

الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة سعيدة - د. الطاهر مولاي -

Université Saïda - Dr. Tahar Moulay -
Faculté des Mathématiques, Informatique et Télécommunications



MEMOIRE

Présenté pour l'obtention du Diplôme de MASTER en Télécommunications

Spécialité : Réseaux et Télécommunications

Par : Mr. AMMOUR El-Haithem

Mr. ADDADI Oussama

Algorithmes K-NN: Applications aux Communications Optique

Soutenu, le 17/06/ 2026, devant le jury composé de :

Dr. A. Harzellaoui	MCB	Président
Dr. E. Ouis	MCB	Encadreur
Dr. S. Belhadj	MCB	Examineur

2025/2026

REMERCIEMENTS

Avant tout, nous remercions Dieu le Tout-Puissant de nous avoir donné la force, la patience et le courage nécessaires pour mener à bien ce modeste travail.

Nous tenons à exprimer notre profonde gratitude et nos remerciements les plus sincères à notre encadrante, **Dr. OUIS Esma**, nous la remercions chaleureusement pour son soutien indéfectible, sa grande disponibilité, ses précieux conseils et ses corrections rigoureuses tout au long de ce projet. Sa rigueur scientifique, sa bienveillance et ses orientations constantes ont été pour nous une véritable source de motivation et d'apprentissage.

Nos remerciements s'adressent également à **Pr. GARADI Ahmed**, nous lui exprimons notre vive reconnaissance pour son aide précieuse, son implication remarquable et le temps qu'il nous a consacré. Ses remarques constructives, son accompagnement sur le terrain et sa manière de nous guider face à nos erreurs ont été essentiels à l'accomplissement de ce mémoire.

Nous tenons aussi à remercier chaleureusement les membres du jury qui ont accepté d'évaluer notre travail et de l'honorer par leur jugement critique.

Enfin, ce parcours n'aurait jamais été possible sans le soutien inconditionnel de nos chères familles. Un grand merci à nos parents pour leurs sacrifices, leurs prières quotidiennes, leur amour et leur encouragement continu. Sans eux, nous ne serions pas là aujourd'hui. Que ce travail soit le fruit de leur dévouement.

À tous ceux qui ont contribué de près ou de loin à la réalisation de ce projet, merci infiniment.

DÉDICACES

Aucun mot ne saurait exprimer l'immensité de l'amour, du respect et de la gratitude que nous portons à nos très chers parents. Vous avez été les piliers inébranlables de notre réussite. À **nos chers parents**, qui nous ont appris la valeur du sacrifice et de la persévérance, et à nos chères mères, sources de tendresse et de prières pures qui nous ont portés tout au long de nos doutes. Ce travail est le fruit de votre dévouement quotidien. Que Dieu vous accorde santé, longue vie et bonheur.

À **nos chères familles**, frères, sœurs et proches, qui ont partagé nos moments de joie comme de stress. Votre présence bienveillante, vos encouragements constants et votre soutien indéfectible ont créé autour de nous l'environnement serein et motivant nécessaire pour mener à bien ce long parcours universitaire. Vous êtes notre fierté.

À **nos fidèles amis** et compagnons de route, avec qui nous avons partagé des moments inoubliables, des nuits de révision et des défis. Votre solidarité, votre complicité et votre humour ont rendu ce voyage académique plus léger et plus beau. Merci d'avoir toujours été là, dans les réussites comme dans les moments difficiles.

Haïthem et Oussama

RÉSUMÉ

Les communications optiques en espace libre (FSO) sont une technologie de base des réseaux de communication sans fil de nouvelle génération, fournissant des taux élevés de transfert de données similaires à la fibre optique sans avoir besoin de licences de spectre. Cependant, son état de préparation est fortement affecté par les perturbations des canaux atmosphériques et les conditions climatiques défavorables.

Dans ce travail, une approche prédictive intelligente basée sur l'apprentissage automatique a été proposée afin de diagnostiquer et de classer le statut des canaux FSO. Pour y parvenir, la chaîne de transmission du FSO a été modélisée et simulée à l'aide de l'environnement OptiSystem dans le but de créer une base de données physique qui reflète différents scénarios climatiques. Le facteur de qualité (Q-Factor) a été extrait comme principal facteur prédictif. Ensuite, l'algorithme du plus proche voisin (K-NN) a été appliqué en utilisant MATLAB pour la classification automatisée de l'état des liens. Les résultats obtenus ont démontré une grande précision et une excellente robustesse du modèle face à la turbulence, confirmant l'efficacité de l'intégration de l'IA dans le développement des systèmes auto-adaptatifs de FSO et leur capacité à fonctionner efficacement dans des conditions météorologiques dégradées.

Mots-clés : Apprentissage Automatique, Algorithme K-NN, FSO, Facteur de Qualité, OptiSystem, MATLAB.

ABSTRACT

Free-space optical communications (FSO) is a core technology of next generation wireless communication networks, providing high data transfer rates similar to fiber optics without the need for spectrum licenses. However, its state of preparation is strongly affected by the disturbances of atmospheric channels and unfavorable climatic conditions.

In this work, an intelligent predictive approach based on machine learning has been proposed to diagnose and classify the status of FSO channels. To achieve this, the FSO transmission chain was modeled and simulated using the OptiSystem environment in order to create a physical database that reflects different climate scenarios. The quality factor (Q-Factor) was extracted as the main predictive factor. Next, the nearest neighbor algorithm (K-NN) was applied using MATLAB for automated link state classification. The results obtained demonstrated high accuracy and excellent robustness of the model in the face of turbulence, confirming the effectiveness of AI integration in the development of self-driving systems adaptive FSOs and their ability to operate effectively in degraded weather conditions.

Keywords : Machine Learning, K-NN Algorithm, FSO, Quality Factor, OptiSystem, MATLAB.

ملخص

تُعدّ تقنية الاتصالات البصرية في الفضاء الحر (FSO) من التقنيات الأساسية لشبكات الاتصالات اللاسلكية من الجيل الجديد، حيث توفر معدلات نقل بيانات عالية مماثلة للألياف البصرية دون الحاجة إلى تراخيص الطيف الترددي. ومع ذلك، فإن جاهزيتها تتأثر بشكل كبير باضطرابات القناة الجوية والظروف المناخية غير الملائمة.

في هذا العمل، تم اقتراح مقارنة تنبؤية ذكية تعتمد على التعلم الآلي من أجل تشخيص وتصنيف حالة قناة ال-FSO. ولتحقيق ذلك، تم نمذجة ومحاكاة سلسلة إرسال FSO باستخدام بيئة OptiSystem بهدف إنشاء قاعدة بيانات فيزيائية تعكس سيناريوهات مناخية مختلفة. وقد تم استخراج معامل الجودة (Q-Factor) باعتباره المعامل التنبؤي الرئيسي. بعد ذلك، تم تطبيق خوارزمية أقرب الجيران (K-NN) باستخدام MATLAB من أجل التصنيف الآلي لحالة الوصلة. أظهرت النتائج المتحصل عليها دقة عالية ومتانة ممتازة للنموذج في مواجهة الاضطرابات، مما يؤكد فعالية دمج الذكاء الاصطناعي في تطوير أنظمة FSO ذاتية التكيف وقدرتها على العمل بكفاءة تحت الظروف الجوية المتدهورة.

الكلمات المفتاحية: تعلم الآلة، خوارزمية أقرب الجيران (K-NN)، الاتصالات البصرية في الفضاء الحر (FSO)، معامل الجودة، برنامج أوبتي سيستم (OptiSystem)، برنامج ماتلاب (MATLAB).

TABLE DES MATIÈRES

Sommaire.....	I
Remerciements.....	II
Dédicaces.....	III
Résumé.....	IV
Abstract.....	V
ملخص.....	VI
Table des Matières.....	VII
Liste des Figures.....	XII
Liste des Tableaux.....	XIII
Listes des abréviations.....	XIII
Introduction générale.....	1

CHAPITRE I : L'Apprentissage Automatique K-NN

I.1 Introduction.....	4
I.2 Fondaments de l'apprentissage automatique.....	5
I.2.1 Évolution conceptuelle et historique.....	5
I.2.2 Processus opérationnel et domaines d'application.....	6
I.2.3 Analyse critique : avantages et inconvénients.....	7
I.3 Typologie des paradigmes d'apprentissage.....	8
I.3.1 Les trois grandes catégories d'apprentissage.....	8
I.3.2 Synthèse des classifieurs supervisés courants.....	12
I.4 Formation mathématique et évaluation.....	13
I.4.1 Cadre formel et concepts de base.....	13
I.4.2 Formalisation d'un problème de l'apprentissage.....	15
I.4.3 Défis et limites actuels de l'apprentissage automatique.....	16
I.4.4 Les métriques d'évaluation des performances.....	17

I.5 L'algorithme des K plus proches voisins (K-NN).....	19
I.5.1 Principes fondamentaux et mécanisme.....	19
I.5.2 Applications numériques et études de cas.....	21
I.6 Conclusion.....	24

CHAPITRE II : La liaison optique en espace libre FSO

II.1 Introduction.....	26
II.2 Présentation de la technologie FSO.....	26
II.3 Principe de fonctionnement des liaisons FSO.....	27
II.3.1	
L'émetteur.....	28
II.3.1.1 La modulation d'intensité.....	28
II.3.1.2 La modulation externe.....	28
II.3.1.3 Les diodes électroluminescentes (LED).....	29
II.3.1.4 Les diodes Laser.....	30
II.3.2 Canal de transmission.....	30
II.3.2.1 La fibre optique.....	31
II.3.3 Le Récepteur.....	32
II.3.3.1 L'optique de collecte.....	33
II.3.3.2 L'photodétecteur.....	34
II.3.3.3 Le traitement du signal.....	34
II.3.3.4 Techniques de photo-détection.....	34
II.4 Les avantages et les inconvénients de la technologie FSO.....	36
II.4.1 Les avantages.....	36
II.4.2 Les inconvénients.....	37
II.5 Applications de la technologie FSO.....	37
II.6 Classification des réseaux FSO.....	38
II.6.1 Réseau satellitaire optique sans fil.....	38

II.6.2 Réseaux sans fil optiques terrestres.....	38
II.6.3 Réseau domestique optique sans fil.....	39
II.7 Les conditions atmosphériques sur les liaisons FSO.....	39
II.7.1 Atténuation par le brouillard.....	39
II.7.2 Atténuation par la neige.....	40
II.7.3 Atténuation par la pluie.....	40
II.7.4 Atténuation par les gaz atmosphériques (Absorption).....	41
II.7.5 Turbulence atmosphérique (Scintillation).....	41
II.7.6 Atténuation par les aérosols (Fumée et Poussière).....	41
II.8 Valeurs typiques d'atténuation atmosphérique pour les liaisons FSO.....	42
II.9 Applications de l'apprentissage machine sur les liaisons FSO.....	43
II.9.1 Prédiction de l'atténuation atmosphérique.....	43
II.9.2 Atténuation de la scintillation et des turbulences.....	44
II.9.3 Reconnaissance automatique des formats de modulation.....	44
II.9.4 Optimisation du pointage et de la poursuite.....	45
II.10 Conclusion.....	46

CHAPITRE III : Simulation de la liaison FSO et classification par K-NN

III.1 Introduction.....	48
III.2 Environnement de simulation et outils de travail.....	48
III.2.1 Le logiciel OptiSystem.....	48
III.2.2 Le logiciel MATLAB.....	49
III.3 Conception du système de transmission FSO.....	49
III.3.1 Bloc Émetteur	50
III.3.2 Canal de transmission.....	50
III.3.3 Bloc Récepteur	51
III.3.4 Outils d'analyse et de mesure.....	53
III.4 Méthodologie de génération du Dataset.....	53

III.4.1 Scénarios de simulation et paramètres variables.....	53
III.4.2 Extraction des caractéristiques (Features Extraction).....	54
III.5 Traitement des données et Classification sous MATLAB.....	54
III.5.1 Prétraitement et Normalisation des données.....	54
III.5.2 Application de l’algorithme Coarse K-NN.....	55
III.6 Analyse des résultats et discussion.....	57
III.6.1 Analyse de l'exactitude	57
III.6.2 Étude de la matrice de confusion.....	57
III.6.3 Analyse de la courbe ROC	58
III.7 Conclusion	59
Conclusion générale.....	61
Références bibliographiques.....	63

LISTE DES FIGURES

CHAPITRE I : L'Apprentissage Automatique K-NN

Figure I.1 : Relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.....	4
Figure I.2 : Fonctionnement de l'apprentissage automatique	5
Figure 1.3 : Illustration des principaux domaines de l'apprentissage automatique	7
Figure I.4 : Les trois types d'apprentissage	8
Figure I.5 : Classification vs Régression	9
Figure I.6 : Comparaison entre l'apprentissage supervisé et non supervisé	10
Figure I.7 : Diagramme de la boucle d'interaction agent-environnement	11
Figure I.8 : Schéma général de la formalisation d'un problème d'apprentissage supervisé.....	15

CHAPITRE II : La liaison optique en espace libre FSO

Figure II.1 : Schéma simplifié d'un système FSO	26
Figure II.2 : Schéma bloc d'une liaison FSO terrestre	27
Figure II.3 : Principe de fonctionnement d'un modulateur de Mach-Zehnder.....	29
Figure II.4 : Architecture d'une liaison FSO terrestre point à point	31
Figure II.5 : Structure d'une fibre optique	32
Figure II.6 : Schéma du récepteur optique	33
Figure II.7 : Système de collecte optique (télescope)	33
Figure II.8 : Détection directe vs détection cohérente	35
Figure II.9 : Effet de la scintillation atmosphérique	44
Figure II.10 : Architecture d'un réseau de neurones convolutionnel (CNN)	45

CHAPITRE III : Simulation de la liaison FSO et classification par K-NN

Figure III.1 : Architecture globale du système FSO.....	49
Figure III.2 : Bloc émetteur	50
Figure III.3 : Bloc du canal FSO.....	50
Figure III.4 : Bloc récepteur dans OptiSystem	51

Figure III.5 : Paramètre du canal de propagation FSO	52
Figure III.6 : Environnement de classification et analyse des performances sous MATLAB	56
Figure III.7 : Matrice de confusion du modèle Coarse K-NN	57
Figure III.8 : Courbe ROC et performance de classification du système FSO.....	58

LISTE DES TABLEAUX

Tableau II.1 : Comparaison entre les sources LED et LASER	30
Tableau II.2 : Classification des conditions atmosphériques selon l'atténuation	42
Tableau III.1 : Paramètre de configuration du canal FSO	53
Tableau III.2 : Configuration des paramètres du canal FSO dans OptiSystem	54

LISTE DES ABRÉVIATIONS

<i>AMC</i>	<i>Automatic Modulation Classification</i>
<i>APD</i>	<i>Avalanche Photodiode</i>
<i>AUC</i>	<i>Area Under Curve</i>
<i>BER</i>	<i>Bit Error Rate</i>
<i>CNN</i>	<i>Convolutional Neural Network</i>
<i>CW</i>	<i>Continuous Wave</i>
<i>DD</i>	<i>Direct Detection</i>
<i>FN</i>	<i>False Negatives</i>
<i>FOV</i>	<i>Field of View</i>
<i>FP</i>	<i>False Positives</i>
<i>FSO</i>	<i>Free Space Optics</i>
<i>Gbps</i>	<i>Gigabits per second</i>
<i>IA</i>	<i>Intelligence Artificielle</i>
<i>IM</i>	<i>Intensity Modulation</i>
<i>ISL</i>	<i>Inter-Satellite Links</i>
<i>K-NN</i>	<i>K-Nearest Neighbors</i>
<i>LASER</i>	<i>Light Amplification by Stimulated Emission of Radiation</i>
<i>LED</i>	<i>Light Emitting Diodes</i>
<i>Li-Fi</i>	<i>Light Fidelity</i>
<i>LOS</i>	<i>Line of Sight</i>
<i>LSTM</i>	<i>Long Short-Term Memory</i>

<i>ML</i>	<i>Machine Learning</i>
<i>MSE</i>	<i>Mean Squared Error</i>
<i>MZM</i>	<i>Mach-Zehnder Modulator</i>
<i>OOK</i>	<i>On-Off Keying</i>
<i>OSNR</i>	<i>Optical Signal to Noise Ratio</i>
<i>OWHN</i>	<i>Optical Wireless Home Network</i>
<i>OWSN</i>	<i>Optical Wireless Satellite Network</i>
<i>OWTN</i>	<i>Optical Wireless Terrestrial Networks</i>
<i>PIN</i>	<i>Positive-Intrinsic-Negative</i>
<i>PSK</i>	<i>Phase Shift Keying</i>
<i>QAM</i>	<i>Quadrature Amplitude Modulation</i>
<i>RF</i>	<i>Random Forest</i>
<i>RMSE</i>	<i>Root Mean Squared Error</i>
<i>RNN</i>	<i>Recurrent Neural Networks</i>
<i>ROC</i>	<i>Receiver Operating Characteristic</i>
<i>SNR</i>	<i>Signal to Noise Ratio</i>
<i>SVM</i>	<i>Support Vector Machine</i>
<i>TN</i>	<i>True Negatives</i>
<i>TP</i>	<i>True Positives</i>
<i>WDM</i>	<i>Wavelength Division Multiplexing</i>

Introduction générale

Introduction Générale

Avec l'évolution rapide des technologies de télécommunications et l'augmentation continue des besoins en transmission de données à haut débit, les systèmes de communication sans fil occupent aujourd'hui une place essentielle dans les réseaux modernes. Cependant, les technologies radiofréquences conventionnelles présentent plusieurs limitations, notamment la saturation spectrale et les interférences électromagnétiques. Dans ce contexte, les communications optiques en espace libre (Free Space Optics – FSO) apparaissent comme une solution prometteuse grâce à leur grande capacité de transmission, leur large bande passante et leur faible coût de déploiement.

