REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE YAHIA FARES DE MEDEA

Faculté de **T**echnologie

Département du **G**énie **É**lectrique

Mémoire de fin d'études de Master

Filière : Automatique

Spécialité : Automatique et système

Étude d'un système intelligent pour la détection des attaques (intrusion) dans les bâtiments intelligent

Présentés par :

Berkani Mohamed Akram

Korteby Si Mohamed

Proposés et Dirigés par :

Mr. Rebiai Mohamed

Mr. Berkani Mohamed Rafik Aymene

Promotion 2025

Remerciement

Avant de commencer, louange et gratitude à Allah le Tout-Puissant, de nous avoir accordé la volonté, la santé, la patience et le courage nécessaires pour mener à bien ce travail.

Il nous est particulièrement agréable d'adresser nos plus vifs et sincères remerciements à **Mr BERKANI Mohamed Rafik Aymene** et **Mr REBIAI Mohamed**, nos encadreurs, pour leur encadrement de qualité, leur disponibilité constante et leurs précieux conseils tout au long de ce travail. Leur rigueur scientifique, leur sens de l'écoute, ainsi que leur soutien technique et moral, ont largement contribué à la réussite de ce mémoire. Ils trouvent ici l'expression de notre profonde reconnaissance et de notre respect.

Nos sincères remerciements s'adressent également à :

- **Mr Lachenani Sidahmed**, pour avoir bien voulu honorer de sa présence le jury en tant que président.
- **Mr Essaid Billel**, pour avoir accepté d'examiner ce mémoire et de participer activement aux travaux du jury.

Nous tenons également à remercier toutes les personnes, de près ou de loin, qui ont contribué à la réalisation de ce travail par leurs encouragements, leurs remarques constructives ou leur aide.

Nos pensées les plus reconnaissantes vont enfin à nos familles respectives, en particulier à nos chers parents, à nos frères, sœurs et ami(e)s, pour leur soutien moral constant, leur bienveillance, et l'atmosphère chaleureuse qu'ils ont su créer autour de nous durant cette période exigeante.

Dédicace

À mes parents, Je vous dédie ce mémoire avec une profonde reconnaissance et un amour infini. Votre soutien inébranlable, vos sacrifices silencieux, votre foi constante en mes capacités et vos encouragements indéfectibles m'ont accompagné à chaque étape de ce parcours. Vous m'avez transmis les valeurs de persévérance, d'humilité et de rigueur. Sans vous, je ne serais pas arrivé(e) là où je suis aujourd'hui.

À ma mère, douce lumière de ma vie, pour ton amour inconditionnel, tes prières silencieuses et ton courage exemplaire. Tu m'as appris à croire en mes rêves, même dans les moments les plus incertains.

À mon père, pour ta patience, ton soutien discret mais infaillible, et les valeurs de travail, d'honneur et d'intégrité que tu incarnes chaque jour et qui me servent de guide.

À toute ma famille, Merci d'avoir été présents, chacun à votre manière, pour vos paroles réconfortantes, vos gestes affectueux et votre confiance indéfectible. Votre amour m'a porté dans les moments de doute, de stress et de solitude.

Une dédicace toute particulière à mon frère, ami et encadreur, **Mr BERKANI Mohamed Rafik Aymene**, pour son encadrement bienveillant, son appui constant et sa précieuse présence tout au long de cette aventure.

À mes frères et sœurs : Bouchra, Oussama, Islam et Malak, Vous êtes mes piliers, ma source de joie et de stabilité. Merci pour votre patience face à mes absences, vos attentions, vos sourires et votre présence constante, même silencieuse.

À mon binôme, Mohamed Korteby, pour ton esprit de collaboration, et ton soutien précieux.

À ma future épouse (S), pour ta patience, ta tendresse et ta foi en moi. Ton amour est une source de motivation inestimable.

À mes amis les plus proches : Rafik, Riyad, Wail, Saber, Yacin, Fethi, Bilal, Mohamed, Adoula, Fiko, Aymen, Faycel, Abdelrahman, Moutie, Islam, Arslan, Hani, Ayoub, Oussama, Merci pour votre écoute bienveillante, vos conseils sincères, votre bonne humeur et votre amitié authentique. Vous avez su apporter un équilibre salutaire entre sérieux et légèreté tout au long de ce parcours.

À toutes celles et ceux qui, de près ou de loin, ont contribué à la réalisation de ce mémoire, Vos encouragements, votre confiance et vos conseils m'ont été infiniment précieux. Vous vous reconnaîtrez, et je vous remercie du fond du cœur.

BERKANI Mohamed Akram

Dédicace

A ma chère mère,

A mon cher père,

Qui n'ont jamais cessé, de formuler des prières à mon égara, de me soutenir et de m'épauler pour que je puisse atteindre mes objectif.

A mes frères, Mustapha, Riyad, Yassine, Billel, Rimi.

A mes chères sœur, Lamis, Serine, Sihem, Lydia, Sabrine.

Pour ses soutiens moraux et leurs conseils précieux tout au long de mes études.

A mes chères tantes, Hanane, Nadira, Wahiba.

Qui m'aidé et supporté dans les moments difficiles.

A mes neveux, Djawad, Ismail, Djazil, Samy, Yanis

A mon cher oncle.

A ma chère grand-mère,

Qui je souhaite une bonne santé.

A mon chère binôme, Mohamed Akram,

Pour son entente et sa sympathie.

Pour leurs indéfectibles soutiens et leurs patiences infinies.

A mes chères ami(e)s, Luffy, Pitri, Fouad, Aymen, Kassimo, Amine, Walid, Amani, Wiaam

Pour leurs aides et supports dans les moments difficiles.

A toute ma famille,

A tous mes autres

Si Mohamed Korteby

ملخص

هذا العمل المتواضع، يقدم المفاهيم الأساسية لإنترنت الأشياء (IoT)، وأنواع الهجمات في شبكات الاستشعار اللاسلكية (WSN)، بالإضافة إلى الذكاء الاصطناعي وتطبيقاته المتعددة في المباني الذكية. نقترح نظامًا دقيقًا يعتمد على التعلم العميق للكشف عن الهجمات وتصنيفها، لا سيما هجمات حجب الخدمة (DoS)، وذلك باستخدام شبكة عصبية التفافية أحادية البعد (CNN-1D). أظهرت النتائج التجريبية، التي تم الحصول عليها باستخدام قاعدة البيانات WSN-DS، أن نموذج CNN-1D يحقق أداءً مميزًا، حيث بلغ معدل الدقة 99.22%، والدقة الإحصائية 99.22%، والاسترجاع 99.20%، ودرجة F1 99.21%. كما تم تنفيذ مرحلة تفسير لقرارات النموذج باستخدام تقنيات الذكاء الاصطناعي التفسيري (XAI)، من خلال طريقة التفسير المحلي LIME. تؤكد هذه النتائج على متانة النموذج المقترح وفعاليته في تعزيز أمان شبكات الاستشعار اللاسلكية.

الكلمات المفتاحية: إنترنت الأشياء، الذكاء الاصطناعي، الهجمات، WSN، CNN، ANN، LSTM، الذكاء الاصطناعي التفسيري، LIME، التطبيع.

Résumé

Ce modeste travail présente les concepts fondamentaux de l'Internet des Objets (IoT), les attaques dans les réseaux de capteurs sans fil, ainsi que l'intelligence artificielle et ses diverses applications dans les bâtiments intelligents. Nous proposons un système précis basé sur l'apprentissage profond pour détecter et classifier les attaques, en particulier les attaques par déni de service (DoS), dans les réseaux de capteurs sans fil à l’aide d’un réseau de neurones convolutif unidimensionnel (CNN-1D). Les résultats expérimentaux, obtenus à partir de la base de données WSN-DS, indiquent que le modèle CNN-1D atteint des performances remarquables, avec une accuracy de 99,22 %, une precision également de 99,22 %, un recall de 99,20 %, et un F1-score de 99,21 %. Une phase explicative des décisions prises par le modèle proposé a ensuite été mise en œuvre en utilisant l'intelligence artificielle explicative (XAI), via la méthode d'interprétation locale LIME. Ces résultats confirment la robustesse et la pertinence du modèle proposé pour sécuriser les réseaux de capteurs sans fil.

Mots clés : IoT, Intelligence Artificielle, Attaques, WSN, CNN, ANN, LSTM, XAI, LIME, Normalisation.

Abstract

This modest work represents fundamental concepts of the Internet of Things (IoT), attacks in wireless sensor networks (WSNs), as well as artificial intelligence and its various applications in smart buildings. We propose an accurate system based on deep learning for detecting and classifying attacks, particularly Denial-of-Service (DoS) attacks, in WSN environments using a one-dimensional convolutional neural network (CNN-1D). Experimental results achieved using the WSN-DS dataset show that the CNN-1D model achieves remarkable performance, with an accuracy of 99.22%, a precision of 99.22%, a recall of 99.20%, and an F1-score of 99.21%. Furthermore, an explanatory phase of the model's decisions has been implemented using explainable artificial intelligence (XAI) through the local interpretation method LIME. These results validate the robustness and relevance of the proposed model for securing wireless sensor networks.

Keywords: IoT, Artificial Intelligence, Attacks, WSN, CNN, ANN, LSTM, XAI, LIME, Normalization.

Sommaire

Remerciement II
Dédicace III
Dédicace IV
ملخص V
Résumé V
Abstract VI
Sommaire VII
Liste des Figures X
Liste des Tableaux XII
Liste des abréviations XIII
Introduction Générale 1
Chapitre I : Généralités sur les Technologies Intelligentes et les Attaques dans les Bâtiments Intelligents 2
I.1-Introduction 3
I.2-Internet des Objets 3
I.2.1-Concept Générale 3
I.2.2 Architecture et Fonctionnement de l'IoT 3
I.2.3 Domaines d'Application de l'IoT 4
I.2.4- Défis Techniques et Perspectives de Recherche 4
I.3-Batiments Intelligents 4
I.3.1-Caractéristiques des Bâtiments Intelligents 5
I.3.2-Les avantages des bâtiments intelligents 6
I.3.3-Défis de sécurité des bâtiments intelligents 6
I.4-Réseaux capteur sans fil 7
I.4.1-Contraintes et Défis des WSN 7
I.4.2-Architecture Fonctionnelle d'un Capteur 8
I.5-Les attaques dans les réseaux capteur 9
I.5.1-Les types des attaques 9
I.5.2-L'attaque DOS 10
I.5.2.1-Les types d'attaque DOS 10

I.6-Intelligence Artificielle dans le domaine des bâtiments intelligents et la sécurité des WSN....11
I.6.1-Intelligence artificielle dans les bâtiments intelligents....11
I.6.2-Intelligence artificielle pour la Détection d'Attaques dans les WSN....11
I.7-Traveaux connexes....12
8-Conclusion....14
Chapitre II : Intelligence Artificielle Moderne : Deep Learning et XAI....15
II.1-Introduction....15
II.2-Intelligence Artificielle....15
II.2.1-Les domaines d'applications d'intelligence artificielle....16
II.2.2-Type d'intelligence artificielle....18
II.3-Apprentissage automatique (Machine Learning)....19
II.3.1-Algorithmes d'apprentissage automatique....21
II.4-Apprentissage profond (Deep Learning)....21
II.5-Les Types d'apprentissage....22
II.5.1-Apprentissage Supervisé....22
II.5.2-Apprentissage non-supervisé....23
II.5.3-Apprentissage Semi-Supervisé....24
II.5.4-Apprentissage par renforcement....24
II.4-Comparaison entre l'apprentissage automatique et profond....25
II.5-Réseaux Neurones....26
II.5.1-Neurone Biologique....27
II.5.2-Neurone artificielle....27
II.5.3-Architectures des réseaux neurones....28
II.6-Les fonction d'activation....30
II.7-Les couches du CNN....32
II.8-Intelligence Artificielle générative....35
II.8.1-Architectures utilisées en IA générative....35
II.9-Intelligence Artificielle Explicative....36
II.9.1-Les Méthodes Principales....36
II.9.2-Applications....37
II.10-Conclusion....37
Chapitre III : Système de Détection Intelligent et Evaluation des Performances....38
III.1-Introduction....38
III.2-Matériel utiliser....38

III.2.1-Python 38
III.3-Bibliothèque Utilisé 38
III.3.1-Tensorflow 38
III.3.2-Keras 39
III.3.3-Matplotlib 39
III.3.4-Sklearn 39
III.3.5-Pandas 39
III.3.6-Numpy 39
III.4-Méthodologie proposé 40
III.4.1-Prétraitment 41
III.4.2-Décomposition des données 42
III.4.3-Architecture du Modèle Proposé 42
III.5-Expérimentation et discussion des résultats 44
III.5.1-Description de la base des données 44
III.5.2-Les Métriques d'évaluation 44
III.5.3-Discussion des Résultats 45
III.5.4-XAI 51
III.5.5-Comparison 53
III.6-Conclusion 54
Conclusion Générale 55
Bibliographie 56

Liste des Figures

Figure I.1 : Bâtiments intelligents et IoT----------06

Figure I.2 : Architecture fonctionnel d'un capteur----------09

Figure II.1 : Intelligence Artificielle----------16

Figure II.2 : Domaine d'application de l'IA----------18

Figure II.3 : Les types d'IA----------19

Figure II.4 : Les types d'apprentissage----------20

Figure II.5 : Schéma explicative d'apprentissage supervisé----------23

Figure II.6 : Schéma significative d'apprentissage non-supervisé----------23

Figure II.7 : Apprentissage semi-supervisé----------24

Figure II.8 : Schéma significative d'apprentissage par renforcement----------25

Figure II.9 : Schéma significative de la différence entre l'apprentissage automatique et profond----------26

Figure II.10 : Neurone biologique----------27

Figure II.11 : Neurone artificiel et architecture ANN----------28

Figure II.12 : Architecture de réseau neurone artificielle----------29

Figure II.13 : Principes de calcul des neurones artificiels----------31

Figure II.14 : Couche de convolution----------32

Figure II.15 : Fonction d'activation ReLU----------33

Figure II.16 : Regroupement maximum et moyenne----------33

Figure II.17 : Couche d'aplatissement----------34

Figure II.18 : Couche entièrement connecter----------34

Figure II.19 : Couche de dropout----------35

Figure III.1 : Langage de programmation et les bibliothèques utiliser----------40

Figure III.2 : Schéma du système proposé----------40

Figure III.3 : Architecture du modèle CNN-1D proposée----------43

Figure III.4 : Distribution de la base des données----------44

Figure III.5 : Graphe de précision et des pertes----------47

Figure III.6 : Graphe de ROC----------48

Figure III.7 : Matrice de confusion de validation----------50

Figure III.8 : Matrice de confusion de test----------51

Figure III.9 : Explication des caractéristiques par classe selon la méthode LIME----------53

Liste des Tableaux

Tableau I.1: Résumé des travaux connexes sur la détection des attaques dans WSN----------14

Tableau II.1 : Comparaison entre l'apprentissage automatique et profond----------25

Tableau II.2 : Les fonctions d'activation les plus couramment utilisées----------31

Tableau III.1 : Performance des modèles développé----------45

Tableau III.2 : Performance du modèle CNN-1D sur les ensembles des données----------46

Tableau III.3 : Rapport de classification----------48

Tableau III.4 : Tableau de comparaison avec des travaux récente----------53

Liste des abréviations

AI	Intelligence Artificielle
ANN	Réseau de Neurones Artificiels
API	Interface de Programmation d'Applications
CNN	Réseau de Neurones Convolutif
DoS	Déni de Service
GAN	Réseau Adversarial Génératif
GUI	Interface Graphique Utilisateur
IDS	Système de Détection d'Intrusion
IoT	Internet des Objets
K-NN	Méthode des k plus Proches Voisins
LIME	Explications Locales Interprétables Indépendantes du Modèle
LSTM	Mémoire Long Terme à Court Terme
ReLU	Unité Linéaire Rectifiée
ROC	Caractéristique de Fonctionnement du Récepteur
RNN	Réseau de Neurones Récurrent
SHAP	Explications Additives de Shapley
SVM	Machine à Vecteurs de Support
WSN	Réseau de Capteurs Sans Fil
WSN-DS	Base des Données des Réseaux de Capteurs Sans Fil
XAI	Intelligence Artificielle Explicable

Introduction Générale

Avec le développement des nouvelles technologies, nos bâtiments deviennent de plus en plus intelligents. Grâce à l'Internet des objets (IoT) et aux réseaux de capteurs sans fil (WSN), il est désormais possible de gérer automatiquement l'énergie, la sécurité et le confort au sein de nos espaces de vie ou de travail. Mais cette modernisation, aussi prometteuse soit-elle, s'accompagne de nouvelles vulnérabilités, notamment face aux attaques. Parmi les menaces les plus répandues, les attaques par déni de service (DoS) représentent un réel danger. Elles peuvent désactiver des systèmes entiers en saturant les réseaux, mettant ainsi en péril la sécurité et le bon fonctionnement des bâtiments connectés. Face à cette réalité, il devient indispensable de développer des solutions intelligentes capables de détecter rapidement ces attaques et d'y réagir efficacement.

Dans ce contexte que s'inscrit notre projet de fin d'études. Notre objectif est de développer un système intelligent capable de détecter les attaques DoS suivie par une explication, en nous appuyant sur les outils puissants de l'intelligence artificielle, en particulier le deep learning et XAI.

Dans le premier chapitre, nous posons les bases du sujet en explorant l'univers des bâtiments intelligents, les principes de fonctionnement des réseaux de capteurs sans fil, ainsi que les différentes formes d'attaques auxquelles ces systèmes peuvent être exposés. Nous mettons également en lumière les travaux liés à la sécurité de ces environnements connectés.

Le deuxième chapitre nous parlons dans le monde de l'intelligence artificielle. Nous y abordons les grandes familles d'apprentissage supervisé, non supervisé, semi-supervisé, et par renforcement ainsi que les architectures de réseaux de neurones les plus couramment utilisées. Nous parlons aussi sur des approches récentes comme l'IA explicable (XAI), qui permettent de mieux comprendre les décisions prises par les modèles.

Enfin, dans le troisième chapitre, nous présentons notre contribution : un modèle de réseau de neurones convolutif (CNN). Nous détaillons les étapes de mise en œuvre, de la préparation des données à l'analyse des résultats, en passant par la construction du modèle. Les tests réalisés montrent que notre approche permet de détecter les attaques DoS avec une grande précision, tout en conservant un bon niveau d'interprétabilité avec une bon explication des décision prises par le modèle CNN.

Chapitre I : Généralités sur les Technologies Intelligentes et les Attaques dans les Bâtiments Intelligents

I.1-Introduction

Les réseaux de capteurs sans fil (WSN) jouent un rôle clé dans les systèmes intelligents, mais leur nature ouverte et leurs ressources limitées les rendent vulnérables aux attaques. Parmi celles-ci, les attaques par déni de service (DoS) représentent une menace majeure. La détection efficace de ces attaques est essentielle pour assurer la fiabilité et la sécurité des communications dans ces réseaux. Ce chapitre se concentre principalement sur l'internet des objets (IoT) avec son concept et l'architecture, avec quelque domaine d'application. Il détails les bâtiments intelligents et les WSN dans les bâtiments intelligents. Il met la lumière sur les attaques dans les WSN et les types d'attaque DoS. Enfin, nous présenterons brièvement l'intelligence artificielle et son role dans les différents domaines tel que la sécurité notamment pour la détection des attaques, en concluant par une comparaison des travaux connexes.

I.2-Internet des Objets

I.2.1-Concept Générale

L'Internet des objets (IoT) désigne l'interconnexion via Internet d'objets physiques capables de collecter, transmettre et traiter des données sans intervention humaine directe. Ces objets intelligents incluent des dispositifs domestiques, des équipements industriels et des infrastructures publiques. L'objectif principal de l'IoT est de fusionner le monde physique et numérique pour permettre une collecte automatique de données, un traitement contextuel et une prise de décision intelligente en temps réel [1].

I.2.2 Architecture et Fonctionnement de l'IoT

L'architecture de l'IoT repose sur plusieurs couches :

- **Couche de perception :** Intègre des capteurs et actionneurs embarqués sur les objets physiques pour collecter des données environnementales.
- **Couche de réseau :** Assure la transmission des données collectées vers les plateformes de traitement via divers protocoles de communication.
- **Couche de traitement :** Implique le stockage, l'analyse et le traitement des données dans des centres de données ou avec le cloud computing, souvent à l'aide de technologies d'intelligence artificielle.
- **Couche d'application :** Fournit des services personnalisés à l'utilisateur final, tels que la domotique intelligente ou la gestion énergétique [1].

