital pour enfants St. Hedwig, situé à Ratisbonne, en Allemagne. Le jeu de données regroupe des informations variées, notamment les résultats d'analyses biologiques, les constatations issues d'examens cliniques, des scores diagnostiques tels que les scores d'Alvarado et de l'appendicite pédiatrique, ainsi que des observations échographiques effectuées par des experts. Au total, la base comprend 782 échantillons, chacun décrit par 53 caractéristiques.

Ce jeu de données permet d'aborder la classification selon deux variables cibles distinctes : le diagnostic (**présence ou absence d'appendicite**), la prise en charge (**conservatrice /Primary surgical /Secondary surgical /Simultaneous appendectomy**).

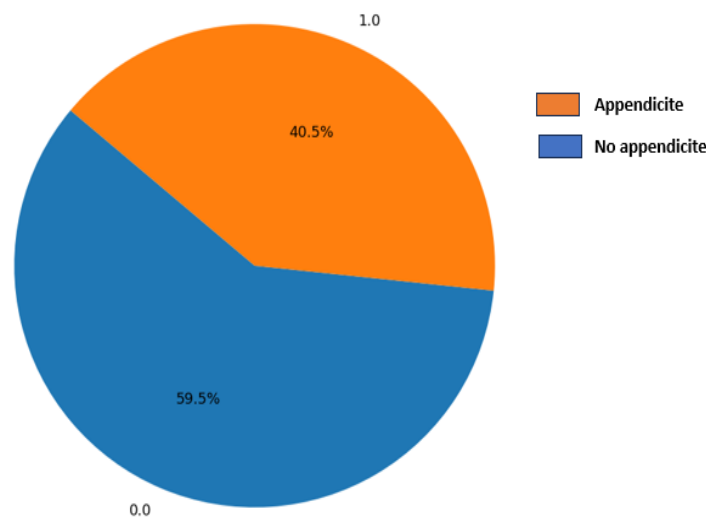


Figure 24 – Répartition des classes dans Diagnostic

3.2 Prétraitement de la base de données

Dans le cadre de cette étude, la phase de prétraitement des données a joué un rôle fondamental pour garantir l'efficacité des modèles de classification basés sur le Machine Learning (ML) et le Deep Learning (DL). Cette étape assure la transformation des données brutes en un format exploitable, en améliorant leur qualité, leur cohérence et leur pertinence, ce qui conditionne directement la performance des algorithmes d'apprentissage.

La base de données Regensburg, utilisée dans ce travail, est structurée sous forme tabulaire et a été importée depuis un fichier Excel. Elle contient des variables de types numérique et catégoriel. Comme le pipeline de classification mis en place requiert exclusivement des données numériques, un ensemble d'opérations de traitement a été appliqué. Ces opérations incluent la gestion des valeurs manquantes (en les remplaçant par la moyenne des colonnes concernées et par changement de type), et la transformation des variables catégorielles via des techniques d'encodage appropriées par exemple :

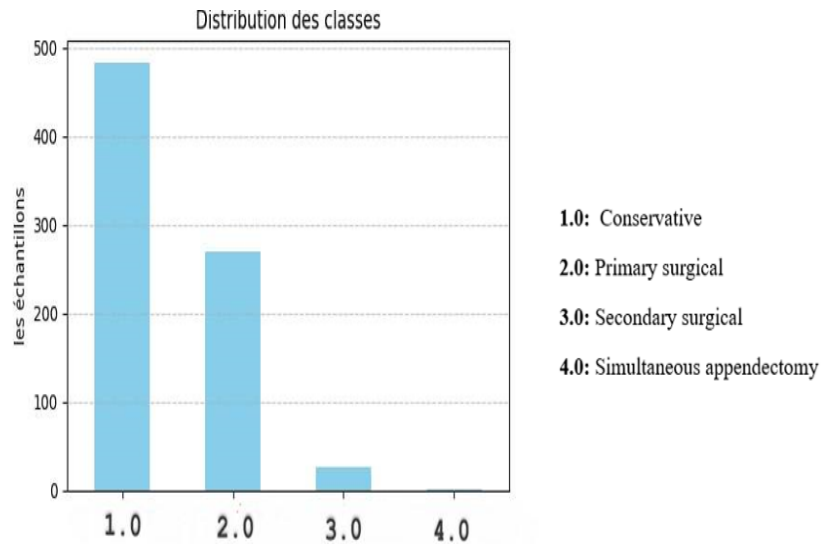


Figure 25 – Répartition des classes dans Management

Colonnes	Classes	Encodage
Diagnosis	Appendicitis	1
	No Appendicitis	0
Management	Conservative	1
	Primary surgical	2
	Secondary surgical	3
	Simultaneous appendectomy	4
Severity	Uncomplicated	0
	Complicated	1

Table 2 – Encodage des variables catégorielles

Une fois les différentes étapes de prétraitement effectuées, la base de données est rendue compatible avec les algorithmes d'apprentissage. Nous passons alors à la phase de classification, au cours de laquelle les données ainsi préparées sont utilisées pour entraîner les modèles d'apprentissage automatique et d'apprentissage profond.

3.3 Classification de l'appendicite : approches proposées

La phase de classification s'appuie sur les données numériques obtenues à l'issue du prétraitement, afin de permettre l'apprentissage d'un modèle capable de distinguer les cas d'appendicite des cas non pathologiques, ainsi que les classes de management

Algorithm 1

Étape 1 : Importation des bibliothèques

Importation des bibliothèques nécessaires à la manipulation des données, à la modélisation et à l'évaluation des performances

Étape 2 : Chargement des données

- Chargement d'un fichier CSV contenant des données médicales relatives à l'appendicite.
- Ce fichier comprend généralement des colonnes représentant des caractéristiques cliniques, ainsi qu'une colonne cible ('Diagnosis').

Étape 03 : Séparation des caractéristiques et de la cible

- X : ensemble des variables explicatives (toutes les colonnes sauf la cible).
- y : variable cible, c'est-à-dire ce que l'on souhaite prédire.

Étape 04 : Division du jeu de données

- 80 % des données sont utilisées pour l'entraînement ($X - train, y - train$).
- 20 % sont réservées aux tests ($X - test, y - test$).

Étape 05 : Initialisation et entraînement du modèle

- Création du modèle avec ses paramètres spécifiques (voir Tableau 3).
- Entraînement du modèle à partir des données d'apprentissage.

Étape 06 : Prédiction

- Prédiction de la variable cible à partir des données de test.

Étape 07 : Évaluation des performances

- *accuracy - score* : mesure la précision globale du modèle.
- *f1 - score* : équilibre entre la précision et le rappel.
- *recall - score* : capacité du modèle à détecter toutes les occurrences positives.
- *precision - score* : proportion de vraies positives parmi les prédictions positives.

Étape 08 : Affichage des résultats

- Affichage des différentes métriques d'évaluation pour analyser les performances du modèle.
-

3.3.1 Contribution 01 : Classification binaire

Dans cette section, nous exposons notre contribution à travers la conception et l'implémentation d'algorithmes de classification basés sur ML et DL, en mettant en avant les méthodes appliquées, les architectures développées dans le cadre d'un problème de classification binaire.

A/ Application des modèles classiques

Nous avons appliqué des modèles d'apprentissage automatique supervisé tels que Le SVM, XGBoost, La forêt aléatoire, le perceptron multicouche (MLP), la régression logistique et Le MLB sur la variable cible 'Diagnosis', traitée comme une classe binaire. Voici le tableau présentant chacun des modèles classiques utilisés, ainsi que les paramètres associés pour chaque modèle :

Modèle	Paramètres de création
Régression logistique	<code>random-state=0,solver='lbfgs',max-iter=5000, multi-class='ovr'</code>
XGBoost	<code>n_estimators=600,objective='binary:logistic', learning_rate=0.05, max_depth=5, min_child_weight=3</code>
SVM	<code>model = SVC()</code> (aucun paramètre précisé par défaut)
MLP (Perceptron multicouche)	<code>hidden_layer_sizes=(100), max_iter=1000, random_state=42</code>
Random Forest	<code>n_estimators=200,max_depth=3, random_state=0</code>
MultinomialNB	<code>MLB = MultinomialNB()</code> (aucun paramètre précisé par défaut)

Table 3 – Paramètres de création des modèles classiques

B/ Application de L'apprentissage en profondeur

Nous avons appliqué des modèles d'apprentissage profond, à savoir Le CNN et Le RNN, à la classification binaire basée sur la colonne 'Diagnosis'

Architecture du CNN Dans notre projet, nous avons implémenté un réseau de neurones convolutif (CNN) séquentiel pour effectuer une classification binaire à partir de données numériques normalisées. Le modèle a été construit avec l'API Sequential de Keras.

L'architecture comprend :

- Deux couches de convolution 2D suivies de MaxPooling2D pour extraire automatiquement les caractéristiques pertinentes des variables tabulaires,
- Une couche Flatten pour transformer les cartes de caractéristiques en vecteur,
- Une couche dense cachée avec Dropout pour prévenir le surapprentissage,
- Une couche de sortie avec activation sigmoïde, adaptée à la classification binaire.

Ce modèle CNN séquentiel a permis de modéliser efficacement les relations entre les variables d'entrée et de prédire la variable cible binaire avec précision.