La technologie FSO permet de transmettre des données à travers l'atmosphère en utilisant un faisceau lumineux sans support physique guidé. Malgré ses nombreux avantages, cette technologie reste fortement influencée par les conditions atmosphériques telles que le brouillard, la pluie et la turbulence, qui peuvent dégrader considérablement la qualité de la liaison optique. Ainsi, l'évaluation et la prédiction de l'état du canal FSO deviennent nécessaires afin d'assurer des communications fiables et performantes.

Par ailleurs, l'intelligence artificielle et particulièrement les techniques d'apprentissage automatique (Machine Learning) connaissent un développement important dans le domaine des télécommunications intelligentes. Ces techniques permettent d'analyser les données issues des systèmes de transmission afin d'améliorer leurs performances et de faciliter la prise de décision automatique. Parmi les algorithmes les plus utilisés, l'algorithme des K-Plus Proches Voisins (K-NN) se distingue par sa simplicité et son efficacité dans les problèmes de classification.

Dans ce travail, nous proposons une approche intelligente basée sur l'apprentissage automatique pour classifier l'état d'une liaison FSO. Une chaîne de transmission FSO a été conçue

et simulée sous l'environnement OptiSystem afin de générer une base de données correspondant à différents scénarios atmosphériques. Le facteur de qualité (Q-Factor) a ensuite été exploité comme paramètre principal pour l'implémentation de l'algorithme K-NN sous MATLAB dans le but de classifier automatiquement l'état du canal.

Ce mémoire est structuré en trois chapitres. Le premier chapitre est consacré aux concepts de l'apprentissage automatique ainsi qu'à l'étude de l'algorithme des *K plus proches voisins* (K-NN). Le deuxième chapitre présente les notions fondamentales des systèmes de communication à espace libre optique (FSO) et leurs principales caractéristiques. Enfin, le troisième chapitre décrit la conception de la chaîne FSO, la génération du jeu de données (*Dataset*), l'implémentation du modèle K-NN ainsi que l'analyse des résultats obtenus.

Chapitre I

L'Apprentissage Automatique K-NN

I.1 Introduction :

L'Intelligence Artificielle (IA) est définie comme l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. Elle représente aujourd'hui un domaine de recherche majeur dont l'objectif est de permettre aux systèmes informatiques de percevoir, de raisonner et d'agir de manière autonome face à des problèmes complexes [1].

Dans ce chapitre, nous explorons les fondements de cette discipline en nous concentrant sur l'apprentissage automatique [2] et ses différentes catégories [3]. Nous détaillerons plus particulièrement le fonctionnement, les métriques de distance et les limites de l'algorithme des K-Plus Proches Voisins (K-NN) [4], afin d'établir une base théorique solide pour son application ultérieure dans l'analyse et la détection des perturbations au sein des réseaux optiques sans fil (FSO).

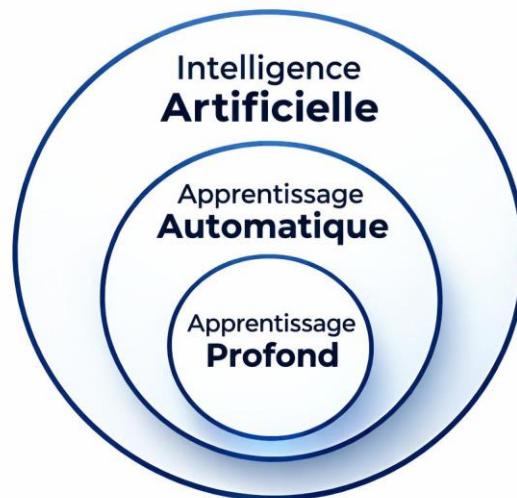


Figure I.1 : Relation entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond [4].

I.2 Fondements de l'apprentissage automatique :

I.2.1 Évolution conceptuelle et historique :

▪ **Définition :** L'apprentissage automatique est une sous-discipline de l'intelligence artificielle qui repose sur des méthodes statistiques pour permettre aux machines d'améliorer leurs performances dans l'exécution de tâches spécifiques [3]. Le processus consiste à utiliser un algorithme pour extraire des motifs à partir d'un ensemble de données d'entraînement, afin de construire un modèle capable de réaliser des prédictions ou de prendre des décisions sur de nouvelles données d'entrée [2] [5].

▪ **Historique :** Le domaine de l'apprentissage automatique ou en anglais *Machine Learning (ML)* a émergé officiellement en 1959, lorsque Arthur Samuel a défini ce champ comme la capacité des ordinateurs à apprendre sans être explicitement programmés [6]. Par la suite, en 1997, Tom Mitchell a fourni une définition formelle et plus technique, largement adoptée aujourd'hui : « Un programme informatique est dit apprenant d'une expérience E , par rapport à une tâche T et une mesure de performance P , si sa performance dans la tâche T , mesurée par P , s'améliore avec l'expérience E » [7].

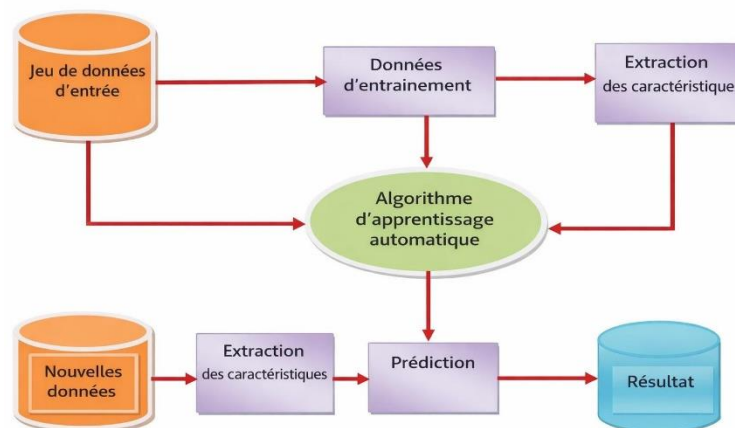


Figure I.2 : Fonctionnement de l'apprentissage automatique [5].

▪ **La différence entre l'intelligence artificielle et l'apprentissage automatique :** Bien que souvent utilisés de manière interchangeable, l'IA et l'apprentissage automatique désignent des concepts distincts. L'intelligence artificielle est le concept large visant à créer des machines capables de simuler l'intelligence humaine, incluant des approches basées sur des règles logiques prédéfinies. En revanche, l'apprentissage automatique est une sous-catégorie de l'IA qui permet au système d'évoluer de manière autonome grâce à l'analyse de données massives. Contrairement aux systèmes d'IA traditionnels, les modèles d'apprentissage automatique ne sont pas limités par des règles statiques ; ils s'améliorent continuellement en apprenant de l'expérience [1] [2].

I.2.2 Processus opérationnel et domaines d'application :

▪ Le fonctionnement de l'apprentissage automatique s'articule autour de deux phases fondamentales. La première est la phase d'apprentissage (entraînement), durant laquelle l'algorithme analyse un ensemble de données historiques afin d'ajuster ses paramètres internes et d'optimiser une fonction de coût [3] [5]. Une fois le modèle stabilisé, la seconde phase, dite de déploiement ou d'inférence, consiste à exploiter ce modèle pour générer des prédictions à partir de nouvelles données d'entrée [2]. L'objectif final est d'assurer une capacité de généralisation optimale, permettant au modèle de maintenir une haute performance face à des données non rencontrées lors de l'entraînement [3].

▪ Grâce à sa capacité à traiter des données complexes, l'apprentissage automatique trouve des applications dans divers domaines stratégiques comme :

- **Télécommunications** : Optimisation des réseaux, détection d'anomalies et filtrage intelligent des signaux [2].
- **Santé** : Aide au diagnostic médical et analyse d'images radiologiques [5].
- **Finance** : Détection de fraudes bancaires et prévisions boursières [3].

- **Reconnaissance vocale et vision par ordinateur** : Identification de visages et traduction automatique en temps réel [4].



Figure 1.3 : Illustration des principaux domaines de l'apprentissage automatique [2].

I.2.3 Analyse critique : avantages et inconvénients :

Il transforme de nombreux aspects de notre quotidien, de la médecine à la sécurité financière. Toutefois, il présente des avantages majeurs et certains risques critiques [8].

▪ **Avantages :**

- **Diminution des coûts opérationnels** : L'automatisation des tâches réduit le besoin d'intervention humaine constante, diminuant ainsi les coûts globaux [8].
- **Efficacité et précision** : Les modèles peuvent accomplir des tâches spécifiques avec une précision extrême et une rapidité supérieure à celle des humains [3].
- **Amélioration des connaissances** : Capacité d'identifier des tendances cachées dans de vastes ensembles de données [2].

▪ **Inconvénients :**

- **Besoin de données de qualité** : La performance dépend strictement de la qualité des données d'entrée [5].

- **Complexité** : La mise en place nécessite une expertise technique élevée et des ressources de calcul importantes [4].
- **Interprétabilité** : Certains modèles (boîtes noires) sont difficiles à expliquer, rendant la compréhension des décisions complexe [4].
- **Sécurité et confidentialité** : Vulnérabilité aux cyberattaques et risques liés à la protection des données sensibles [7] [8].

L'apprentissage automatique joue un rôle fondamental dans la transformation des réseaux de télécommunications modernes. Dans le contexte des liaisons optiques sans fil de type FSO, cette technologie est exploitée pour la prédiction proactive des atténuations atmosphériques, la compensation des turbulences et l'optimisation de la qualité du signal, permettant ainsi d'assurer une stabilité de service accrue malgré les défis environnementaux.

I.3 Typologie des paradigmes d'apprentissages :

I.3.1 Les trois grandes catégories d'apprentissage :

Il existe plusieurs approches permettant à une machine d'apprendre à partir de données, chacune possédant ses propres caractéristiques architecturales. Le choix d'une méthode dépend étroitement de la nature des problématiques rencontrées, de la structure des données disponibles et des objectifs décisionnels visés [3]. On distingue principalement trois catégories d'apprentissage

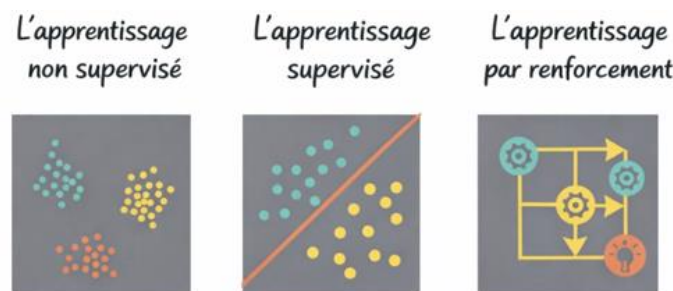


Figure I.4 : Les trois types d'apprentissage [3].

▪ L'apprentissage supervisé :

L'apprentissage supervisé est le mode le plus courant dans l'apprentissage automatique. Dans ce paradigme, l'algorithme est entraîné à l'aide d'un ensemble de données d'entrée associées à des

étiquettes ou des sorties correspondantes souhaitées [3] [5]. Le système apprend à établir une correspondance entre les entrées et les sorties afin de prédire correctement le résultat pour de nouvelles données invisibles. Ce type d'apprentissage se divise en deux grandes tâches principales [2] [3] :

- **Classification :**

La classification est une tâche où la donnée d'entrée doit être affectée à une catégorie ou une classe discrète. L'objectif est d'apprendre un modèle capable d'attribuer une étiquette représentant une classe à chaque nouvelle instance (par exemple : prédire si un e-mail est "Spam" ou "Non-Spam"). Lorsque le nombre de classes est limité à deux, on parle de classification binaire [2] [5].

- **Régression :**

Contrairement à la classification, la régression traite des variables de sortie continues ou numériques. Par exemple, prédire le prix d'une maison, la température ou la valeur d'un signal dans un réseau de communication. Ici, l'objectif est de trouver une fonction capable de prédire une valeur numérique précise à partir des caractéristiques fournies en entrée [3].

Exemple : un système de filtrage de courriers indésirables utilise l'apprentissage supervisé. En analysant des milliers d'e-mails déjà marqués par l'utilisateur comme "Spam", l'algorithme extrait des caractéristiques (mots-clés, expéditeurs) pour classer automatiquement les nouveaux messages reçus [8].

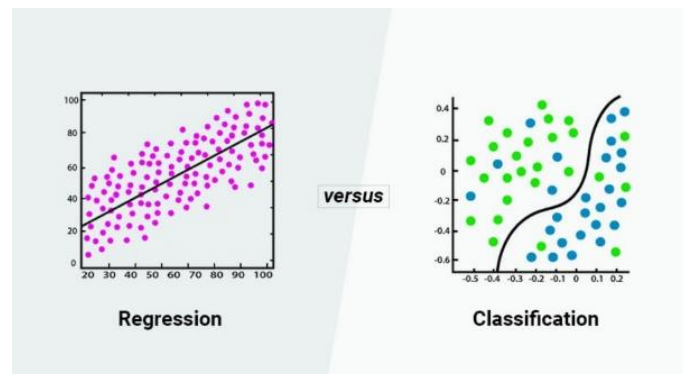


Figure I.5 : Classification vs Régression [5].

▪ L'apprentissage non supervisé :

L'apprentissage non supervisé est un paradigme où l'algorithme travaille sur des données d'entrée qui ne sont pas étiquetées [2]. Contrairement au mode supervisé, aucun résultat attendu n'est fourni au système. L'algorithme doit donc explorer les données de manière autonome pour y détecter des structures cachées, des similitudes ou des groupements naturels [3] [7].

Ce type d'apprentissage est principalement utilisé pour les tâches suivantes :

- **Le Partitionnement** : ou *clustering en anglais* consiste à regrouper des données similaires dans des clusters distincts (par exemple, segmenter des clients selon leurs comportements d'achat) [2] [5].
- **Réduction de dimensionnalité** : permet de simplifier les données complexes en ne gardant que les caractéristiques les plus importantes [3].

Exemple : dans un réseau WDM (Multiplexage en longueur d'onde) en anglais : *Wavelength Division Multiplexing*, l'apprentissage non supervisé permet de collecter les données de performance des canaux optiques (puissance, bruit) et de les étiqueter automatiquement selon leur état (normal, interférence, ou défaillance). Cela aide à entraîner ensuite des modèles supervisés pour la maintenance prédictive [9].

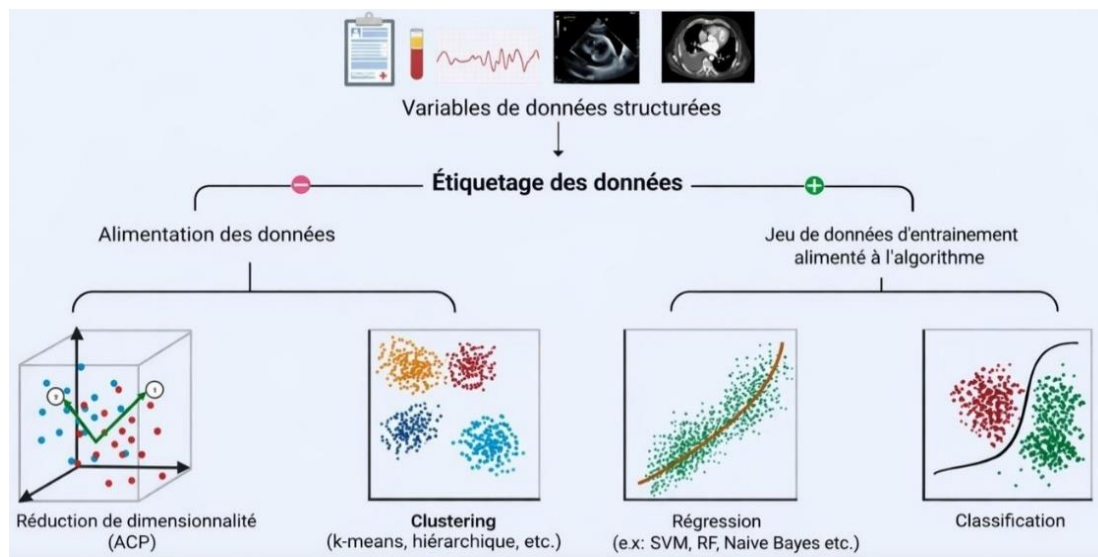


Figure I.6 : Comparaison entre l'apprentissage supervisé et non supervisé [3].

- **Focus sur l'algorithme K-means** : Introduit par MacQueen en 1967, le k-means est un algorithme itératif majeur du clustering [10]. Son fonctionnement repose sur les étapes suivantes :

- **Initialisation** : Définition préalable du nombre de clusters k et placement de leurs centres.
- **Assignment** : Rapprochement de chaque donnée vers le centre le plus proche.
- **Actualisation** : Recalcul de la position des centres en fonction de la moyenne des points assignés.
- **Convergence** : Répétition du processus jusqu'à l'obtention d'une configuration stable.

La performance du modèle est fortement sensible à la position initiale des centres et au choix de la valeur k .

▪ L'apprentissage par renforcement :

L'apprentissage par renforcement est une forme d'apprentissage basée sur l'interaction dynamique entre un agent autonome et son environnement [1]. Le but de l'agent est d'apprendre à prendre les meilleures décisions possibles en fonction de son état actuel pour maximiser une récompense globale à long terme [3].