Un exemple concret est celui d'une voiture connectée qui, à l'approche du domicile, envoie sa position géographique au thermostat intelligent de la maison, lequel ajuste automatiquement la température selon les préférences de l'utilisateur.

I.2.3 Domaines d'Application de l'IoT

L'IoT trouve des applications dans de nombreux secteurs :

- **Grand public (IoT domestique) :** Maisons intelligentes avec des systèmes de chauffage et d'éclairage automatisés, appareils de santé portables, véhicules connectés offrant des fonctionnalités avancées pour le confort et la sécurité.
- **IoT industriel (IIoT) et d'entreprise :** Suivi en temps réel des patients dans le secteur de la santé, surveillance des lignes de production dans l'industrie, gestion intelligente des stocks et des livraisons dans la logistique.
- **Villes intelligentes et agriculture connectée :** Utilisation de capteurs pour réguler le trafic, surveiller la pollution, optimiser l'éclairage public, et en agriculture, suivi des conditions environnementales pour améliorer les rendements [1].

I.2.4- Défis Techniques et Perspectives de Recherche

Le déploiement massif de l'IoT soulève plusieurs défis techniques et éthiques :

- **Sécurité :** La multiplication des points d'accès augmente les risques de cyberattaques.
- **Confidentialité des données :** Nécessité de mettre en place des mécanismes de protection des données personnelles.
- **Interopérabilité :** Standardisation des protocoles pour permettre la communication entre différents dispositifs.

Malgré ces défis, l'IoT demeure un pilier fondamental de la transformation numérique actuelle, notamment dans les bâtiments intelligents, les systèmes de surveillance et la gestion énergétique.

I.3-Batiments Intelligents

Les bâtiments intelligents, ou "smart buildings", désignent des structures intégrant des technologies numériques avancées, telles que l'Internet des objets (IoT), l'intelligence artificielle (IA) et les systèmes automatisés de contrôle, dans le but d'améliorer la performance énergétique, la sécurité, le confort et la gestion des ressources. Cette approche s'appuie sur le concept du réseau électrique intelligent ("smart grid") permettant une interconnexion des équipements électriques pour une gestion efficace et décentralisée de l'énergie [2, 3].

La domotique, précurseur des bâtiments intelligents, évolue depuis plus de quarante ans en intégrant progressivement l'automatisation et la connectivité dans les espaces résidentiels et professionnels [2]. Actuellement, les Nouvelles Technologies de l'Information et de la Communication (NTIC) permettent de concevoir des infrastructures intelligentes adaptées aux nouveaux usages sociétaux et environnementaux [3].

Bâtiment intelligent combine des systèmes de gestion énergétique actifs (automatisation) et passifs (conception thermique, orientation, isolation), avec pour objectif de réduire la consommation d'énergie tout en garantissant le confort et la sécurité des usagers [3].

I.3.1-Caractéristiques des Bâtiments Intelligents

- Automatisation des équipements, de l'éclairage et de la sécurité.
- IoT : communication temps réel entre les appareils connectés pour un pilotage centralisé.
- Gestion énergétique : surveillance, optimisation et prévision des consommations à l'aide d'outils analytiques.
- Sécurité avancée : systèmes de détection de mouvement, caméras, accès biométriques.
- Interopérabilité : compatibilité entre des dispositifs hétérogènes issus de différents fabricants.
- Maintenance prédictive : grâce aux données collectées, les systèmes anticipent les défaillances et améliorent la durée de vie des équipements.
- Adaptabilité dynamique : le bâtiment peut ajuster ses fonctions en fonction du comportement des usagers ou des conditions environnementales en temps réel.
- Accessibilité intelligente : aménagements technologiques facilitant l'accès pour les personnes à mobilité réduite (ascenseurs automatisés, portes connectées, etc.).

Les bâtiments connectés se distinguent par l'intégration d'éléments actifs (capteurs, actionneurs, objets connectés, systèmes cloud, alertes intelligentes) et passifs (matériaux isolants, conception bioclimatique, orientation) pour améliorer leur efficacité globale [9].

La Figure I.1 illustre le système interconnecté d’un bâtiment intelligent, dans lequel les dispositifs communiquent de manière fluide via des connexions filaires et des réseaux Wi-Fi sans fil. Cette intégration permet des interactions harmonieuses, telles que les thermostats intelligents fonctionnant de concert avec les systèmes d’éclairage et de sécurité afin d’optimiser le contrôle climatique et d’améliorer la sécurité, notamment en l’absence des occupants. Le schéma catégorise les dispositifs en systèmes fonctionnels : confort thermique pour la régulation du climat, éclairage intelligent pour une illumination adaptable, détection de gaz

pour la surveillance de la qualité de l'air, capteurs intelligents pour la collecte de données, détection d'anomalies pour l'identification des intrusions, surveillance pour un suivi en temps réel, communication réseau pour la connectivité et le contrôle à distance, efficacité énergétique pour la gestion des appareils, et santé pour le suivi des données biomédicales et le soutien en cas d'urgences médicales. Ensemble, ces systèmes transforment les bâtiments en environnements plus sûrs, plus efficients et de meilleure qualité, en tirant parti de l'automatisation et de l'analyse des données pour améliorer le quotidien moderne grâce à un échange d'informations fluide et à une coordination fonctionnelle des dispositifs [3].

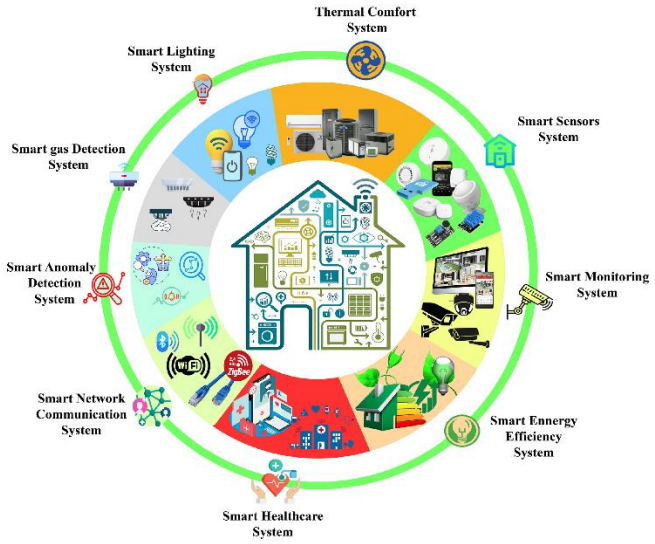


Figure I.1 : Bâtiments intelligents et IoT [3].

I.3.2-Les avantages des bâtiments intelligents

- Réduction de la consommation d'énergie.
- Amélioration du confort des usagers.
- Renforcement de la sécurité des informations personnel.
- Maintenance prédictive des équipements.
- Meilleure gestion des ressources comme eau, électricité, et chauffage.
- Amélioration de la productivité et du bien-être des occupants via la qualité de l'air, de l'éclairage et de la température.
- Capacité d'adaptation aux besoins évolutifs des utilisateurs et aux contraintes environnementales [4].

I.3.3-Défis de sécurité des bâtiments intelligents

Détection des attaques et cybersécurité dans les bâtiments intelligents s'appuient sur une connectivité constante entre dispositifs, ce qui les expose à divers types d'attaques, telles que les intrusions réseau, les attaques par déni de service (DoS), et le détournement de capteurs. La

sécurité des données personnelles, souvent essentielles au fonctionnement des services intelligents, est un enjeu majeur. Les systèmes doivent intégrer des mécanismes de détection d'intrusion (IDS), de chiffrement des communications, et de segmentation réseau pour limiter les risques. En réponse, les organismes comme l'ANSSI et la CNIL proposent des lignes directrices pour la sécurisation des systèmes connectés [5].

I.4-Réseaux capteur sans fil

Les réseaux de capteurs sans fil (RCSF/ Wireless Sensor Networks, (WSN)) représentent une technologie clé dans le déploiement de systèmes intelligents, facilitant la surveillance distribuée et l'acquisition de données dans des environnements variés. Leurs performances ont été rendues possibles par l'évolution des technologies de microélectronique et de communication sans fil, qui ont réduit le coût, la taille et la consommation énergétique des nœuds capteurs [6][7]. Les WSN sont aujourd'hui largement employés dans des domaines aussi divers que :

- **Surveillance environnementale :** suivi de la pollution atmosphérique et aquatique, détection des incendies et prévention des inondations.
- **Contrôle des écosystèmes :** surveillance de la faune, des forêts, de la croissance des plantes, et des réserves naturelles.
- **Applications militaires :** détection d'intrusions, surveillance périmétrique, reconnaissance sur les zones de conflit.
- **Santé et biomédical :** surveillance des patients, capteurs portables pour le suivi du diabète, de la fréquence cardiaque ou de la pression artérielle [8, 9, 10].

I.4.1-Contraintes et Défis des WSN

La conception et le déploiement des WSN sont confrontés à plusieurs défis techniques majeurs, notamment dans le domaine de la cybersécurité, qui est au cœur de cette recherche. Les WSN sont particulièrement vulnérables en raison de leurs ressources limitées, de la nature sans fil de la communication, et de l'absence d'une infrastructure de sécurité centralisée. Voici les principaux défis liés à la sécurité :

- Authentification et contrôle d'accès : assurer que seuls les nœuds autorisés puissent accéder au réseau et échanger des données. L'authentification légère est souvent nécessaire en raison des contraintes énergétiques.
- Confidentialité des données : les communications dans un réseau WSN peuvent être interceptées. Le chiffrement des données est indispensable, mais les algorithmes doivent être adaptés aux capacités restreintes des nœuds.

- Intégrité et non-répudiation : garantir que les données ne soient pas altérées lors de la transmission et que leur origine soit vérifiable.
- Détection des intrusions (IDS) : développement de systèmes légers de détection d'intrusions adaptés aux réseaux distribués comme déni par service et contraints en énergie, les WSN sont vulnérables à divers types d'attaques, ce qui impose des protocoles de sécurité spécifiques à faible coût.
- Mécanismes de résilience : capacité du réseau à maintenir ses fonctions même en cas de compromission partielle, via la redondance et les techniques de reconfiguration automatique [11].
- Gestion de l'énergie : les capteurs fonctionnent souvent sur batteries et dans des zones difficiles d'accès, ce qui nécessite des mécanismes d'économie d'énergie (communication asynchrone, mise en veille prolongée, etc.).
- Fiabilité de la communication : les perturbations radio, les collisions de données et les pertes de paquets peuvent compromettre la qualité du service.
- Scalabilité et tolérance aux fautes : les réseaux doivent être capables de s'adapter dynamiquement à l'ajout ou la perte de nœuds [12].

I.4.2-Architecture Fonctionnelle d'un Capteur

Un nœud capteur, est un dispositif embarqué compact intégrant plusieurs composants essentiels:

- Unité de traitement : composée d'un microcontrôleur (processeur), de mémoires RAM et Flash pour le traitement des données et le stockage des programmes.
- Module de communication radio : assure la transmission bidirectionnelle sans fil grâce à une antenne intégrée.
- Unité de détection (capteurs) : permet la mesure de grandeurs physiques ou chimiques (température, humidité, gaz, etc.).
- Système d'alimentation : batterie ou source énergétique externe (par exemple, panneaux solaires) assurant l'autonomie du dispositif.
- Unité de contrôle et d'interface : LED, ports I/O, et composants facilitant le déploiement et l'intégration dans un réseau [13].

La Figure I.2 présente un schéma typique de l'architecture fonctionnelle d'un capteur sans fil.

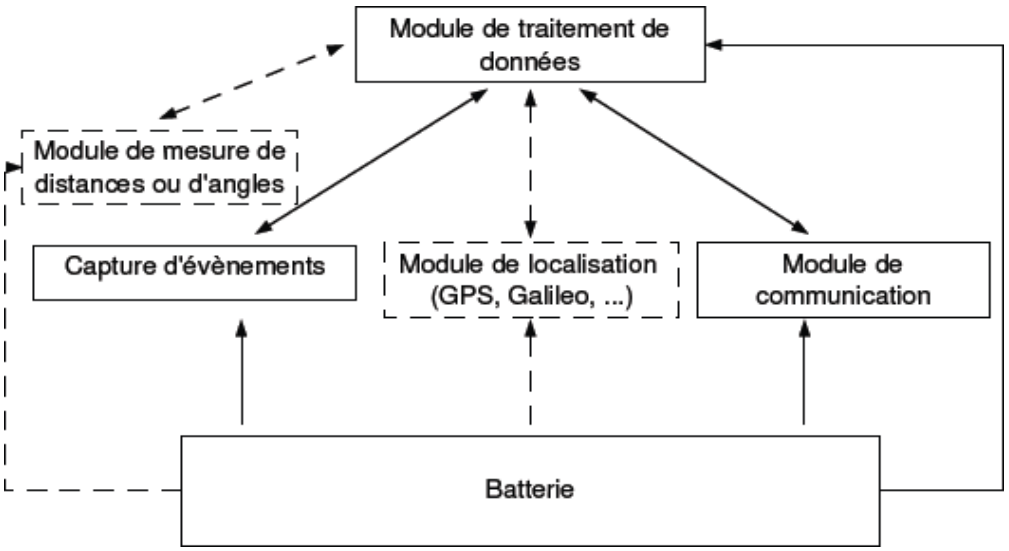


Figure I.2 : Architecture fonctionnelle d'un capteur.

I.5-Les attaques dans les réseaux capteur

Les réseaux de capteurs sans fil (WSN) sont particulièrement vulnérables à un large éventail d'attaques en raison de leurs ressources limitées (énergie, mémoire, capacité de traitement), de la nature ouverte du canal de communication et de l'absence d'une infrastructure centralisée. Les attaques peuvent affecter différentes couches du protocole réseau, allant de la couche physique à la couche application, compromettant ainsi la confidentialité, l'intégrité, la disponibilité et l'authenticité des données [11].

I.5.1-Les types des attaques

Les WSN sont exposés à une variété d'attaques en raison de leurs ressources limitées, de la nature sans fil des communications et de leur déploiement souvent non sécurisé. Les attaques les plus courantes dans ce domaine incluent :

- Attaques par Déni de Service (DoS) : surchargent un nœud ou le réseau afin de rendre les services indisponibles [15].
- Attaques par Déni de Service Distribué (DDoS) : plusieurs nœuds compromis agissent de manière coordonnée pour submerger une cible, rendant l'attaque plus difficile à détecter et à contrer [17].
- Attaque Sybil : un nœud usurpe plusieurs identités pour manipuler les processus de routage ou de vote dans les réseaux distribués [18].
- Attaque Wormhole (vortex) : deux attaquants coopèrent en créant un tunnel entre deux points pour rediriger le trafic réseau [16].
- Attaque Sinkhole (puits) : un nœud malveillant se fait passer pour une route optimale afin d'attirer le trafic [18].

- Jamming (brouillage) : l'attaquant émet des interférences radio pour bloquer ou perturber les transmissions [14].
- Attaque Replay (par rejeu) : des paquets interceptés sont réinjectés dans le réseau à un moment ultérieur pour tromper les systèmes [11].
- Attaque Selective Forwarding (transfert sélectif) : un nœud intercepte des messages et ne retransmet qu'une partie des paquets, provoquant une perte d'information subtile [11].

I.5.2-L'attaque DOS

L'attaque par DoS vise à rendre un réseau ou un service indisponible pour ses utilisateurs légitimes. Dans les WSN, ce type d'attaque est particulièrement nuisible car il peut épuiser les ressources critiques des nœuds capteurs, entraînant leur panne ou leur isolement.

I.5.2.1-Les types d'attaque DOS

Les attaques DoS et leurs variantes peuvent prendre plusieurs formes dans les réseaux de capteurs sans fil. Voici une liste détaillée des types les plus courants :

- Attaque Jamming : Émission de signaux radio de forte puissance pour perturber les communications sur la couche physique, empêchant ainsi les capteurs de transmettre leurs données.
- Attaque Exhaustion : Sollicitation excessive des ressources d'un nœud (répétition de requêtes, redémarrages) jusqu'à vider sa batterie, rendant le capteur inopérant.
- Attaque Hello Flood : Utilisation de messages de type "Hello" émis à haute puissance pour tromper les nœuds sur la proximité de l'attaquant et provoquer une redirection du trafic.
- Attaque Sybil : Un seul nœud adopte plusieurs identités pour influencer les décisions de routage, d'élection ou de stockage dans le réseau.
- Attaque Blackhole : Variante du sinkhole, où l'attaquant capture tous les paquets sans jamais les retransmettre, causant une perte totale d'information.
- Attaque Grayhole : Le nœud malveillant choisit de filtrer sélectivement certains paquets, introduisant une perte partielle et difficilement détectable.
- Attaque Flooding : L'attaquant envoie un grand nombre de messages de requêtes pour saturer les files d'attente mémoire et épuiser les ressources des capteurs.
- TDMA-Based DoS : Exploitation des créneaux temporels (dans les protocoles TDMA) pour envoyer des messages en dehors des plages allouées, causant des collisions ou un épuisement des ressources [19].

I.6-Intelligence Artificielle dans le domaine des bâtiments intelligents et la sécurité des WSN

L'intelligence artificielle (IA) joue un rôle fondamental dans la transformation numérique des bâtiments intelligents et dans le renforcement de la sécurité des WSN. En combinant des capacités d'analyse avancée, d'apprentissage automatique (ML) et de prise de décision autonome, l'IA permet d'optimiser l'efficacité énergétique, de surveiller en temps réel les systèmes, et de détecter de manière proactive les anomalies ou les attaques dans les environnements connectés [20].

I.6.1-Intelligence artificielle dans les bâtiments intelligents

IA dans les Bâtiments Intelligents Dans les infrastructures intelligentes, l'IA est utilisée pour modéliser le comportement des occupants, anticiper leurs besoins, et adapter dynamiquement les ressources énergétiques et environnementales. Les principales applications comprennent :

- Optimisation énergétique : des algorithmes de régression, réseaux de neurones ou systèmes de flou adaptent le chauffage, l'éclairage et la ventilation selon l'usage réel du bâtiment [21, 3].
- Maintenance préventive : des techniques de ML analysent les séries temporelles provenant de capteurs IoT pour anticiper les défaillances et planifier les interventions [23].
- Confort et sécurité des occupants : les systèmes IA apprennent les habitudes des utilisateurs et détectent des situations inhabituelles, améliorant la sûreté et l'expérience utilisateur [24, 3].
- Détection des fumées et d'incendies : l'apprentissage profond (DL) est utilisé sur les flux vidéo pour identifier des anomalies visuelles indiquant un départ de feu [22].

I.6.2-Intelligence artificielle pour la Détection d'Attaques dans les WSN

L'IA est particulièrement prometteuse pour renforcer la sécurité dans les réseaux de capteurs sans fil, en raison de sa capacité à analyser des comportements complexes et à détecter des attaques discrètes.

- Systèmes de détection d'intrusion (IDS) intelligents : des modèles d'apprentissage supervisé (SVM, KNN, RF, XGBoost) ou non supervisé (clustering, auto-encodeurs) permettent de classer le trafic comme normal ou anormal [25, 26].
- Approches fédérées : l'apprentissage fédérée (FL) permet de former des modèles de détection localement sur les nœuds capteurs tout en préservant la confidentialité des données, essentiel dans les architectures distribuées [3, 22].

- Détection d'attaques DoS : les réseaux CNN ou les modèles LSTM sont capables d'apprendre des séquences temporelles de comportement réseau pour identifier des motifs d'attaque [27].
- Systèmes hybrides : combinaisons d'algorithmes heuristiques (algorithmes génétiques, ACO) et de réseaux neuronaux pour la détection multicritère et l'adaptation aux attaques évolutives [28].

I.7-Traveaux connexes

N. Singh et al [29], ont proposé un système basé sur des règles floues nommé FzMAI (Fuzzified Methodology to Avert Intrusion) pour prévenir les intrusions dans les WSN. Contrairement aux approches conventionnelles de détection, leur méthodologie empêche les nœuds malveillants d'intégrer le réseau. La méthode proposée fonctionne en trois étapes : extraction des caractéristiques, calcul des valeurs d'appartenance et application des règles floues, afin de classer les nœuds comme malveillants, suspects ou normaux. Testé sur l'ensemble de données WSN-DS, le système a atteint une précision de 98,29 %, surpassant ainsi les méthodes existantes fondées sur la logique floue et démontrant son efficacité pour renforcer la sécurité du réseau de manière proactive.