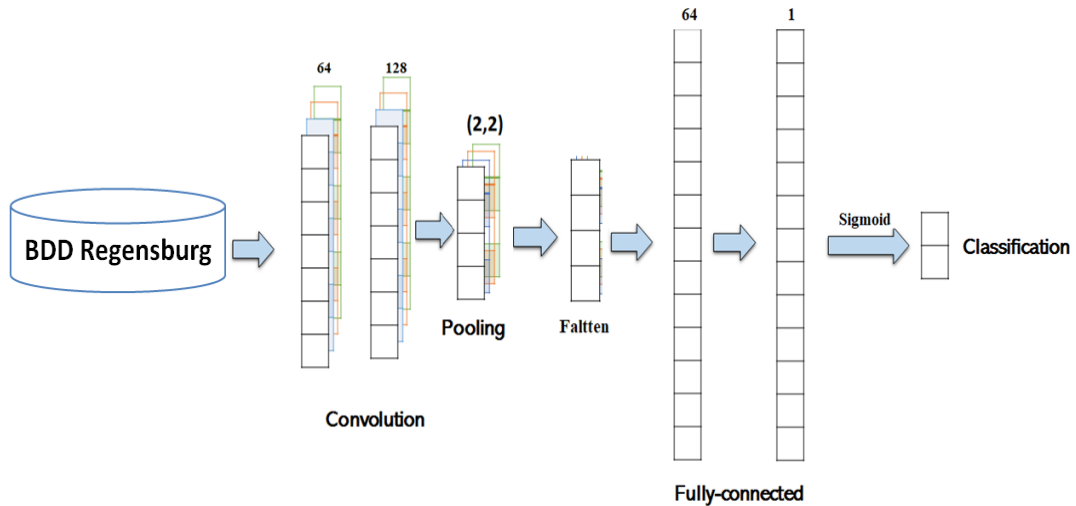


Figure 26 – Architecture du CNN pour la classification binaire

Architecture du RNN Dans notre travail, nous avons également exploré l'utilisation d'un réseau de neurones récurrents (RNN) séquentiel pour résoudre notre problème de classification binaire. Le modèle a été développé à l'aide de l'API Sequential de Keras, qui permet de construire des architectures couche par couche de manière simple et ordonnée. L'architecture comprend :

- Une couche SimpleRNN avec 50 unités récurrentes, conçue pour traiter des séquences de données, même si nos entrées ne sont pas temporelles à l'origine. Elle permet d'apprendre des relations entre les variables sur l'ensemble de la séquence,
- Une couche Dropout pour réduire le risque de surapprentissage en désactivant aléatoirement certaines connexions pendant l'entraînement,
- Une couche finale Dense avec une activation sigmoïde afin de produire une sortie binaire.

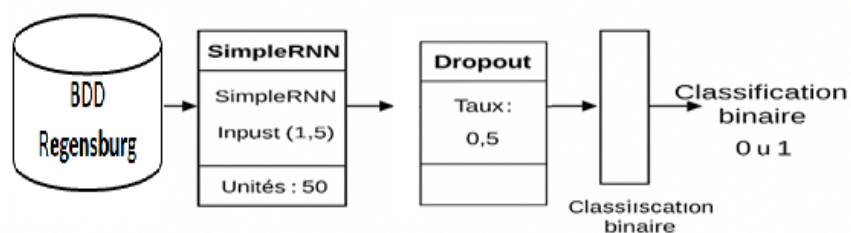


Figure 27 – Architecture du RNN pour la classification binaire

3.3.2 Contribution 02 : Classification multi classe

Cette section met en lumière notre contribution à travers le développement d'approches de classification multi classe, en combinant des techniques issues de ML et du DL. Nous y décrivons les stratégies adoptées, les architectures implémentées.

A/ Application des modèles classiques Nous avons mis en œuvre les mêmes modèles d'apprentissage supervisé utilisés dans la première contribution afin de prédire la

variable cible « Management », considérée comme une variable à classes multiples. Le tableau ci-dessous présente chacun des modèles classiques utilisés ainsi que les paramètres spécifiques appliqués à chaque modèle :

Modèle	Paramètres de création
Régression logistique	<code>random-state=0,solver='lbfgs',max-iter=5000, multi-class='ovr'</code>
XGBoost	<code>n_estimators=600,objective='multi:softmax' (pour multi classe), learning_rate=0.05, max_depth=5, min_child_weight=3,num_classe=4</code>
SVM	<code>model = SVC() (aucun paramètre précisé par défaut)</code>
MLP (Perceptron multi-couche)	<code>hidden_layer_sizes=(100), max_iter=1000, random_state=42</code>
Random Forest	<code>n_estimators=200,max_depth=3, random_state=0</code>
MultinomialNB	<code>MLB = MultinomialNB() (aucun paramètre précisé par défaut)</code>

Table 4 – Paramètres des modèles classiques pour la classification multi classe

B/ Application de L'apprentissage en profondeur

Nous avons utilisé les modèles d'apprentissage profond, notamment le CNN et le RNN, pour effectuer une classification multi classe sur la variable cible « Management ».

Architecture du CNN

Dans cette partie du travail, nous avons implémenté un réseau de neurones convolutif (CNN) pour effectuer une classification multiclasse. L'architecture et les paramètres utilisés sont similaires à ceux du modèle CNN précédemment appliqué à la classification binaire. La principale différence réside dans la fonction d'activation de la couche de sortie : nous avons utilisé la fonction softmax au lieu de la fonction sigmoïde, afin d'adapter le modèle au contexte multiclasse.

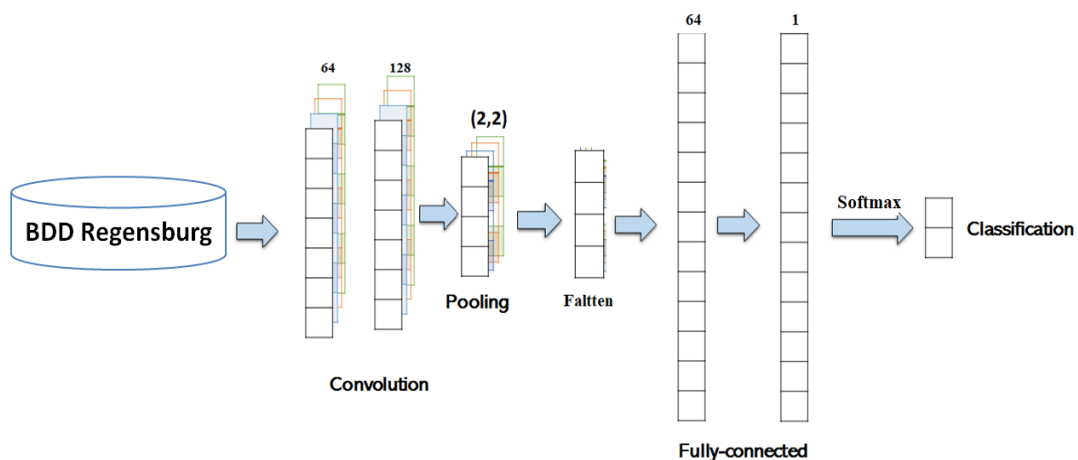


Figure 28 – Architecture du CNN pour la classification multi classe

Architecture du RNN

nous avons implémenté le modèle RNN séquentiel pour effectuer une classification multi classe (4 classes). Après avoir normalisé les données, nous avons remodelé les entrées pour qu'elles soient compatibles avec une architecture RNN.

Le modèle contient une couche SimpleRNN avec 50 neurones, suivie d'un Dropout de 0.5 pour limiter le sur apprentissage. La couche de sortie utilise une activation softmax adaptée à la classification multi classe. Le modèle est entraîné sur 100 époques.

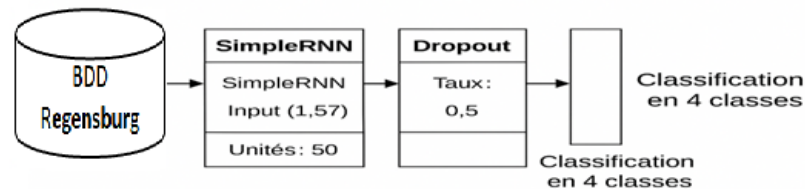


Figure 29 – Architecture du RNN pour la classification multi classe

4 Résultats et discussion

Dans cette section, nous présentons les résultats obtenus à l'issue de l'entraînement et de l'évaluation des différents modèles de classification, qu'ils soient issus de l'apprentissage automatique ou de l'apprentissage profond, dans le cadre des deux contributions. Nous avons comparé leurs performances à l'aide des métriques standards présentées dans la section suivante. L'objectif est d'identifier le ou les meilleurs modèles de classification pour la détection de l'appendicite.

Ces résultats sont ensuite analysés afin de mieux comprendre le comportement des modèles, d'identifier leurs points forts et leurs limites, et d'en tirer des conclusions pertinentes au regard de notre problématique de classification..

4.1 Critères d'évaluation

Voici les définitions des critères d'évaluation que nous avons utilisés dans notre projet pour la classification de l'appendicite :

1. **Taux de classification** : Le taux de classification est la proportion de prédictions correctes par rapport au nombre total de prédictions effectuées. Il est calculé comme suit :

$$\text{Taux de classification} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}} \quad (3.1)$$

Un taux de classification élevé indique que le modèle effectue beaucoup de bonnes prédictions, mais il peut être trompeur si les classes sont déséquilibrées.