Contrairement aux autres types d'apprentissage, l'agent ne reçoit pas de données étiquetées en entrée. À chaque action effectuée, l'environnement lui renvoie un signal sous forme de récompense (positive ou négative). À travers des itérations répétées, l'agent développe une stratégie (appelée politique) pour équilibrer l'exploitation de ses connaissances actuelles et l'exploration de nouvelles actions plus performantes [2] [7].

Exemple : dans le domaine des jeux vidéo, comme FIFA, un joueur virtuel doit décider en temps réel entre faire une passe, tirer ou dribbler. En fonction de l'issue de l'action (marquer un but ou perdre le ballon), le système ajuste son comportement futur pour atteindre la solution optimale [2].

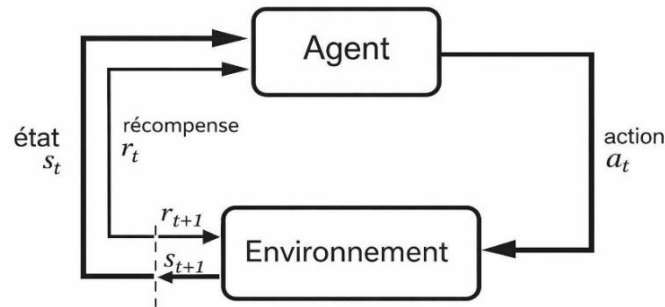


Figure I.7 : Diagramme de la boucle d'interaction agent-environnement [8].

I.3.2 Synthèse des classifieurs supervisés courants :

L'apprentissage supervisé consiste à entraîner un algorithme sur un ensemble de données étiquetées afin de construire un modèle capable de prédire les classes de nouvelles données [7]. Ce processus repose sur l'identification de corrélations entre les variables d'entrée et les étiquettes de sortie fournies par un expert ou un système de référence [2].

Voici les approches de classification les plus couramment utilisées :

▪ La régression logistique :

Une technique utilisée pour prédire la probabilité qu'une observation appartienne à l'une des deux catégories possibles [9]. Bien qu'elle soit de nature linéaire, elle utilise une fonction sigmoïde pour transformer la sortie en une valeur de probabilité comprise entre 0 et 1 [2].

▪ Les arbres de décision :

Cette méthode décompose un jeu de données complexe en sous-ensembles de plus en plus petits tout en développant un arbre de décision associé [3]. Le modèle final est une structure hiérarchique où les nœuds internes représentent des tests sur les caractéristiques et les feuilles représentent les décisions finales [7].

▪ **Les machines à vecteurs de support (SVM) :**

Le SVM cherche à définir un hyperplan de séparation optimal qui maximise la distance entre les classes dans un espace de grande dimension [2]. Grâce à l'utilisation de fonctions noyaux, il peut traiter efficacement des données qui ne sont pas linéairement séparables dans leur espace d'origine [3].

▪ **Le classifieur bayésien naïf (Naïve Bayes) :**

Ce classifieur est basé sur le théorème de Bayes et suppose une indépendance totale entre toutes les caractéristiques des données [9]. Il est particulièrement performant pour les tâches de classification à grande échelle comme le filtrage de courrier indésirable ou l'analyse de sentiments [7]. La formule de Bayes s'exprime par :

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad eq.01$$

I.4 Formation mathématique et évaluation :

I.4.1 Cadre formel et concepts de base :

L'implémentation de solutions basées sur l'intelligence artificielle nécessite la maîtrise de plusieurs piliers méthodologiques qui structurent le cycle de vie d'un projet de l'apprentissage automatique [3].

▪ **Dataset :**

- Le jeu de données représente la matière première indispensable à l'apprentissage, sans laquelle aucun projet ne peut être initié.
- Dans un contexte supervisé, il se compose de paires d'entrées-sorties (X, Y), où X constitue le vecteur de caractéristiques et Y la réponse attendue.
- À l'inverse, l'apprentissage non supervisé exploite des données X uniquement, laissant l'algorithme identifier seul les structures cachées.

▪ **Modèle et ses paramètres :**

- Un modèle est défini comme une représentation mathématique, linéaire ou complexe, capable de transformer des données d'entrée en prédictions.
- Les paramètres sont les variables internes du modèle, tels que les poids ou les coefficients, qui sont ajustés de manière itérative durant la phase d'entraînement.
- L'objectif de cet ajustement est de minimiser l'écart entre les prédictions du modèle et les annotations réelles fournies dans le dataset.

▪ **Les hyperparamètres :**

- Contrairement aux paramètres, les hyperparamètres sont des configurations externes définies par l'utilisateur avant le lancement de l'apprentissage.
- Ils incluent des éléments cruciaux tels que le nombre d'itérations, la constante de régularisation ou encore la valeur de la graine aléatoire (seed) pour garantir la reproductibilité.
- Le choix optimal de ces valeurs influence directement la vitesse de convergence et la précision finale du modèle.

▪ **Fonction Coût :**

- La fonction coût quantifie la performance du modèle en calculant l'erreur globale entre les résultats prédits et les valeurs réelles du dataset.
- Le processus d'optimisation vise à minimiser cette fonction pour réduire les erreurs, notamment les faux positifs et les faux négatifs.
- Une faible valeur de cette fonction indique que le modèle a réussi à identifier une configuration performante pour les données traitées.

▪ **Généralisation :**

- La généralisation désigne l'aptitude d'un modèle entraîné à fournir des prédictions exactes sur des données qu'il n'a jamais rencontrées auparavant.

- Pour valider cette capacité et éviter le surapprentissage, le modèle est testé sur un échantillon externe, garantissant ainsi sa fiabilité en conditions réelles.
- Ce processus est essentiel pour l'industrialisation des systèmes intelligents, car il confirme la robustesse du modèle face à la variabilité des flux de données.

Donc , Le flux de travail commence par la préparation d'un échantillon représentatif incluant prédictors et annotations. Le modèle est ensuite calibré via ses hyperparamètres avant de subir un entraînement rigoureux. Une phase de validation, s'appuyant sur une fonction coût, permet d'ajuster les performances. Enfin, la robustesse du système est vérifiée par un test de généralisation sur des données indépendantes [3].

I.4.2 Formalisation d'un problème de l'apprentissage :

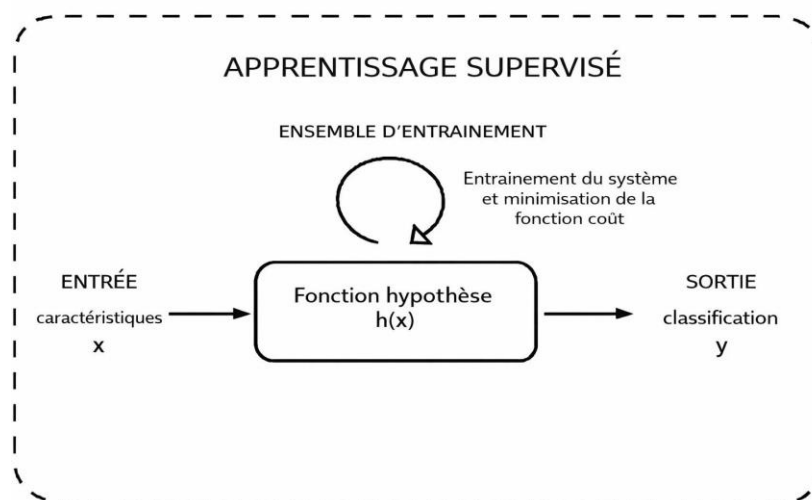


Figure I.8 : Schéma général de la formalisation d'un problème d'apprentissage supervisé [7].

L'apprentissage automatique est défini formellement comme la recherche d'une fonction de prédiction h , nommée hypothèse, capable de mapper des entrées vers des sorties cibles [7]. L'objectif principal de cette formalisation est de découvrir une règle de décision qui minimise l'erreur de prédiction sur des données non observées [5].

▪ **Donnée à interpréter ($x \in X$) :**

- Chaque instance x est représentée sous la forme d'un vecteur de caractéristiques au sein d'un espace multidimensionnel X [7].
- Ces variables d'entrée peuvent être des mesures physiques complexes, telles que la puissance du signal ou le niveau de bruit. [5]

▪ **Sortie attendue ou prédiction ($y \in Y$) :**

- La sortie y constitue la réponse ou l'étiquette que le modèle doit attribuer à chaque vecteur d'entrée x [7].
- La nature de l'espace Y détermine le type de tâche, qu'il s'agisse d'une classification pour un ensemble fini ou d'une régression pour un domaine continu [5].

▪ **Exemple annoté (x_i, y_i) :**

- Un exemple annoté est une paire unique constituée d'une observation et de sa valeur de vérité terrain correspondante [7].
- On postule que ces paires sont extraites de façon indépendante d'une distribution de probabilité $\mathcal{D}$ inconnue [5].

▪ **Ensemble d'apprentissage ($D = \{ (x_i, y_i) \}$) :**

- Le dataset D représente la base de connaissances finie utilisée par l'algorithme pour ajuster ses paramètres [7].
- La capacité du modèle à généraliser dépend directement de la qualité et de la diversité des exemples contenus dans cet ensemble [5].

1.4.3 Défis et limites actuels de l'apprentissage automatique :

Malgré les avancées technologiques, le déploiement de l'apprentissage automatique se heurte à des contraintes de qualité et de quantité des données, car un modèle entraîné sur un échantillon bruité ou insuffisant perdra inévitablement en fiabilité [7]. Un autre défi majeur réside

dans le phénomène de sur-apprentissage en anglais *overfitting*, où l'algorithme finit par mémoriser les particularités et le bruit du jeu d'entraînement au lieu de comprendre la logique sous-jacente, ce qui entraîne une chute brutale des performances face à de nouvelles situations réelles [5].

1.4.4 Les métrique d'évaluation des performances :

L'évaluation constitue une phase déterminante pour valider l'utilité pratique d'un algorithme et sa capacité à traiter des données qu'il n'a pas rencontrées lors de sa conception. Le choix d'une métrique de performance est dicté par la nature spécifique de la tâche (classification, régression ou clustering) et par l'impact relatif des différents types d'erreurs sur l'application finale [7]. Une évaluation rigoureuse permet non seulement de comparer plusieurs modèles entre eux, mais aussi de s'assurer que le système répond aux exigences de fiabilité nécessaires à une mise en production [5].

▪ Pour les problèmes de classification :

Dans un contexte de classification, les modèles attribuent des étiquettes discrètes aux instances en fonction des frontières de décision apprises. L'analyse de performance repose généralement sur la confrontation entre les prédictions et les étiquettes réelles, souvent synthétisée par les métriques suivantes [7]:

- **Exactitude** : Elle représente le taux global de réussite du modèle en calculant le ratio des prédictions correctes sur la totalité des échantillons testés [7].

$$Exactitude = \frac{TP + TN}{TP + TN + FP + FN} \quad eq. 02$$

- **Précision** : Cette métrique mesure la fiabilité du modèle lorsqu'il prédit la classe positive, en évaluant la proportion de vrais positifs parmi tous les exemples déclarés positifs [7].

$$Précision = \frac{TP}{TP + FP} \quad eq. 03$$

- **Rappel** : Il quantifie l'aptitude du système à identifier l'intégralité des exemples appartenant réellement à la classe positive, évitant ainsi les omissions critiques [7].

$$Rappel = \frac{TP}{TP + FN} \quad eq.04$$

- **Score F1** : Il s'agit de la moyenne harmonique de la précision et du rappel, offrant un indicateur unique et équilibré, particulièrement utile lorsque les classes du dataset sont déséquilibrées [7].

$$F1 = 2 \times \frac{précision \times Rappel}{précision + Rappel} \quad eq.05$$

- **Courbe ROC et AUC** : La courbe ROC (Courbe de performance du récepteur) illustre dynamiquement le compromis entre le taux de vrais positifs et celui des faux positifs, tandis que l'AUC (Aire Sous la Courbe) fournit une mesure globale de la capacité du modèle à discriminer les différentes classes [11] [12].

▪ Pour les problèmes de régression :

Pour les modèles dont la sortie est une valeur numérique continue, la performance est évaluée par l'ampleur de l'écart (l'erreur) entre la valeur prédite $\hat{y}$ et la valeur réelle y [7]:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad eq.06$$

- **Erreur quadratique moyenne (MSE)** : en anglais *Mean Squared Error* Elle calcule la moyenne des carrés des erreurs, ce qui a pour effet de pénaliser plus lourdement les écarts importants [7].
- **Racine de l'erreur quadratique moyenne (RMSE)** : en anglais *Root Mean Squared Error* en prenant la racine carrée de la MSE, on obtient une mesure d'erreur exprimée dans la même unité que la variable cible, facilitant ainsi son interprétation physique [5].

$$RMSE = \sqrt{MSE} \quad eq.07$$

▪ Pour les problèmes de clustering (non supervisé) :

En l'apprentissage non supervisé, où les étiquettes sont absentes, l'évaluation repose sur la structure intrinsèque des groupes formés par l'algorithme [7]:

- Indice de Rand ajusté : Cette métrique compare la similarité entre deux partitions en tenant compte des regroupements fortuits, avec un score proche de 1 indiquant une correspondance parfaite [7].
- Indice de Silhouette : Il mesure la qualité du partitionnement en évaluant si chaque point est suffisamment proche de son propre cluster (cohésion) et suffisamment éloigné des clusters voisins (séparation) [13].

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad eq.08$$

I.5 L'algorithme des K plus proches voisins (k-NN) :

I.5.1 Principes fondamentaux et mécanisme :

L'algorithme k-NN (k-Nearest Neighbors) est une méthode d'apprentissage supervisé non paramétrique qui classe une observation en se basant sur la majorité statistique de ses voisins les plus proches dans l'espace des caractéristiques [14]. Contrairement aux modèles paramétriques, il ne nécessite pas de phase d'entraînement explicite, ce qui lui permet de s'adapter dynamiquement à de nouveaux échantillons sans recalculer l'intégralité du système [15].

▪ Où utiliser k-NN ?

Le k-NN est particulièrement robuste pour les tâches de classification et de régression dans des environnements où la distribution des données est complexe ou non linéaire [14]. En milieu industriel, il est privilégié pour sa simplicité de mise en œuvre et sa capacité à fournir des prédictions rapides lorsque la taille du jeu de données permet une recherche de voisinage efficace [15].

▪ Le fonctionnement de l'algorithme k-NN :

Le mécanisme opérationnel du k-NN repose sur une logique géométrique simple : pour chaque nouvelle instance, l'algorithme calcule sa distance par rapport à tous les points de la base d'apprentissage [14]. Après avoir sélectionné les k points présentant les distances les plus faibles, une règle de vote majoritaire (pour la classification) ou une moyenne pondérée (pour la régression) est appliquée pour générer la prédiction finale [15].

L'algorithme k-NN est une méthode d'apprentissage non paramétrique qui classe une nouvelle instance en fonction de la majorité des classes de ses voisins les plus proches [8]. Contrairement aux autres modèles, il ne nécessite pas de phase d'apprentissage explicite, car il stocke simplement tous les cas d'entraînement en mémoire pour la phase de prédiction [3]. La performance du k-NN dépend fortement du choix de la mesure de distance (comme la distance euclidienne) et de la valeur du paramètre k [8].

▪ Avantages et inconvénients de K-NN :

▫ Les avantages :

- **Capacité d'adaptation aux nouvelles données :** En raison de sa nature non paramétrique et de son approche d'apprentissage paresseux (*lazy learning*), le k-NN intègre immédiatement toute nouvelle observation sans nécessiter une phase de réentraînement complexe [8].
- **Interprétabilité des résultats :** Le processus de décision est transparent et intuitif, car chaque classification peut être justifiée par l'analyse directe des voisins les plus proches ayant influencé le résultat final [8].

▫ Les inconvénients :

- **Temps de classification élevé :** Comme l'algorithme ne construit pas de modèle préalable, la charge de calcul est transférée à l'étape de prédiction, ce qui peut ralentir le système lors du traitement de bases de données volumineuses.

- **Sensibilité à la dimensionnalité** : Les performances se dégradent fortement avec l'augmentation du nombre de caractéristiques non pertinentes, un phénomène connu sous le nom de "malédiction de la dimension".
- **Dépendance au choix du paramètre k** : La stabilité de la prédiction est étroitement liée à la valeur de k ; un k trop petit rend le modèle sensible au bruit, tandis qu'un k trop grand peut masquer les structures locales des classes.
- **Choix du bon critère de distance** : L'efficacité de l'algorithme repose sur la pertinence de la métrique de distance choisie, laquelle doit être adaptée à la distribution statistique des données.

▪ **Choix du paramètre K dans l'algorithme K-NN :**

La sélection de la valeur optimale de K représente un compromis crucial entre le biais et la variance. Une valeur de K trop faible peut conduire à un sur-apprentissage (*overfitting*) en capturant le bruit des données, tandis qu'une valeur trop élevée risque de provoquer un sous-apprentissage (*underfitting*) en simplifiant excessivement la frontière de décision. En pratique, des méthodes comme la validation croisée sont employées pour identifier le K qui maximise la capacité de généralisation du modèle sur des échantillons inconnus [16].

I.5.2 Application numériques et études de cas :

L'application pratique de l'algorithme k-NN repose sur une séquence de calculs rigoureux permettant de transformer des mesures brutes en une décision de classification ou de régression.