I Almomani et al [30], ont proposé WSN-DS, un ensemble de données spécialisé pour la détection d'intrusions dans les WSN. Leur approche s'est concentrée sur quatre types d'attaques DoS : Blackhole, Grayhole, Flooding et Scheduling, en utilisant le protocole de routage LEACH. Le jeu de données, composé de 23 attributs extraits des simulations réalisées avec Network Simulator 2 (NS-2), a été évalué à l'aide des réseaux de neurones artificiels (ANN), atteignant des taux élevés de précision allant jusqu'à 97.54 % et 98.52% en utilisant Kfold cross validation (10 folders) pour la détection des attaques. L'étude a démontré l'efficacité du base des données pour détecter et classer précisément ces attaques, renforçant ainsi les mécanismes de détection d'intrusions dans les WSN.

S. Ismail et al. [31] ont proposé un système léger à plusieurs couches, basé sur ML, pour détecter les cyberattaques dans les WSN. Leur approche combine deux modèles de ML : un algorithme Naïve Bayes (NB) dans la première couche pour la classification binaire, et un algorithme LightGBM dans la deuxième couche pour une classification multiclasse. Ce cadre de détection hiérarchique, complété par un robot mobile pour accélérer la détection et l'acheminement des données, permet d'identifier efficacement quatre types d'attaques par déni de service (DoS) au niveau réseau (Blackhole, Grayhole, Flooding et Scheduling), en utilisant la base des données WSN-DS. Les résultats ont montré une performance élevée, atteignant des précisions de

détection allant jusqu'à 97.25% (NB) et 99,3% (LightGBM), tout en réduisant considérablement la latence de détection et la surcharge de communication au sein des réseaux à ressources limitées.

Alsulaiman et al [32] ont évalué l'efficacité de cinq algorithmes d'apprentissage automatique NB, réseaux de neurones (ANN), machines à vecteurs de support (SVM), arbres de décision (J48) et forêt aléatoire (Random Forest) (RF) pour détecter les attaques DoS dans les WSN en utilisant la base des données WSN-DS. Leur analyse expérimentale a démontré que le classificateur RF surpassait les autres techniques, atteignant une précision de détection maximale de 99,72 %. Cette approche s'est avérée particulièrement efficace pour identifier des types d'attaques spécifiques tels que Blackhole et Grayhole, soulignant ainsi son adéquation aux tâches de détection d'intrusion dans des réseaux sans fil à ressources limitées.

Salmi et al [33] ont proposé et évalué un ensemble de systèmes de détection d'intrusion légers basés sur DL pour détecter les différents types d'attaques DoS dans les WSN. En utilisant la base des données WSN-DS, ils ont implémenté et comparé quatre architectures : réseaux neurones profond (DNN), réseaux neurones convolutif (CNN), réseaux neurones récurrent (RNN) et un modèle hybride CNN+RNN. Les modèles ont été testés sur la classification de quatre types d'attaques DoS ; Blackhole, Grayhole, Flooding et Scheduling. Parmi eux, le modèle CNN a obtenu la meilleure précision avec 98,79 %, suivi du DNN (97,04 %), du CNN+RNN (96,86 %) et du RNN (96,48 %), démontrant la capacité supérieure du CNN à identifier avec précision des modèles d'attaque complexes dans des environnements WSN à ressources limitées.

Saleh et al. [34] ont développé SG-IDS, un système de détection d'intrusion basé sur l'apprentissage automatique, conçu pour les WSN. Leur approche intègre des techniques de sélection de caractéristiques l'analyse en composantes principales (PCA) et la décomposition en valeurs singulières (SVD) afin de réduire la dimensionnalité des données et d'améliorer la performance de détection. Le système utilise les classificateurs Gaussian Naive Bayes (GNB) et Stochastic Gradient Descent (SGD), et a été évalué sur l'ensemble de données WSN-DS. SG-IDS a atteint une précision de 98%, surpassant plusieurs méthodes de pointe. Ces résultats démontrent son efficacité pour détecter avec précision divers types d'intrusions réseau, tout en conservant des performances en temps réel et une faible consommation de ressources.

Tableau I.1 : Résumé des travaux connexes sur la détection des attaques dans WSN.

Travail	Année	Prétraitement	Modèle	Base des données	Précision (%)
N. Singh et al [29]	2020	Sélection des caractéristiques (CFS), Calcul des fonctions d'appartenance	FzMAI	WSN-DS	98.29
I.Almomani et al [30]	2016	Extraction de caractéristiques (23 attributs extraits à partir de LEACH)	ANN	WSN-DS	97.54 et 98.52 (cross validation)
S. Ismail et al [31]	2022	Sélection de caractéristiques par information mutuelle (Mutual Information)	NB, Light GBM	WSN-DS	97.25 (NB) 99.3 (Light GBM)
L. Alsulaiman et al [32]	2021	/	RF	WSN-DS	99.72
S. Salmi et al [33]	2023	Normalisation	DNN	WSN-DS	97.04
			CNN		98.85
			RNN		96.48
			CNN+RNN		96.86
H M. Saleh et al [34]	2024	SVD + PCA15	GNB	WSN-DS	98

8-Conclusion

Ce chapitre a présenté une vue d'ensemble des technologies clés liées à l'Internet des Objets, aux bâtiments intelligents, et aux WSN. Il a mis en lumière les défis de sécurité, notamment les attaques DoS, et a introduit l'apport de l'intelligence artificielle comme solution innovante pour la détection proactive des intrusions. Ces éléments constituent une base solide pour comprendre les problèmes actuels en matière de cybersécurité dans les environnements intelligents.

Chapitre II : Intelligence Artificielle Moderne : Deep Learning et XAI

II.1-Introduction

Ce chapitre présente les concepts fondamentaux de l'intelligence artificielle, ses domaines d'application, ainsi que les types d'apprentissage : supervisé, non supervisé, semi-supervisé et par renforcement. Il détaille les algorithmes d'apprentissage, les architectures profondes (CNN, RNN, LSTM, GAN, Transformers), les fonctions d'activation, et les couches constitutives des réseaux convolutifs. Il aborde également les paradigmes récents d'IA générative et explicable (XAI), utilisés pour la génération de données et l'interprétation des modèles complexes.

II.2-Intelligence Artificielle

L'intelligence artificielle (IA) constitue un domaine central de l'informatiques, qui vise à la fois un intérêt scientifique croissant et un défi conceptuel majeur. Cette difficulté repose en grande partie sur l'absence de définition précise et universellement acceptée de ce qu'est l'intelligence artificielle, ainsi que sur la complexité à distinguer clairement ses caractéristiques de celles de l'intelligence naturelle. En termes généraux, l'IA désigne un ensemble de systèmes informatiques capables de simuler ou de reproduire certaines fonctions cognitives humaines, telles que l'apprentissage, le raisonnement logique, la résolution de problèmes, la prise de décision ou encore la compréhension du langage naturel, et l'adaptation à de nouvelles situations [35].

Le terme **'artificielle'** de cette intelligence réside dans le fait qu'elle n'émerge pas de processus biologiques évolutifs, comme c'est le cas chez les êtres vivants, mais résulte plutôt de la conception et de la programmation humaine. Dans l'IA englobe toute machine ou système démontrant des capacités associées à l'intelligence humaine, notamment la généralisation, l'apprentissage à partir de l'expérience, l'identification de patterns, ou encore l'adaptation à de nouveaux environnements [35, 37].

Malgré les progrès technologiques significatives aient permis d'augmenter considérablement la puissance de calcul et la capacité de stockage des systèmes informatiques, les modèles actuels restent encore loin d'égaler la souplesse cognitive humaine, en particulier dans les domaines nécessitant des connaissances contextuelles. Toutefois, dans des tâches spécifiques et bien délimitées, plusieurs systèmes d'IA ont atteint, voire surpassé, les niveaux de performance d'experts humains. On retrouve ainsi des applications performantes d'IA dans des domaines variés tels que le diagnostic médical, la reconnaissance vocale, l'analyse d'images, les systèmes de recommandation ou les moteurs de recherche [38].

L'IA est une science interdisciplinaire, combinant des approches de l'informatique, des mathématiques, des sciences cognitives, de la linguistique, de la philosophie et de l'éthique.

Les avancées récentes en apprentissage automatique (machine learning), et plus particulièrement en apprentissage profond (deep learning), ont contribué à un véritable changement de paradigme dans le domaine, impactant profondément la recherche et l'innovation dans de nombreux secteurs industriels et sociétaux [35,36].

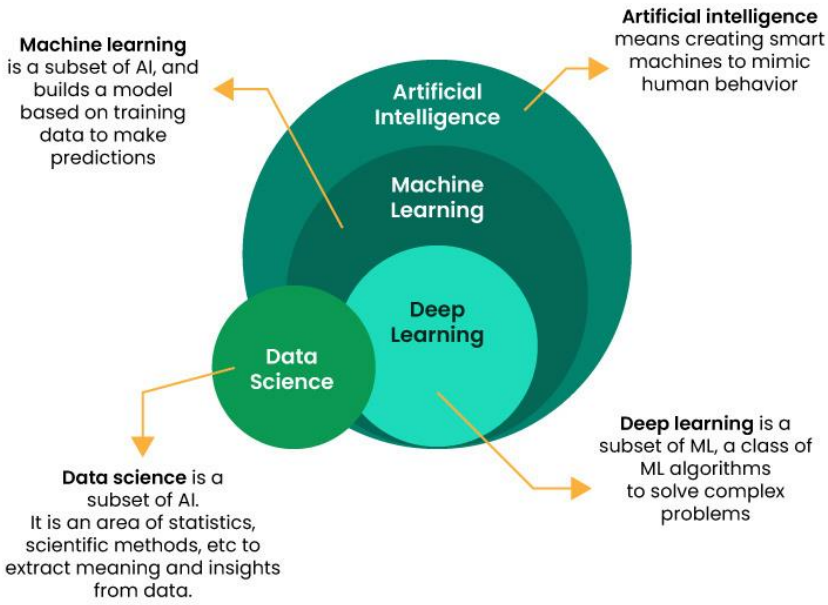


Figure II.2 : Intelligence Artificielle

II.2.1-Les domaines d'applications d'intelligence artificielle

L'IA est exploitée dans de nombreux secteurs spécialisés, où elle apporte des solutions concrètes à des problématiques complexes. Les domaines d'application suivants illustrent la diversité et la précision des usages actuels de l'IA :

- **Systèmes experts :** Programmes capables de reproduire des raisonnements spécialisés pour exécuter des tâches ciblées avec un haut degré de précision.
- **Réseaux de neurones :** Modèles computationnels inspirés du cerveau humain, fondés sur des neurones artificiels interconnectés, utilisés pour la reconnaissance de motifs, la classification et la prédiction.
- **Calcul formel :** Traitement symbolique des expressions mathématiques, distinct du calcul numérique, avec des applications en algèbre, en vérification formelle et en résolution d'équations.

- **Représentation des connaissances :** Encodage structuré de savoirs sous forme symbolique pour permettre leur traitement logique et leur exploitation par des systèmes intelligents.
- **Modélisation du raisonnement :** Implémentation de schémas logiques tels que les logiques modales et temporelles, visant à simuler les mécanismes du raisonnement humain.
- **Traitement automatique du langage naturel :** Analyse, compréhension et génération du langage humain à des fins de traduction, résumé, extraction d'information ou interaction homme-machine.
- **Résolution automatique de problèmes :** Application de méthodes heuristiques ou algorithmiques pour résoudre des situations complexes, notamment dans les jeux ou la planification.
- **Reconnaissance vocale et manuscrite :** Identification automatique de la parole et de l'écriture, appliquée à la transcription, à la commande vocale et à l'indexation de documents.
- **Reconnaissance faciale :** Détection et identification automatisées des visages dans des contextes de sécurité, de contrôle d'accès ou de surveillance.
- **Robotique intelligente :** Intégration de l'IA dans les systèmes robotiques afin d'améliorer la perception, la planification des mouvements et la prise de décision autonome.
- **Systèmes complexes adaptatifs :** Modèles évolutifs combinant des algorithmes génétiques et des mécanismes de vie artificielle pour simuler des dynamiques de populations ou des comportements collectifs.
- **Médecine :** Soutien au diagnostic, à la prédiction de maladies, à l'analyse d'imagerie médicale et à la personnalisation des traitements.
- **Ingénierie :** Optimisation de la conception, de la maintenance prédictive et du contrôle de systèmes complexes par l'intégration de solutions intelligentes.
- **Automatisation industrielle :** Réduction de l'intervention humaine dans les tâches répétitives par l'implémentation de systèmes autonomes basés sur l'IA.
- **Maison intelligent :** Régulation automatisée de l'environnement domestique (éclairage, sécurité, énergie) en fonction du comportement et des préférences des occupants.
- **Villes intelligentes :** Amélioration des services urbains (mobilité, énergie, sécurité, gestion des déchets) à travers l'analyse prédictive et la gestion en temps réel des données [36, 37].

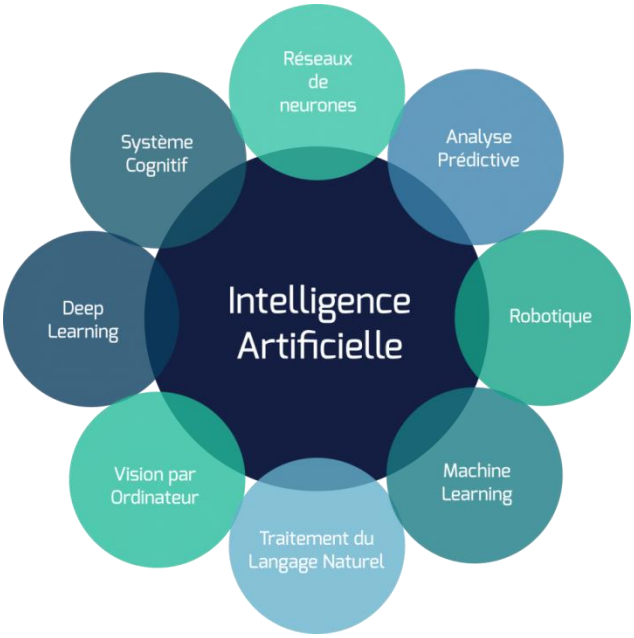


Figure II.2 : Domaine d'application de l'IA

II.2.2-Type d'intelligence artificielle

L'IA est classifiée selon deux axes principaux : son niveau fonctionnel, basé sur la complexité de ses capacités cognitives simulées, et son stade évolutif, défini par sa capacité à percevoir, mémoriser et interagir avec son environnement.

II.2.2.1-Niveaux fonctionnels

- **IA étroite (ANI) :** Cette forme d'IA est conçue pour exécuter une tâche spécifique, dans un cadre strictement délimité. Elle ne dispose ni de polyvalence ni de capacité à transférer ses connaissances à d'autres domaines. Elle est largement déployée dans les systèmes de reconnaissance faciale, les moteurs de recommandation, les assistants vocaux ou encore les outils de détection d'anomalies.
- **IA générale (AGI) :** Encore à l'état de recherche, l'AGI vise à reproduire la flexibilité cognitive humaine. Elle serait capable d'apprendre, de raisonner et de s'adapter à des situations multiples, sans être limitée à un domaine précis.
- **IA superintelligente (ASI) :** Entièrement théorique, cette forme représenterait une intelligence dépassant celle de l'humain dans l'ensemble des domaines cognitifs, incluant la créativité, la résolution de problèmes complexes, et la compréhension émotionnelle. Elle soulève des enjeux éthiques majeurs [45].

II.2.2.2-Stades évolutifs

- **Machines réactives :** Ces systèmes réagissent uniquement à des entrées immédiates et ne disposent d'aucune mémoire. Ils ne peuvent pas améliorer leur comportement sur la base d'expériences passées. L'ordinateur Deep Blue, conçu pour jouer aux échecs, illustre ce type d'architecture.
- **IA à mémoire limitée :** Capables d'intégrer des données passées à court terme, ces systèmes adaptent leurs décisions en fonction d'expériences récentes. Les véhicules autonomes s'appuient sur ce modèle pour analyser l'environnement et ajuster leur conduite en temps réel.
- **Théorie de l'esprit :** Ce niveau, encore non réalisé, correspond à une IA capable de représenter les états mentaux d'autrui (intentions, croyances, émotions), ouvrant la voie à des interactions homme-machine contextuellement intelligentes.
- **IA consciente de soi :** Représentant l'ultime niveau de développement, ce stade supposerait une conscience réflexive de l'agent artificiel. Il impliquerait la capacité à modéliser sa propre identité, ses états internes et son influence sur l'environnement [45].

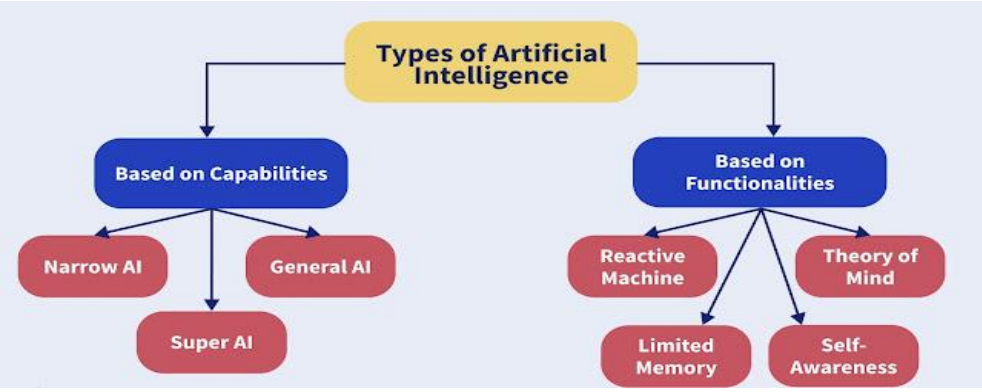


Figure II.3 : Les types d'IA

II.3-Apprentissage automatique (Machine Learning)

L'apprentissage automatique constitue un sous-domaine essentiel de l'IA, fondé sur la capacité d'un système informatique à extraire des connaissances à partir de données et à accomplir des tâches spécifiques sans être explicitement programmé pour chaque situation. Le principe basé sur la construction de modèles statistiques ou algorithmiques capables de détecter des motifs réguliers dans des ensembles de données massifs, afin de formuler des prédictions, classifications ou décisions.

Ces modèles sont générés à partir d'un ensemble de données d'entraînement, représentant des observations annotées du problème ciblé. L'algorithme ajuste ses paramètres internes en

minimisant une fonction de coût, puis évalue sa performance sur un ensemble de données de test indépendant, garantissant sa capacité de généralisation.

Le processus d'apprentissage nécessite une phase préparatoire impliquant plusieurs interventions humaines, notamment la sélection des caractéristiques pertinentes, le choix de l'algorithme approprié, la normalisation des données, et l'évaluation des performances à l'aide de métriques quantitatives.

L'apprentissage automatique permet d'exécuter des tâches complexes telles que la classification de signaux, la détection d'anomalies, la prédiction de tendances, l'estimation de probabilités conditionnelles, et l'adaptation continue à de nouveaux flux de données [45].

Deux types principaux d'ensembles de données sont requis pour l'apprentissage :

- **Données d'entraînement :** Utilisées pour ajuster les paramètres du modèle en exposant l'algorithme à des exemples connus.
- **Données de test :** Permettent de mesurer objectivement les performances du modèle sur des cas non vus, assurant ainsi sa capacité de généralisation.

Les algorithmes d'apprentissage automatique se répartissent en trois grandes catégories (voir Figure II-4) :

- Apprentissage supervisé, exploitant des données étiquetées.
- Apprentissage non supervisé, pour découvrir des structures cachées sans étiquettes.
- Apprentissage par renforcement, optimisant les actions à partir de récompenses [35, 38].

