

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة أبي بكر بلقايد - تلمسان

Université Aboubakr Belkaïd – Tlemcen –

Faculté de TECHNOLOGIE



MEMOIRE

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

En : Télécommunications

Spécialité : Réseaux et télécommunications

Par : Ghellai Maria & Guenineche Djawida

Sujet

Explicabilité des modèles de reconnaissance de modulation des signaux radio

Soutenu publiquement, le 22/06/2026, devant le jury composé de :

M MOUSSAOUI Djilali
Mme BENLALDJ Lamia
M HADJILA Mourad

MCA
MCB
PR

Université de Tlemcen
Université de Tlemcen
Université de Tlemcen

Président
Examinateur
Encadreur

Remerciements

Avant d'entamer la présentation de ce travail, il nous tient à cœur d'exprimer nos remerciements les plus sincères à toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire.

Nous remercions tout d'abord Allah, le Tout-Puissant et le Très Miséricordieux, de nous avoir accordé la santé, la force, la patience et la persévérance nécessaires à l'aboutissement de ce projet. Sans Sa guidance, rien n'aurait été possible.

Nos remerciements vont également à l'ensemble du corps enseignant et administratif de Université Aboubakr Belkaid Tlemcen, pour la qualité de leur encadrement tout au long de notre formation. Leur disponibilité, leur rigueur et leur engagement ont grandement enrichi notre parcours académique.

Nous adressons notre plus profonde gratitude à Mr HADJILA Mourad, notre encadrant, pour sa confiance, sa bienveillance, ses conseils précieux, ainsi que pour son accompagnement tout au long de la réalisation de ce mémoire. Son expertise et sa rigueur scientifique ont été d'un apport inestimable.

Nous exprimons aussi notre reconnaissance envers nos familles, pour leur amour inconditionnel, leur soutien moral et leur patience tout au long de notre parcours universitaire. Leur présence et leurs encouragements ont été essentiels dans chaque étape de notre chemin.

Enfin, nous remercions les membres du jury pour l'intérêt porté à notre travail, ainsi que toutes les personnes, connues ou anonymes, qui ont contribué d'une manière ou d'une autre à la réussite de ce projet.

Trouvez dans ces mots l'expression de notre profonde reconnaissance.

Dédicaces

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Ce mémoire représente bien plus qu'un simple travail académique. Il est le fruit d'un parcours partagé, jalonné de défis, d'efforts communs, de remises en question, de nuits blanches et de moments de doute, mais également d'apprentissage, de satisfaction et de petites victoires. C'est avec un profond respect et une sincère reconnaissance que nous le dédions à toutes celles et à tous ceux qui ont contribué, de près ou de loin, à sa réalisation.

À nos très chers parents, pour leur amour indéfectible, leurs sacrifices, leurs encouragements constants et leur soutien inconditionnel tout au long de notre parcours. Aucun mot ne saurait exprimer notre profonde gratitude envers eux,

À nos frères et sœurs, pour leur affection, leur présence rassurante et leur soutien permanent.

À nos belles-sœurs, pour leur chaleureux accueil au sein de la famille, leur sympathie et leurs encouragements réguliers.

À nos amis et collègues, pour leur solidarité, leur aide spontanée et les instants partagés durant ces années d'études.

À tous les professeurs qui ont marqué nos parcours, par leur passion, leur rigueur et leur bienveillance. Merci pour les connaissances transmises, les échanges inspirants et les moments marquants que nous n'oublierons jamais.

Enfin, nous remercions toutes les personnes qui nous ont soutenues et encouragées à avancer.

تتناول هذه المذكرة التصنيف التلقائي للتضمين (AMC) باستخدام بنية هجينة تجمع بين الشبكات العصبية التلافيفية (CNN) والشبكات التلافيفية الزمنية (TCN)، وذلك بالاعتماد على مجموعة البيانات RadioML 2016.10a. ويهدف العمل إلى تحسين تصنيف الإشارات الراديوية في ظروف ضوضاء مختلفة من خلال دمج نسبة الإشارة إلى الضوضاء (SNR) ضمن عملية التصنيف.