2. **Précision** : La précision mesure la proportion de vraies positives parmi toutes les prédictions positives. Elle est calculée comme suit :

$$\text{Précision} = \frac{TP}{TP + FP} \quad (3.2)$$

Une précision élevée indique que peu de fausses alertes sont émises par le modèle. C'est particulièrement important dans les contextes où les faux positifs ont des conséquences significatives.

3. **Rappel** : Le rappel, également appelé sensibilité ou taux de vrais positifs, mesure la proportion de vraies positives parmi toutes les instances réellement positives. Il est défini comme :

$$\text{Rappel} = \frac{TP}{TP + FN} \quad (3.3)$$

Un rappel élevé indique que le modèle identifie correctement la plupart des cas positifs. Cela est crucial dans des contextes où il est essentiel de détecter autant de cas positifs que possible.

4. **F1 Score** : Le F1 score est la moyenne harmonique entre la précision et le rappel, fournissant un équilibre entre les deux. Il est particulièrement utile lorsque les classes sont déséquilibrées. Il se calcule ainsi :

$$F1 = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (3.4)$$

Un F1 score élevé indique que le modèle a à la fois une bonne précision et un bon rappel. C'est un bon indicateur de la performance globale, notamment lorsqu'il faut équilibrer faux positifs et faux négatifs.

5. **Matrice de confusion** : La matrice de confusion est un outil d'évaluation utilisé en apprentissage supervisé, notamment pour les problèmes de classification. Elle permet de visualiser les performances d'un modèle en comparant les prédictions aux valeurs réelles (ou vraies étiquettes) des données de test.
6. **Courbe ROC** : La courbe ROC est un outil d'évaluation des performances d'un modèle de classification binaire. Elle trace le taux de vrais positifs (TPR) en fonction du taux de faux positifs (TFP) pour différents seuils de classification. Pour chaque seuil de probabilité (entre 0 et 1), on calcule :

$$TPR = \text{Sensibilit} = \frac{TP}{TP + FN}, \quad TFP = \frac{FP}{FP + TN} \quad (3.5)$$

En faisant varier ce seuil, on obtient une courbe ROC. Plus la courbe est proche du coin supérieur gauche, meilleur est le modèle.

4.2 Contribution 01 : Classification binaire

Le tableau suivant présente les performances des différents modèles de machine ML et de DP appliqués à la classification binaire sur notre base de données.

Table 5 – Résultats expérimentaux des modèles de ML et de DL pour la classification binaire

Catégorie	Modèle	Accuracy	Précision	Rappel	F1-score
Machine Learning	SVM	0.90	0.90	0.90	0.90
	XGBoost	0.97	0.97	0.95	0.97
	Random Forest	0.92	0.92	0.92	0.92
	MLB	0.70	0.70	0.71	0.70
	Régression Logique	0.92	0.92	0.92	0.92
	MLP	0.89	0.89	0.89	0.89
Deep Learning	RNN	0.89	0.89	0.89	0.89
	CNN	0.96	0.96	0.96	0.96

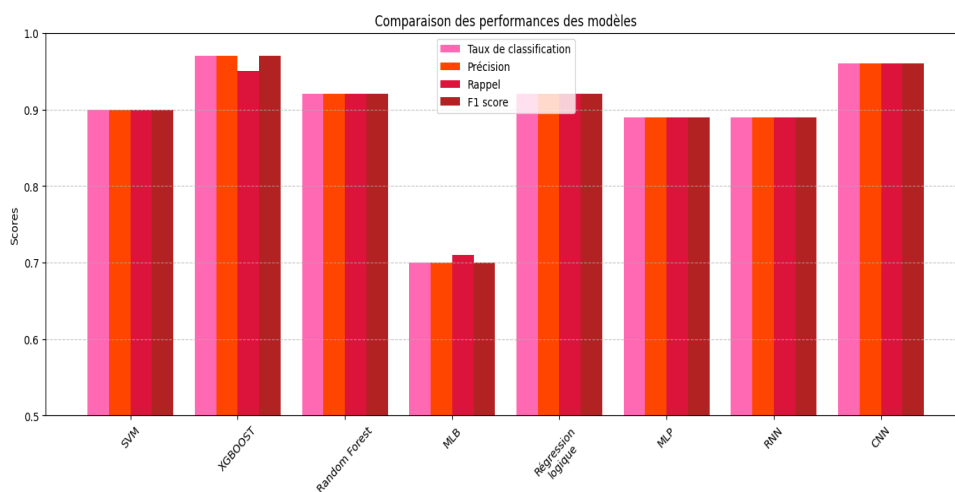


Figure 30 – Analyse comparative des performances des modèles ML et DL en classification binaire

L'analyse comparative des performances entre les modèles de ML et de DP révèle que le modèle XGBoost se distingue nettement avec les meilleures performances globales, atteignant un taux de classification, une précision et un F1-score de 0.97. Le modèle CNN arrive en seconde position avec un rappel remarquable de 0.96, soulignant sa capacité à identifier efficacement les cas positifs. Les modèles classiques tels que SVM, Random Forest et la Régression Logique présentent également des résultats solides (autour de 0.90 à 0.92) et peuvent être considérés comme des alternatives fiables. En revanche, le modèle MLB affiche des performances nettement inférieures (0.70), indiquant une inadéquation avec la tâche de classification étudiée. Enfin, bien que le modèle RNN issu du DP offre des résultats corrects (0.89), il reste moins performant que XGBoost et CNN dans ce contexte. Ainsi, XGBoost apparaît comme le modèle le plus adapté à cette problématique de classification.

La matrice de confusion ci-dessous illustre les performances du meilleur modèle XGBoost :

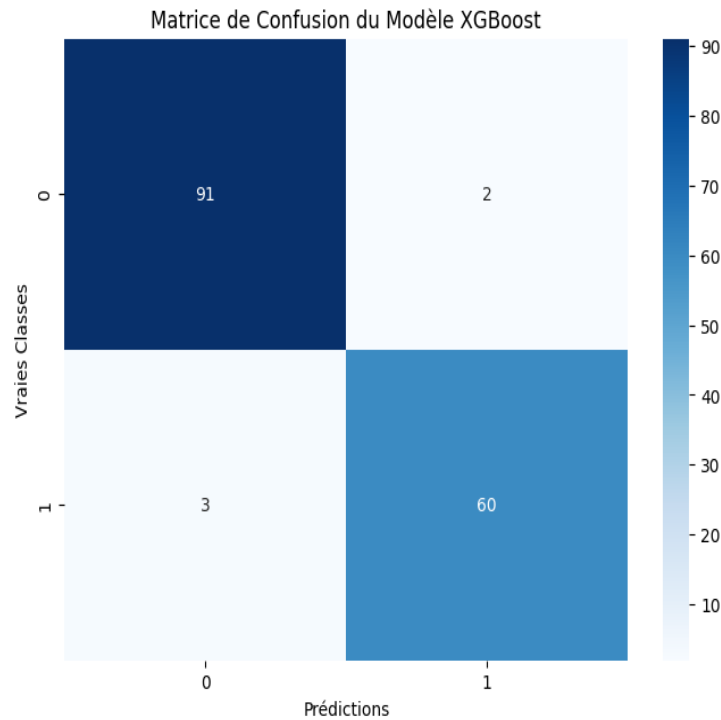


Figure 31 – Répartition de matrice de confusion du XGBoost en classe Binaire

La matrice de confusion du modèle XGBoost montre une bonne performance globale pour la tâche de classification binaire.

- **Pour la classe 0**, le modèle a correctement prédit **91 instances**, avec seulement 2 erreurs, ces dernières ayant été classées à tort comme appartenant à la classe 1.
- **Pour la classe 1**, **60 instances** ont été correctement identifiées, tandis que 3 ont été incorrectement classées comme étant de la classe 0.

Ces résultats indiquent que le modèle distingue globalement bien les deux classes, bien que quelques confusions subsistent, probablement liées à des cas limites. Le taux de faux positifs (2) et de faux négatifs (3) reste faible, ce qui traduit un **bon équilibre entre la précision et le rappel** pour les deux classes.

4.3 Contribution 02 : Classification multi classe

Le tableau suivant présente les performances des différents modèles de machine learning et de deep learning appliqués à la classification multiclasse sur notre base de données.