▪ **Étude de cas 01 : application sur le jeu de données "Iris" [16]**

Le jeu de données "Iris" est utilisé ici pour illustrer le mécanisme de décision de l'algorithme K-NN dans un espace multidimensionnel. L'objectif est de classer une nouvelle instance $Q(4.0 ; 1.2)$ en se basant sur un échantillon de référence composé de quatre points (P1, P2, P3, P4).

1. Calcul des distances euclidiennes :

Pour déterminer la proximité, nous calculons la distance entre le point inconnu Q et chaque point du dataset en utilisant la formule suivante :

$$d(P, Q) = \sqrt{(X_q - X_p)^2 + (Y_q - Y_p)^2}$$

Échantillon	Coordonnées (x; y)	Calcul de la distance $d(Q, P)$	Range de proximité
P1 (Setosa)	(1.4 ; 0.2)	$\sqrt{(4.0 - 1.4)^2 + (1.2 - 0.2)^2} = 2.78$	4
P2 (Setosa)	(1.5 ; 0.2)	$\sqrt{(4.0 - 1.5)^2 + (1.2 - 0.2)^2} = 2.69$	3
P3 (Versicolor)	(4.5 ; 1.3)	$\sqrt{(4.0 - 4.5)^2 + (1.2 - 1.3)^2} = 0.51$	1
P4 (Virginica)	(4.8 ; 1.8)	$\sqrt{(4.0 - 4.8)^2 + (1.2 - 1.8)^2} = 1.00$	2

2. Analyse du voisinage et classification (K=3) :

En fixant le paramètre de l'apprentissage automatique à $K=3$, nous sélectionnons les trois points ayant les distances les plus courtes. Les voisins retenus sont P3 (Versicolor), P4 (Virginica) et P2 (Setosa).

Bien que le voisinage soit hétérogène, les points P3 et P4 présentent la plus grande proximité spatiale avec l'instance Q . En appliquant la règle de la majorité pondérée par la distance, l'algorithme valide la classification de l'échantillon inconnu Q comme appartenant à la classe Versicolor [16].

▪ Étude de cas 02 : classification par calcul manuel (approche numérique) [3]

Cet exemple, basé sur la méthodologie décrite dans le document [3], démontre précisément comment l'algorithme traite les coordonnées spatiales pour classer une nouvelle instance Q dans un repère bidimensionnel.

Données d'apprentissage: Nous disposons d'un jeu de données composé de quatre points référencés:

P1 (1, 2): Classe A / P2 (2, 3): Classe A / P3 (4, 4): Classe B / P4 (6, 5): Classe B

Objectif: Déterminer la classe du point $Q(3, 3)$ avec un paramètre $K = 3$.

1.Calcul des distances euclidiennes: nous calculons la distance entre $Q(3, 3)$ et chaque point P du dataset [3]:

$$d(P, Q) = \sqrt{(3-1)^2 + (3-2)^2} = \sqrt{2^2 + 1^2} = \sqrt{5} = 2.23$$

$$d(P, Q) = \sqrt{(3-2)^2 + (3-3)^2} = \sqrt{1^2 + 0^2} = \sqrt{1} = 1.00$$

$$d(P, Q) = \sqrt{(3-4)^2 + (3-4)^2} = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2} = 1.41$$

$$d(P, Q) = \sqrt{(3-6)^2 + (3-5)^2} = \sqrt{(-3)^2 + (-2)^2} = \sqrt{13} = 3.60$$

2.Tri et sélection des $K = 3$ voisins les plus proches :

En classant les distances par ordre croissant, nous identifions les trois voisins immédiats :

1. 1er voisin: P (Distance = 1.00) $\rightarrow$ Classe A

2. 2ème voisin: P (Distance = 1.41) $\rightarrow$ Classe B

3. 3ème voisin: P (Distance = 2.23) $\rightarrow$ Classe A

3.Donc : Sur les trois voisins sélectionnés, la classe A est majoritaire (2 occurrences contre 1 pour la Classe B). Par conséquent, en application des principes de l'apprentissage automatique , le point Q est affecté à la Classe A [3].

▪ **Analyse critique : limites de K-NN en contexte réel :**

Bien que l'algorithme K-NN se distingue par sa simplicité, son application en contexte réel se heurte à des contraintes techniques significatives. Il manifeste une sensibilité accrue à la dimensionnalité élevée, où l'augmentation du nombre de variables dégrade la pertinence des mesures de distance, un phénomène identifié comme la « malédiction de la dimensionnalité ». En outre, sa complexité computationnelle est élevée car il impose un parcours intégral du jeu d'entraînement à chaque phase de prédiction, ce qui limite son efficacité sur les bases de données massives. Enfin, l'absence d'un modèle explicite rend l'algorithme vulnérable aux données bruitées ou déséquilibrées, rendant l'interprétation des résultats plus complexe par rapport aux méthodes paramétriques [9].

I.6 Conclusion :

Dans ce chapitre, nous avons présenté les concepts fondamentaux de l'IA et du l'apprentissage automatique, en mettant en évidence leurs relations hiérarchiques et leurs domaines d'application [16]. Une attention particulière a été portée à l'algorithme K-Nearest Neighbors (K-NN), une méthode d'apprentissage supervisé simple mais puissante, utilisée notamment pour des tâches de classification et de régression [3].

Nous avons détaillé le fonctionnement de l'algorithme K-NN, ses avantages, ses limites, ainsi que l'importance du choix du paramètre K, qui influence directement la précision des résultats obtenus [3]. Dans le prochain chapitre, nous allons introduire les principes fondamentaux des réseaux de communication optique sans fil FSO afin d'appliquer l'algorithme K-NN pour classifieur et prédire des phénomènes liés aux performances de la liaison, notamment en présence de perturbations atmosphériques et d'atténuations du signal.

Chapitre II

La Liaison Optique En Espace Libre

FSO

II.1 Introduction :

La technologie de communication optique en espace libre ou FSO en anglais (*Free Space Optics*) s'est imposée comme une solution de connectivité à haut débit capable de répondre à la demande croissante de bande passante, notamment pour résoudre le problème du "dernier kilomètre"[18]. Contrairement aux liaisons radiofréquences traditionnelles, le FSO utilise des faisceaux laser pour transmettre des données à travers l'atmosphère, offrant ainsi une alternative sécurisée, sans licence spectrale et facile à déployer dans divers environnements urbains et militaires [17].

Cependant, la performance de ces liaisons est intrinsèquement liée aux conditions climatiques du canal de propagation. Les phénomènes atmosphériques tels que le brouillard, la pluie et les turbulences provoquent des atténuations sévères et des fluctuations du signal optique, limitant ainsi la portée et la disponibilité du système [17] [18]. Ce chapitre examine les principes de fonctionnement du FSO et analyse l'impact des défis environnementaux, tout en explorant comment l'apprentissage machine peut être exploité pour optimiser la résilience de ces communications.

II.2 Présentation de la technologie FSO :

La technologie FSO est un système de communication optique point à point qui transmet des signaux de données (voix, vidéo ou données informatiques) à travers l'atmosphère. Contrairement à la fibre optique classique qui guide la lumière à l'intérieur d'un cœur de verre [9], le FSO utilise l'air comme milieu de propagation [18].

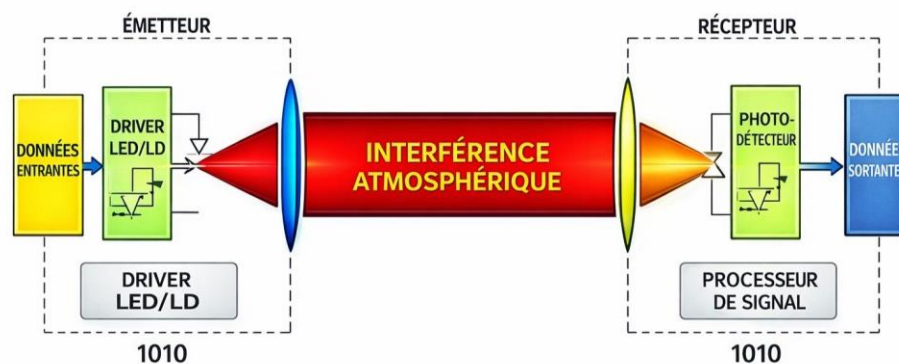


Figure II.1 : Schéma simplifié d'un système FSO [18].

Cette technologie repose sur l'utilisation de faisceaux laser de faible puissance, opérant généralement dans les fenêtres de transmission infrarouges (785 nm à 1550 nm), ce qui permet d'atteindre des débits binaires comparables à ceux de la fibre optique, dépassant souvent les 10 Gbps [19]. Le succès d'une liaison FSO dépend impérativement de l'existence d'une ligne de vue directe (LOS : *Line of Sight*) dégagée entre l'émetteur et le récepteur [17].

II.3 Principe de fonctionnement des liaisons FSO :

Le principe de fonctionnement d'une liaison FSO est analogue à celui des communications par fibre optique, à la différence près que le signal n'est pas guidé, mais propagé de manière directive dans l'air. Le processus commence par la conversion des données numériques en signaux électriques, lesquels pilotent une source lumineuse [17].

Le système se compose de trois sous-ensembles principaux : l'émetteur, le canal de transmission (l'atmosphère) et le récepteur. Au niveau de l'émetteur, le message est modulé sur une porteuse optique. Une fois émis, le faisceau traverse le canal atmosphérique où il subit diverses atténuations dues à l'absorption et à la diffusion par les molécules de l'air [19]. Enfin, au niveau du récepteur, l'énergie optique collectée par une lentille ou un télescope est reconvertie en signal électrique via un photodétecteur, avant d'être démodulée pour retrouver l'information originale [21].

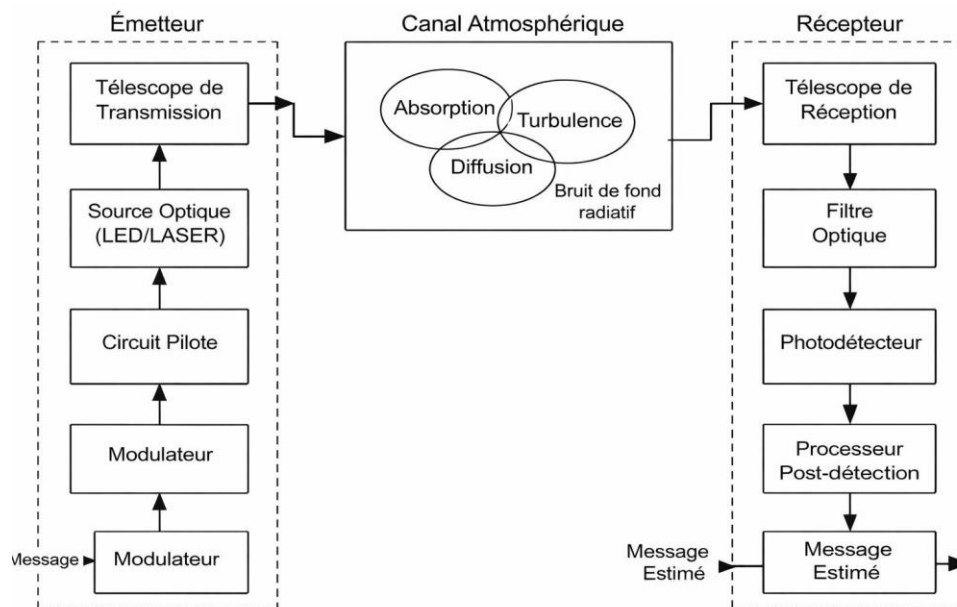


Figure II.2 : Schéma bloc d'une liaison FSO terrestre [17].

II.3.1 L'émetteur :

L'émetteur est le premier bloc critique de la chaîne de transmission FSO. Sa fonction principale est de convertir le signal électrique issu de la source d'information en un signal optique modulé et de le projeter vers le récepteur distant [18]. Il se compose généralement d'un driver (modulateur), d'une source lumineuse (Laser ou LED) et d'un système optique de mise en forme du faisceau (lentilles de collimation) [17].

Le choix de la source lumineuse est crucial car il détermine la portée de la liaison, le débit de données et la sécurité oculaire. Les systèmes FSO modernes privilégient les sources opérant dans le spectre infrarouge pour minimiser l'atténuation atmosphérique et garantir une transmission à haute performance [20].

II.3.1.1 La modulation d'intensité :

Dans la majorité des systèmes FSO actuels, on utilise la technique de modulation d'intensité (IM) en anglais (*Intensity Modulation*) couplée à une détection directe (DD) [21]. Ce procédé consiste à faire varier directement la puissance optique de la source lumineuse en fonction de l'amplitude du signal électrique.

Le format le plus simple et le plus utilisé est le OOK en anglais (*On-Off Keying*), où la présence d'une impulsion lumineuse représente un "1" logique et son absence représente un "0" logique [19]. Bien que simple à implémenter, cette méthode nécessite une gestion rigoureuse du seuil de décision au niveau du récepteur pour contrer les fluctuations de puissance causées par la turbulence atmosphérique [17].

II.3.1.2 La modulation externe:

Contrairement à la modulation directe où le courant de la source lumineuse est modulé, la modulation externe utilise un composant séparé, appelé modulateur externe, placé juste après la source lumineuse qui fonctionne en mode continu (CW : *Continuous Wave*) [21]. Cette technique est privilégiée pour les liaisons FSO à très haut débit (supérieur à 2.5 Gbps) car elle permet d'éviter

le phénomène de “Chirp” (variation de la longueur d’onde) qui peut survenir lors d’une commutation rapide du laser [17].

Le modulateur externe le plus répandu est le modulateur Mach-Zehnder (MZM). Ce dispositif divise le faisceau lumineux en deux bras ; l’un des bras subit un déphasage induit par un signal électrique avant que les deux faisceaux ne soient recombinaés [22]. Cette interférence permet de moduler l’intensité lumineuse de manière très précise, offrant ainsi une meilleure qualité de signal et une réduction des erreurs binaires (BER) sur de longues distances [18].

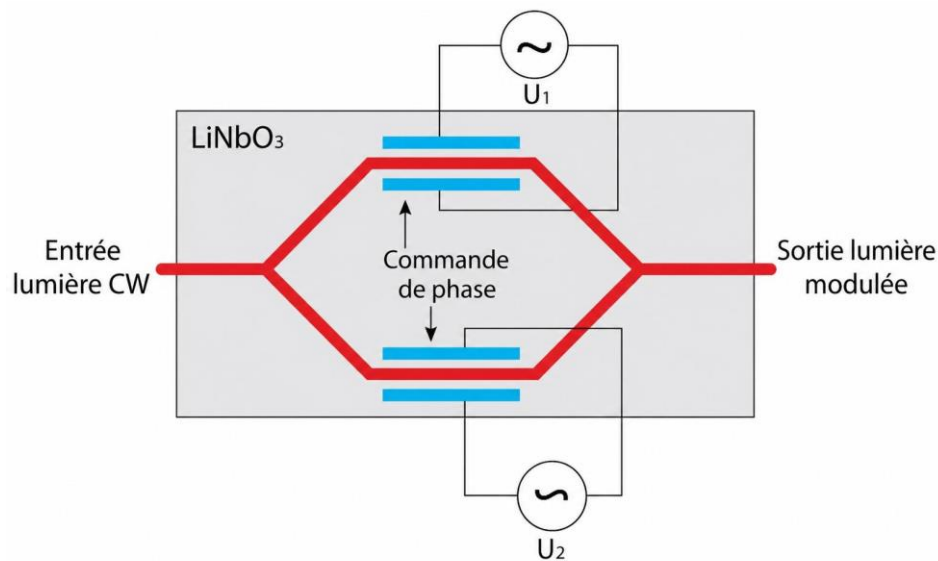


Figure II.3 : Principe de fonctionnement d’un modulateur de Mach-Zehnder [17].

II.3.1.3 Les diodes électroluminescentes (LED) :

Les LED en anglais (*Light Emitting Diodes*) constituent une option économique pour les sources lumineuses dans les systèmes FSO à courte portée ou pour des applications de communication optique à basse consommation [19]. Elles émettent une lumière incohérente avec un spectre relativement large, ce qui limite leur utilisation à des débits de données modérés.

L’avantage principal des LED réside dans leur robustesse, leur longue durée de vie et leur faible coût de production par rapport aux diodes laser [20]. Cependant, en raison de la divergence importante de leur faisceau, elles ne sont généralement pas adaptées aux liaisons FSO terrestres de longue distance qui exigent une grande directivité pour traverser l’atmosphère [21].

II.3.1.4 Les diodes Laser :

Les diodes Laser, en anglais (*Light Amplification by Stimulated Emission of Radiation*) sont les sources de prédilection pour les systèmes FSO performants. Elles produisent une lumière cohérente, monochromatique et extrêmement directionnelle, permettant de concentrer l'énergie lumineuse en un faisceau étroit capable de parcourir plusieurs kilomètres [17].

Caractéristique	LED	LASER
Puissance de sortie	Faible	Élevée
Vitesse de modulation	Limitée (jusqu'à 1 Gbps)	Très élevée (> 10 Gbps)
Directivité du faisceau	Large (Divergent)	Très étroite (Directif)
Portée de transmission	Courte (quelques mètres)	Longue (plusieurs kilomètres)
Coût	Faible	Élevée
Complexité du système	Simple	Complexe (nécessite un alignement)

Tableau II.1 : Comparaison entre les sources LED et LASER

Ce tableau montre que bien que les LED soient moins coûteuses, le Laser reste la source privilégiée pour les liaisons FSO à longue distance en raison de sa forte directivité et de sa capacité à supporter des débits très élevés [17].