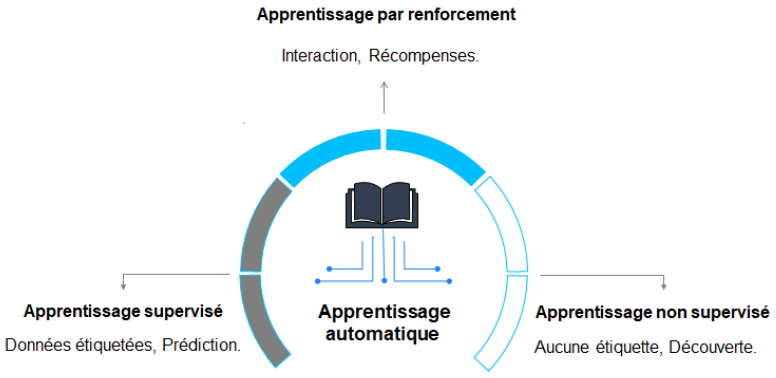


Figure II.4 : Les Types d'apprentissage

II.3.1-Algorithmes d'apprentissage automatique

Les algorithmes d'apprentissage automatique constituent le socle des systèmes intelligents, permettant l'extraction de connaissances à partir de données. Les plus couramment utilisés incluent :

- **Régression linéaire :** Modèle de prédiction d'une variable continue à partir de variables indépendantes. Il s'applique aux tâches de régression simples.
- **Régression logistique :** Adaptée aux problèmes de classification binaire, elle estime la probabilité d'appartenance à une classe donnée.
- **Arbres de décision :** Modèles arborescents prenant des décisions selon des règles conditionnelles. Bien qu'interprétables, ils sont sensibles au surapprentissage.
- **Forêts aléatoires (Random Forest) :** Ensemble d'arbres de décision entraînés sur des sous-ensembles aléatoires, réduisant la variance et le surapprentissage (overfitting).
- **Machines à vecteurs de support (SVM) :** Algorithmes de classification efficaces pour la séparation de classes, avec possibilité de généralisation à des cas non linéaires via les noyaux.
- **k-plus proches voisins (k-NN) :** Méthode non paramétrique classant les observations en fonction de la distance à leurs voisins. Efficace mais coûteuse pour de grands volumes de données.
- **Naïve Bayes :** Classifieur probabiliste basé sur le théorème de Bayes, supposant l'indépendance conditionnelle entre les attributs. Particulièrement utile pour la classification de textes.
- **Réseaux de neurones artificiels :** Modèles multi couches capables d'apprendre des relations complexes. Ils constituent la base des approches en apprentissage profond [39, 45].

Le choix de l'algorithme dépend du type de tâche (régression, classification), de la nature des données, des contraintes d'interprétabilité, ainsi que des exigences en matière de précision et de coût computationnel [38, 39].

II.4-Apprentissage profond (Deep Learning)

L'apprentissage profond (Deep Learning), constitue une branche avancée de l'IA, dérivée de l'apprentissage automatique. Il repose sur l'utilisation de réseaux de neurones artificiels profonds, organisés en couches hiérarchiques successives, permettant la modélisation de relations non linéaires complexes dans les données.

Contrairement aux approches classiques fondées sur des règles explicites, cette méthode exploite de grandes quantités de données pour ajuster automatiquement ses paramètres, en extrayant des représentations de plus en plus abstraites au fil des couches du réseau. Ce mécanisme hiérarchique est particulièrement efficace pour les tâches nécessitant l'identification de structures complexes [40].

L'apprentissage profond a démontré une efficacité remarquable dans plusieurs domaines, notamment :

- **Vision par ordinateur :** détection et classification d'objets, reconnaissance faciale, segmentation d'images.
- **Reconnaissance vocale :** transcription automatique, assistants vocaux.
- **Traduction automatique :** systèmes multilingues en temps réel.
- **Génération de contenu :** production de texte, d'images ou de séquences audio.

Grâce à sa capacité à traiter des données massives et à apprendre sans supervision explicite, le Deep Learning s'impose comme une technologie de référence dans le traitement de l'information complexe. Il continue d'évoluer à travers des innovations architecturales (CNN, RNN, Transformers) et des avancées en optimisation computationnelle [38, 45].

II.5-Les Types d'apprentissage

II.5.1-Apprentissage Supervisé

L'apprentissage supervisé repose sur l'utilisation d'un ensemble de données d'entraînement constitué de couples (x_i, y_i), où x_i représente l'entrée et y_i la sortie attendue (target). Le modèle apprend une fonction de prédiction $f : X \rightarrow Y$ en ajustant ses paramètres pour minimiser une fonction de perte mesurant l'écart entre les prédictions $f(x_i)$ et les valeurs cibles y_i.

Ce paradigme est particulièrement adapté aux tâches de classification et de régression. Par exemple, en classification d'images, le modèle est entraîné à associer chaque image à une étiquette parmi un ensemble de classes prédéfinies.

Cependant, l'apprentissage supervisé présente certaines contraintes, notamment la nécessité de disposer de jeux de données volumineux et étiquetés, ainsi que la difficulté d'application dans les contextes où la définition explicite de la sortie cible est imprécise ou coûteuse [40].

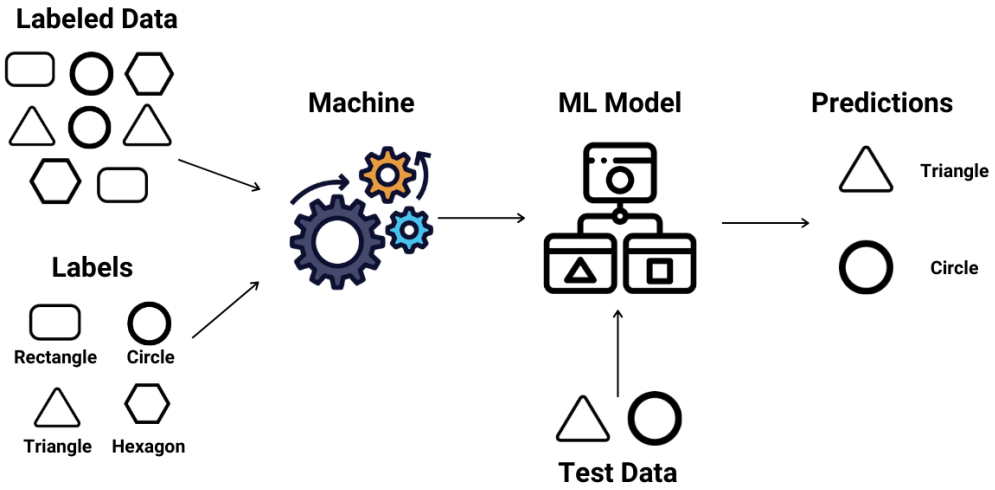


Figure II.5 : Schéma explicative d'apprentissage supervisé

II.5.2-Apprentissage non-supervisé

L'apprentissage non supervisé basé sur l'analyse de données d'entrée non étiquetées, dans le but d'extraire des structures latentes ou des régularités statistiques. Le modèle tente d'organiser ces données en groupes cohérents (clustering) ou de réduire leur dimensionnalité tout en préservant leur variance informative (réduction de dimension).

Par exemple, les algorithmes de regroupement tels que K-means ou DBSCAN permettent d'identifier automatiquement des clusters de points présentant des similarités intrinsèques, sans connaissance préalable des classes.

L'absence de supervision rend cependant l'évaluation plus complexe, car il n'existe pas de vérité terrain explicite. De plus, l'interprétation des résultats dépend fortement du contexte d'application et de la structure des données [41, 45].

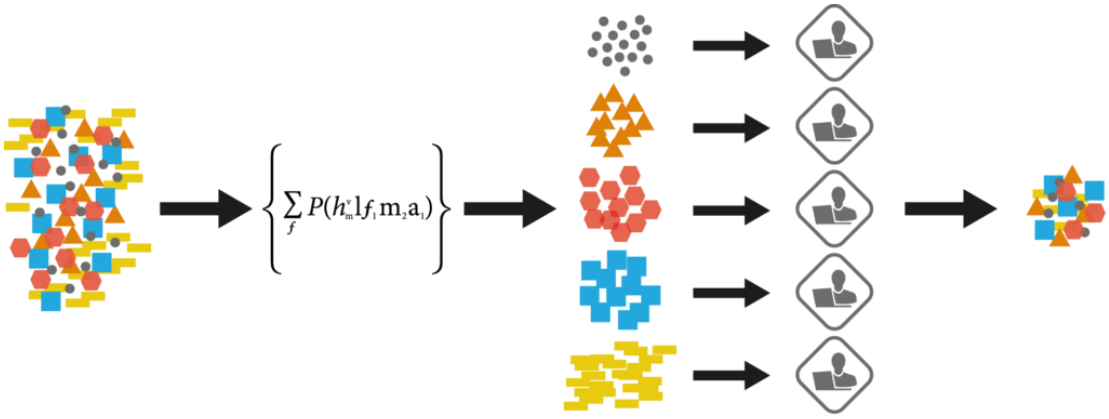


Figure II.6 : Schéma significative d'apprentissage non-supervisé

II.5.3-Apprentissage Semi-Supervisé

L'apprentissage semi-supervisé s'appuie simultanément sur un sous-ensemble de données étiquetées et un ensemble plus large de données non étiquetées. Le modèle est d'abord entraîné sur les exemples étiquetés, puis utilise la structure apprise pour inférer des régularités et améliorer la prédiction sur les données non annotées.

Cette approche est particulièrement utile dans les contextes où l'étiquetage manuel est coûteux ou partiellement disponible. Par exemple, en analyse de sentiments sur les réseaux sociaux, quelques messages annotés peuvent guider l'algorithme à classifier automatiquement de grands volumes de contenu non étiqueté.

Cependant, son efficacité dépend fortement de la qualité des annotations disponibles et de la représentativité des données non supervisées [41].

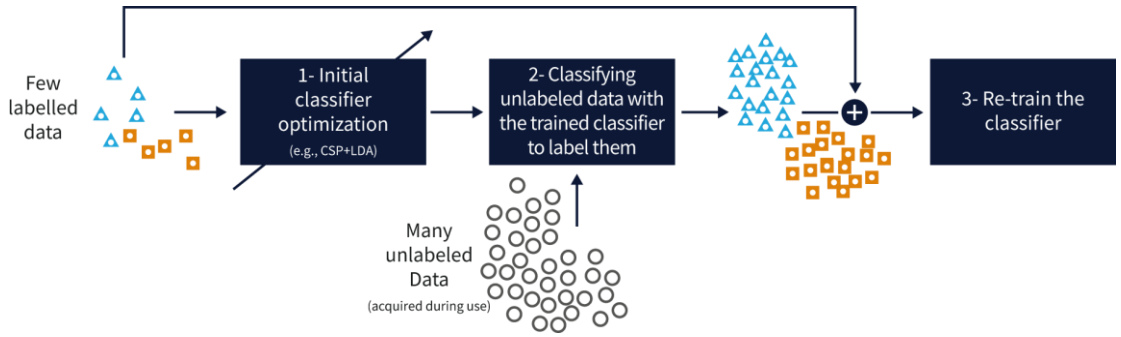


Figure II.7 : Schéma significative d'apprentissage semi-supervisé

II.5.4-Apprentissage par renforcement

L'apprentissage par renforcement repose sur l'interaction entre un agent intelligent et un environnement. À chaque étape, l'agent effectue une action et reçoit un retour sous forme de récompense numérique, reflétant la qualité de cette action par rapport à un objectif défini. L'agent ajuste ainsi sa stratégie dans le but de maximiser la récompense cumulative sur le long terme.

Cette approche est particulièrement efficace pour l'apprentissage séquentiel de politiques décisionnelles, comme démontré dans les jeux stratégiques (échecs, Go) ou la robotique autonome. Toutefois, elle requiert des ressources computationnelles importantes et présente des limitations dans les environnements à retour de signal limité, retardé ou non stationnaire [41].

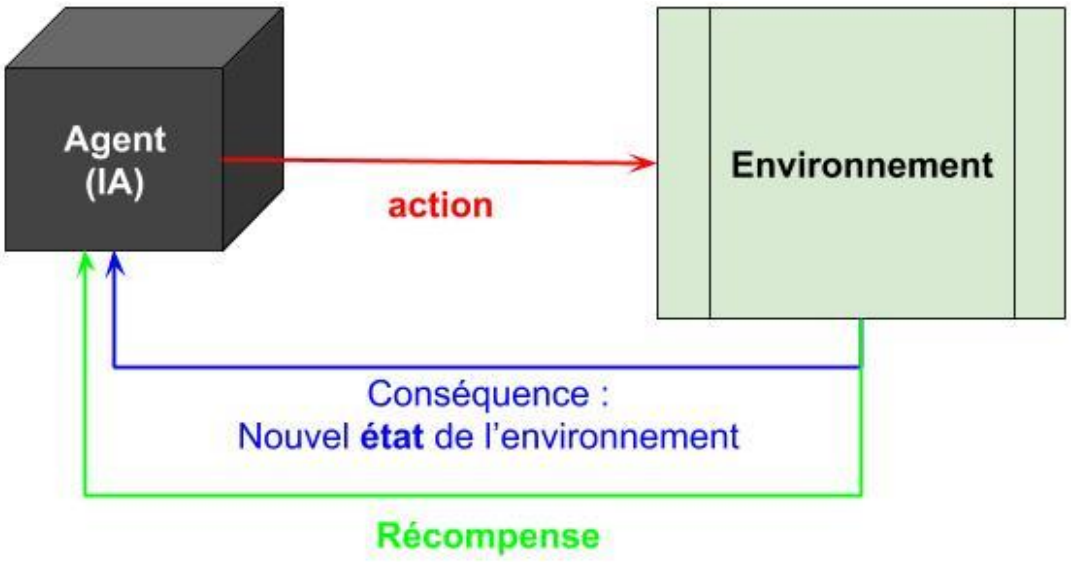


Figure II.8 : Schéma significative d'apprentissage par renforcement

II.4-Comparaison entre l'apprentissage automatique et profond

L'apprentissage automatique et l'apprentissage profond diffèrent principalement par la nature des données traitées, l'architecture des modèles et les techniques de modélisation employées [36, 42]. La comparaison entre les deux type d'apprentissage dans le tableau II.1.

Tableau II.1 : Comparaison entre l'apprentissage automatique et profond [42].

	Apprentissage Automatique	**Apprentissage Profond**
Type de données	Données structurées tableaux, ou vecteurs.	Données non structurées comme des images, texte, audio, signaux, vidéos
Extraction de caractéristiques	Extraction des caractéristiques manuelle	Extraction des caractéristiques automatique avec les couches du modèle
Architecture du modèle	Architecture simple	Architecture complexer avec réseaux profonds avec plusieurs couches (CNN, RNN, LSTM, DNN, etc.)
Taille des données	Efficace avec une taille moyenne	Une grandes quantités de données
Temps d'entraînement	Moins coûteux en calcul	Temps long, dépendance aux GPU
Techniques	Régression, SVM, arbres de décision, k-NN, etc.	DNN, CNN, LSTM, RNN, Transformers

Applications	Analyse prédictive, classification, regression.	Vision par ordinateur, reconnaissance vocale, NLP
Expertise humaine	Forte intervention humaine	Moins d'intervention humaine

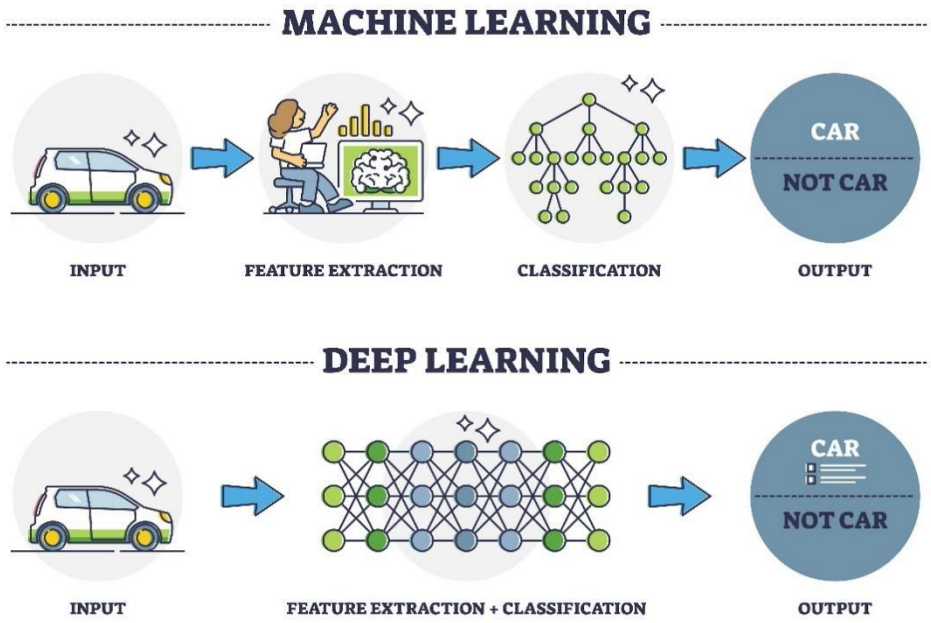


Figure II.9 : Schéma significative de la différence entre l'apprentissage automatique et profond [10, 45].

II.5-Réseaux Neurones

Les réseaux de neurones artificiels sont des modèles computationnels inspirés du fonctionnement du cerveau humain, où un grand nombre d'unités de calcul simples (neurones) opèrent en parallèle, sans qu'une unité centrale de contrôle ne coordonne explicitement leurs actions [40].

L'apprentissage s'effectue principalement par l'ajustement des pondérations (poids synaptiques) entre les neurones, ces poids constituant le principal mécanisme de mémorisation à long terme des informations dans le réseau. La mise à jour de ces pondérations, généralement à travers des algorithmes de rétropropagation, permet au réseau d'intégrer de nouvelles connaissances à partir des données d'entraînement [35].

La dynamique d'un réseau de neurones dépend fortement de son architecture, définie par plusieurs paramètres clés : le nombre de neurones, le nombre de couches (couches d'entrée, cachées et de sortie), ainsi que la nature des connexions entre ces couches (entièrement connectées, convolutives, récurrentes, etc.). Ces choix structurels influencent directement la

capacité du réseau à apprendre des représentations complexes et à généraliser à de nouveaux jeux de données [35, 40, 43].

II.5.1-Neurone Biologique

Le cerveau humain est constitué d'un réseau complexe de plus de 10 milliards de neurones interconnectés, chacun établissant des connexions synaptiques avec environ 10 000 autres neurones. Chaque neurone comprend un corps cellulaire, ou soma, qui reçoit les signaux via des prolongements appelés dendrites, et transmet les signaux via un unique prolongement appelé axone. Ce dernier se ramifie en plusieurs terminaisons et assure la communication avec d'autres neurones à travers des jonctions spécialisées appelées synapses (voir Figure II.10).

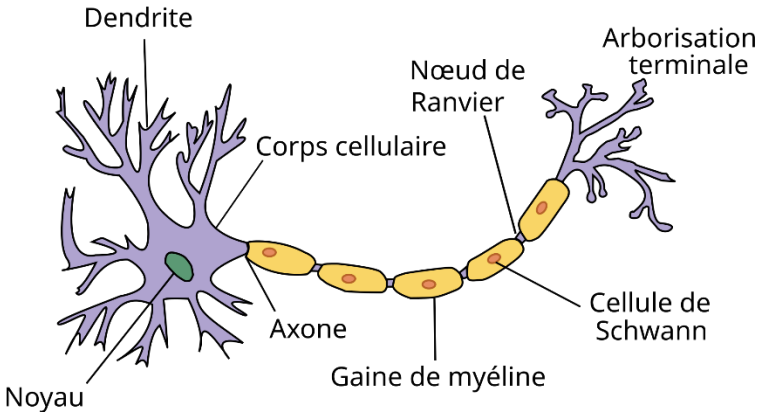


Figure II.10 : Neurone Biologique.

Les dendrites, formant une structure arborescente, convergent vers le soma, où est localisé le noyau cellulaire. L'axone, issu du soma, se termine par des ramifications qui établissent des connexions synaptiques avec les dendrites d'autres neurones. La transmission synaptique repose sur un mécanisme biochimique par lequel des neurotransmetteurs sont libérés à partir de la terminaison présynaptique. Ces substances chimiques modifient le potentiel électrique de la membrane postsynaptique. Lorsque ce potentiel dépasse un certain seuil, un potentiel d'action est généré et propagé le long de l'axone, déclenchant ainsi l'activation du neurone récepteur.

Ce mécanisme constitue la base du traitement et de la transmission de l'information dans le système nerveux, et il inspire la structure et le fonctionnement des réseaux de neurones artificiels en intelligence artificielle [35, 36].