Table 6 – Résultats expérimentaux des modèles de ML et de DL pour la classification multi classe

Catégorie	Modèle	Accuracy	Précision	Rappel	F1-score
Machine Learning	SVM	0.90	0.90	0.90	0.90
	XGBoost	0.97	0.97	0.97	0.96
	Random Forest	0.95	0.92	0.95	0.93
	MLB	0.59	0.70	0.59	0.62
	Régression Logique	0.93	0.91	0.93	0.92
	MLP	0.91	0.90	0.91	0.91
Deep Learning	RNN	0.92	0.90	0.92	0.91
	CNN	0.93	0.91	0.93	0.92

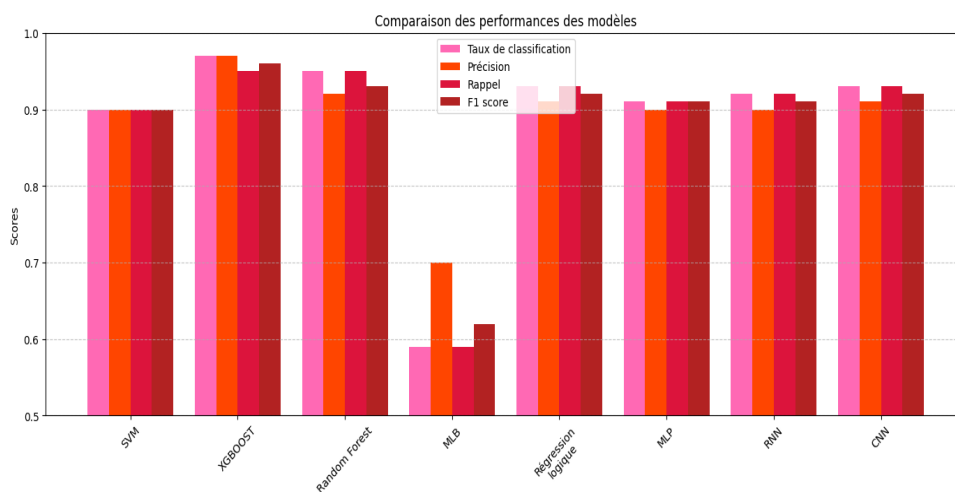


Figure 32 – Analyse comparative des performances des modèles ML et DL en classification multi classe

Le modèle XGBOOST se distingue nettement par ses performances élevées et équilibrées sur l'ensemble des critères : un taux de classification, une précision et un rappel de 0.97, et un F1-score de 0.96. Ces résultats font de XGBOOST le meilleur modèle de l'ensemble analysé, démontrant une capacité remarquable à classifier les données de manière fiable et cohérente.

Parmi les approches de DL, les modèles CNN et RNN montrent également de très bonnes performances avec des F1-scores de 0.92 et 0.91, respectivement.

Les modèles classiques comme Random Forest et Régression logistique montrent une efficacité correcte (F1-scores de 0.93 et 0.92 respectivement), tandis que SVM offre une performance stable mais légèrement en retrait (F1-score de 0.90). MLP, quant à lui, affiche des performances légèrement inférieures avec un taux de classification de 0.91, une précision de 0.90, et un F1-score de 0.91, mais reste néanmoins compétitif. En revanche, le modèle MLB présente des performances nettement inférieures (F1-score de 0.62), ce qui le rend inadapté à cette tâche dans les conditions testées.

La matrice de confusion ci-dessous illustre les performances du meilleur modèle XGBoost pour la classification multi classe.

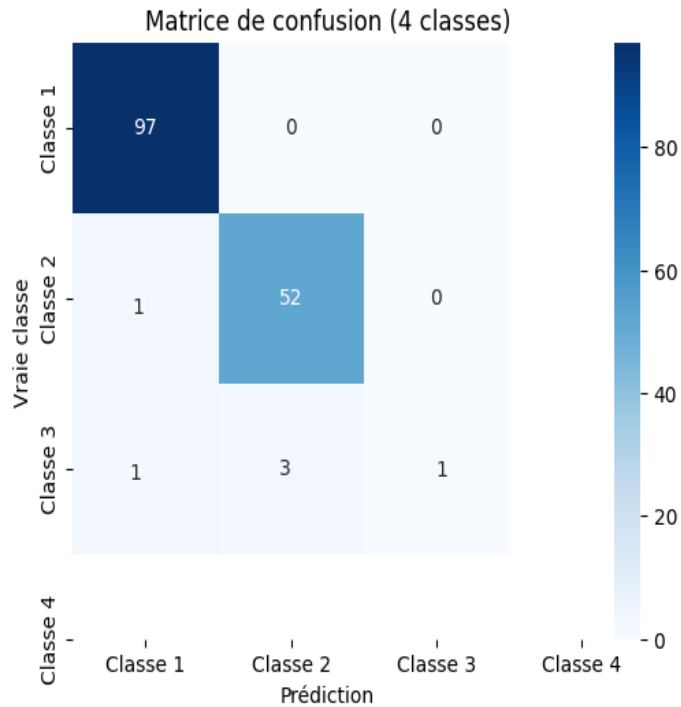


Figure 33 – Répartition de matrice de confusion du XGBoost en multi classe

La matrice de confusion du modèle XGBoost indique une excellente performance pour les classes 1 et 2, avec respectivement 97 et 52 prédictions correctes, et un taux d'erreur très faible. La classe 1 est parfaitement identifiée, sans aucune confusion. La classe 2 présente une légère confusion avec la classe 1, mais reste bien reconnue. En revanche, la classe 3 est mal classifiée : seule une instance sur plusieurs a été correctement identifiée, les autres étant confondues principalement avec les classes 1 et 2. La classe 4 ne figure pas dans la matrice, probablement en raison d'un nombre d'exemples insuffisant dans le jeu de test, ce qui empêche le modèle de l'apprendre efficacement. Cette absence reflète un déséquilibre important qui peut affecter la capacité du modèle à généraliser correctement.

Ci-dessous, la courbe De ROC du modèle XGBoost en classification binaire :

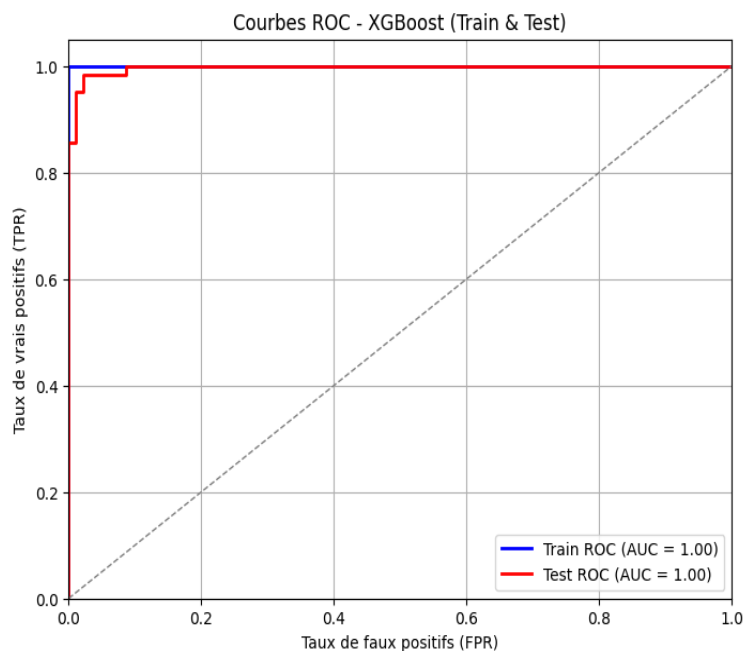


Figure 34 – La courbe de ROC du modèle XGBoost en classification binaire

La courbe ROC du modèle XGBoost montre une AUC de 1.00 sur les ensembles d'entraînement et de test, traduisant une séparation parfaite entre les deux classes. Bien que ces résultats indiquent une excellente performance, ils peuvent également suggérer un surapprentissage.

Ci-dessous, la courbe De ROC du modèle XGBoost en classification multi classe :

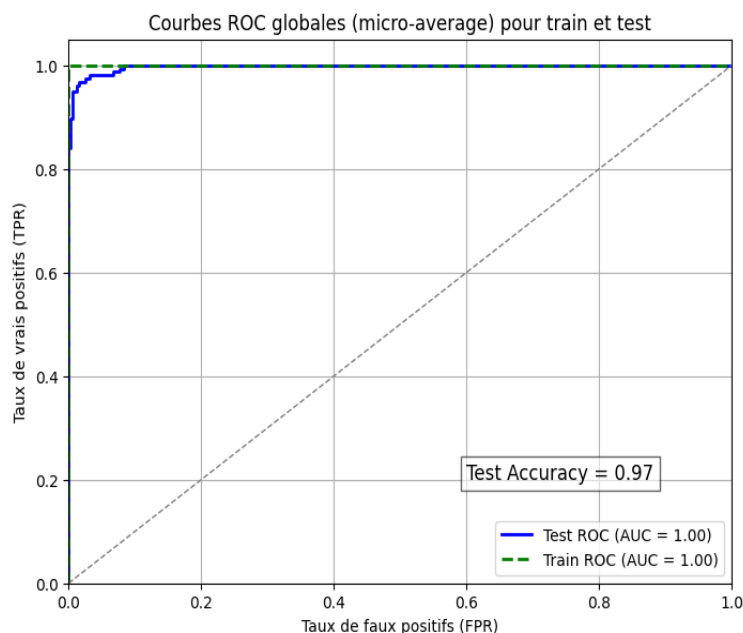


Figure 35 – La moyenne de courbe de ROC du modèle XGBoost en classification multi-classe

La courbe ROC micro-average du modèle XGBoost indique une excellente capacité

de discrimination avec une AUC de 1.00 sur les ensembles d'entraînement et de test, et une précision de 97% sur les données de test. Ces résultats suggèrent une bonne généralisation du modèle. Toutefois, une telle performance idéale nécessite une analyse approfondie.