II.3.2 Canal de transmission :

Le canal de transmission dans un système FSO est l'atmosphère terrestre. Contrairement à la fibre optique où le milieu est contrôlé et protégé, le canal atmosphérique est un milieu variable et imprévisible [18]. La performance de la liaison dépend directement des interactions entre les photons du faisceau laser et les particules présentes dans l'air (molécules de gaz, aérosols, gouttes d'eau, etc.) [17].

Lors de la propagation, le signal optique subit deux phénomènes majeurs : l'atténuation (perte de puissance) et la scintillation (fluctuation de l'intensité) [20]. Ces dégradations sont causées par l'absorption moléculaire, la diffusion (Mie et Rayleigh) et les turbulences optiques dues aux variations de température et de pression de l'air [23]. La compréhension de ce canal est essentielle pour concevoir des systèmes capables de maintenir une connectivité stable malgré les aléas climatiques [19].

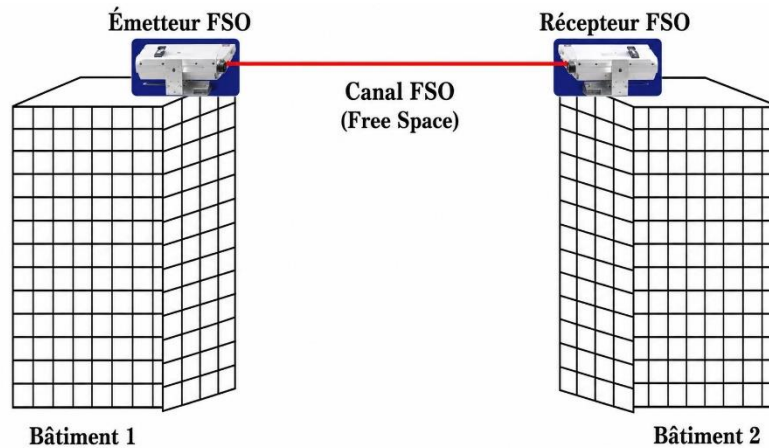


Figure II.4 : Architecture d'une liaison FSO terrestre point à point entre deux bâtiments [18].

II.3.2.1 La fibre optique :

L'une des comparaisons les plus pertinentes dans le domaine des communications optiques est celle entre la transmission en espace libre (FSO) et la transmission par fibre optique [9]. Une fibre optique peut être définie comme un guide d'onde diélectrique de forme cylindrique, conçu pour transporter des signaux lumineux sur de longues distances. Elle est structurellement composée de trois couches concentriques essentielles : le cœur (*Core*), qui constitue le milieu de propagation du signal lumineux ; la gaine (*Cladding*), qui entoure le cœur pour confiner la lumière par réflexion ; et le revêtement de protection (*Coating*), qui assure la résistance mécanique et la protection contre les agressions extérieures [22].

Le principe de fonctionnement de la fibre optique repose sur le phénomène de la réflexion totale interne. Contrairement au canal FSO où le faisceau se propage librement et subit une divergence naturelle dans l'air, la fibre assure un confinement physique de l'énergie lumineuse. Ce guidage est rendu possible par la différence de densité optique, ou indice de réfraction, entre le

matériau du cœur et celui de la gaine. Cette propriété physique force les rayons lumineux à rebondir continuellement à l'intérieur du cœur, permettant ainsi de transporter l'information avec des pertes d'énergie extrêmement faibles, même sur des distances s'étendant sur des dizaines de kilomètres [18] [21].

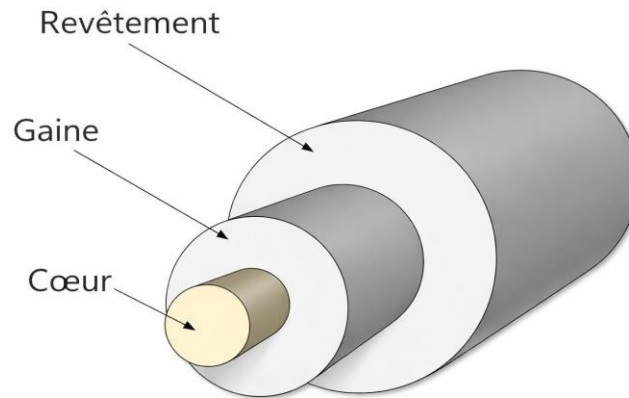


Figure II.5 : Structure d'une fibre optique. [22]

Dans le contexte d'une infrastructure réseau, la technologie FSO est souvent qualifiée de “ fibre optique sans fil ” (*Fiber optics without fiber*) [17]. Tandis que la fibre optique nécessite une installation physique lourde, incluant des travaux de génie civil et de câblage parfois impossibles dans des zones urbaines denses ou des terrains accidentés, le FSO offre une alternative de connectivité rapide et flexible. Cependant, la distinction majeure réside dans la protection du milieu : là où la fibre est isolée des perturbations climatiques par ses couches protectrices, le lien FSO doit traverser une atmosphère changeante et faire face à des obstacles tels que le brouillard, la pluie et la turbulence, ce qui constitue son principal défi technique [19].

II.3.3 Le Récepteur :

Le récepteur est l'élément terminal de la liaison FSO. Sa fonction principale est d'extraire l'information transmise par le faisceau laser après sa propagation à travers le canal atmosphérique [17]. Étant donné que le signal reçu est souvent affaibli par l'atténuation et perturbé par le bruit de fond (lumière ambiante, rayonnement solaire), le récepteur doit posséder une sensibilité élevée et un rapport signal sur bruit (SNR) optimal pour garantir la fiabilité de la communication [18].

Le bloc de réception se compose généralement de trois éléments majeurs : l'optique de collecte, le photodétecteur et l'électronique de traitement du signal [21].

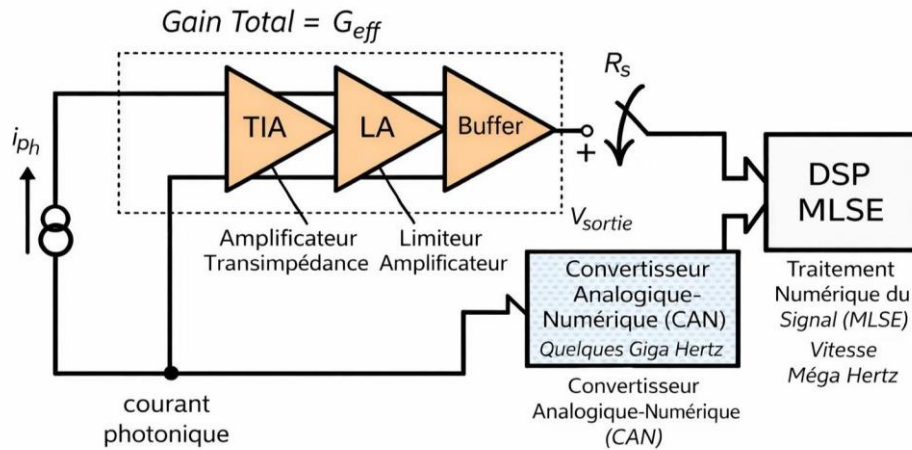


Figure II.6 : Schéma du récepteur optique [21].

II.3.3.1 L'optique de collecte :

L'optique de réception, souvent constituée de lentilles ou d'un télescope, joue le rôle de "collecteur de photons". Sa tâche est de concentrer le faisceau laser incident sur la surface active du photodétecteur [19]. Plus la surface de l'optique de collecte est grande, plus le système peut capter d'énergie, ce qui permet de compenser une partie de l'atténuation atmosphérique [20].

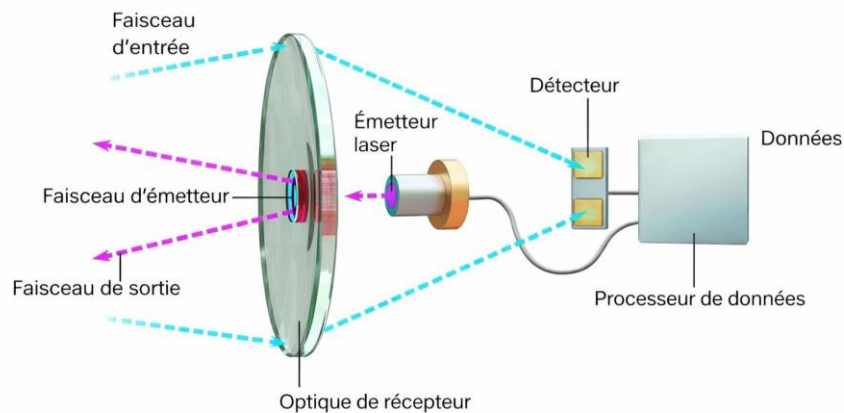


Figure II.7 : Système de collecte optique (télescope) [19].

Cependant, un champ de vision (FOV en anglais *Field of View*) trop large peut laisser passer une quantité importante de bruit optique parasite. Pour y remédier, des filtres optiques passe-bande sont installés pour ne laisser passer que la longueur d'onde spécifique du laser émetteur (par exemple 1550 nm), améliorant ainsi la sélectivité du système [17].

II.3.3.2 Le photodétecteur :

Le photodétecteur est le cœur du récepteur. Il assure la conversion optoélectronique en transformant les photons reçus en un courant électrique proportionnel à l'intensité lumineuse [21]. Dans les systèmes FSO, on utilise principalement deux types de photodiodes :

- **La photodiode PIN** : Elle est appréciée pour sa stabilité, son faible coût et sa facilité de mise en œuvre, bien qu'elle ne possède pas de gain interne [22].
- **La photodiode à avalanche (APD)** : Elle est beaucoup plus sensible que la diode PIN grâce à un processus d'ionisation par impact qui génère un gain interne. Elle est donc privilégiée pour les liaisons de longue distance où le signal reçu est très faible [9] [21].

II.3.3.3 Le traitement du signal :

Une fois le courant électrique généré par le photodétecteur, il passe par un amplificateur de transimpédance pour convertir le courant en une tension exploitable. [18] Ensuite, le signal est filtré électroniquement pour éliminer le bruit résiduel, puis dirigé vers un circuit de décision qui régénère les bits originaux (0 et 1) [19]. Enfin, le signal est démodulé pour restituer l'information initiale (données, voix ou vidéo) avec le minimum d'erreurs possible [17].

II.3.3.4 Techniques de photo-détection :

La photo-détection est le processus par lequel l'énergie optique collectée est convertie en un signal électrique. Dans les systèmes de communication optique en espace libre, deux techniques principales sont mises en œuvre : la détection directe (*Direct Detection*) et la détection cohérente (*Coherent Detection*) [17] [21].

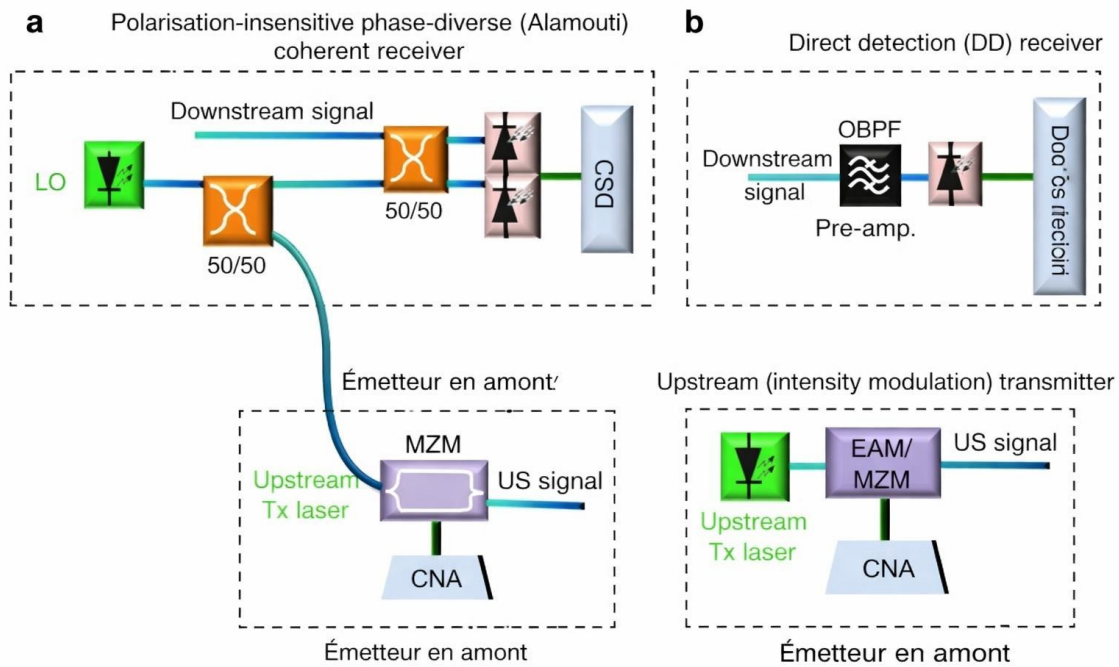


Figure II.8 : Détection directe vs détection cohérente [21]

A. Détection Directe (DD) :

La détection directe est la technique la plus simple et la plus largement déployée dans les systèmes FSO actuels, particulièrement avec la modulation d'intensité (IM/DD). Dans ce schéma, le photodétecteur agit comme un dispositif quadratique qui répond uniquement aux variations de la puissance optique (l'intensité) du faisceau reçu [18].

- Avantages : Simplicité de conception, faible coût, et robustesse vis-à-vis des fluctuations de phase causées par la turbulence atmosphérique.
- Inconvénients : Sensibilité limitée, car elle ne permet pas de récupérer les informations contenues dans la phase ou la fréquence de l'onde porteuse [19].

B. Détection Cohérente :

La détection cohérente est une technique plus avancée où le signal optique reçu est mélangé à un signal provenant d'un oscillateur local (un laser interne au récepteur) avant d'atteindre le

photodétecteur [21]. Ce mélange permet de transposer le signal optique vers une fréquence intermédiaire électrique plus basse.

- **Les avantages** : Une sensibilité nettement supérieure (permettant des liaisons plus longues), une meilleure réjection du bruit de fond solaire, et la capacité d'utiliser des formats de modulation complexes (PSK, QAM) en exploitant la phase et la fréquence de la lumière [21].
- **Les inconvénients** : Grande complexité technologique, coût élevé, et nécessité de dispositifs de compensation pour corriger les distorsions de phase induites par l'atmosphère [17].

II.4 Les avantages et les inconvénients de la technologie FSO :

La technologie FSO s'est imposée comme une solution de connectivité optique sans fil prometteuse, capable de combler le fossé entre les réseaux à fibre optique et les systèmes radiofréquences (RF) conventionnels [18] . Cependant, comme toute technologie de transmission, elle présente un ensemble d'atouts majeurs et de limitations techniques qu'il convient d'analyser [17].

II.4.1 Les avantages :

Les points forts de la technologie FSO résident principalement dans sa capacité à offrir des performances optiques avec la flexibilité du sans-fil :

- **Large bande passante et hauts débits** : Grâce à l'utilisation de fréquences optiques très élevées (térahertz), le FSO permet d'atteindre des débits binaires de l'ordre du Gigabit par seconde (Gbps), comparables à ceux de la fibre optique [21].
- **Absence de licence spectrale** : Contrairement aux systèmes RF (Wi-Fi, 4G/5G) qui nécessitent des autorisations gouvernementales coûteuses, le spectre optique est libre d'utilisation partout dans le monde [19].
- **Sécurité intrinsèque** : Le faisceau laser est extrêmement étroit et directionnel, ce qui rend l'interception du signal presque impossible sans interrompre physiquement la liaison (détection d'intrusion immédiate) [17].

- Immunité aux interférences électromagnétiques : Le signal optique n'est pas perturbé par les ondes radio environnantes, ce qui permet au FSO de fonctionner sans interférence dans des zones saturées en signaux RF [18].
- Faible coût et rapidité de déploiement : L'installation ne nécessite pas de travaux de génie civil (creusement de tranchées), ce qui réduit considérablement le temps et le coût de mise en service [20].

II.4.2 Les inconvénients :

Malgré ses avantages, la propagation en espace libre impose des contraintes physiques non négligeables :

- Sensibilité aux conditions météorologiques : Le brouillard est l'ennemi principal du FSO, car il provoque une atténuation massive du signal par diffusion. La pluie et la neige affectent également la disponibilité du lien [23].
- Turbulences atmosphériques : Les variations de température de l'air créent des poches de réfraction variables (scintillation), provoquant des fluctuations rapides de l'intensité du signal reçu (*fading*) [17].
- Alignement et pointage : La directivité du faisceau exige un alignement parfait entre l'émetteur et le récepteur. Les mouvements des bâtiments dus au vent ou aux vibrations peuvent rompre la liaison [19].
- Obstacles physiques : Une liaison FSO nécessite une ligne de vue directe et dégagée. Tout obstacle temporaire (oiseaux, arbres, nouveaux bâtiments) interrompt immédiatement la transmission [18].
- Sécurité oculaire : La puissance d'émission est limitée par les normes de sécurité internationales pour éviter tout dommage aux yeux des personnes qui pourraient traverser le faisceau par inadvertance [21].

II.5 Applications de la technologie FSO :

La technologie FSO est polyvalente et s'adapte à divers scénarios de communication là où la fibre optique est difficile à installer. Ses principales applications incluent : [17] [18]

- Le “Dernier Kilomètre” (*Last Mile Access*) : Connecter les utilisateurs finaux au réseau central sans avoir à creuser des tranchées pour la fibre optique [19].
- Interconnexion de réseaux LAN-to-LAN : Relier des bâtiments d’entreprises ou des campus universitaires à haut débit [21].
- Récupération après sinistre (*Disaster Recovery*) : Déploiement rapide de liaisons de secours en cas de rupture de câbles physiques lors de catastrophes naturelles [17].
- Communications militaires et spatiales : Utilisation pour des liaisons sécurisées entre navires, avions ou satellites grâce à l’étroitesse du faisceau laser [20].