II.5.2-Neurone artificielle

Les réseaux de neurones artificiels (ANN) sont des modèles computationnels inspirés du fonctionnement des neurones biologiques. Ils se composent de neurones artificiels organisés en couches et interconnectés par des liens pondérés. Chaque neurone calcule une sortie en

appliquant une fonction d'activation sur la somme pondérée de ses entrées. Les poids synaptiques représentent la mémoire du réseau et sont ajustés au cours de l'apprentissage pour minimiser l'erreur entre les sorties prédictives et les valeurs attendues. Cette capacité d'ajustement permet aux ANN d'exécuter diverses tâches telles que la classification, la prédiction, la reconnaissance de formes, ou la modélisation de comportements complexes.

Le traitement de l'information s'effectue de manière hiérarchique, depuis les couches d'entrée jusqu'à la sortie, avec propagation du signal et adaptation des poids via des algorithmes d'optimisation tels que la rétropropagation. Ces architectures sont largement utilisées dans les domaines de l'ingénierie, de l'analyse de données et de l'intelligence artificielle [36].

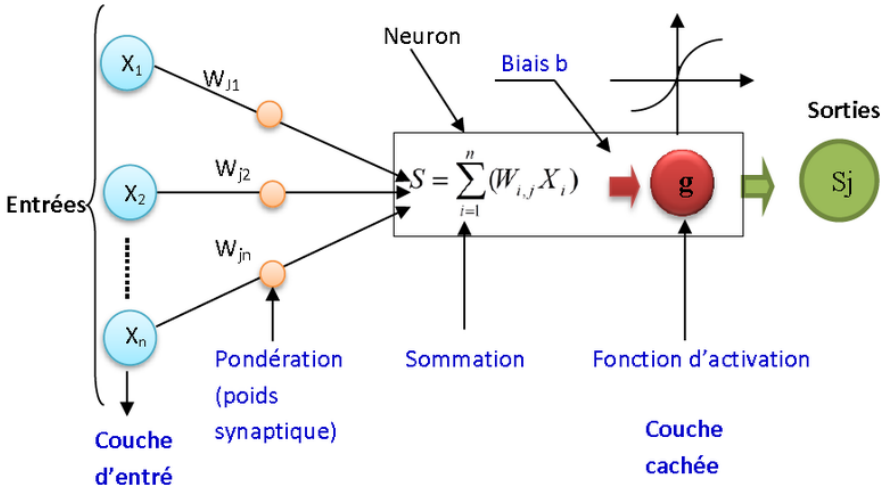


Figure II.11 : Neurone Artificielle.

II.5.3-Architectures des réseaux neurones

Un réseau de neurones artificiels est structuré en trois types de couches fonctionnelles :

- **Couche d'entrée :** Elle reçoit les vecteurs d'attributs décrivant les observations. Chaque neurone de cette couche correspond à une variable explicative du jeu de données.
- **Couches cachées :** Situées entre l'entrée et la sortie, elles assurent la transformation des données à travers des fonctions non linéaires. La profondeur du réseau dépend du nombre de ces couches empilées, permettant la modélisation de relations complexes.
- **Couche de sortie :** Elle produit le vecteur de sortie final du réseau, généralement sous forme d'une probabilité ou d'une classe prédite, en fonction du type de tâche (classification, régression, etc.) [35].

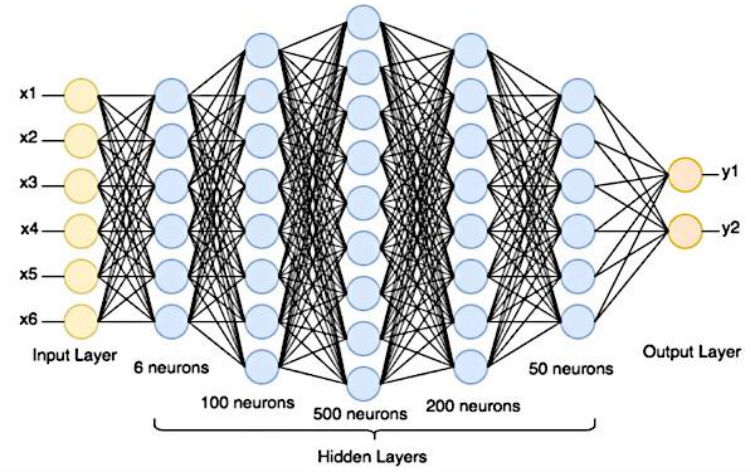


Figure II.12 : Architecture de réseau neurone artificielle.

II.5.3.1-Les différents réseaux neurones

II.5.3.1.1-Réseaux convolutifs (CNN)

CNNs sont des architectures neuronales conçues pour le traitement de données disposées en grille, telles que les images. Contrairement aux réseaux entièrement connectés, les CNN utilisent des opérations de convolution pour extraire automatiquement des caractéristiques locales (bords, textures, formes) avec des filtres glissants. Cette structure permet une réduction significative du nombre de paramètres tout en préservant les relations spatiales.

Les CNN sont largement utilisés dans :

- Classification des images.
- Détection des objets.
- Segmentation.
- Reconnaissance faciale.
- Diagnostic assisté par imagerie médicale.
- Détection des attaques.
- Traitements des données [35].

II.5.3.1.2-Réseaux récurrents (RNN)

Les RNNs sont adaptés à l'analyse de séquences où l'ordre des données est crucial, comme dans le traitement du langage naturel et les signaux audio. Leur boucle interne permet de conserver un état temporel, ce qui facilite l'exploitation du contexte dans les données séquentielles. Cependant, ils sont souvent remplacés ou complétés par des architectures comme les LSTM ou les Transformers pour mieux gérer les longues dépendances [45].

II.5.3.1.3-Réseaux à mémoire long terme (LSTM)

LSTM est une variante des réseaux neuronaux récurrents (RNN), spécifiquement développée pour surmonter les limitations liées à l'oubli d'informations dans les longues séquences. Leur architecture repose sur une cellule mémoire régulée. Les LSTM sont performants pour les séries temporelles, la reconnaissance vocale, l'analyse de séquences biologiques et la traduction automatique.

II.5.3.1.4-Réseaux génératifs adverses (GAN)

Les GANs associent deux réseaux : un générateur qui produit de nouvelles données à partir de bruit aléatoire, et un discriminateur qui tente de distinguer les données réelles de celles générées. Ce mécanisme compétitif permet au générateur d'apprendre à créer des données réalistes. Les GAN sont utilisés dans la synthèse d'images, la restauration de données, la génération de visages, et la traduction d'image à image [44].

II.5.3.1.5-Transformers

Les Transformers sont des modèles fondés sur un mécanisme d'attention, introduits pour traiter efficacement des séquences longues sans recourir à la récursivité. Leur capacité à modéliser les dépendances globales dans une séquence les rend particulièrement adaptés au traitement du langage naturel, à la traduction automatique, ainsi qu'à la génération de texte. Ils constituent l'architecture de base des modèles modernes comme BERT, GPT et T5 [44].

II.6-Les fonction d'activation

Pour déterminer la sortie d'un neurone artificiel, chaque entrée x_i est multipliée par un poids synaptique w_i, reflétant son influence relative. La somme pondérée obtenue, généralement notée $z = \sum_i x_i w_i + b$, est ensuite soumise à une fonction d'activation $f(z)$, introduisant une non-linéarité nécessaire à l'approximation de fonctions complexes. Le rôle de la fonction d'activation est d'ajuster dynamiquement la réponse du neurone en fonction de l'intensité du signal. Plusieurs fonctions sont couramment utilisées, chacune adaptée à des architectures ou à des contextes d'optimisation particuliers [45]. Les plus fréquentes incluent :

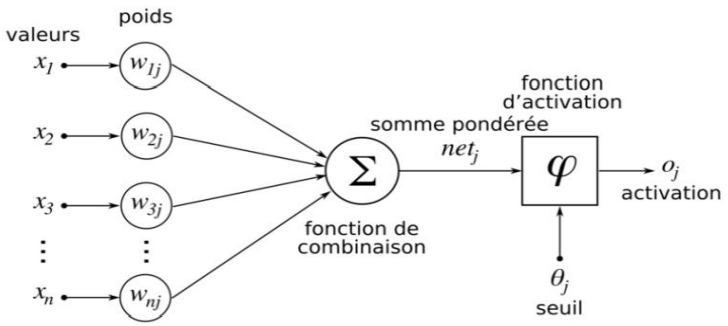
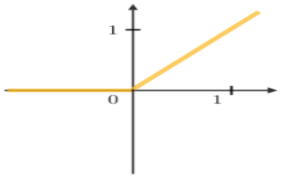
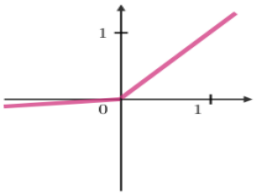
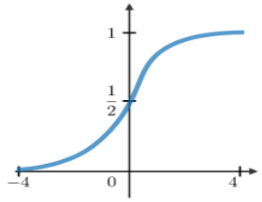
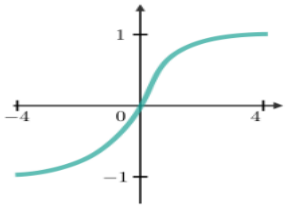
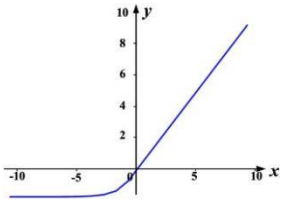
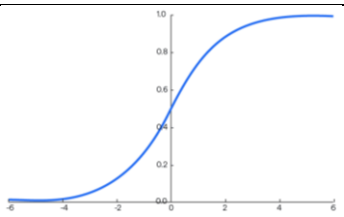


Figure II.13 : Principes de calcul des neurones artificiels.

Tableau II.2 : Les fonctions d'activation les plus couramment utilisées [35, 44].

Fonction	Equation	Description	Graph
Rectified Linear Unit (ReLU)	$\begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases}$	Introduit une activation nulle pour les entrées négatives, facilitant l'apprentissage profond grâce à sa simplicité et sa rapidité de convergence.	
Leaky ReLU (LReLU)	$\begin{cases} 0 & x \geq 0 \\ ax & Ailleurs \end{cases}$	Variante de ReLU, elle évite la désactivation totale des neurones en conservant un faible gradient pour les entrées négatives.	
Sigmoïd	$\frac{1}{1+e^{-x}}$	Utilisée pour des sorties probabilistes, surtout dans les couches de sortie des classifieurs binaires	
Tangente Hyperbolique (Tanh)	$\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	Centrée sur zéro, elle atténue les valeurs extrêmes tout en conservant la non-linéarité.	
ELU	$\begin{cases} x & x > 0 \\ \alpha(e^{x}-1) & x \leq 0 \end{cases}$	Introduit une croissance exponentielle pour les valeurs négatives, ce qui améliore la stabilité et accélère la convergence pendant l'apprentissage	
Softmax	$\frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}}$	Utilisée dans la couche de sortie pour la classification	

		multi-classes, elle convertit un vecteur de scores en une distribution de probabilité, chaque sortie étant comprise entre 0 et 1 et la somme des sorties égale à 1	

II.7-Les couches du CNN

Les CNN sont des architectures multicouches composées de couches spécifiques assurant l'extraction hiérarchique des caractéristiques et la prise de décision. Ils se composent notamment de couches de convolution, d'activation, de regroupement (pooling), d'aplatissement (flattening), de couches entièrement connectées et d'une couche de sortie.

- **Couche d'entrée :** Elle reçoit l'image d'entrée sous forme de matrice de pixels, où chaque élément encode l'intensité d'un pixel. Cette représentation matricielle constitue la base de traitement des couches suivantes.
- **Couche de convolution :** Cette couche applique des filtres (ou noyaux) glissants sur l'image pour extraire des motifs locaux (bords, textures, formes). Chaque filtre produit une carte de caractéristiques (feature map), et l'empilement de plusieurs filtres permet d'apprendre différentes représentations de l'image. Le processus est répété sur plusieurs couches, conduisant à une abstraction progressive des informations visuelles [35].

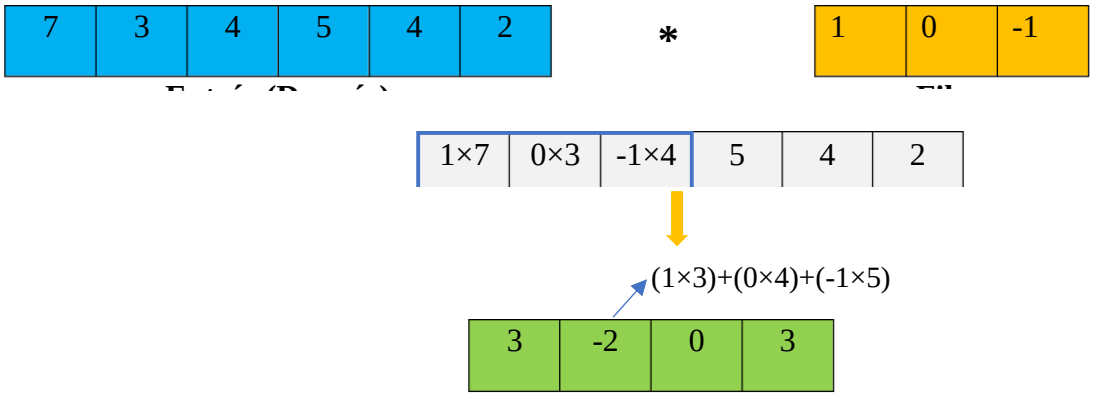


Figure II.14 : Couche de convolution

- **Couche d'activation ReLU :** Après chaque opération de convolution, une fonction d'activation non linéaire, généralement la ReLU, est appliquée. Cette fonction remplace les valeurs négatives par zéro, permettant d'introduire de la non-linéarité tout en conservant les activations positives [35].

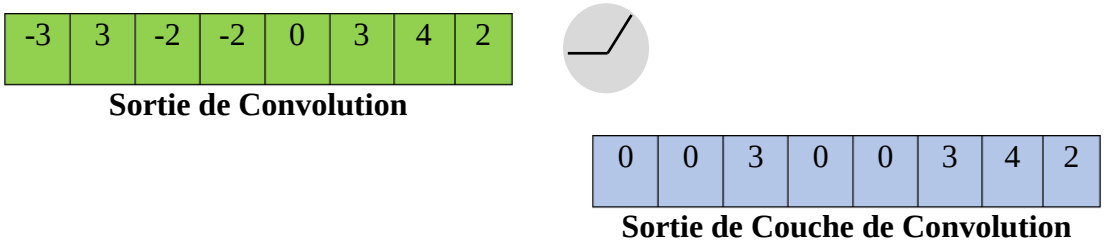


Figure II.15 : Fonction d'activation ReLU

- **Le pas (Stride) :** Le pas (stride), désigne la distance de déplacement du filtre convolutif entre deux positions successives lors de l'application de l'opération de convolution. Une foulée de 1 implique un déplacement du filtre pixel par pixel, permettant une couverture dense de l'image ou vecteur. En revanche, un pas supérieure à 1 induit un sous-échantillonnage spatial, réduisant la résolution de la carte des caractéristiques. Cette approche est utile pour diminuer la complexité computationnelle et extraire des représentations plus globales des entrées [35].
- **Couche de regroupement (Pooling) :** Elle réduit les dimensions spatiales des cartes de caractéristiques tout en conservant les informations les plus significatives. Les deux techniques les plus utilisées sont le maximum local (Max Pooling) et la moyenne locale (Average Pooling). Cela permet de diminuer la complexité computationnelle et de limiter le surapprentissage [35].

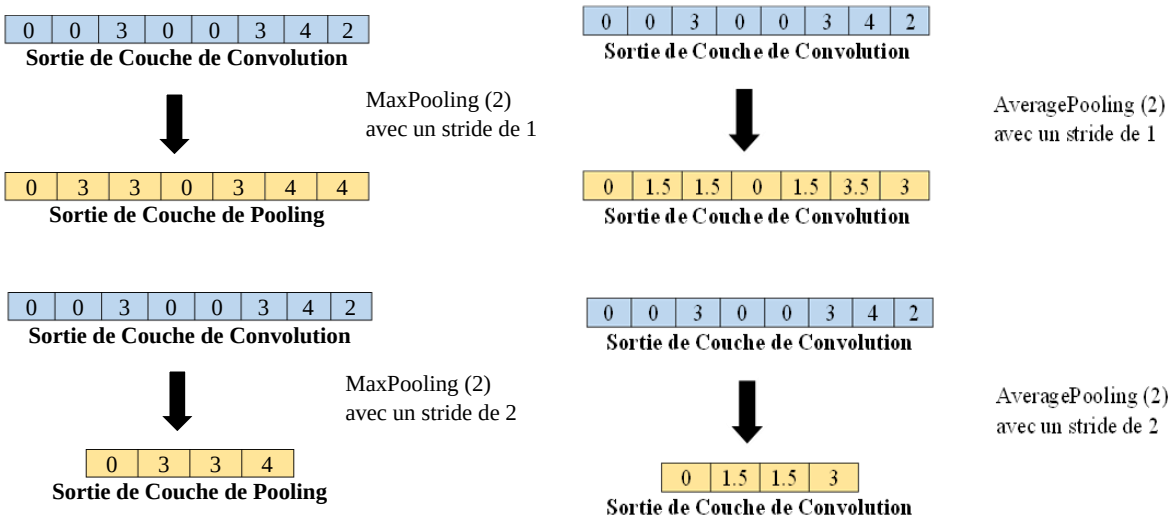


Figure II.16 : Regroupement maximum et moyenne

- **Couche d'aplatissement (Flatten Layer) :** Cette opération transforme les cartes de caractéristiques multidimensionnelles en un vecteur unidimensionnel. Elle constitue

une passerelle entre les couches de traitement spatial (convolutives et pooling) et les couches denses [35].

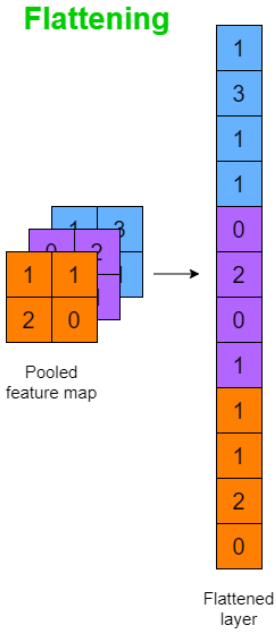


Figure II.17 : Couche d'aplatissement

- **Couche entièrement connectée (Fully Connected Layer) :** Tous les neurones de cette couche sont connectés à l'ensemble des activations de la couche précédente. Elle permet une exploitation globale des caractéristiques extraites afin d'effectuer une classification ou une régression. Cette couche est souvent utilisée en phase finale du modèle pour agréger les informations issues des couches précédentes [35].

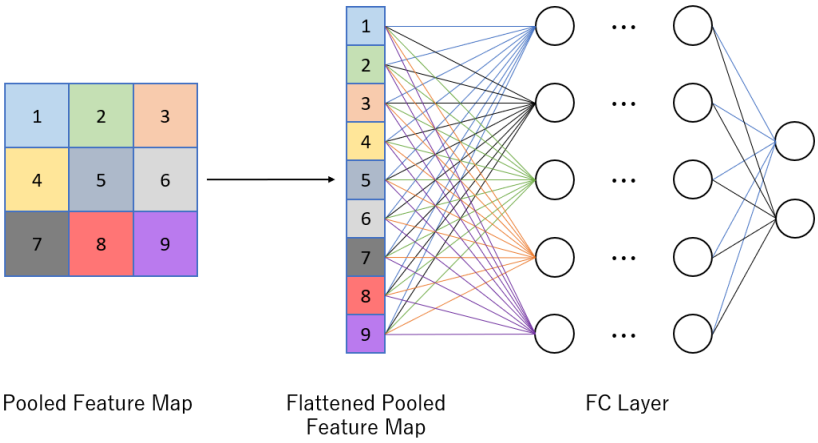


Figure II.18 : Couche entièrement connecter.

- **Couche de sortie (Output Layer) :** Elle fournit la prédiction finale du modèle. Dans le cadre d'une classification, cette couche contient autant de neurones que de classes cibles, et applique généralement la fonction d'activation Softmax pour produire une distribution de probabilité.