4.4 Discussion des résultats

L'évaluation comparative des performances entre les approches de ML et de DL appliquées à la base de données Regensburg permet de dégager plusieurs constats significatifs. En premier lieu, les modèles de ML, notamment XGBoost, montrent une supériorité claire sur les modèles de DL dans ce contexte. En effet, XGBoost atteint des performances remarquables en classification binaire (accuracy de 97 %, F1-score de 96 %) et conserve cette excellence en classification multi classe (F1-score de 0.97), démontrant ainsi une grande capacité à modéliser efficacement les interactions complexes dans les données numériques. Cette performance s'explique par les mécanismes de boosting séquentiel propres à XGBoost, qui permettent d'optimiser la classification en corrigeant progressivement les erreurs des arbres précédents.

Les méthodes de DL, comme le CNN et le RNN, obtiennent des résultats compétitifs mais légèrement inférieurs, en particulier dans un contexte où les données ne présentent pas une structure spatiale ou temporelle exploitable. Le CNN, bien que performant (F1-score de 0.96 en multi classe), n'atteint pas le niveau de XGBoost, car il n'est pas spécifiquement optimisé pour les données tabulaires.

Les architectures telles que les réseaux de neurones sont puissantes, mais elles nécessitent généralement de grandes quantités de données et des représentations complexes (images, séquences), ce qui n'est pas dans notre cas ici. De plus, leur entraînement est plus coûteux en temps et en ressources.

Il est également important de noter que les algorithmes classiques comme le MLB montrent des limites évidentes face à des modèles plus sophistiqués, avec des performances nettement inférieures (F1-scores autour de 0.70 ou moins). Cela confirme que l'utilisation de modèles avancés est essentielle pour obtenir de bons résultats sur des problèmes de classification réels.

5 Analyse comparative avec les travaux existant

Avant d'évaluer la pertinence de notre méthode, il est crucial de la comparer aux travaux précédents dans ce domaine. Cette section propose une analyse comparative visant à souligner les points communs, les différences, ainsi que les possibles améliorations par rapport aux approches antérieures.

Plusieurs études ont utilisé la base de données Regensburg, réputée pour la qualité et la richesse de ses données cliniques structurées. Le tableau présenté met en évidence les résultats obtenus par ces recherches, les méthodes employées, et leur comparaison avec notre propre approche.

Les résultats présentés dans ce tableau montrent clairement que notre méthode basée sur XGBoost surpasse les autres approches en termes de performance, avec un taux de classification de 0,97, supérieur à celui des autres modèles évalués. Alors que des

Table 7 – comparaison des résultats avec les résultats des travaux utilisés la BDD Regensburg

Auteurs	Méthodologie	Performance (%)
Abhilash Pati et al. (2023)	ADABOOST	0.92
Pedro Roig Aparicio et al. (2021)	Risk SLIM (modèle linéaire à coefficients entiers), comparé à Random Forest.	Performances comparables aux modèles complexes, avec des scores simples adaptés à l'usage clinique.
Ricards Marcinkevics et al. (2021)	Random Forest	0.96 pour la classification binaire 0.94 pour la classification multi-classes
Sotirios C. Messinis et al. (2024)	Apprentissage fédéré avec confidentialité différentielle (DPS-GAT)	0,82 dans les meilleurs scénarios, et reste stable au-dessus de 0,7 quelle que soit la contrainte de confidentialité différentielle appliquée.
Nos contributions	XGBoost	0.97 pour la classification binaire 0.97 pour la classification multi classe

techniques robustes comme Random Forest et ADABOOST atteignent déjà des scores élevés (entre 0,92 et 0,96), notre méthode améliore encore cette précision, ce qui traduit une meilleure capacité à exploiter les données cliniques complexes de la base Regensburg. De plus, même si l'apprentissage fédéré avec confidentialité différentielle constitue une avancée importante pour la protection des données, sa performance reste inférieure à la nôtre (0,82 au meilleur). Ainsi, notre approche combine à la fois une grande précision et une adaptabilité pratique, faisant d'elle une solution prometteuse pour une meilleure détection et gestion de l'appendicite pédiatrique.

6 Interface interactive de classification

Dans le but de faciliter l'utilisation du modèle de classification XGBoost par un utilisateur non technique, une interface interactive a été développée à l'aide de la bibliothèque ipywidgets dans un environnement Google colab. Cette interface permet de charger un fichier de données, d'entraîner un modèle, puis de tester la classification sur un échantillon de l'ensemble de test.



Figure 36 – Interface interactive de classification avec XGBoost

6.1 Architecture de l'interface

L'implémentation de l'interface repose sur plusieurs modules Python :

- pandas pour la manipulation des données,
- scikit-learn pour le prétraitement (normalisation) et la séparation des jeux de données,
- xgboost pour l'entraînement du modèle,
- ipywidgets pour la création des éléments interactifs.

L'interface repose sur quatre composants principaux :

a) Zone de chargement des données

- Bouton dédié à l'import de fichiers CSV
- Aperçu automatique des premières lignes après chargement

b) Module de classification

- Lancement de l'entraînement du modèle XGBoost

c) Sélection des échantillons

- Liste déroulante dynamique des indices testés.
- Mise à jour automatique après classification
- Affichage du nombre d'échantillons disponibles

d) Zone de prédiction

- Affichage détaillé des caractéristiques de l'échantillon
- Prédiction binaire (0 ou 1) pour la variable (Diagnosis)
- Comparaison avec la valeur réelle extraite de la base de données

6.2 Fonctionnalités de l'interface

L'interface comprend les éléments suivants :

1. **Chargement de la base de données** : un bouton Charger la BDD permet à l'utilisateur de charger un fichier (.csv). Dès que le fichier est chargé, un aperçu des premières lignes est affiché pour vérifier son contenu.

Charger la BDD (1)

Classifier

Échantillon... Prédire l'échantillon...

Base de données chargée avec succès ! Aperçu :

Unnamed: 0	Age	BMI	Sex	Height	Weight	Length_of_Stay	Management	Severity	Diagosis_Presumptive	...	Appendicular_Abscess	Abscess_Location	Pathological
0	0	12.68	16.9	1.0	148.0	17.0	1.0	1.0	0.0	0.0	...	0.0	0.0
1	1	14.10	11.9	0.0	147.0	69.5	2.0	1.0	0.0	0.0	...	0.0	0.0
2	2	14.14	21.1	1.0	161.0	62.0	4.0	1.0	0.0	0.0	...	0.0	0.0
3	3	16.17	20.6	1.0	165.0	56.0	1.0	1.0	0.0	0.0	...	0.0	0.0
4	4	11.08	16.9	1.0	161.0	45.0	1.0	1.0	0.0	0.0	...	0.0	0.0

5 rows x 58 columns

Figure 37 – Affichage des premières lignes de La base de données.

- Classification des données :** après chargement, un bouton Classifier entraîne notre modèle XGBoost sur les données en prenant comme variable cible la colonne Diag-nosis.

Charger la BDD (1)

Classifier

Échantillon... 597 Prédire l'échantillon...

156 échantillons disponibles pour le test
 /usr/local/lib/python3.11/dist-packages/xgboost/core.py:158: UserWarning: [20:20:36] WARNING: /workspace/src/learner.cc:740: Parameters: { "use_label_encoder" } are not used.

warnings.warn(msg, UserWarning)

Classification terminée avec XGBoost :

	precision	recall	f1-score	support
0	0.97	0.98	0.97	91
1	0.97	0.95	0.96	65
accuracy			0.97	156
macro avg	0.97	0.97	0.97	156
weighted avg	0.97	0.97	0.97	156

Figure 38 – Rapport de classification de notre modèle XGBoost.

- Sélection d'un échantillon de test :** une liste déroulante est générée dynamiquement à partir des indices de l'ensemble de test. L'utilisateur peut sélectionner un échantillon pour le tester individuellement.

596

596

588

208

291

174

292

711

210

266

331

192

738

234

626

768

155

Échantillon... 711 Prédire l'échantillon...

Figure 39 – Sélection d'un échantillon de test.

- Prédiction sur un échantillon unique :** un bouton Prédire l'échantillon sélectionné permet d'effectuer une prédiction sur l'échantillon choisi et de comparer la sortie du modèle avec la valeur réelle de la classe cible.

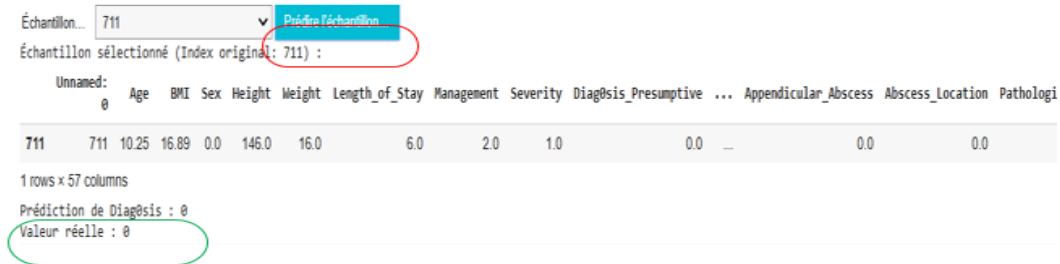


Figure 40 – Affichage détaillé des caractéristiques de L'échantillon, de La prédiction de L'échantillon et sa valeur réelle du diagnostic

6.3 Avantages pédagogiques et pratiques

Cette interface présente plusieurs avantages :

- Elle rend le processus d'apprentissage automatique plus accessible.
- Elle permet de tester rapidement l'efficacité du modèle sur des cas réels.
- Elle peut être étendue ou intégrée dans une application plus complète à visée médicale. Cette interface représente un pont entre l'analyse computationnelle et la pratique clinique, offrant un cadre rigoureux mais accessible pour l'évaluation des modèles prédictifs en médecine.