II.6 Classification des réseaux FSO :

Les systèmes de communication optique en espace libre (FSO) peuvent être classés en plusieurs catégories en fonction de leur portée, de l’environnement de déploiement et des applications spécifiques [18] . On distingue généralement trois grandes familles de réseaux :

II.6.1 Réseau satellitaire optique sans fil :

Le réseau OWSN en anglais (*Optical Wireless Satellite Network*) concerne les communications optiques au-delà de l’atmosphère terrestre [17]. Ce domaine inclut les liaisons inter-satellites (ISL) ainsi que les communications entre satellites et stations au sol [21]. En l’absence d’atmosphère dans l’espace, ces réseaux ne subissent pas d’atténuation météorologique, ce qui permet d’atteindre des débits extrêmement élevés sur des distances de plusieurs milliers de kilomètres [17].

II.6.2 Réseaux sans fil optiques terrestres :

Les réseaux OWTN en anglais (*Optical Wireless Terrestrial Networks*) représentent l’application la plus courante de la technologie FSO. Ils consistent à établir des liaisons point à point entre des bâtiments ou des infrastructures fixes au sol à travers l’atmosphère. Ces réseaux sont principalement utilisés pour résoudre la problématique du “Dernier Kilomètre” en anglais (*Last Mile*), qui désigne la partie finale du réseau de télécommunications qui relie le nœud de distribution principal (généralement le réseau de fibres optiques à haut débit) à l’abonné final

(entreprise, domicile ou bâtiment). Bien que le nom suggère une distance d'un kilomètre, il représente en réalité toute liaison de courte distance nécessaire pour acheminer les données vers l'utilisateur.

Le FSO permet ainsi de franchir ce goulot d'étranglement sans nécessiter de travaux de génie civil coûteux, tout en offrant des débits de l'ordre du Gigabit par seconde. Cependant, ces liaisons terrestres sont très sensibles aux conditions météorologiques locales, telles que le brouillard et la pluie, qui peuvent dégrader la qualité du signal.

II.6.3 Réseau domestique optique sans fil :

Le réseau OWHN en anglais (*Optical Wireless Home Network*) s'applique aux communications de courte portée à l'intérieur des bâtiments (Indoor) [21]. Cette classification inclut des technologies comme le Li-Fi (*Light Fidelity*), où l'éclairage LED est utilisé pour transmettre des données vers des appareils mobiles [18]. Contrairement aux réseaux terrestres extérieurs, les réseaux domestiques ne sont pas affectés par la météo, mais doivent faire face aux obstacles physiques et à la lumière ambiante [21].

II.7 Les conditions atmosphériques sur les liaisons FSO :

La performance et la disponibilité d'une liaison FSO sont intrinsèquement liées aux caractéristiques du canal de propagation atmosphérique [18]. L'atmosphère est un milieu dynamique composé de divers gaz, de particules d'eau et d'aérosols qui interagissent avec le faisceau laser [17]. Ces interactions provoquent une dégradation du signal optique, principalement à travers des phénomènes d'absorption et de diffusion, ce qui réduit la puissance reçue au niveau du détecteur [21].

II.7.1 Atténuation par le brouillard :

Le brouillard est considéré comme l'obstacle le plus sévère pour les communications FSO terrestres [19]. Il est composé de fines gouttelettes d'eau dont le diamètre est comparable à la longueur d'onde des lasers utilisés (proche de 1550 nm), ce qui déclenche un phénomène de diffusion de Mie [17] [23].

La diffusion de Mie est un phénomène physique qui se produit lorsque la lumière interagit avec des particules en suspension dont la taille est approximativement égale ou supérieure à sa propre longueur d'onde. Dans le contexte des systèmes FSO, ce processus survient principalement lorsque le faisceau laser traverse des milieux denses comme le brouillard ou les nuages.

- Ce phénomène disperse l'énergie lumineuse dans toutes les directions, entraînant une chute drastique de la puissance du signal [18].
- Contrairement à la pluie, le brouillard peut persister pendant de longues périodes, provoquant des interruptions prolongées de la liaison si la marge de puissance du système est insuffisante [20].

II.7.2 Atténuation par la neige :

La neige influence la transmission FSO de manière complexe en raison de la grande variété de formes et de tailles des flocons [21].

- Les flocons de neige, étant plus volumineux que les gouttelettes de brouillard, provoquent une atténuation géométrique importante [18].
- On distingue généralement la "neige sèche", qui cause une diffusion modérée, et la "neige mouillée", qui, en raison de sa teneur élevée en eau, peut provoquer des pertes de signal extrêmement élevées dépassant parfois celles de la pluie battante [17] [23].

II.7.3 Atténuation par la pluie :

Bien que la pluie soit souvent perçue comme un obstacle majeur, son impact sur le FSO est généralement moins sévère que celui du brouillard [19].

- Les gouttes de pluie sont beaucoup plus grandes que la longueur d'onde du faisceau laser, ce qui limite l'effet de diffusion de Mie au profit d'une atténuation par blocage spatial (diffusion non sélective) [21].
- L'atténuation par la pluie est directement proportionnelle à l'intensité des précipitations (mesurée en mm/h) [18]. Bien qu'elle réduise la portée de la liaison, elle conduit rarement à une interruption totale du service, sauf en cas d'averses torrentielles exceptionnelles [20].

II.7.4 Atténuation par les gaz atmosphériques (Absorption) :

Même par temps clair, l'atmosphère n'est pas totalement transparente au rayonnement laser. Elle contient divers gaz, tels que la vapeur d'eau (H_2O), le dioxyde de carbone (CO_2) et l'ozone (O_3), qui absorbent une partie de l'énergie photonique du faisceau [18]. Ce phénomène d'absorption dépend fortement de la longueur d'onde utilisée ; les molécules de gaz capturent l'énergie lumineuse et la transforment en chaleur, réduisant ainsi l'intensité du signal reçu [21]. Pour minimiser ces pertes, les systèmes FSO sont conçus pour fonctionner dans des "fenêtres de transmission atmosphérique" spécifiques, comme la bande des 1550 nm, où l'absorption moléculaire est naturellement plus faible, permettant ainsi une propagation sur de plus longues distances [17] [19].

II.7.5 Turbulence atmosphérique (Scintillation) :

La turbulence atmosphérique est l'un des défis les plus imprévisibles pour les liaisons FSO, résultant des variations dynamiques de l'indice de réfraction de l'air dues aux changements de température et de pression [23]. Ce phénomène crée des poches d'air de tailles différentes, appelées cellules de turbulence, qui agissent comme des lentilles mouvantes déviant le faisceau laser [17].

- Ce processus engendre la "scintillation", qui se manifeste par des fluctuations rapides et aléatoires de l'intensité lumineuse au niveau du récepteur, similairement au scintillement des étoiles [19].
- Outre la perte de puissance, la turbulence peut provoquer un élargissement du faisceau (*beam broadening*) ou un déplacement du centre du faisceau (*beam wander*), rendant le maintien de l'alignement entre l'émetteur et le récepteur extrêmement difficile sans systèmes de suivi actifs ou algorithmes de correction intelligents [18] [21].

II.7.6 Atténuation par les aérosols (Fumée et Poussière) :

En milieu urbain ou industriel, la présence d'aérosols tels que la fumée, la poussière et les polluants atmosphériques constitue une source supplémentaire d'atténuation [19]. Ces particules fines agissent de manière similaire aux gouttelettes de brouillard en provoquant une diffusion de

la lumière [18]. Bien que leur impact soit généralement localisé, une concentration élevée de poussière, notamment lors de tempêtes de sable dans certaines zones géographiques, peut entraîner une dégradation significative de la visibilité et, par conséquent, une augmentation brutale du taux d'erreur binaire (BER) de la liaison optique [20] [23].

II.8 Valeurs typiques d'atténuation atmosphérique pour les liaisons FSO :

L'impact des conditions atmosphériques sur la propagation du faisceau laser se traduit par une perte de puissance exprimée en décibels par kilomètre (dB/km) [18]. Ces valeurs varient considérablement selon l'opacité du milieu et la densité des particules présentes dans le canal de transmission [17] . Le tableau suivant résume les ordres de grandeur de l'atténuation pour les conditions les plus courantes rencontrées dans la littérature technique : [19] [21]

Condition atmosphérique	Atténuation typique (dB/km)
Ciel clair (<i>Clear Air</i>)	0.2 à 3
Pluie faible (<i>Light Rain</i>)	4 à 15
Pluie forte (<i>Heavy Rain</i>)	20 à 30
Brouillard léger (<i>Haze</i>)	5 à 40
Brouillard dense (<i>Thick Fog</i>)	40 à 300+
Neige faible (<i>Light Snow</i>)	30 à 50
Neige forte (<i>Heavy Snow</i>)	50 à 200+
Pollution et Fumée	2 à 10
Turbulences (Scintillation)	Fluctuations (2 à 10+)

Tableau II.2 : Classification des conditions atmosphériques selon l'atténuation [17].

Il est important de noter que ces valeurs sont indicatives et peuvent fluctuer en fonction de la visibilité locale et de la longueur d'onde utilisée [21] . Par exemple, dans des conditions de brouillard extrêmement dense, l'atténuation peut dépasser les 400 dB/km, ce qui peut provoquer une rupture totale de la liaison si la marge de puissance n'est pas suffisante [20] [23].

Cette variabilité extrême et souvent imprévisible du canal atmosphérique souligne la nécessité d'intégrer des solutions intelligentes [17]. L'apprentissage machine permet non seulement de prédire ces pertes de puissance en temps réel, mais aussi d'adapter dynamiquement les paramètres de transmission (tels que la puissance d'émission ou le format de modulation) afin de garantir une connectivité résiliente face aux aléas climatiques [19] .

II.9 Applications de l'apprentissage machine sur les liaisons FSO :

L'intégration de l'apprentissage machine (ML) dans les systèmes FSO est devenue une nécessité pour surmonter l'imprévisibilité du canal atmosphérique [17]. Ces algorithmes permettent de transformer les données brutes issues du récepteur en informations prédictives permettant d'optimiser la liaison en temps réel [19]. Les paragraphes suivants détaillent les principales contributions du ML dans ce domaine:

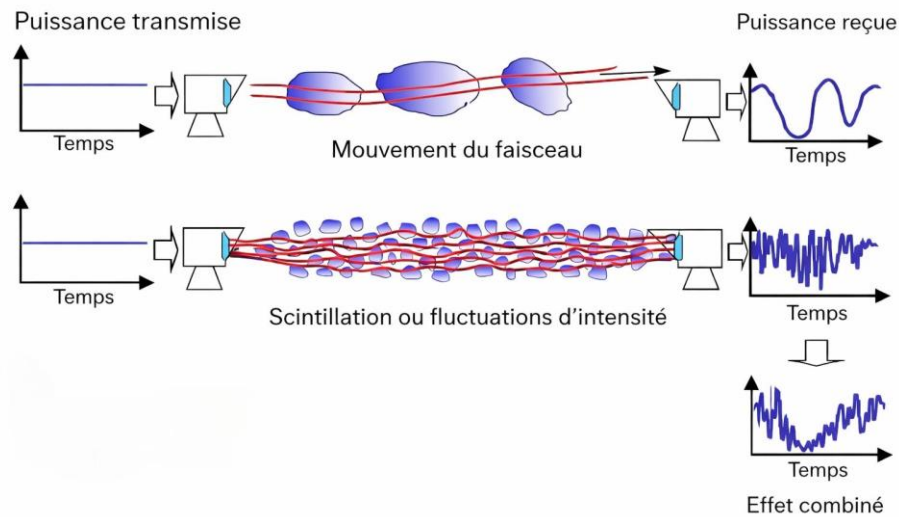
II.9.1 Prédiction de l'atténuation atmosphérique :

Dans le domaine des communications FSO, plusieurs travaux de recherche ont exploité les réseaux de neurones de type LSTM (Long Short-Term Memory) afin de prédire l'évolution de l'atténuation atmosphérique à partir de données météorologiques historiques. Le principe consiste à alimenter le modèle par une série temporelle de paramètres tels que la visibilité, l'humidité relative, la température et l'intensité des précipitations. Grâce à sa capacité à mémoriser les dépendances temporelles sur de longues périodes, le réseau LSTM apprend la relation existante entre les variations climatiques et les pertes de puissance observées sur la liaison optique. Après la phase d'apprentissage, le modèle est capable d'estimer les futures valeurs d'atténuation avec une précision élevée, permettant ainsi d'anticiper les dégradations du canal et d'adapter les paramètres de transmission avant l'apparition d'une rupture de la liaison. Cette approche contribue à améliorer la disponibilité et la fiabilité des systèmes FSO dans des environnements atmosphériques variables [17].

II.9.2 Atténuation de la scintillation et des turbulences :

Les turbulences atmosphériques induisent des variations rapides et aléatoires de l'indice de réfraction de l'air, provoquant un phénomène de scintillation qui dégrade le rapport signal sur bruit (SNR) [23]. Pour contrer cet effet, des algorithmes d'apprentissage profond en anglais (*Deep Learning*) sont utilisés comme des égaliseurs de canal adaptatifs performants. En s'entraînant sur des échantillons de signaux perturbés, ces modèles apprennent à identifier les motifs de distorsion et à prédire la fonction de transfert du canal en temps réel. Cette approche permet de compenser dynamiquement les fluctuations d'amplitude et de phase, stabilisant ainsi le flux de données reçu même dans des conditions de turbulence forte [17] [19].

Figure II.9 : Effet de la scintillation atmosphérique [23].



II.9.3 Reconnaissance automatique des formats de modulation :

Dans les réseaux FSO modernes et intelligents, le format de modulation doit s'adapter dynamiquement à l'état du canal pour maximiser l'efficacité spectrale [21]. Le ML joue un rôle clé au niveau du récepteur à travers la classification automatique des modulations en anglais (*Automatic Modulation Classification – AMC*). Les réseaux de neurones à convolution (CNN) sont souvent exploités pour extraire des caractéristiques complexes des constellations de signaux reçus afin de les classer sans aucune information préalable. Cette capacité d'auto-reconnaissance permet

de basculer de manière fluide entre des modulations de haut niveau et des modulations plus robustes en fonction de la qualité du lien optique [19].

Le CNN (Réseau de Neurones à Convolution) est une architecture d'apprentissage profond (Deep Learning) inspirée du cortex visuel humain, initialement conçue pour la reconnaissance d'images [2]. Dans le domaine du FSO, les CNN sont utilisés pour extraire des caractéristiques complexes à partir des représentations temporelles ou fréquentielles des signaux reçus (comme les diagrammes de constellation) afin de classer les modulations avec une précision supérieure aux méthodes statistiques classiques [17] [19].

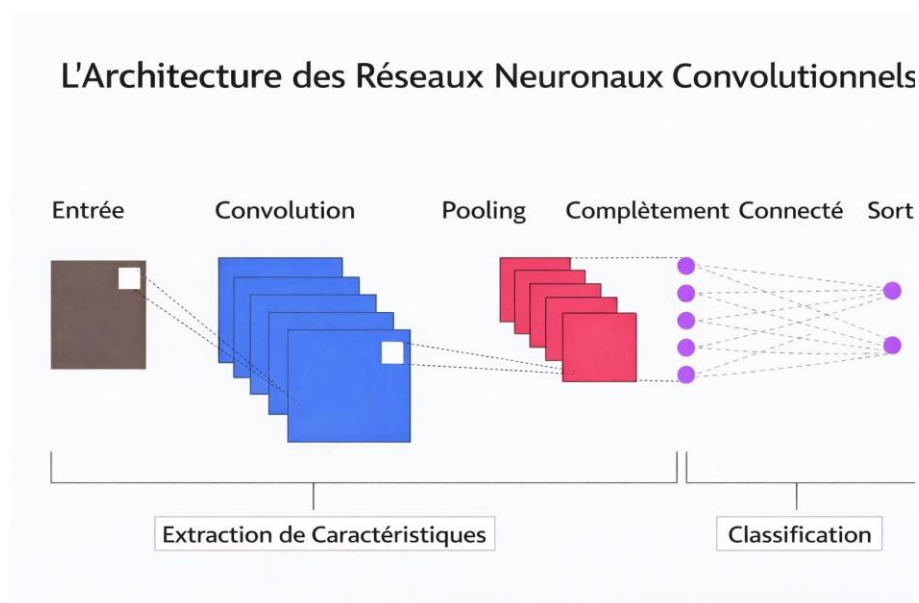


Figure II.10 : Architecture d'un réseau de neurones convolutionnel (CNN) [2].

II.9.4 Optimisation du pointage et de la poursuite :

Le maintien d'un alignement parfait entre l'émetteur et le récepteur est l'un des défis majeurs du FSO, car de faibles erreurs de pointage peuvent entraîner des pertes de signal massives. L'utilisation de l'apprentissage par renforcement permet de concevoir des systèmes de poursuite intelligents capables de s'auto-ajuster. Ces modèles apprennent par interactions successives à corriger les déviations du faisceau causées par les micro-vibrations des bâtiments ou les effets du vent. En optimisant continuellement l'orientation des miroirs ou des lentilles, ces algorithmes

réduisent significativement les erreurs de pointage et garantissent une réception stable sur de longues distances [17] [21].