- **Dropout :** Cette couche implémente une stratégie de régularisation consistant à désactiver aléatoirement un certain pourcentage de neurones pendant l'entraînement. Cela empêche la co-adaptation excessive entre neurones et réduit le surapprentissage (overfitting). Typiquement, une probabilité de 0.5 est utilisée, ce qui signifie que la moitié des activations sont ignorées à chaque itération d'entraînement [35].

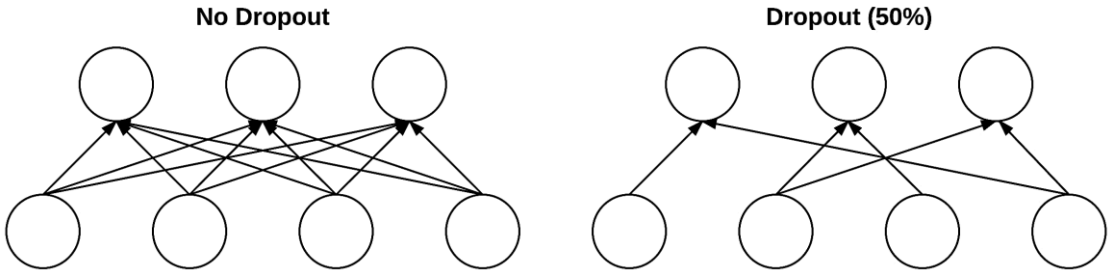


Figure II.19 : Couche de dropout.

II.8-Intelligence Artificielle générative

L'IA générative constitue un sous-domaine de l'IA qui vise à produire de nouveaux contenus, tels que des images, du texte, de l'audio ou de la musique, à partir de modèles d'apprentissage automatique. Contrairement aux approches discriminatives centrées sur des tâches de classification ou de prédiction, l'IA générative apprend la distribution des données d'entraînement afin de générer des échantillons synthétiques présentant des caractéristiques similaires à celles observées [44].

II.8.1-Architectures utilisées en IA générative

II.8.1.1-Réseaux Génératifs Adverses (GAN)

Les Generative Adversarial Networks reposent sur l'interaction entre deux réseaux neuronaux : un générateur G, chargé de produire des données synthétiques, et un discriminateur D, chargé de distinguer les données réelles des données générées. Le processus d'apprentissage est formulé comme un jeu à somme nulle, où le générateur tente d'induire le discriminateur en erreur. Cette structure permet la génération de contenus hautement réalistes. Les GAN sont utilisés dans des tâches telles que la génération d'images photoréalistes, la restauration d'images basse résolution, la colorisation d'images noir et blanc, ou la synthèse de visages artificiels [44].

II.8.1.2-Réseaux Neuronaux Génératifs (GNN)

GNN a ne pas confondre avec les Graph Neural Networks, les GNN génératifs font référence à des architectures capables de modéliser explicitement la distribution de probabilité des données

d'entrée. Contrairement aux GAN, ces modèles ne reposent pas sur une structure compétitive entre deux réseaux, mais utilisent une seule architecture, souvent fondée sur des autoencodeurs variationnels (VAE) ou des modèles autoregressifs (comme PixelRNN ou GPT), pour générer directement des données cohérentes avec la distribution observée [44].

II.9-Intelligence Artificielle Explicative

L'intelligence artificielle explicable (XAI) regroupe un ensemble de méthodes et de techniques permettant d'interpréter, de justifier et de comprendre les décisions prises par des modèles d'IA, en particulier ceux considérés comme des « boîtes noires », tels que les réseaux de neurones profonds. Face à la complexité croissante des modèles d'apprentissage automatique, la transparence devient un enjeu crucial dans des domaines critiques comme la santé, la finance, ou la cybersécurité.

L'objectif principal de la XAI est de fournir des explications compréhensibles par l'humain concernant les prédictions du modèle, tout en conservant un haut niveau de performance. Elle répond à des exigences réglementaires, éthiques, et de confiance. Une explication peut concerner des aspects globaux du modèle (comportement général) ou locaux (explication d'une prédiction particulière) [46].

II.9.1-Les Méthodes Principales

Les approches explicatives peuvent être classées en deux grandes catégories :

- **Méthodes intrinsèques :** Utilisent des modèles naturellement interprétables comme les arbres de décision, les règles d'association, ou les modèles linéaires. Leur structure même permet une compréhension directe du processus décisionnel.
- **Méthodes post-hoc :** Fournissent des explications a posteriori, après l'entraînement du modèle. Elles sont adaptées aux modèles complexes. Les plus répandues incluent :
  - **Model Interprétable a Locale Explication (LIME) :** Méthode locale qui approxime un modèle complexe par un modèle linéaire localement interprétable autour d'un échantillon donné [46].
  - **Explanations additives de Hapley (SHAP) :** Basée sur la théorie des données, cette méthode attribue à chaque caractéristique une contribution additive à la prédiction en respectant des axiomes de cohérence et de localité [46].
  - **Cartographie de l'activation des classes pondérée par gradient (Grad-CAM) :** Utilisée pour l'interprétation visuelle des CNN, elle génère des cartes de chaleur indiquant les zones de l'image les plus influentes dans la prise de décision [46].

II.9.2-Applications

La XAI est largement utilisée dans :

- Le diagnostic médical assisté par l'IA (interprétation de décisions cliniques),
- Les systèmes de détection d'intrusion en cybersécurité (justification des alertes),
- L'audit algorithmique (évaluation de biais et de discriminations),
- Les modèles de confiance dans les véhicules autonomes.

II.10-Conclusion

Ce chapitre fournissent les fondations nécessaires pour comprendre et concevoir des modèles intelligents adaptés à des tâches complexes. Les sections traitant de l'apprentissage profond, des architectures neuronales, de l'IA générative et explicable préparent à l'élaboration de solutions performantes et interprétables qui seront exploitées dans les étapes expérimentales suivantes

Chapitre III : Système de Détection Intelligent et Evaluation des Performances

III.1-Introduction

Ce chapitre expose une approche innovante de détection d'attaques dans les réseaux capteur de bâtiments intelligents, fondée sur un modèle de réseau de CNN-1D. L'objectif est de concevoir un modèle d'apprentissage profond capable de classifier différents types d'intrusions à partir de données de trafic réseau. Ce travail détaille les principales étapes expérimentales, à commencer par la préparation et la normalisation des données, l'entraînement et la validation du modèle, jusqu'à l'analyse des performances. Le chapitre intègre également une analyse explicative à l'aide de la méthode LIME, permettant d'interpréter les prédictions du modèle. Enfin, les résultats obtenus sont comparés à ceux d'approches existantes de l'état de l'art, afin de mettre en évidence la pertinence de l'architecture proposée.

III.2-Matériel utiliser

III.2.1-Python

Python est un langage de programmation interprété, dynamique et polyvalent, largement reconnu pour sa syntaxe claire, sa lisibilité et sa communauté de développement particulièrement active. Il est couramment employé dans divers domaines, notamment le développement web, la science des données, l'apprentissage automatique, ainsi que l'automatisation de processus. L'un de ses principaux atouts réside dans la richesse de son écosystème, incluant un vaste ensemble de bibliothèques et de frameworks spécialisés, ce qui en fait une plateforme de choix pour le développement d'applications fonctionnelles et scientifiques [1].

III.3-Bibliothèque Utilisé

Dans le cadre de cette étude, plusieurs bibliothèques spécialisées dans l'intelligence artificielle (IA) et le calcul numérique ont été mobilisées pour le développement, l'entraînement des modèles, ainsi que pour la visualisation et l'analyse des résultats.

III.3.1-Tensorflow

TensorFlow est une plateforme open source de bout en bout, spécifiquement conçue pour le développement des modèles d'apprentissage profond (Deep Learning). Elle offre un ensemble modulaire et extensible d'outils, de bibliothèques et de ressources communautaires permettant d'accélérer la recherche et l'implémentation de solutions intelligentes. Elle est également adaptée au déploiement rapide de modèles sur des environnements variés, allant du cloud aux dispositifs embarqués [35].

III.3.2-Keras

Keras est une bibliothèque open source de haut niveau développée en Python, qui fournit une interface conviviale pour la conception, l'entraînement et l'évaluation de réseaux de neurones artificiels. Intégrée à TensorFlow, elle permet de réaliser rapidement du prototypage, ce qui en fait un outil privilégié pour les chercheurs et praticiens du Deep Learning [35].

III.3.3-Matplotlib

Matplotlib est une bibliothèque dédiée à la visualisation de données en Python. Elle permet de produire un large éventail de visualisations graphiques allant de simples tracés linéaires à des représentations scientifiques complexes pour facilitant l'interprétation des résultats. Elle est couramment utilisée dans les domaines de la recherche, de la science des données et de l'ingénierie [35].

III.3.4-Sklearn

Scikit-learn (ou sklearn) est une bibliothèque Python destinée à l'apprentissage automatique. Elle propose une implémentation robuste de nombreux algorithmes d'apprentissage supervisé et non supervisé, ainsi que des outils pour le prétraitement des données, l'évaluation des performances, et la sélection de modèles. Elle constitue un cadre de référence pour le développement de solutions intelligentes dans des domaines variés tels que la classification, la régression, ou encore la réduction de dimensionnalité [35].

III.3.5-Pandas

Pandas est une bibliothèque incontournable pour la manipulation et l'analyse de données structurées. Elle introduit des structures de données efficaces, comme les DataFrame, facilitant le traitement, le nettoyage, l'exploration et la visualisation des ensembles de données. Pandas est largement utilisée dans les projets d'analyse de données, de business intelligence et de modélisation statistique [35].

III.3.6-Numpy

Numpy constitue une bibliothèque fondamentale pour le calcul scientifique en Python. Elle propose des objets puissants, notamment les tableaux multidimensionnels, ainsi qu'un ensemble de fonctions performantes pour les opérations arithmétiques, statistiques et algébriques. Numpy est un pilier du calcul numérique, utilisé dans les domaines du traitement du signal, de l'analyse mathématique, et des applications scientifiques et techniques [35].

Figure III-1 : Langage de programmation et les bibliothèques utiliser.

III.4-Méthodologie proposé

Cette section présente un aperçu technique détaillé des étapes séquentielles impliquées dans le prétraitement des données, la décomposition des données et l'entraînement du modèles CNN proposés, à partir d'une base de données constituée par des flux de trafique des attaques dans les WSNs. La démarche adoptée débute par une phase de prétraitement des données, comprenant notamment la normalisation. Ces opérations visent à améliorer la qualité et la diversité des données d'entrée, favorisant ainsi une meilleure capacité de généralisation des modèles.

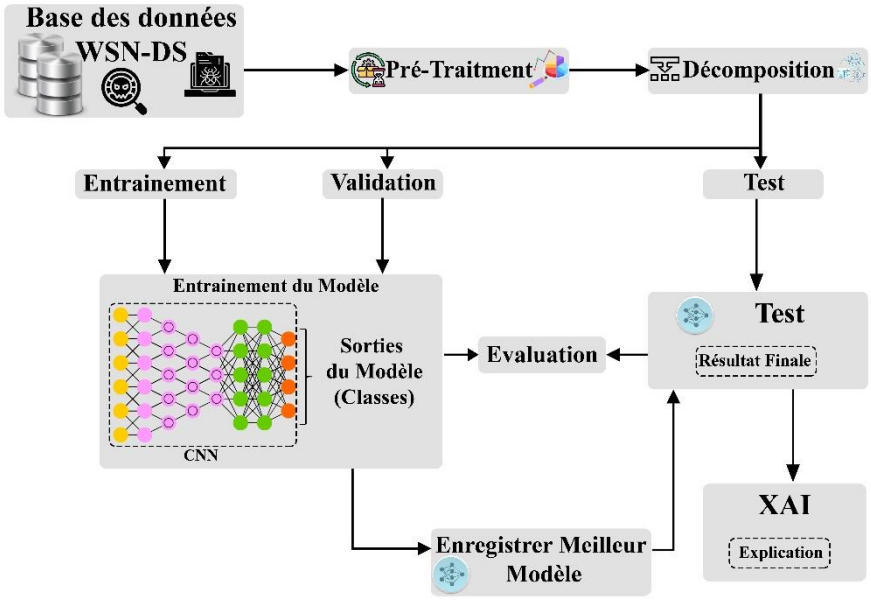


Figure III.2 : Schéma du système proposé.

Les données prétraitées sont ensuite divisées de manière rigoureuse en ensembles distincts dédiés à l'entraînement, à la validation et aux tests. Cette partition méthodique permet de garantir une évaluation fiable des performances du modèle tout en évitant les fuites de données entre les phases d'apprentissage et de test.

Les données ainsi préparées servent à entraîner le modèle proposés CNN, lui permettant d'apprendre des représentations hiérarchiques et d'extraire des caractéristiques discriminantes essentielles pour la détection des attaques dans les WSNs dans des bâtiments intelligents.

Le processus se termine par une phase finale de test et de généralisation par une partie des données qui ne sont pas exploité par le modèle durant la phase d'entrainement ou de validation pour valider et assurer la robustesse des performances du modèle proposée, suivie par une phase explicative a l'aide du XAI avec la méthode locale LIME pour expliquer les décision pris par le modèle CNN proposé.

III.4.1-Prétraitment

III.4.1.1-Normalisation

La normalisation constitue une étape fondamentale dans le prétraitement des données, visant à fournir aux modèles d'intelligence artificielle des entrées cohérentes et de haute qualité. Elle contribue à améliorer l'efficacité de l'apprentissage automatique en harmonisant les échelles des différentes caractéristiques, ce qui permet une convergence plus rapide et une réduction des risques liés à la disparition du gradient. Parmi les techniques couramment utilisées figurent la mise à l'échelle min-max et la normalisation z-score.

Dans cette étude, nous adoptons la mise à l'échelle min-max, qui consiste à transformer les valeurs d'un vecteur de caractéristiques afin qu'elles soient comprises dans un intervalle défini entre 0 et 1. Cette méthode permet d'uniformiser les données tout en préservant les relations proportionnelles entre les valeurs d'origine. La formulation mathématique de la mise à l'échelle min-max est donnée par :

$$X_n = \frac{X - \text{Min}\ (X)}{\text{Max}(X) - \text{Min}(X)}$$

Où X représente le vecteur d'origine, et min(x) et max(x) sont respectivement les valeurs minimale et maximale de ce vecteur. Le vecteur normalisé est noté X_n.

III.4.1.2-Reshape

Les données normalisées sont ensuite restructurées afin de garantir leur compatibilité avec les exigences d'entrée du notre modèle proposé, facilitant l'entraînement du modèle. Ce processus consiste à adapter la structure des données en les transformant en un vecteur bidimensionnel de forme (18,1), correspondant à la représentation des caractéristiques.

III.4.2-Décomposition des données

Une fois les données normalisées et restructurées, elles sont d'abord mélangées de manière aléatoire afin de garantir une distribution homogène des. Cette étape de permutation aléatoire est essentielle pour assurer la robustesse du processus d'apprentissage. Par la suite, les données sont divisées en trois sous-ensembles distincts, chacun a un rôle fondamental dans le développement, l'ajustement et l'évaluation du modèle. Cette décomposition suit une méthodologie bien définie, détaillée comme suit :

III.4.2.1-Ensemble d'entrainement

70 % des données sont allouées à l'apprentissage du modèle. Cet ensemble est utilisé pour ajuster les paramètres du modèle au cours du processus d'optimisation.

III.4.2.2-Ensemble de Validation

Sur les 30 % restants, 20 % sont consacrés à la validation du modèle. Ces données permettent de surveiller les performances pendant l'apprentissage et d'ajuster les hyperparamètres afin de prévenir le surapprentissage (overfitting).

III.4.2.3-Ensemble de Test et généralisation

Les 10 % finaux sont utilisés exclusivement pour tester le modèle après la phase d'entraînement. Ces données de test permet d'évaluer la capacité de généralisation du modèle sur des données encore jamais vues.

III.4.3-Architecture du Modèle Proposé

L'architecture proposée, illustrée à la Figure III.2, repose sur un modèle CNN-1D spécifiquement conçu pour la détection d'attaques dans les bâtiments intelligents à partir de l'analyse des flux de trafic réseau. Cette architecture a été méthodiquement élaborée en tenant compte de plusieurs facteurs critiques affectant ses performances, notamment la taille et la complexité de la base de données, les contraintes en ressources de calcul, ainsi que les exigences spécifiques associées à la détection d'intrusions dans les environnements intelligents.

L'objectif principal de cette architecture est de fournir un modèle à la fois performant, précis et interprétable. À cet effet, l'approche XAI, avec l'utilisation de la méthode LIME, est intégrée afin de permettre une interprétation des prédictions du modèle, renforçant ainsi la confiance dans le processus de détection et d'analyse des anomalies.

L'entrée du modèle est constituée d'un vecteur de caractéristiques de dimension (18,1), représentant les attributs extraits du trafic réseau. La première couche de traitement est une couche convolution 1D avec 64 filtres, de taille 3×1, et utilisant la fonction d'activation ReLu.

Cette fonction introduit une non-linéarité en ne conservant que les valeurs positives et en remplaçant les valeurs négatives par zéro. Un remplissage par zéros (padding) est appliqué pour préserver les dimensions d'entrée, ce qui permet une extraction efficace des caractéristiques en bordure de séquence. Cette couche est suivie par une couche de normalisation par lot (Batch Normalization), utilisée pour stabiliser et accélérer l'apprentissage, tout en améliorant la qualité de l'extraction des caractéristiques. Une deuxième couche convolution 1D est ensuite appliquée, comprenant 32 filtres de taille 3×1, avec une fonction d'activation ReLu. Elle est suivie par une couche de max pooling 1D de taille 2×1, qui réduit la dimension des cartes de caractéristiques en conservant la valeur maximale de chaque segment de deux éléments, ce qui permet de diminuer la complexité computationnelle tout en conservant l'information discriminante.

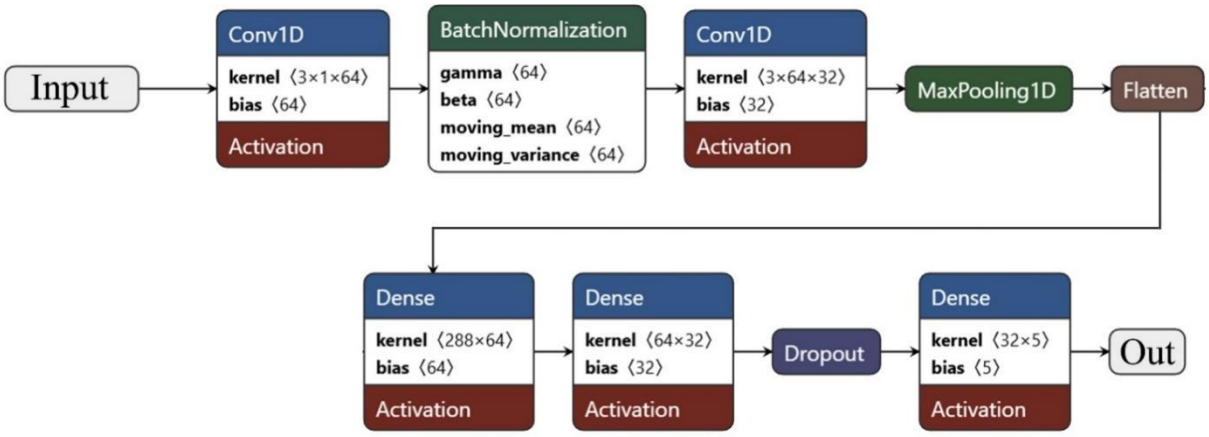


Figure III.3 : Architecture du modèle CNN-1D proposée.

Après l'extraction des caractéristiques, une couche de Flatten est utilisée pour transformer les cartes de caractéristiques multidimensionnelles en un vecteur unidimensionnel compatible avec les couches entièrement connectées. Ce vecteur est ensuite traité par deux couches denses comprenant respectivement 64 et 32 neurones, activées par la fonction d'activation ReLu. Une couche de régularisation Dropout, avec un taux de 0.4, est insérée après ces couches pour atténuer le surapprentissage (overfitting) durant la phase d'entraînement. La couche de sortie est constituée de 5 neurones, chacun correspondant à l'une des classes étudier. Une fonction d'activation Softmax est utilisée pour générer des probabilités normalisées associées à chaque classe.