7 Conclusion

Les résultats obtenus à l'issue de l'évaluation de plusieurs modèles de classification, tant en Machine Learning qu'en Deep Learning, mettent en évidence la pertinence des approches automatisées pour le diagnostic de l'appendicite pédiatrique. En particulier, les algorithmes XGBoost, Random Forest et CNN ont affiché des performances remarquables sur la base de données Regensburg, avec des taux de classification et des scores F1 très élevés.

La comparaison avec les études récentes démontre que notre approche est hautement compétitive, voire supérieure dans tous les cas étudiés. Parmi tous les modèles testés, XGBoost s'est distingué comme le meilleur classifieur, offrant les résultats les plus précis, aussi bien pour la classification binaire que multiclasse. Son efficacité s'explique par sa capacité à gérer des données numériques structurées, ce qui correspond parfaitement à la nature des variables présentes dans notre base.

Le modèle CNN, bien qu'ayant fourni de bons résultats, s'est montré légèrement moins performant que XGBoost. Cela s'explique par le fait que les réseaux de neurones convolutifs sont généralement mieux adaptés aux données non structurées, telles que les images ou les textes, plutôt qu'aux données tabulaires numériques.

En définitive, cette étude souligne l'intérêt de XGBoost comme solution performante, stable et interprétable pour le diagnostic automatisé de l'appendicite chez l'enfant. Son intégration dans des outils cliniques pourrait contribuer à une prise de décision médicale plus rapide, plus fiable et fondée sur des données objectives.

Conclusion générale

Ce projet de fin d'études a démontré le potentiel significatif de l'Intelligence Artificielle (IA), et plus spécifiquement des algorithmes de Machine Learning (ML) et de Deep Learning (DL), pour améliorer la précision du diagnostic de l'appendicite aiguë pédiatrique. En abordant une problématique clinique majeure, ce travail a permis de concevoir et d'évaluer un système d'aide au diagnostic basé sur une base de données numérique, offrant des perspectives prometteuses pour la prise en charge des jeunes patients.

Les résultats expérimentaux ont clairement mis en évidence les performances supérieures de certains modèles. Le XGBoost et les Réseaux de Neurones Convolutionnels (CNN) se sont distingués comme les plus efficaces, tant pour la classification binaire (appendicite vs. non-appendicite) que multiclasse (stades de l'appendicite). Néanmoins, une observation clé a été que XGBoost a globalement surpassé le CNN dans ce contexte spécifique. Cette performance s'explique principalement par la nature de notre base de données, qui était composée de données numériques tabulaires. Les CNN excellent généralement dans l'analyse d'images et de textes, où ils peuvent tirer parti de structures spatiales ou séquentielles complexes. Pour des données structurées en tableau, les algorithmes de ML basés sur les arbres de décision, comme XGBoost, sont souvent plus adaptés et efficaces.

Ce projet ouvre la voie à plusieurs perspectives d'amélioration et de recherche futures. Il serait crucial de mettre en œuvre une validation croisée robuste pour optimiser la généralisabilité des modèles et pallier certaines limitations. En particulier, la classe 4 de notre jeu de données, ne contenant qu'un seul patient qui s'est retrouvé dans le jeu d'entraînement, n'a pas pu être représentée dans la matrice de confusion. Une validation croisée permettrait de s'assurer que chaque échantillon est utilisé à la fois pour l'entraînement et la validation, garantissant ainsi une évaluation plus complète et représentative. Par ailleurs, un équilibrage des données pour les classes sous-représentées, notamment les classes 3 et 4, par des techniques d'augmentation ou de suréchantillonnage, permettrait d'améliorer la capacité des modèles à identifier ces cas rares mais cliniquement importants. Ces améliorations contribueraient à renforcer davantage la fiabilité et la pertinence clinique de tels systèmes d'aide au diagnostic dans le futur.

Bibliographie

- [1] Elsan. (s.d.). *Appendicite : définition, causes et traitements*. <https://www.elsan.care/fr/pathologie-et-traitement/maladie-digestive/appendicite-definition-causes-traitement>. Consulté le 12/01/2025.
- [2] GBD 2021. *Trends and levels of the global, regional, and national burden of appendicitis between 1990 and 2021*. ScienceDirect.
- [3] Ran He, Jianxiong Lai, Ou Jiang, Jian Li. (2021). *The incidence and temporal trend of appendicitis in children : An analysis from the Global Burden of Disease Study 2021*.
- [4] Linjing Guan, Zhen Liu, Guangdong Pan, Bulin Zhang, Yongrong Wu, Tao Gan, Guoqing Ouyang. (2019). *The global, regional, and national burden of appendicitis in 204 countries and territories, 1990–2019 : a systematic analysis from the Global Burden of Disease Study 2019*. National Library of Medicine.
- [5] Appendicite. (s.d.). *Wikimédica*. <https://wikimedi.ca/wiki/Appendicite>. Consulté le 12 janvier 2025.
- [6] AP-HP. *Anatomie de l'appendice*. [En ligne] : <https://chirurgie-digestive-sat.aphp.fr/pathologies/appendicite/anatomie-de-lappendice/>. Consulté le 12 janvier 2025.
- [7] Belkind-Gerson, J. (2023, novembre). *Appendicite chez l'enfant*. MSD Manuals – Version grand public. <https://www.msmanuals.com/fr/accueil/probl%C3%A8mes-de-sant%C3%A9-des-enfants/troubles-gastro-intestinaux-chez-lenfant/appendicite-chez-lenfant>
- [8] Dictionnaire médical de l'Académie de Médecine – ancienne version 2020. Consulté le 14 avril 2025.
- [9] Andersson, R. E. (2004). *Meta-analysis of the clinical and laboratory diagnosis of appendicitis*. *British Journal of Surgery*, 91(1), 28–37.
- [10] Yu, C. W., Juan, L. I., Wu, M. H., Shen, C. J., Wu, J. Y., & Lee, C. C. (2013). *Systematic review and meta-analysis of the diagnostic accuracy of procalcitonin, C-reactive protein and white blood cell count for suspected acute appendicitis*. *Academic Emergency Medicine*, 20(6), 595–606.

- [11] Simon, L., Gauvin, F., Amre, D. K., Saint-Louis, P., & Lacroix, J. (2004). *Serum procalcitonin and C-reactive protein levels as markers of bacterial infection : A systematic review and meta-analysis*. *Clinical Infectious Diseases*, 39(2), 206–217.
- [12] Sand, M., Bechara, F. G., Holland-Letz, T., Sand, D., Mehnert, G., & Mann, B. (2009). *Diagnostic value of hyperbilirubinemia as a predictive factor for appendiceal perforation in acute appendicitis*. *The American Journal of Surgery*, 198(2), 193–198.
- [13] Duke, E., Kalb, B., Arif-Tiwari, H., Patel, B. K., & Martin, D. R. (2016). *A systematic review and meta-analysis of diagnostic performance of MRI for evaluation of acute appendicitis*. *AJR American Journal of Roentgenology*, 206(3), 508–517.
- [14] Alvarado, A. (1986). *A practical score for the early diagnosis of acute appendicitis*. *Annals of Emergency Medicine*, 15(5), 557–564.
- [15] Samuel, M. (2002). *Pediatric appendicitis score*. *Journal of Pediatric Surgery*, 37(6), 877–881.
- [16] Bhangu, A., Søreide, K., Di Saverio, S., Assarsson, J. H., & Drake, F. T. (2015). *Acute appendicitis : modern understanding of pathogenesis, diagnosis, and management*. *The Lancet*, 386(10000), 1278–1287.
- [17] Snyder, M. J., Guthrie, M., & Cagle, S. (2018). *Acute appendicitis : Efficient diagnosis and management*. *American Family Physician*, 98(1), 25–33.
- [18] MedHelpsis. (s.d.). *Appendicite gangréneuse : Symptômes, diagnostic et traitement*. Récupéré de MedHelpsis.
- [19] Apollo Hospitals. (2025, 18 février). *Quels sont les signes de l'appendicite ?* <https://www.apollohospitals.com/fr/health-library/what-are-the-signs-of-appendicitis>. Consulté le 14 avril 2025.
- [20] Boumas, N., Mindze, D., Nzue Nguema, H., Abegue, M., & Orema Ntyale, P. (2019). *Appendicites aiguës de l'enfant au CHU Mère Enfant Fondation Jeanne Ebori : Aspects épidémiologiques, diagnostiques et thérapeutiques*. *Health Sciences and Disease*. Récupéré de HSD.
- [21] Hadj Kaddour, A., & Derrad, M. (2018). *L'appendicite aiguë chez l'enfant de moins de 7 ans*. Université de Tlemcen.
- [22] Humes, D. J., & Simpson, J. (2006). *Acute appendicitis*. *BMJ*, 333(7567), 530–534.
- [23] Graffeo, C. S., & Counselman, F. L. (2016). *Appendicitis*. *Emergency Medicine Clinics of North America*, 34(2), 327–337.
- [24] Snyder, K. B., Hunter, C. J., & Buonpane, C. L. (2024). *Perforated appendicitis in children : Management, microbiology, and antibiotic stewardship*. *Seminars in Pediatric Surgery*, 30(3), 151051.
- [25] Almaramhy, H. H. (2017). *Acute appendicitis in young children less than 5 years : Review article*. *Italian Journal of Pediatrics*, 43, Article 15.
- [26] Tulin-Silver, S., Babb, J., Pinkney, L., Strubel, N., Lala, S., Milla, S. S., Tomita, S., & Fefferman, N. R. (2015). *[Titre manquant – compléter selon la source originale]*.
- [27] Swenson, D. W., Ayyala, R. S., Sams, C., & Lee, E. Y. (2018). *Practical imaging strategies for acute appendicitis in children*. *American Journal of Roentgenology*, 211(4), 901–909.
- [28] Radiological Society of North America. (s.d.). *Pediatric MRI for Appendicitis*. <https://www.radiologyinfo.org>. Consulté le 19 avril 2025.