II.10 Conclusion :

En conclusion, ce chapitre a présenté les fondements et les défis majeurs des systèmes FSO [18]. Bien que cette technologie offre des performances exceptionnelles en ciel clair, sa sensibilité aux aléas climatiques reste un obstacle critique à sa fiabilité constante [17]. L'intégration de l'apprentissage machine apparaît alors comme une évolution nécessaire pour transformer ces liaisons en systèmes intelligents capables d'anticiper et de compenser les perturbations du canal atmosphérique. Cette étude théorique constitue le socle indispensable pour le développement de modèles prédictifs performants dans la suite de ce travail.

Chapitre III

Simulation De La Liaison FSO Et

Classification Par K-NN

III.1 Introduction :

Ce chapitre constitue la phase pratique de notre étude, où nous faisons le lien entre les concepts théoriques de la transmission optique sans fil (FSO) et l'intelligence artificielle représentée par l'algorithme des K-Plus Proches Voisins (K-NN). L'objectif principal est de concevoir un modèle capable de diagnostiquer l'état du canal de propagation en se basant sur les performances physiques du signal.

Pour ce faire, nous avons structuré notre démarche en trois étapes clés : d'abord, la simulation d'une liaison FSO robuste sous l'environnement OptiSystem pour générer une base de données représentative, ensuite le prétraitement et la normalisation de ces données via MATLAB et enfin l'entraînement et l'évaluation du classifieur K-NN. Cette approche nous permettra d'automatiser la détection des dégradations atmosphériques et d'évaluer la fiabilité de l'apprentissage automatique dans des conditions climatiques variables.

III.2 Environnement de simulation et outils de travail :

La réalisation de ce projet a nécessité l'utilisation de deux plateformes logicielles complémentaires (OptiSystem / Matlab); chacune intervenant à un stade précis de la chaîne de traitement.

III.2.1 Le logiciel OptiSystem :

OptiSystem est un logiciel de conception de systèmes de communication optique développé par Optiwave. Il est largement reconnu dans le milieu académique et industriel pour sa capacité à simuler des couches physiques complexes.

- Nous avons utilisé ce logiciel pour modéliser la liaison FSO. Il nous permet de définir les paramètres de l'émetteur (Laser), les caractéristiques du canal atmosphérique (Atténuation, Distance) et la sensibilité du récepteur.
- La force d'OptiSystem réside dans ses outils d'analyse visuelle, tels que le BER Analyzer et l'Eye Diagram, qui fournissent des mesures précises du facteur de qualité (Q-Factor), essentielles pour constituer notre Dataset.

III.2.2 Le logiciel MATLAB :

MATLAB (Matrix Laboratory) est une plateforme de calcul numérique et de programmation indispensable pour la manipulation de données massives. Dans notre étude MATLAB intervient après la phase de simulation pour deux fonctions critiques :

1. Prétraitement (Scripting) : Pour importer les 4000 valeurs générées et appliquer l'algorithme de normalisation Z-Score afin d'unifier l'échelle des données.
2. Classification (App Machine Learning) : L'utilisation de l'application Classification Learner a permis d'implémenter la variante Coarse K-NN, de réaliser l'entraînement et de générer les métriques de performance comme la matrice de confusion et les courbes ROC.

III.3 Conception du système de transmission FSO :

Afin d'analyser les performances de la liaison optique et de collecter les données nécessaires à notre étude, nous avons conçu un modèle de simulation complet sous l'environnement OptiSystem. Ce modèle reproduit une chaîne de transmission FSO typique, intégrant les blocs essentiels allant de la génération du signal optique jusqu'à sa détection et son analyse en réception.

Dans cette section, nous détaillons l'architecture de la liaison FSO simulée. Le système est composé de trois blocs principaux et la figure ci-dessous (Figure III.1) illustre l'interface globale de notre projet ainsi que l'architecture du système mis en œuvre :

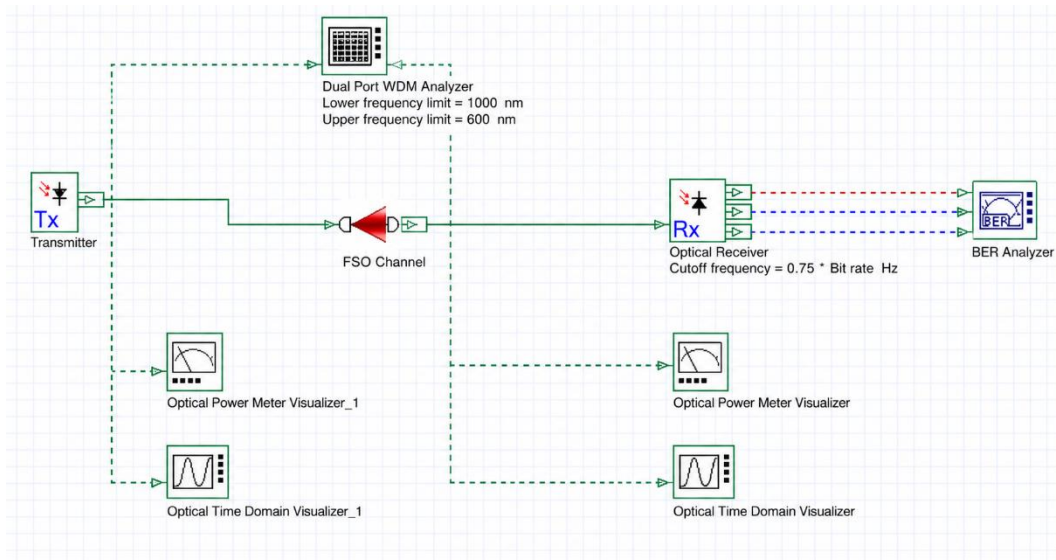


Figure III.1 : Architecture globale du système FSO.

III.3.1 Bloc Émetteur (Transmitter) :

Le bloc émetteur sert à transformer le signal électrique contenant l'information en un signal optique, il se compose de :

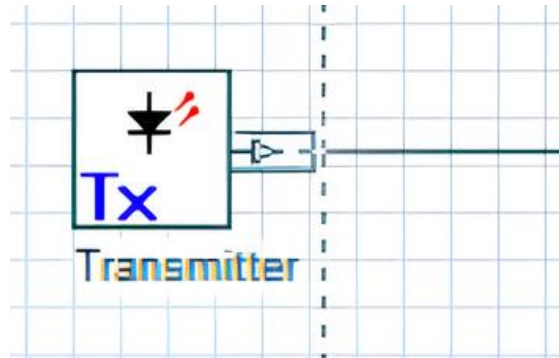


Figure III.2 : Bloc émetteur.

- Pseudo-Random Bit Sequence Generator : Génère une suite binaire aléatoire représentant les données à transmettre.
- NRZ Pulse Generator : Code les bits en impulsions électriques selon le format “Non-Return-to-Zero”.
- CW Laser : La source optique fonctionnant à une longueur d’onde de 785 nm. Cette fenêtre est choisie pour sa faible atténuation et sa compatibilité avec les standards de télécommunication.
- Mach-Zehnder Modulator (MZM) : Un modulateur externe qui imprime le signal électrique sur le faisceau laser porteur.

III.3.2 Canal de transmission (FSO Channel) :

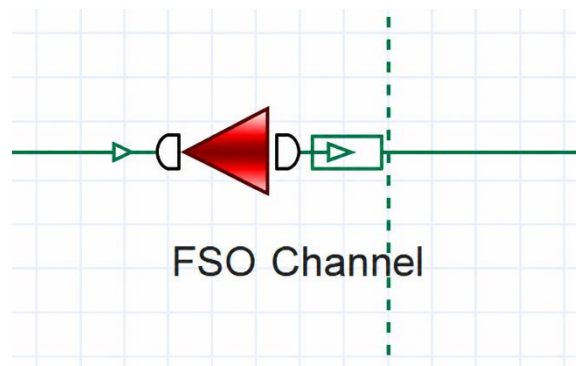


Figure III.3 : Bloc du canal FSO.

Le canal FSO représente l'espace libre (l'atmosphère) à travers lequel le faisceau laser se propage. Dans OptiSystem, ce composant simule les effets de l'environnement sur le signal optique. Deux phénomènes majeurs y sont modélisés :

- L'atténuation géométrique : liée à la divergence du faisceau laser avec l'augmentation de la distance.
- L'atténuation atmosphérique : provoquée par l'interaction du signal avec les molécules d'air et les hydrométéores (pluie, brouillard, neige).

Le canal est ainsi caractérisé par sa capacité à basculer entre des conditions de « Ciel clair » (faible atténuation) et des conditions de « Brouillard ou air turbide » (forte atténuation), ce qui impacte directement la puissance reçue et le facteur de qualité (Q-Factor).

III.3.3 Bloc Récepteur (Receiver) :



Figure III.4 : Bloc récepteur dans OptiSystem.

Le bloc récepteur a pour mission de récupérer le signal optique atténué par le canal atmosphérique et de le convertir à nouveau en un signal électrique exploitable. Ce bloc est le point le plus critique pour l'extraction de nos données. Il se compose des éléments suivants :

- Aperture du Récepteur (Receiver Aperture) : Elle capte le faisceau lumineux en provenance de l'espace libre. Sa taille est optimisée pour collecter le maximum de photons malgré l'élargissement du faisceau.
- PIN Photodetector : Il assure la conversion opto-électronique. Nous avons opté pour une photodiode PIN en raison de sa grande bande passante et de sa sensibilité adaptée à la longueur d'onde de 785 nm.
- Low Pass Bessel Filter : Ce filtre permet de supprimer le bruit de grenaille (shot noise) et le bruit thermique tout en conservant l'intégrité de la forme d'onde du signal.

- 3R Regenerator : Ce composant est essentiel pour la synchronisation. Il effectue la réamplification, la mise en forme et le resynchronisation du signal électrique.
- Optical Time Domain BER Analyzer : Comme illustré dans la Figure III.4, cet instrument est relié à la sortie du régénérateur. Il nous permet de visualiser le diagramme de l'œil et d'extraire le Facteur Q, qui constitue la donnée brute de notre étude.

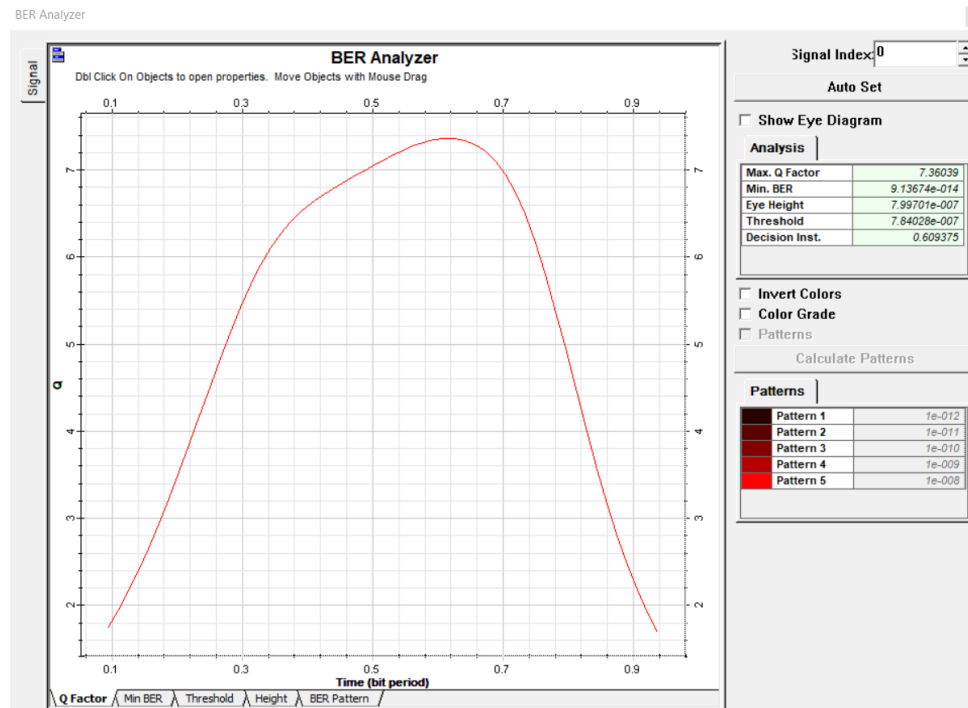


Figure III.5 : Paramètre du canal de propagation FSO.

L'analyse des résultats affichés dans la Figure III.5 montre des performances exceptionnelles du système de transmission. Le facteur de qualité atteint une valeur extrêmement élevée ($Q \approx 7.36$) associée à un taux d'erreur binaire (BER min = $9.1 \cdot 10^{-14}$).

Cette performance idéale s'explique par les conditions spécifiques de cette simulation, où nous avons configuré une distance courte (Range = 100 m) et une faible atténuation atmosphérique (0.7 dB/km). Dans ces conditions, la puissance du signal reste intacte, garantissant une communication sans erreur.

Il est important de noter que dans le domaine des communications optiques, une liaison est considérée comme réussie dès que le facteur Q dépasse le seuil de 6, ce qui correspond à un BER inférieur à 10^{-9} . Nos résultats dépassent largement ces normes, ce qui confirme la validité de notre conception avant d'entamer la variation des paramètres pour le dataset.

III.3.4 Outils d'analyse et de mesure :

Pour évaluer la qualité de la transmission et collecter les données nécessaires de notre étude, nous avons intégré plusieurs instruments de mesure virtuels :

- Dual Port WDM Analyzer : Cet outil permet de visualiser le spectre du signal optique et de mesurer la puissance du signal ainsi que le rapport signal sur bruit optique (OSNR) avant et après la transmission.
- Optical Power Meter : Utilisé pour mesurer la puissance optique précise en différents points de la liaison, ce qui nous aide à calculer les pertes totales subies dans le canal FSO.
- Optical Time Domain BER Analyzer : C'est l'outil le plus critique de notre simulation. Il permet d'extraire le Facteur Q et le BER (Bit Error Rate), et de visualiser le diagramme de l'œil (Eye Diagram), ce qui constitue la base de notre dataset.

Paramètre	Valeur	Unité
Longueur d'onde	785	nm
Puissance d'émission	0.03	dBm (mW)
Débit binaire	10	Gbps
Distance (Range)	100 à 10 095	m
Pas de variation (Step)	5	m
Atténuation (Ciel clair)	0.7	dB/km
Atténuation (Brouillard)	2.5	dB/km

Tableau III.1 : Paramètre de configuration du canal FSO.

III.4 Méthodologie de génération du dataset :

Pour construire un modèle de classification robuste, il est impératif de disposer d'une base de données riche et variée. Notre stratégie de simulation repose sur la variation systématique de la distance de propagation sous deux conditions atmosphériques distinctes.

III.4.1 Scénarios de simulation et paramètres variables :

Pour construire notre dataset, nous avons exploité la flexibilité du canal FSO en faisant varier ses paramètres de manière systématique :

1. Plage de distance (Range) : Nous avons configuré un balayage (Sweep) allant de 100 m à 10 095 m, avec un pas de 5 m, afin d'observer la dégradation du signal sur de longues distances.
2. Modélisation des classes :
 - Classe 0 (Ciel clair) : Simulée avec une atténuation fixe de 0.7 dB/km. Ce scénario a permis de générer 2000 échantillons représentatifs d'une liaison stable.
 - Classe 1 (Conditions dégradées) : Simulée avec une atténuation de 2.5 dB/km (brouillard léger). Ce scénario a généré 2000 autres échantillons pour capturer les moments de défaillance du système.

Disp	Name	Value	Units	Mode
☒	Range	100	m	Normal
☒	Attenuation	0.7	dB/km	Normal
☒	Attenuation	2.5	dB/km	Normal

Tableau III.2 : Configuration des paramètres du canal FSO dans OptiSystem.

III.4.2 Extraction des caractéristiques (Features Extraction) :

À la fin de chaque itération, le logiciel extrait automatiquement le Facteur Q (Q-Factor). Ce paramètre a été choisi comme variable d'entrée principale (Feature) pour notre algorithme K-NN, car il reflète directement la fiabilité de la liaison. Le résultat final est un fichier contenant 4000 observations, prêt à être traité sous MATLAB.

III.5 Traitement des données et classification sous MATLAB

Après avoir généré les 4000 échantillons de données via OptiSystem, l'étape suivante consiste à préparer ces données pour l'apprentissage machine. Cette phase est cruciale car la performance de l'algorithme K-NN dépend fortement de la qualité et de l'échelle des données d'entrée.

III.5.1 Prétraitement et normalisation des données :

Dans notre étude, nous utilisons le Facteur Q comme caractéristique principale (Feature). Cependant, les valeurs du Facteur Q peuvent varier considérablement en fonction de la distance et des conditions atmosphériques. Pour assurer une convergence rapide et éviter que les grandes

valeurs ne dominent le calcul de la distance dans l'algorithme K-NN, nous avons appliqué une technique de normalisation appelée Z-Score Normalization.

A. Algorithme de normalisation (Z-Score) :

Le principe de cette méthode est de transformer les données pour qu'elles aient une moyenne μ de 0 et un écart-type σ de 1. La formule mathématique appliquée est la suivante :

$$Z = \frac{x - \mu}{\sigma} \quad eq. 01$$

x : la valeur brute du Facteur Q.

μ : la moyenne de l'ensemble des données.

σ : l'écart-type de l'ensemble des données.

B. Implémentation sous MATLAB :

Comme illustré dans le script MATLAB que nous avons développé, nous avons automatisé ce processus :

- Chargement : Lecture du fichier Excel généré par OptiSystem.
- Calcul : Calcul de la moyenne et de l'écart-type pour les 4000 valeurs.
- Transformation : Application de la formule Z-Score à chaque échantillon pour obtenir de nouvelles valeurs normalisées.
- Exportation : Création d'un nouveau fichier contenant les données prêtes pour la classification.