III.5-Expérimentation et discussion des résultats

III.5.1-Description de la base des données

Dans cette étude, nous avons utilisé la base de données proposée par Almomani et al. [1]. Elle a été générée à l'aide du simulateur NS-2 en simulant un réseau de capteurs sans fil (WSN) basé sur le protocole LEACH [47]. La base des données comprend 374 661 enregistrements étiquetés, représentant à la fois du trafic normal et malveillant, incluant quatre types d'attaques DoS : blackhole, grayhole, flooding et scheduling. Chaque enregistrement est caractérisé par 19 attributs, tels que l'identifiant du nœud, le nombre de paquets envoyés et reçus, ainsi que des informations topologiques du réseau.

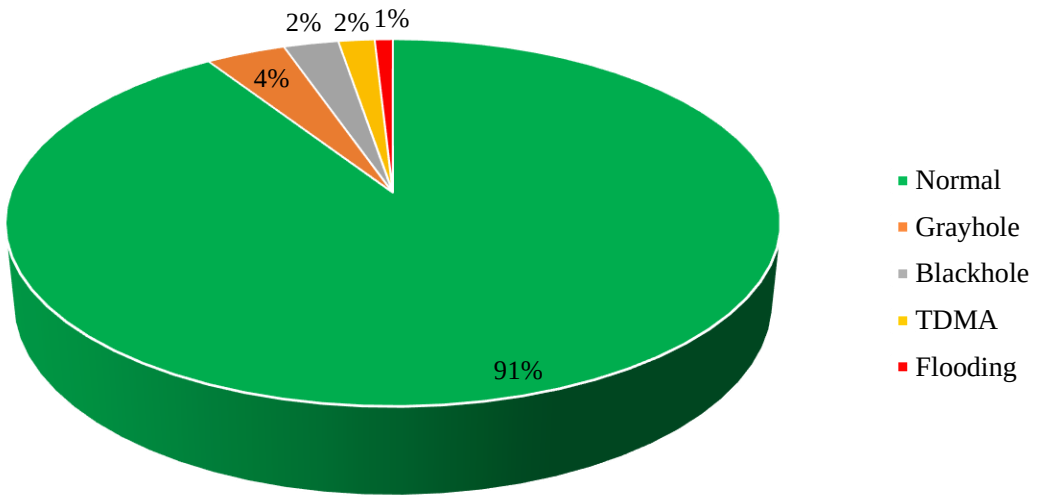


Figure III.4 : Distribution de la base des données.

III.5.2-Les Métriques d'évaluation

L'évaluation de l'efficacité du modèle proposé repose sur l'utilisation de métriques de performance standards, à savoir : l'accuracy, la précision, le recall et le F1-score, définies formellement par les équations suivantes. Ces indicateurs fournissent une évaluation quantitative et exhaustive des performances prédictives du modèle, en mesurant l'accord entre les étiquettes prédites et les annotations de vérité terrain.

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

$$Precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 \times \frac{Precision \times recall}{Precision + recall}$$

Où :

- TP : Nombre d'instances positives correctement classées par le modèle.
- TN : Nombre d'instances négatives correctement identifiées.
- FP : Nombre d'instances négatives incorrectement classées comme positives.
- FN : Nombre d'instances positives incorrectement classées comme négatives.

III.5.3-Discussion des Résultats

Le Tableau III.1 résume les performances du système proposé pour la détection des attaques, qui exploite les modèles ANN, LSTM, CNN, et CNN-1D.

Le modèle CNN-1D a obtenu la meilleure précision, atteignant 99.22% sur l'ensemble des métriques de performance. Les modèles ANN, LSTM, et CNN ont également donné de bons résultats, avec des précisions respectives de 91.31%, 97.70%, et 98.15%.

Ces résultats indiquent que le modèle CNN-1D indique une efficacité et robustesse du système de détection proposé, comparativement à des autres modèles.

Tableau III.1 : Performance des modèles développé

Modèle	Accuracy (%)	Précision (%)	Recall (%)	F1-Score (%)
ANN	91.31	91.29	91.27	91.27
LSTM	97.70	97.70	97.67	97.68
CNN	98.15	98.14	98.11	98.12
CNN-1D	**99.22**	**99.22**	**99.20**	**99.21**

Le Tableau III.2 présente les performances du modèle CNN-1D sur les trois ensembles de données : entraînement, validation, et test. Les résultats obtenus démontrent la stabilité et la robustesse du modèle à travers les différentes phases de l'expérimentation.

Lors de l'apprentissage, le modèle a atteint une accuracy de 99.35 %, accompagnée de valeurs identiques pour la précision, le recall et le F1-score. Cette homogénéité entre les métriques suggère une convergence stable du modèle, sans indication de surapprentissage. La valeur de perte relativement faible 0,019 reflète une optimisation efficace du processus d'apprentissage.

Sur l'ensemble de validation, les performances restent élevées avec une accuracy de 99.22 % et des valeurs de précision, recall et F1-score très proches, témoignant de la capacité du modèle à généraliser au-delà des données d'entraînement.

Enfin, l'évaluation sur l'ensemble de test confirme cette tendance, avec des performances identiques à celles obtenues lors de la validation. Cette constance indique une excellente capacité de généralisation du modèle face à des données inédites.

Tableau III.2 : Performance du modèle CNN-1D sur les ensembles des données

	Accuracy (%)	**Précision (%)**	**Recall (%)**	**F1-Score (%)**	**Perte**
Entrainement	99.35	99.35	99.35	99.35	0.019
Validation	99.22	99.22	99.20	99.21	0.024
Test et Généralisation	99.22	99.22	99.20	99.21	0.024

Les résultats obtenus valident l'efficacité du modèle CNN-1D proposée pour la détection d'anomalies dans les réseaux de bâtiments intelligents. Le modèle se distingue par une précision élevée, une stabilité inter-ensembles, et une faible perte, le rendant adapté à des scénarios réels et contraints**.**

III.5.3.1-Courbe de précision et des pertes

La Figure III.3 illustre les courbes d'évolution de la accuracy et de perte pour les ensembles d'entraînement et de validation en fonction du nombre d'époques. On observe que la précision augmente proportionnellement de 90% à 99.22% avec le nombre d'époques pour les deux ensembles, ce qui indique une amélioration progressive de la capacité du modèle à classer correctement les instances. Inversement, les valeurs de perte diminuent de 0.35 à 0.024 au fil des époques, traduisant une meilleure concordance entre les prédictions du modèle et les étiquettes réelles. Ces résultats témoignent d'un apprentissage efficace du modèle, accompagné d'un ajustement adaptatif des paramètres, conduisant à une réduction des erreurs et à une amélioration continue des performances.

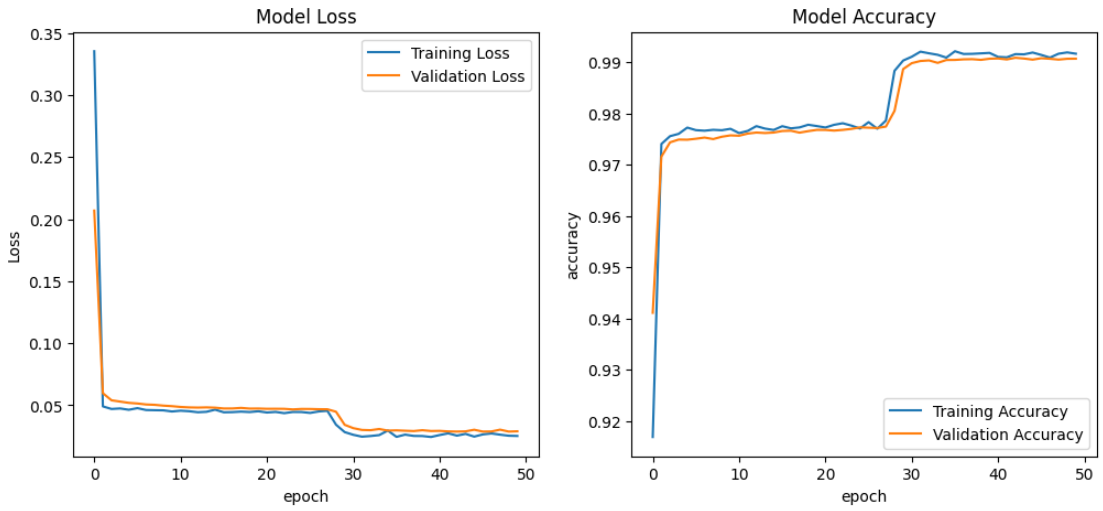


Figure III.5 : Graphe de précision et des pertes.

III.5.3.2-Courbe de ROC

Cette étude utilise les courbes ROC (Receiver Operating Characteristic) ainsi que l'aire sous la courbe ROC (AUC) comme principales métriques d'évaluation. La courbe ROC trace le taux de vrais positifs (TPR) en fonction du taux de faux positifs (FPR), offrant ainsi une représentation visuelle des performances diagnostiques du modèle selon différents seuils de décision. Dans le cas d'un modèle idéal, la courbe ROC se rapproche du coin supérieur gauche (TPR = 1, FPR = 0). L'AUC permet de quantifier ces performances, avec des valeurs allant de 0.5 (aléatoire) à 1 (classification parfaite).

La Figure III.4 illustre les courbes ROC obtenues, mettant en évidence les solides performances du modèle sur l'ensemble des classes étudiées. Les courbes s'orientent clairement vers le coin supérieur gauche, indiquant un compromis optimal entre sensibilité et spécificité. Dans un contexte de classification multi-classe, la courbe ROC et l'AUC s'avèrent être des outils analytiques particulièrement pertinents. Comme le montre la Figure III.4, les valeurs d'AUC obtenues pour les quatre classes varient entre 0.98 et 1.00, attestant de la robustesse globale du modèle et de son excellente capacité à distinguer avec précision les différentes catégories.

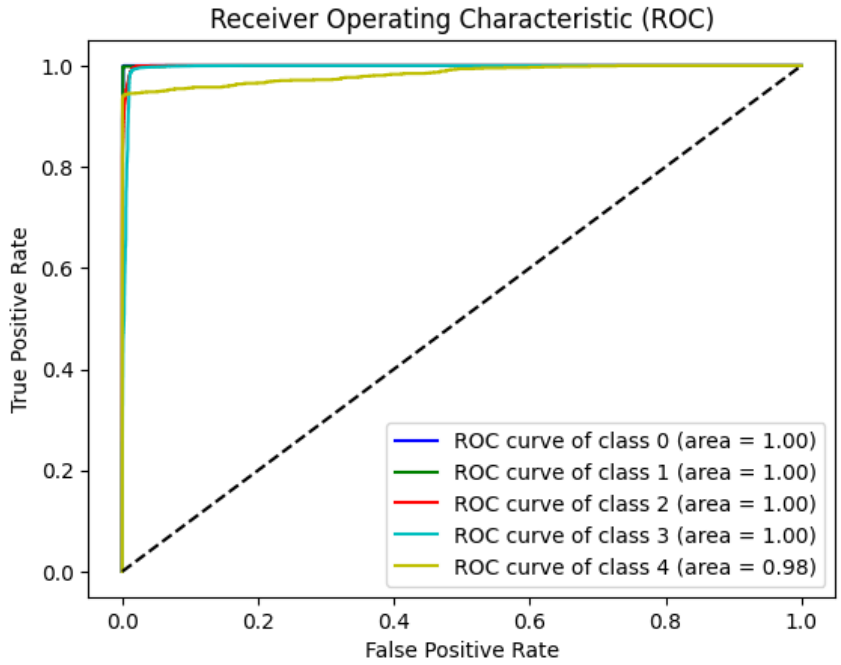


Figure III.6 : Graphe de ROC.

III.5.3.3-Rapport de Classification

Le Tableau III.3 présente un rapport de classification détaillé, fournissant une évaluation approfondie des performances du modèle d'apprentissage profond proposé à travers des métriques clés : précision, rappel, F1-score et support (nombre d'échantillons par classe). L'analyse met en évidence des performances remarquables :

- **Précision :** Le modèle atteint une précision de 100% pour les classes Normal et Flooding, 99% pour Blackhole, 93% pour TDMA, et 91% pour Grayhole, traduisant une faible proportion de faux positifs.
- **Recall :** Les scores de rappel sont tout aussi élevés, avec 100% pour Normal, 96% pour Blackhole, 94% pour Grayhole, 97% pour Flooding, et 94% pour TDMA, indiquant une forte capacité de détection des instances pertinentes.
- **F1-Score :** Les scores F1 obtenus 100% pour Normal, 97% pour Blackhole, 93% pour Grayhole, 98% pour Flooding, et 96% pour TDMA témoignent d'un bon équilibre entre précision et rappel sur l'ensemble des classes.

Tableau III.3 : Rapport de classification.

	Précision	Recall	F1-Score	Support
Normal	1.00	1.00	1.00	68096
Blackhole	0.99	0.96	0.97	2045

Grayhole	0.91	0.94	0.93	2796
Flooding	1.00	0.97	0.98	766
TDMA	0.93	0.98	0.96	1230
Accuracy			0.99	74933
Macro avg	0.97	0.97	0.97	74933
Weighted avg	0.99	0.99	0.99	74933

Ces résultats soulignent la robustesse et la fiabilité du modèle CNN-1D proposé face aux données testées. Sa capacité à maintenir des performances constantes, même dans des conditions complexes ou variables, indique une excellente généralisation sur des données non vues. Cela démontre que le modèle a correctement appris les caractéristiques discriminantes de chaque classe. Par conséquent, cette architecture est particulièrement bien adaptée à des scénarios réels de détection d'attaques dans les bâtiments intelligents, où la précision, la stabilité et la sensibilité sont essentielles à la protection des systèmes critiques. Ces résultats démontrent que peut être qu'il y a une similarité entre les classes.

III.5.3.4-Matrice de confusion

La matrice de confusion, présentée à la Figure III.5, constitue un outil analytique fondamental pour évaluer la performance du modèle CNN-1D en classification multi-classes. Elle permet d'analyser la correspondance entre les prédictions du modèle et les étiquettes réelles, en distinguant les classifications correctes, situées sur la diagonale principale, des erreurs de classification, représentés par les éléments hors diagonale.

L'analyse de cette matrice représente une performance globale satisfaisante, avec des taux de classification correcte atteignant 99.24% pour la classe normal, 99.86% pour blackhole, 91.10% pour grayhole, 99,66% pour flooding, et 93,42% pour TDMA. Toutefois, des erreurs subsistent, notamment une confusion entre certaines classes proches comme, 0.76% des instances de la classe normal ont été mal classées comme grayhole, et 0.14% des instances de la classe blackhole comme flooding. La classe grayhole présente des erreurs, avec respectivement 3.01%, 5.26% et 0.62% de ses instances incorrectement attribuées aux classes normal, flooding et TDMA. De plus, la classe flooding a été confondue à une mal classifications de 0.11% avec blackhole, 0.22% avec grayhole et 0.01% avec TDMA. Enfin, la classe TDMA a des erreurs de 0,15 % comme grayhole et 6,42% comme flooding.

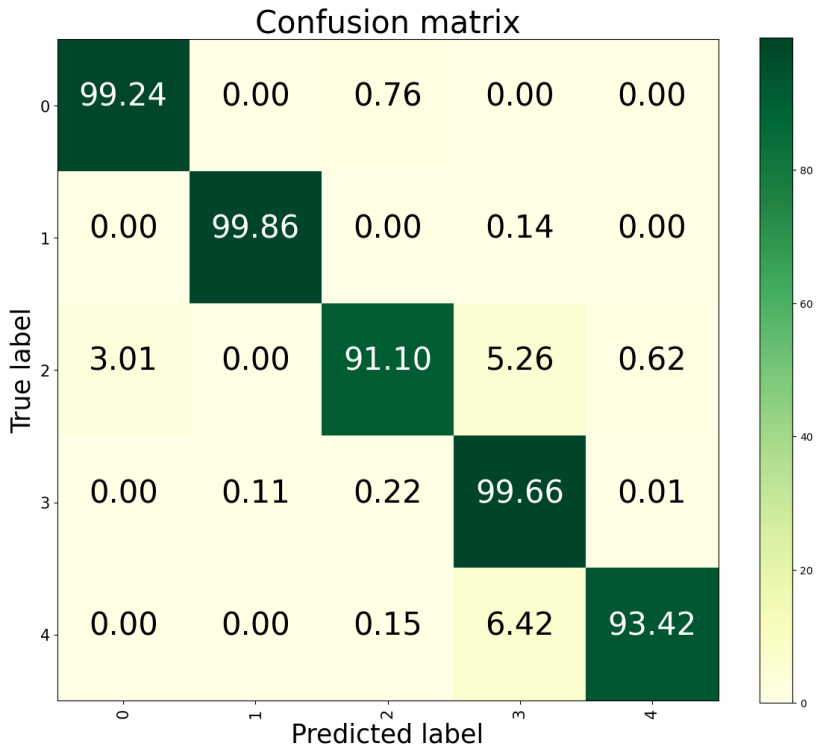


Figure III.7 : Matrice de confusion de validation.

Ces résultats indiquent que, bien que le modèle démontre une forte capacité de généralisation, certaines classes présentant des similarités dans leurs caractéristiques. Cela suggère des pistes d'amélioration potentielles, notamment en affinant les représentations discriminantes ou en enrichissant l'apprentissage par des techniques avancées telles que l'attention ou l'augmentation ciblée des données. Ainsi, le modèle pourrait encore améliorer sa capacité à distinguer avec précision entre des comportements d'attaque subtilement différents dans les environnements intelligents.

III.5.3.4-Généralisation et test

Notre évaluation comprend une phase rigoureuse de test et de généralisation visant à valider la robustesse et la fiabilité du modèle d'apprentissage profond CNN-1D proposé pour la détection d'attaques dans les environnements de bâtiments intelligents. Après l'entraînement initial, le modèle a été soumis à des tests sur des sous ensembles distincts de validation et de généralisation, afin d'évaluer la stabilité de ses performances sur des données non vues.

Les résultats de test et de généralisation dans le Tableau III.2, montrent un haut niveau de cohérence entre les différentes phases d'évaluation. Plus précisément, le modèle atteint des performances très elevées, avec une accuracy de 99.22%, une précision de 99.22%, un recall de 99.20%, un f1-score de 99.21% et une perte de 0.024. Ces performances confirmant la capacité du modèle à maintenir une prédiction fiable sur des instances inédites. Ces résultats

soulignent la capacité de généralisation du modèle CNN-1D, qui ne montre aucun signe de surapprentissage et maintient des performances constantes indépendamment du sous ensemble évalué. Cette stabilité est essentielle pour une utilisation en contexte réel, où les conditions de déploiement diffèrent souvent de celles rencontrées lors de l'entraînement.

Par ailleurs, l'analyse de la matrice de confusion de test (Figure III.6) vient renforcer cette observation et les résultats obtenus. Le modèle présente une très faible confusion entre les classes, même celles aux comportements similaires, ce qui illustre sa capacité à discriminer finement les différents types d'attaques. En somme, ces résultats confirment que notre approche basée sur le CNN-1D constitue une solution fiable, précise et généralisable pour la détection d'intrusions dans les environnements intelligents.

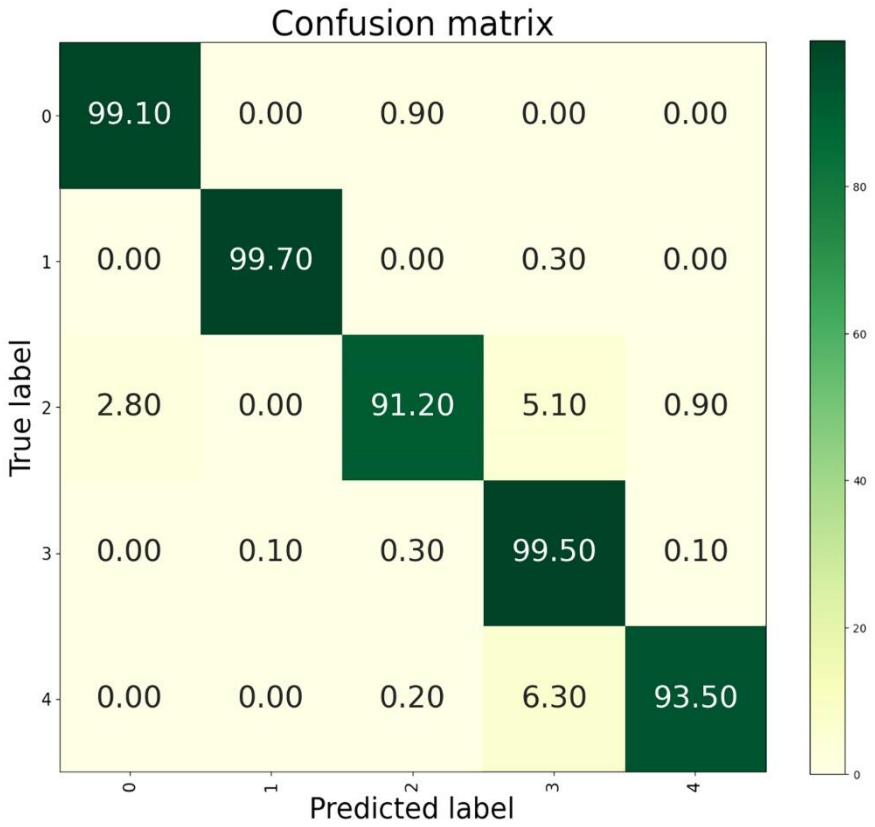


Figure III.8 : Matrice de confusion de test.