- [29] Chen, F., Lu, H. L., & Simonetti, A. (2023). *Practical statistical considerations for the clinical validation of AI/ML-enabled medical diagnostic devices*.
- [30] Beaudin, M., Piché, N., et al. (2025). *Appendectomy versus antibiotics for acute uncomplicated appendicitis in children : an open-label, international, multicentre, randomised, non-inferiority trial*. *The Lancet*.
- [31] Maita, S., Andersson, B., Svensson, J. F., & Wester, T. (2020). *Nonoperative treatment for nonperforated appendicitis in children : a systematic review and meta-analysis*. *Pediatric Surgery International*, 36(3), 261–269.
- [32] M.S. in Artificial Intelligence Engineering - Biomedical Engineering (MSAIE-BME). <http://www.cmu.edu/bme/Academics/artificial-intelligence-engineering/index.html>.
- [33] IBM. (2024, 30 octobre). *Intelligence artificielle*. Cole Stryker, Eda Kavlakoglu. <https://www.ibm.com/fr-fr/topics/artificial-intelligence>.
- [34] RPUBS - SCI1035-CH1. (s.d.). <https://rpubs.com/alexvezeau/SCI1035-CH1>. Visitée en 20/04/2025.
- [35] Robert, J. (2020, 18 novembre). *Machine Learning : Définition, fonctionnement, utilisations*. DataScientest. <https://datascientest.com/machine-learning-tout-savoir>.
- [36] Sekkil, H. M., & Mebrouki, M. (2021). *Étude comparative entre les différentes architectures des réseaux de neurones convolutifs (CNNs) pour la détection de la fatigue du conducteur*. École supérieure en sciences appliquées, Tlemcen.
- [37] What is machine learning ? Understanding types & applications. (2025, 10 mars). Spiceworks Inc. <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/>. Visitée en 20/04/2025.
- [38] Machine learning dans l'épidémiologie (application aux maladies transmissibles et non transmissibles). (s.d.). https://adminbupfe.univ-saida.dz/opac_css/doc_num.php?explnum_id=43. Visitée en 21/04/2025.
- [39] L., B. (2019, 4 février). *Machine Learning et santé : top 5 des applications médicales*. LEBIGDATA.FR. <https://www.lebigdata.fr/machine-learning-top-medicales>. Visitée en 21/04/2025.
- [40] L'apprentissage non supervisé. (s.d.). MATLAB & Simulink. <https://fr.mathworks.com/discovery/unsupervised-learning.html>. Visitée en 07/03/2025.
- [41] Principaux algorithmes d'apprentissage non supervisé | Linedata. (s.d.). <https://fr.linedata.com/principaux-algorithmes-dapprentissage-non-supervise>. Consulté le 07/03/2025.
- [42] Murtagh, F., & Contreras, P. (2011). Algorithms for hierarchical clustering : an overview. *Wiley Interdisciplinary Reviews Data Mining And Knowledge Discovery*, 2(1), 86–97. <https://doi.org/10.1002/widm.53>. Consulté le 19/03/2025.
- [43] Comprendre l'apprentissage semi-supervisé. (2024, 18 octobre). IA School. <https://www.intelligence-artificielle-school.com/ecole/technologies/apprentissage-semi-supervise-definition/>. Consulté le 20/03/2025.

- [44] IONOS editorial team. (2025, 2 avril). What is semi-supervised learning? IONOS Digital Guide. <https://www.ionos.ca/digitalguide/online-marketing/search-engine-marketing/semi-supervised-learning/>. Consulté le 19/04/2025.
- [45] Qu'est-ce que l'apprentissage par transfert? - Explication de l'apprentissage par transfert dans le machine learning - AWS. (s.d.). Amazon Web Services, Inc. <https://aws.amazon.com/fr/what-is/transfer-learning/>. Consulté le 20/03/2025.
- [46] Illustration of reinforcement learning. (s.d.). https://www.researchgate.net/figure/Illustration-of-reinforcement-learning-based-on-4_fig2_357118740. Consulté le 20/03/2025.
- [47] Zhang, H. (2004). The Optimality of Naive Bayes. In *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, 562–567.
- [48] Rish, I. (2001). An Empirical Study of the Naive Bayes Classifier. IBM Research Report.
- [49] Tanguy, C. (2025, 10 avril). La régression linéaire en Machine Learning.
- [50] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley. <https://www.jedha.co/formation-ia/regression-lineaire-machine-learning>. Consulté le 16/03/2025.
- [51] Crochet-Damais, A. (2022, 13 juin). Régression linéaire : définition, fonctionnement et interprétation. Journal du Net. <https://www.journaldunet.fr/intelligence-artificielle/guide-de-l-intelligence-artificielle/1501891-regression-lineaire-definition-fonctionnement-et-interpretation/>. Consulté le 30/03/2025.
- [52] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*.
- [53] Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2e édition). O'Reilly.
- [54] t-Test, khi-deux, ANOVA, Régression, Corrélation... (s.d.-b). <https://datatab.fr/tutorial/logistic-regression>. Consulté le 15/03/2025.
- [55] Robert, J. (2023a, 9 novembre). La régression logistique, qu'est-ce que c'est? Formation Data Science | DataScientest.com. <https://datascientest.com/regression-logistique-quest-ce-que-cest>. Consulté le 30/03/2025.
- [56] Régression logistique. (s.d.). Google For Developers. <https://developers.google.com/machine-learning/crash-course/logistic-regression?hl=fr>. Consulté le 04/04/2025.
- [57] Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1(1), 81–106.
- [58] Robert, J. (2023b, 9 novembre). Qu'est-ce que l'algorithme KNN? DataScientest. <https://datascientest.com/knn>. Consulté le 10/04/2025.
- [59] IBM. (2025b, 30 janvier). Arbre de décisions. IBM. <https://www.ibm.com/fr-fr/think/topics/decision-trees>. Consulté le 04/04/2025.
- [60] Arbre de décision : exemple concret d'utilisation en entreprise (modèle inclus). (2024, 20 mars). Hubspot. <https://blog.hubspot.fr/marketing/arbre-de-decision>. Consulté le 04/04/2025.

- [61] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/bf00994018>.
- [62] Rédaction, L. (2022, 24 janvier). Machine à vecteurs de support (SVM) : définition et cas d'usage. *Journal du Net*. <https://www.journaldunet.fr/intelligence-artificielle/guide-de-l-intelligence-artificielle/1501879-machine-a-vecteurs-de-support-svm-definition-et-cas-d-usage/>.
- [63] Optimization of Machine Learning Algorithms Hyper-Parameters for Improving the Prediction of Patients Infected with COVID-19 - Scientific Figure on ResearchGate. Disponible sur : https://www.researchgate.net/figure/Diagram-representing-the-SVM-classification-of-binary-class-data-Positive-fig1_348484806. Consulté le 15/05/2025.
- [64] Kassel, R. (2023, 9 novembre). SVM : Définition, application et exemples - multi.datascientest.com. Formation Data Science | DataScientest.com. <https://datascientest.com/svm>. Consulté le 10/04/2025.
- [65] Zhou, Z.-H. (2012). Ensemble Methods : Foundations and Algorithms.
- [66] Breiman, L. (2001). Random Forest. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>.
- [67] Machine Learning : Concepts clés, fonctionnement applications pratiques. (2025, 6 janvier). CallMeNewton. <https://www.callmenewton.fr/guide-ia/machine-learning/>. Consulté le 16/04/2025.
- [68] Robert, J. (2023, 9 novembre). Random Forest : Forêt d'arbre de décision - Définition et fonctionnement. Formation Data Science | DataScientest.com. <https://datascientest.com/random-forest-definition>. Consulté le 16/04/2025.
- [69] Belaidi, N. (s.d.-b). Random Forest : comment ça fonctionne? Formation Tech et Data en ligne | Blent.ai. <https://blent.ai/blog/a/random-forest-comment-ca-marche>. Consulté le 17/04/2025.
- [70] Random Forest en Machine Learning : comment ça fonctionne? (2024, 5 février). IA School. <https://www.intelligence-artificielle-school.com/ecole/technologies/random-forest-en-machine-learning/>. Consulté le 20/04/2025.
- [71] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937–1967. <https://doi.org/10.1007/s10462-020-09896-5>. Visitée en 18/04/2025.
- [72] He, Z., Lin, D., Lau, T., & Wu, M. (2019, août 19). Gradient Boosting Machine : a survey. arXiv.org. <https://arxiv.org/abs/1908.06951>.
- [73] Chen, T., & Guestrin, C. (2016). XGBoost : A Scalable Tree Boosting System. arXiv preprint arXiv :1603.02754.
- [74] Henri. (s. d.). Les réseaux neuronaux et leur rôle dans les machines Learning. Phi0. <https://phi0.org/reseaux-neuronaux-role-machine-learning/>.
- [75] Florent, M., & Florent, M. (2024, avril 25). Neurones artificiels et neurones biologiques. Tree Learning. <https://www.tree-learning.fr/plateforme-lms-elearning/neurones-artificiels/>.
- [76] Réseaux de neurones. (s. d.). Google For Developers. <https://developers.google.com/machine-learning/crash-course/neural-networks?hl=fr>. Visitée en 17/04/2025.