III.5.2 Application de l'algorithme Coarse K-NN :

Pour la phase de classification, nous avons utilisé l'application Classification Learner de MATLAB. Cette interface permet d'explorer plusieurs variantes de l'algorithme K-NN.

A. Choix de la variante "Coarse K-NN" :

Nous avons opté pour le modèle Coarse K-NN (K-NN grossier). Contrairement au "Fine K-NN" qui utilise un petit nombre de voisins, le Coarse K-NN se distingue par :

- Nombre de voisins (K) : Généralement fixé à $K = 100$.
- Métrique de distance : Distance Euclidienne simple.
- Pondération (Weight) : Égale (tous les voisins ont la même influence).

Le choix de Coarse K-NN offre des frontières de décision lisses, ce qui le rend plus robuste face au bruit et plus performant pour distinguer globalement entre un “Ciel clair” et un “Environnement perturbé” sans risque de surapprentissage (Overfitting).

B. Processus d’entraînement (Training) :

- Données d’entrée : Les valeurs normalisées du Facteur Q.
- Cibles (Targets) : Classe 0 (Ciel clair) et Classe 1 (Brouillard).
- Validation : Nous avons utilisé la méthode de Validation Croisée (Cross-Validation) à 5 ou 10 plis pour garantir que le modèle est capable de généraliser sur de nouvelles données

Après avoir importé et normalisé notre Dataset, nous avons utilisé l’application Classification learner intégrée à MATLAB. Cette plateforme nous permet d’entraîner plusieurs modèles d’apprentissage supervisé et d’évaluer leur capacité à classer les données. La figure suivante (Figure III.6) illustre l’interface de travail ainsi que la configuration choisie pour l’algorithme Coarse K-NN :

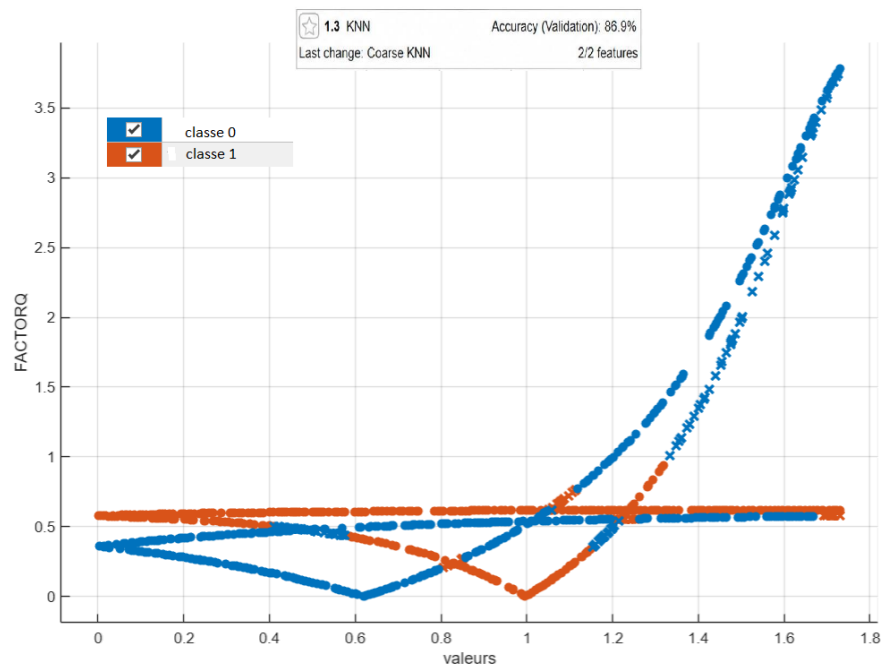


Figure III.6 : Environnement de classification et analyse des performances sous MATLAB.

La Figure III.6 (Scatter Plot) expose la cartographie bidimensionnelle de notre Dataset de 4000 échantillons après normalisation Z-Score, mettant en relation le facteur de qualité (FACTORQ) et la distribution des observations (Valeurs).

Ce graphique met en évidence un phénomène physique critique : le chevauchement ou (tressage) entre les données du ciel clair (Classe 0, en bleu) et celles du brouillard (Classe 1, en rouge). Ce comportement non-linéaire s'explique par l'interaction entre la distance (Range) et l'atténuation. À de courtes distances, le signal sous brouillard reste intense et simule un ciel clair. À l'inverse, à de très longues distances, l'atténuation géométrique affaiblit le signal en ciel clair, réduisant le facteur Q à des niveaux similaires à ceux d'un canal dégradé.

C'est précisément pour résoudre cette zone de confusion globale que l'algorithme Coarse K-NN est pertinent. En configurant un nombre étendu de voisins ($K = 100$), le classifieur génère des frontières de décision lisses et globales. Cette flexibilité géométrique permet de filtrer les ambiguïtés de transition et de stabiliser la prédiction, ce qui justifie l'obtention d'une précision remarquable de 86,9 %.

III.6 Analyse des résultats et discussion :

Dans cette dernière section, nous évaluons la performance du modèle Coarse K-NN à travers les outils d'analyse fournis par MATLAB. L'objectif est de quantifier la capacité de notre algorithme à classifier correctement les états de la liaison FSO.

III.6.1 Analyse de l'exactitude (Accuracy) :

Comme nous l'avons observé dans l'interface de classification (Figure III.6), le modèle a atteint une précision globale de 86,9 %. Ce résultat est jugé très satisfaisant compte tenu de la complexité du canal FSO et de la large plage de distances simulées (jusqu'à 10 km). Une précision de cet ordre démontre que le Facteur Q est un indicateur fiable pour distinguer un ciel clair d'un environnement perturbé par le brouillard.

III.6.2 Étude de la matrice de confusion (Confusion Matrix) :

La matrice de confusion est un outil essentiel pour visualiser les erreurs du modèle. Elle permet de confronter les classes réelles (True Classes) aux classes prédites par l'algorithme.

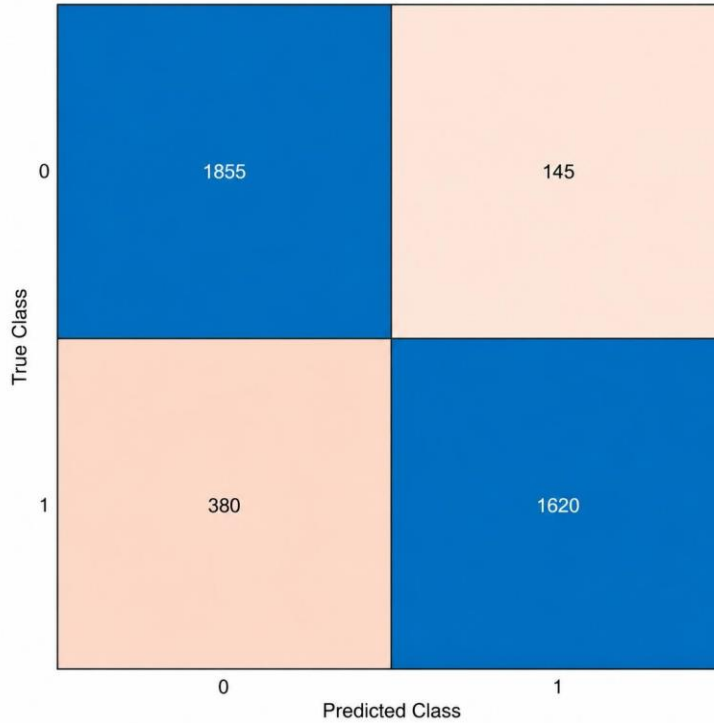


Figure III.7 : Matrice de confusion du modèle Coarse K-NN.

La majorité des échantillons sont correctement classés sur la diagonale principale. On observe toutefois un léger taux d'erreur (environ 13,1 %) qui correspond aux zones de transition, là où la distance est très élevée. À de grandes distances, le signal en "Ciel clair" peut s'affaiblir au point de ressembler techniquement à un signal sous "Brouillard", ce qui explique les quelques confusions du modèle.

Par ailleurs, la matrice de confusion repose sur quatre indicateurs essentiels : les Vrais Positifs (TP), les Vrais Négatifs (TN), les Faux Positifs (FP) et les Faux Négatifs (FN). Ces paramètres permettent de calculer l'exactitude du modèle selon la formule (eq. 02) du chapitre I :

$$Exactitude = \frac{TP + TN}{TP + TN + FP + FN} \times 100 = \frac{1855 + 1620}{1855 + 1620 + 145 + 380} \times 100 = 86,875 \approx 86,9 \%$$

III.6.3 Analyse de la courbe ROC (Receiver Operating Characteristic) :

La courbe ROC (Receiver Operating Characteristic) permet d'évaluer la performance du classifieur indépendamment du seuil de décision.

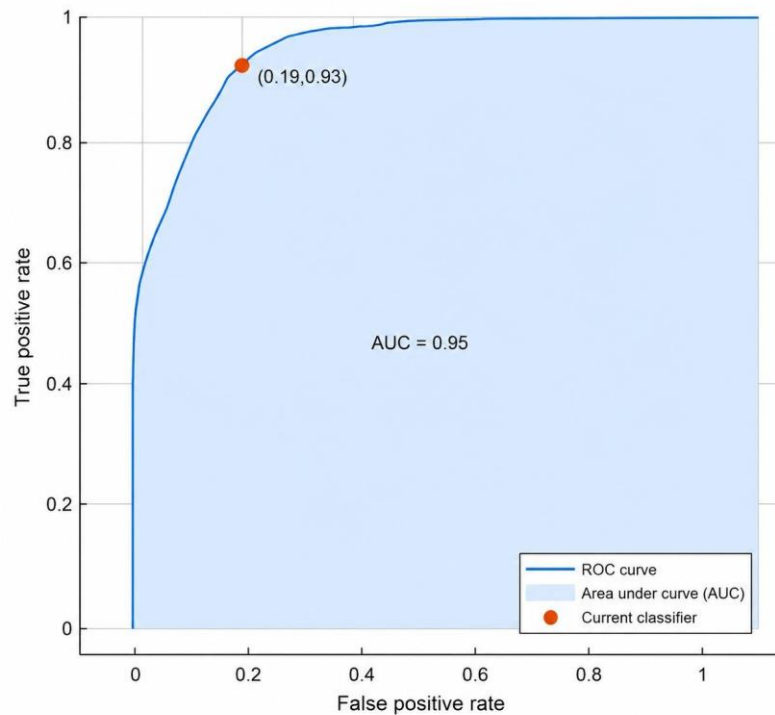


Figure III.8 : Courbe ROC et performance de classification du système FSO.

La courbe s'approche significativement du coin supérieur gauche, ce qui indique un taux de vrais positifs élevé et un faible taux de faux positifs. L'aire sous la courbe (AUC - Area Under Curve) proche de 1 ($AUC = 0.95$) confirme que le modèle Coarse K-NN possède une excellente capacité de discrimination entre les deux états atmosphériques.

III.7 Conclusion :

En conclusion, ce troisième chapitre a permis de valider la méthodologie de diagnostic du canal FSO par l'intelligence artificielle. La simulation sur *OptiSystem* a généré une base de données rigoureuse de 4000 observations (ciel clair et conditions dégradées), basée sur le facteur de qualité (Q-Factor) comme indicateur clé de performance.

Le traitement des données sous *MATLAB*, via la normalisation par le Z-Score, a assuré l'homogénéité nécessaire à une classification optimale. L'application de l'algorithme des K plus

proches voisins (modèle *Coarse K-NN*, $K=100$ distance euclidienne) a donné des résultats probants, validés par la matrice de confusion et les courbes ROC. Le modèle a atteint une précision globale de 86.9 %, démontrant une forte robustesse face aux perturbations atmosphériques, sans aucun phénomène de surapprentissage (*overfitting*).

Ces résultats numériques confirment l'efficacité de l'apprentissage supervisé pour le diagnostic automatisé et instantané des liaisons optiques en espace libre.

Conclusion générale

Conclusion générale

Ce travail a porté sur l'étude des systèmes de communications optiques en espace libre (FSO) et sur l'intégration des techniques d'apprentissage automatique pour améliorer leurs performances. Les systèmes FSO constituent une solution attractive pour les télécommunications modernes grâce à leur grande capacité de transmission, leur large bande passante et leur facilité de déploiement. Néanmoins, leurs performances restent sensibles aux perturbations atmosphériques telles que le brouillard, la pluie et la turbulence.

Dans le cadre de cette étude, une chaîne complète de transmission FSO a été conçue et simulée sous OptiSystem afin d'évaluer l'impact des différentes conditions atmosphériques sur la qualité de la liaison. L'analyse des paramètres de performance, notamment le facteur de qualité (Q-Factor) et le taux d'erreur binaire (BER), a permis de mettre en évidence la dégradation des performances en présence de conditions atmosphériques défavorables.

Les données issues des simulations ont ensuite été exploitées pour construire un dataset destiné à une approche basée sur l'apprentissage automatique. L'algorithme K-Plus Proches Voisins (K-NN), implémenté sous MATLAB, a permis de classifier efficacement l'état de la liaison FSO, démontrant ainsi l'intérêt de l'intelligence artificielle pour le développement de systèmes de communications optiques plus intelligents et adaptatifs.

Les résultats obtenus confirment que l'intégration des techniques de Machine Learning contribue à améliorer la fiabilité et l'adaptation des systèmes FSO face aux variations du canal atmosphérique.

Comme perspectives futures, il serait intéressant d'explorer d'autres algorithmes d'apprentissage automatique et d'apprentissage profond afin d'améliorer davantage la précision de classification et la robustesse du système. De plus, cette étude pourrait être étendue à des conditions atmosphériques plus complexes ainsi qu'à des architectures hybrides combinant les technologies FSO et RF afin d'assurer une meilleure disponibilité et continuité de la liaison.

Références :

- [1] **Jordan, M. I., & Mitchell, T. M.**, "Machine learning: Trends, perspectives, and prospects", *Science*, Vol. 349, No. 6245, 2015.
- [2] **J. McCarthy**, "From Here to Human-Level AI", *Artificial Intelligence*, Vol. 171, No. 18, 2007.
- [3] **Goodfellow, I., Bengio, Y., and Courville, A.**, "Deep Learning", MIT Press, 2016.
- [4] **N. J. Nilsson**, "The Quest for Artificial Intelligence: A History of Ideas and Achievements", *Cambridge University Press*, 2009.
- [5] **S. Russell and P. Norvig**, "Artificial Intelligence: A Modern Approach", *Pearson Education*, 4th Edition, 2020.
- [6] **S. Shalev-Shwartz and S. Ben-David**, "Understanding Machine Learning: From Theory to Algorithms", Cambridge University Press, 2014.
- [7] **C. M. Bishop**, "Pattern Recognition and Machine Learning", Springer, 2006.
- [8] **E. Alpaydin**, "Introduction to Machine Learning", *MIT Press*, 4th Edition, 2020.
- [9] **G. P. Agrawal**, "Fiber-Optic Communication Systems", *John Wiley & Sons*, 4th Edition, 2010.
- [10] **Jain, A. K.**, "Data clustering: 50 years beyond K-means", *Pattern Recognition Letters*, Vol. 31, No. 8, 2010.
- [11] **T. Fawcett**, "An Introduction to ROC Analysis", *Pattern Recognition Letters*, Vol. 27, No. 8, 2006.

- [12] **D. M. W. Powers**, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation", *Journal of Machine Learning Technologies*, Vol. 2, No. 1, 2011.
- [13] **Rendón, E., et al.**, "Internal evaluation metrics for clustering algorithms", *International Journal of Artificial Intelligence and Applications*, Vol. 2, No. 1, 2011.
- [14] **Wu, X., Kumar, V., Quinlan, J. R., et al.**, "Top 10 algorithms in data mining", *Knowledge and Information Systems*, Vol. 14, No. 1, 2008.
- [15] **MathWorks**, "Classification Learner App in Statistics and Machine Learning Toolbox", MATLAB Documentation (The MathWorks, Inc.), 2024.
- [16] **G. James, D. Witten, T. Hastie, and R. Tibshirani**, "An Introduction to Statistical Learning: with Applications in R", *Springer Science & Business Media*, 2013.
- [17] **S. Kaushal and G. Kaddoum**, "Optical Communication in Space: Challenges and Mitigation Techniques", *IEEE Communications Surveys & Tutorials*, Vol. 19, No. 1, 2016.
- [18] **Amirabadi, M. A., & Khaleghi, F.**, "Application of Machine Learning Techniques in Free Space Optical Communication Systems: A Review", *IEEE Access*, Vol. 10, 2022.
- [19] **M. A. Khalighi and M. Uysal**, "Survey on Free Space Optical Communication: A Communication Theory Perspective", *IEEE Communications Surveys & Tutorials*, Vol. 16, No. 4, 2014.
- [20] **A. K. Majumdar**, "Free Space Optical Communications: Theory and Practice", *Springer*, 2015.

- [21] **Z. Ghassemlooy, W. Popoola, and S. Rajbhandari**, "Optical Wireless Communications: System and Channel Modelling with MATLAB®", *CRC Press*, 2nd Edition, 2019.
- [22] **Malik, A., & Singh, P.**, "Free space optics: current applications and future challenges", *International Journal of Optics*, Vol. 2015, 2015.
- [23] **L. C. Andrews and R. L. Phillips**, "Laser Beam Propagation through Random Media", *SPIE Press*, 2nd Edition, 2005.