III.5.4-XAI

Pour compléter l'évaluation de notre modèle CNN-1D, nous avons exploré sa capacité d'explicabilité locale à l'aide de la méthode LIME. Cette technique vise à fournir une compréhension interprétable des décisions du modèle en identifiant les attributs les plus influents pour chaque prédiction, à l'échelle de l'instance.

La Figure III.7 représente les explications générées pour cinq classes cibles correspondant aux différentes attaques détectées dans notre ensemble de données : normal (classe 0), blackhole (classe 1), grayhole (classe 2), flooding (classe 3), et TDMA (classe 4). Pour chaque classe, la technique LIME génère une importance locale des attributs, visualisée sous forme de barres vertes (pour la classe ciblé) et rouges (pour d'autre classes).

Les résultats montrent que certaines variables se révèlent récurrentes et discriminantes dans les décisions du modèle. Comme illustré, les attributs DATA_S, dist_CH_to_BS, Rank et SCH_R apparaissent comme des facteurs dominants pour les classes 0 et 1, indiquant leur forte contribution à la différenciation entre comportements normaux et attaques de type blackhole. Pour la classe grayhole (classe 2 (c)), la variabilité des attributs clés reflète la complexité intrinsèque de cette attaque, caractérisée par des comportements intermittents. Dans le cas de la classe flooding (classe 3 (d)), les attributs Rank, dist_CH_to_BS, et ADV_S sont les plus déterminants, montrant que le modèle se base sur des patterns topologiques et temporels pour détecter cette attaque. Enfin, pour la classe TDMA (classe 4 (e)), les prédictions sont principalement influencées par les valeurs des champs DATA_S, send_code, Expanded Energy, et JOIN_S, ce qui suggère une capacité du modèle à capter des anomalies dans les communications périodiques.

Globalement, cette analyse par LIME confirme que le modèle apprend des représentations significatives et qu'il base ses décisions sur des caractéristiques explicables et cohérentes avec la nature des attaques. Cela renforce la confiance dans le modèle, notamment pour des applications critiques où l'interprétabilité est indispensable, telles que la sécurité des bâtiments intelligents. De plus, l'utilisation de LIME peut également guider l'amélioration future du modèle en identifiant des attributs sur ou sous-exploités.

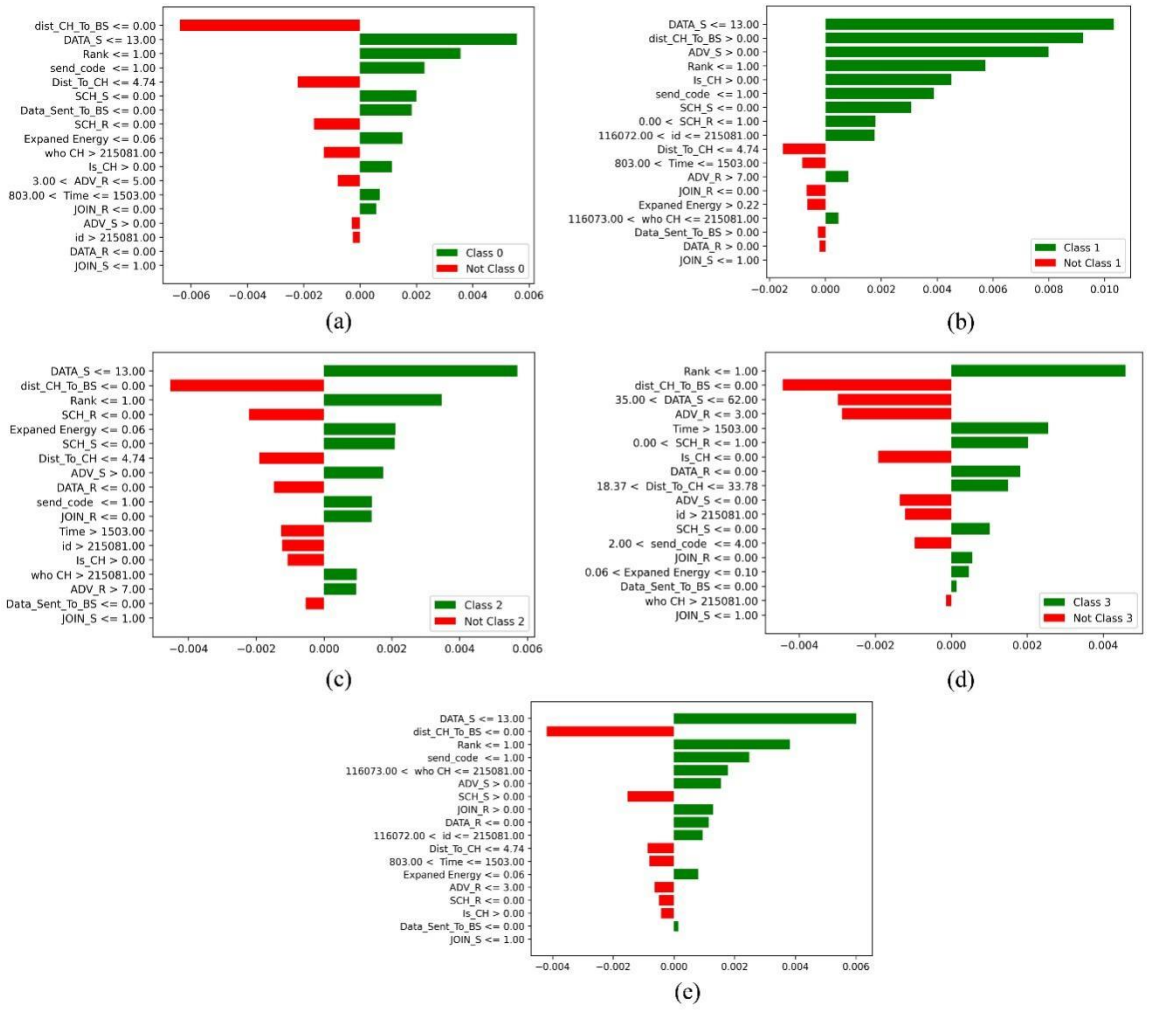


Figure III.9 : Explication des caractéristiques par classe selon la méthode LIME.

III.5.5-Comparison

Le Tableau III.4 présente une évaluation comparative des performances entre notre modèle proposé et plusieurs approches récentes de l'état de l'art, appliquées sur la même base des données WSN-DS. La comparaison repose principalement sur l'accuracy de classification et l'application de XAI, mettant en évidence la robustesse et l'efficacité de l'architecture proposée. Le modèle CNN-1D développé dans cette étude a atteint une accuracy remarquable de 99.22 % avec une explication à l'aide de la méthode LIME de XAI, constituant une amélioration significative par rapport aux d'autre méthodes existantes.

Plus précisément, le modèle surpasse aussi bien les autres modèles utiliser tel que les approches d'apprentissage profond et automatique, notamment : Naive Bayes (NB) avec 97.25%, SVM avec un taux de 97.11%, CNN avec une accuracy de 98.85%, GNB avec une accuracy de 98.00%, et CNN-RNN avec 96.84%.Ces résultats témoignent des capacités supérieures de classification du modèle proposé, et confirment la pertinence des améliorations architecturales

apportées pour traiter la complexité de la détection d'intrusions dans les réseaux de capteurs sans fil.

Tableau III.4 : Tableau de comparaison avec des travaux récente

Travail	Année	Prétraitment	Modèle	Accuracy (%)	XAI	Base des données
[48]	2021	/	SVM	97.11	Non	WSN-DS
[30]	2016	/	ANN	98.52	Non	WSN-DS
[31]	2022	/	NB	97.25	Non	WSN-DS
[33]	2023	/	CNN-RNN	96.84	Non	WSN-DS
[33]	2023	/	CNN	98.85	Non	WSN-DS
[34]	2024	PCA	GNB	98.00	Non	WSN-DS
Notre	2025	Normalisation	ANN	91.31	Oui	WSN-DS
Notre	2025	Normalisation	LSTM	97.70	Oui	WSN-DS
Notre	**2025**	**Normalisation**	**CNN-1D**	**99.22**	**Oui**	**WSN-DS**

Autrement, l'intégration de techniques d'explicabilité XAI a l'aide de la méthode LIME renforce la compréhension et la confiance dans les prédictions du modèle, en offrant une interprétation des décisions prises. Cette approche constitue ainsi une avancée significative dans la conception de systèmes intelligents, à la fois performants et transparents, pour des applications critiques en cybersécurité.

III.6-Conclusion

Dans ce chapitre nous avons démontrer l'efficacité du modèle CNN-1D proposé pour la détection d'attaques dans les réseaux capteurs des bâtiments intelligents. Les résultats expérimentaux ont montré une précision élevée et une stabilité remarquable sur l'ensemble des phases d'entraînement, de validation et de test. De plus, l'intégration de la méthode LIME a apporté une dimension explicative aux décisions du modèle, renforçant ainsi sa transparence et sa fiabilité. L'ensemble de ces éléments confirme la pertinence de l'approche retenue et ouvre des perspectives intéressantes pour des applications réelles en cybersécurité intelligente.

Conclusion Générale

En conclusion, notre étude a mis en évidence la pertinence et l'efficacité des approches basées sur l'intelligence artificielle pour renforcer la sécurité des bâtiments intelligents. En utilisant les réseaux CNN, nous avons conçu un système capable de détecter les attaques DoS avec une grande précision, même dans des environnements contraints en ressources comme les réseaux de capteurs sans fil.

L'approche proposée a montré de hautes performances de classification, confirmées par des résultats expérimentaux solides. Grâce à un traitement automatisé et intelligent du trafic réseau, le système parvient à identifier rapidement les comportements malveillants, ce qui permet une réaction proactive face aux intrusions. Cette solution ouvre ainsi la voie à une cybersécurité plus autonome, évolutive et efficace dans les infrastructures connectées.

Au-delà des résultats techniques, notre projet souligne l'importance croissante des technologies d'IA dans les systèmes critiques. L'intégration de modèles intelligents dans les bâtiments intelligents ne se limite plus au confort ou à la gestion énergétique, mais s'étend désormais à la sécurité et à la résilience face aux attaques. En mettant l'accent sur la détection d'attaques DoS, nous avons ciblé l'un des risques les plus répandus et les plus destructeurs pour les réseaux IoT.

Notre démarche a également montré que l'interprétabilité des modèles est essentielle dans les contextes sensibles. L'utilisation potentielle de techniques d'intelligence artificielle explicable (XAI) constitue un axe prometteur pour garantir la transparence des systèmes et renforcer la confiance des utilisateurs et des opérateurs.

Enfin, notre travail démontre que les architectures CNN a l'aide du XAI constituent une solution prometteuse pour la cybersécurité des environnements intelligents. À l'avenir, ce type de système pourrait être intégré à plus grande échelle dans les bâtiments connectés, les infrastructures critiques ou les villes intelligentes, contribuant à une protection plus fiable, autonome et durable contre les menaces numériques.

Bibliographie

[1] Baccelli, E. (2021). Internet des Objets: Défis sociétaux et domaines de recherche scientifique pour l'Internet des Objets (IoT).

[2] Huseien, G. F., & Shah, K. W. (2022). A review on 5G technology for smart energy management and smart buildings in Singapore. Energy and AI, 7, 100116.

[3] Berkani, M. R. A., Chouchane, A., Himeur, Y., Ouamane, A., Miniaoui, S., Atalla, S., ... & Al-Ahmad, H. (2025). Advances in Federated Learning: Applications and Challenges in Smart Building Environments and Beyond. Computers, 14(4), 124. Doi : 10.3390/computers14040124.

[4] Ejidike, C. C., & Mewomo, M. C. (2023). Benefits of adopting smart building technologies in building construction of developing countries: Review of literature. SN Applied Sciences, 5(2), 52.

[5] Sándor, B., & Rajnai, Z. (2023, September). Smart Building IoT Cybersecurity: A Review of Threats and Mitigation Technique. In 2023 IEEE 21st Jubilee International Symposium on Intelligent Systems and Informatics (SISY) (pp. 000321-000326). IEEE.

[6] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. Computer networks, 38(4), 393-422.

[7] Yick, J., Mukherjee, B., & Ghosal, D. (2008). Wireless sensor network survey. Computer networks, 52(12), 2292-2330.

[8] Hart, J. K., & Martinez, K. (2006). Environmental sensor networks: A revolution in the earth system science?. Earth-Science Reviews, 78(3-4), 177-191.

[9] Arampatzis, T., Lygeros, J., & Manesis, S. (2005, June). A survey of applications of wireless sensors and wireless sensor networks. In Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation Intelligent Control, 2005. (pp. 719-724). IEEE.

[10] Pantelopoulos, A., & Bourbakis, N. G. (2009). A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(1), 1-12.

[11] Ullah, F., Mehmood, T., Habib, M., Ibrahim, M., & Zulfikar, S. (2011, January). SPINS: security protocols for sensor networks. In Proceedings of the International Conference on Communication Engineering and Networks (IPCSIT'11).

[12] Younis, M., & Nadeem, T. (2004). Energy efficient MAC protocols for wireless sensor networks. Book chapter in wireless ad hoc and sensor networks. Dordrecht: Kluwer.

[13] Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R., & Anderson, J. (2002, September). Wireless sensor networks for habitat monitoring. In Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications (pp. 88-97).

[14] Xu, W., Trappe, W., Zhang, Y., & Wood, T. (2005, May). The feasibility of launching and detecting jamming attacks in wireless networks. In Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing (pp. 46-57).

[15] Wood, A. D., & Stankovic, J. A. (2002). Denial of service in sensor networks. computer, 35(10), 54-62.

[16] Karlof and Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures," Ad Hoc Networks, 2003.

[17] Mihajlov, B., & Bogdanoski, M. (2014). Analysis of the WSN MAC Protocols under Jamming DoS Attack. Int. J. Netw. Secur., 16(4), 304-312.

[18] Lakshmi, H. N., Anand, S., & Sinha, S. (2019). Flooding attack in wireless sensor network-analysis and prevention. International Journal of Engineering and Advanced Technology, 8(5), 1792-1796.

[19] Sharma, H. S., Singh, M. M., & Sarkar, A. (2023, January). Machine learning-based DoS attack detection techniques in wireless sensor network: A review. In Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2 (pp. 583-591). Singapore: Springer Nature Singapore.

[20] Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. IEEE Communications Surveys & Tutorials, 20(4), 2923-2960.

[21] Yu, L., Qin, S., Zhang, M., Shen, C., Jiang, T., & Guan, X. (2021). A review of deep reinforcement learning for smart building energy management. IEEE Internet of Things Journal, 8(15), 12046-12063.

[22] Berkani, M. R. A., Chouchane, A., Himeur, Y., Abdennebi, A., Sagiroglu, S., & Amira, A. (2024, December). Federated Learning-Based Intelligent Indoor Smoke and Fire Detection System for Smart Buildings. In 2024 International Conference on Telecommunications and Intelligent Systems (ICTIS) (pp. 1-7). IEEE.

[23] Chen, Z., Xiao, F., & Guo, F. (2023). Similarity learning-based fault detection and diagnosis in building HVAC systems with limited labeled data. Renewable and Sustainable Energy Reviews, 185, 113612.

[24] Sayed, A. N., Himeur, Y., & Bensaali, F. (2022). Deep and transfer learning for building occupancy detection: A review and comparative analysis. Engineering Applications of Artificial Intelligence, 115, 105254.

[25] Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761-768.

[26] Aljawarneh, S., Aldwairi, M., & Yassein, M. B. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. Journal of Computational Science, 25, 152-160.

[27] Salmi, S., & Oughdir, L. (2023). Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network. Journal of Big Data, 10(1), 17.

[28] Davahli, A., Shamsi, M., & Abaei, G. (2020). A lightweight Anomaly detection model using SVM for WSNs in IoT through a hybrid feature selection algorithm based on GA and GWO. Journal of Computing and Security, 7(1), 63-79.

[29] Singh, N., Virmani, D., & Gao, X. Z. (2020). A fuzzy logic-based method to avert intrusions in wireless sensor networks using WSN-DS dataset. International Journal of Computational Intelligence and Applications, 19(03), 2050018.

[30] Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: a dataset for intrusion detection systems in wireless sensor networks. Journal of Sensors, 2016(1), 4731953.

[31] Ismail, S., Dawoud, D., & Reza, H. (2022, January). A lightweight multilayer machine learning detection system for cyber-attacks in WSN. In 2022 IEEE 12th annual computing and communication workshop and conference (CCWC) (pp. 0481-0486). IEEE.

[32] Alsulaiman, L., & Al-Ahmadi, S. (2021). Performance evaluation of machine learning techniques for DOS detection in wireless sensor network. arXiv preprint arXiv:2104.01963.

[33] Salmi, S., & Oughdir, L. (2023). Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network. Journal of Big Data, 10(1), 17.

[34] Saleh, H. M., Marouane, H., & Fakhfakh, A. (2024). Stochastic gradient descent intrusions detection for wireless sensor network attack detection system using machine learning. IEEE Access, 12, 3825-3836.

[35] Berkani, Mohamed Rafik Aymene & Zahra, Achir & Bengherbia, Billel. (2022). Etude et implémentation d'un système connecté intelligent pour la surveillance de l'état de santé des patients souffrant d'arythmie cardiaque. 10.13140/RG.2.2.10581.64489.

[36] Khadidja, SOLTANI & Berkani, Mohamed Rafik Aymene & Djazia, Bendani & Amel, BENCHELLAL & Fatiha, BECHIRI. (2023). Etude d'un Système Intelligent Pour Suivre les Patients Atteints de Maladies Cardiaques. 10.13140/RG.2.2.31658.02245.

[37] Huang, S. C., & Le, T. H. (2021). Principles and labs for deep learning. Academic Press.

[38] Zhang, X.-D., Machine Learning, in A Matrix Algebra Approach to Artificial Intelligence, X.- D. Zhang, Editor 2020, Springer Singapore: Singapore. p. 223-440.

[39] Mohammed, M., Khan, M. B., & Bashier, E. B. M. Machine learning: algorithms and applications. 2016.

[40] A.Chouchane, M.Berkani, O.Alioune, F.Ghernouati, "Etude d'un système intelligent pour la classification des cancers en utilisant le séquençage des ARN", 2023.

[41] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[42] Chauhan, N. K., & Singh, K. (2018, September). A review on conventional machine learning vs deep learning. In 2018 International conference on computing, power and communication technologies (GUCON) (pp. 347-352). IEEE.

[43] Wu, Y. C., & Feng, J. W. (2018). Development and application of artificial neural network. Wireless Personal Communications, 102, 1645-1656.

[44] Nour Houria, Fekir & Farah, Mokhbat & Mohamed Rafik Aymene, Berkani & Mohamed, Rebiai. (2024). Etude d'un Système Intelligent pour la Détections des Maladies Cardiaques basé sur des signaux ECG.

[45] Ilhem, Besseri & Soumia, Tobbal & Mohamed Rafik Aymene, Berkani & Mohamed, Rebiai. (2024). Extraction des Caractéristiques Fréquentiels pour la Classification des Maladies Cardiaques par Phonocardiogrammes.

[46] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.

[47] Almomani, I., Al-Kasasbeh, B., & Al-Akhras, M. (2016). WSN-DS: a dataset for intrusion detection systems in wireless sensor networks. Journal of Sensors, 2016(1), 4731953.

[48] Alsulaiman, L., & Al-Ahmadi, S. (2021). Performance evaluation of machine learning techniques for DOS detection in wireless sensor network. arXiv preprint arXiv:2104.01963.