- [77] Rokhaya Yade. (2022). *Classification bayésienne, apprentissage et réseaux de neurones : application en science des données*. Mémoire, Université du Québec à Trois-Rivières.
- [78] Introduction à L'apprentissage profond (deep learning) de L'intelligence artificielle — Culture Sciences Physique. (s. d.). <https://culturesciencesphysique.ens-lyon.fr/ressource/IA-apprentissage-Rousseau.xml>. Visitée en 16/04/2025.
- [79] Florent, M., & Florent, M. (2024, 25 avril). Neurones artificiels et neurones biologiques. Tree Learning. <https://www.tree-learning.fr/plateforme-lms-elearning/neurones-artificiels/>.
- [80] Kassel, R. (2023, 9 novembre). Deep Neural Network : qu'est-ce qu'un réseau de neurones profond? DataScientest. <https://datascientest.com/deep-neural-network>. Visitée en 22/04/2025.
- [81] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC Press.
- [82] Robert, J. (2024, 31 janvier). Convolutional Neural Network : Tout ce qu'il y a à savoir. DataScientest. <https://datascientest.com/convolutional-neural-network>.
- [83] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based Learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- [84] OpenClassrooms. (s. d.). OpenClassrooms. <https://openclassrooms.com/fr/courses/5801891-initiez-vous-au-deep-learning/5814626-construisez-des-reseaux-profonds-grace-aux-couches-convolutionnelles>. Visitée en 22/04/2025.
- [85] CS 230 - Pense-bête de réseaux de neurones convolutionnels. (s. d.). <https://stanford.edu/~shervine/l/fr/teaching/cs-230/pense-bete-reseaux-neurones-convolutionnels>. Visitée en 20/04/2025.
- [86] Habib, B. L. E. (s. d.). Les réseaux de neurones convolutifs. <https://datasciencetoday.net/index.php/en-us/deep-learning/173-les-reseaux-de-neurones-convolutifs>. Visitée en 20/04/2025.
- [87] Belaidi, N. (s. d.). Réseaux convolutifs (CNN) : comment ça marche? Formation Tech et Data en ligne | Blent.ai. <https://blent.ai/blog/a/cnn-comment-ca-marche>. Visitée en 04/05/2025.
- [88] Chao, W. (2024, 29 juillet). Research on Features Extraction and Classification for Images based on Transformer Learning. PMLR. <https://proceedings.mlr.press/v245/chao24a.html>.
- [89] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- [90] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).
- [91] ELAOUABER Zineb Aziza. (2023). *Système d'aide au diagnostic pour la détection automatique du Glaucome et de la Rétinopathie diabétique*. Thèse de doctorat, Université de Tlemcen.

- [92] Devoteam. (2025, 4 février). Aller plus loin en deep learning avec les réseaux de neurones récurrents (RNNs) | Devoteam. <https://www.devoteam.com/fr/expert-view/aller-plus-loin-en-deep-learning-avec-les-reseaux-de-neurones-recurrents-1> Visitée en 25/04/2025.
- [93] Vue d'ensemble des applications cliniques de l'intelligence artificielle | CDA-AMC. (2018, 27 septembre). <https://www.cda-amc.ca/fr/vue-densemble-des-applications-cliniques-de-lintelligence-artificielle>. Visitée en 04/05/2025.
- [94] Zou, J., et al. (2019). A primer on deep learning in genomics. *Nature Genetics*.
- [95] Dogan, K., & Selcuk, T. (2024). A Novel Deep Learning Approach for the Automatic Diagnosis of Acute Appendicitis. *Journal Of Clinical Medicine*, 13(16), 4949. <https://doi.org/10.3390/jcm13164949>.
- [96] Topol, E. (2019). *Deep Medicine : How Artificial Intelligence Can Make Healthcare Human Again*.
- [97] Gulshan, V., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*.
- [98] What is Python? Executive Summary. (s. d.). <https://www.python.org/doc/essays/blurb/>.
- [99] Abadi, M., et al. (2016). TensorFlow : A system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16).
- [100] Harris, C. R., et al. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362.
- [101] McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.
- [102] Chollet, F. (2015). Keras : Deep Learning library for Python. GitHub repository.
- [103] Bisong, E. (2019). Google Colaboratory. In Building Machine Learning and Deep Learning Models on Google Cloud Platform (pp. 59–64). Apress, Berkeley, CA.
- [104] Grannell, C. (2014). Using Google Drive and Other Cloud Storage Services. Macworld.
- [105] Kaggle (2021). Kaggle : Your Home for Data Science. Kaggle, Inc.
- [106] Marcinkevičs, R., Reis Wolfertstetter, P., Klimiene, U., Chin-Cheong, K., Paschke, A., Zeres, J., Denzinger, M., Niederberger, D., Wellmann, S., Ozkan, E., Knorr, C., & Vogt, J. E. (2023). *Regensburg Pediatric Appendicitis Dataset (1.03)* [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7711412>.
- [107] Pati, A., Panigrahi, A., Nayak, D. S. K., Sahoo, G., & Singh, D. (2023). Predicting pediatric appendicitis using ensemble learning techniques. *Procedia Computer Science*, 218, 1166–1175.
- [108] Roig Aparicio, P., Marcinkevičs, R., Reis Wolfertstetter, P., Wellmann, S., Knorr, C., & Vogt, J. (2021). Learning Medical Risk Scores for Pediatric Appendicitis. In *Proc. IEEE ICMLA 2021*, pp. 1507–1512. <https://doi.org/10.1109/ICMLA52953.2021.00243>

- [109] Marcinkevics, R., Reis Wolfertstetter, P., Wellmann, S., Knorr, C., & Vogt, J. E. (2021). Using machine learning to predict the diagnosis, management and severity of pediatric appendicitis. *Frontiers in Pediatrics*, 9, 662183.
- [110] Messinis, S. C., Protonotarios, N. E., & Doulamis, N. (2024). Differentially Private Client Selection and Resource Allocation in Federated Learning for Medical Applications Using Graph Neural Networks. *Sensors*, 24(16), 5142. <https://doi.org/10.3390/s24165142>
- [111] Akmese, O. F., Dogan, G., Kor, H., Erbay, H., & Demir, E. (2020). The use of machine learning approaches for the diagnosis of acute appendicitis. *Emergency Medicine International*, 2020, 7306435.
- [112] Hariri, M., Aydın, A., Sıbiç, O., Somuncu, E., Yılmaz, S., Sönmez, S., & Avşar, E. (2024). LesionScanNet : Dual-Path convolutional neural network for acute appendicitis diagnosis. *Health Inf Sci Syst*, 13(1), 3. <https://doi.org/10.1007/s13755-024-00321-7>
- [113] Lee, G. P., et al. (2023). Enhancing Disease Classification in Abdominal CT Scans through RGB Superposition Methods and 2D Convolutional Neural Networks : A Study of Appendicitis and Diverticulitis. *Comput Math Methods Med*, 2023, 7714483. <https://doi.org/10.1155/2023/7714483>
- [114] Baştuğ, B. T., Güneri, G., Yıldırım, M. S., Çorbacı, K., & Dandıl, E. (2024). Fully Automated Detection of the Appendix Using U-Net Deep Learning Architecture in CT Scans. *Journal of Clinical Medicine*, 13(19), 5893.
- [115] Dandıl, E., Baştuğ, B. T., Yıldırım, M. S., Çorbacı, K., & Güneri, G. (2024). MaskAppendix : Backbone-Enriched Mask R-CNN Based on Grad-CAM for Automatic Appendix Segmentation. *Diagnostics*, 14(21), 2346. <https://doi.org/10.3390/diagnostics14212346>
- [116] Zhao, Y., Wang, X., Zhang, Y., et al. (2024). Combination of Clinical Information and Radiomics Models for the Differentiation of Acute Simple Appendicitis and Non-Simple Appendicitis on CT Images. *Scientific Reports*, 14(1), 1854. <https://doi.org/10.1038/s41598-024-52390-z>
- [117] Liang, D., Fan, Y., Zeng, Y., et al. (2024). Development and Validation of a Deep Learning and Radiomics Combined Model for Differentiating Complicated From Uncomplicated Acute Appendicitis. *Academic Radiology*, 31(4), 1344–1354. <https://doi.org/10.1016/j.acra.2023.08.018>