ولتعزيز قابلية التفسير، تم استخدام تقنيات الذكاء الاصطناعي القابل للتفسير (XAI)، وبالأخص SHAP و Integrated Gradients، لتحليل تأثير مكونات الإشارة I/Q على قرارات النموذج. أظهرت النتائج أن النموذج المقترح يحقق أداءً جيداً في التصنيف مع توفير فهم أفضل لآلية عمله الداخلية، مما يجمع بين الدقة والمتانة وقابلية التفسير في تطبيقات الاتصالات الراديوية الذكية.

الكلمات المفتاحية: التصنيف التلقائي للتضمين، التعلم العميق، الشبكات العصبية التلافيفية (CNN)، الشبكات التلافيفية الزمنية (TCN)، نسبة الإشارة إلى الضوضاء (SNR)، الذكاء الاصطناعي القابل للتفسير (XAI)، SHAP، التدرجات المتكاملة.

Résumé

Ce mémoire porte sur la classification automatique des modulations radio (AMC) à l'aide d'une architecture hybride CNN-TCN appliquée à la base de données RadioML 2016.10a. L'objectif est d'améliorer la reconnaissance des signaux radio dans différentes conditions de bruit en intégrant l'information du rapport signal sur bruit (SNR) au processus de classification.

Afin d'assurer l'interprétabilité du modèle, des techniques d'intelligence artificielle explicable (XAI), notamment SHAP et Integrated Gradients, sont utilisées pour analyser l'influence des composantes I/Q sur les décisions du réseau.

Les résultats obtenus montrent que le modèle proposé offre de bonnes performances de classification tout en permettant une meilleure compréhension de son fonctionnement interne. Cette approche combine ainsi précision, robustesse et interprétabilité pour les applications de communications radio intelligentes.

Mots-clés : Classification automatique des modulations radio, apprentissage profond, architecture CNN-TCN, intégration du SNR, interprétabilité des modèles, intelligence artificielle explicable, SHAP, gradients intégrés.

Abstract

This thesis investigates Automatic Modulation Classification (AMC) using a hybrid CNN-TCN architecture on the RadioML 2016.10a dataset. The objective is to improve radio signal classification under various noise conditions by incorporating Signal-to-Noise Ratio (SNR) information into the classification process.

To enhance model interpretability, Explainable Artificial Intelligence (XAI) techniques, namely SHAP and Integrated Gradients, are employed to identify the most influential I/Q signal components involved in the classification decisions.

Experimental results show that the proposed model achieves strong classification performance while providing valuable insights into its internal behavior. The approach therefore combines accuracy, robustness, and interpretability for intelligent radio communication applications.

Keywords: Automatic Modulation Classification, Deep Learning, Convolutional Neural Networks (CNN), Temporal Convolutional Networks (TCN), Signal-to-Noise Ratio (SNR), Explainable Artificial Intelligence (XAI), SHAP, Integrated Gradients.

Sommaire

Remerciements	II
Dédicace	III
Résumé	IV
Liste des figures	IX
Liste des abréviations	XII

INTRODUCTION GENERALE..... 1

Chapitre 1 : GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

1.1	Introduction	3
1.2	Communication radio numérique	3
1.2.1	Chaîne de transmission numérique	3
1.2.2	Notions de bande passante	6
1.2.3	Bruit et canal de propagation	8
1.3	La modulation analogique et ses types de base	11
1.3.1	Principe de modulation analogique	11
1.3.2	La modulation d’amplitude AM.....	11
1.3.3	La modulation de fréquence FM	16
1.3.4	La modulation de phase PM.....	19
1.4	La modulation numérique et ses types de base	20
1.4.1	Principe de modulation numérique	20
1.4.2	La modulation par déplacement d’amplitude (M-ASK-Amplitude Shift Keying) ..	21
1.4.3	La modulation par déplacement de fréquence (M-FSK -Frequency Shift Keying) ..	22
1.4.4	La modulation par déplacement de phase (M-PSK-Phase Shift Keying)	23
1.4.5	La modulation d’amplitude en quadrature (QAM-Quadrature Amplitude Modulation)	24
1.5	Reconnaissance automatique des modulations	25
1.5.1	Définition de l’Automatic Modulation Classification.....	25
1.5.2	Domaines d’application	25
1.5.3	Enjeux et contraintes	26
1.6	Méthodes classiques de classification des modulations	28
1.6.1	Approches basées sur les caractéristiques statiques et spectrales	28
1.6.2	Limites en environnement bruité	28
1.7	Conclusion.....	30

Chapitre 2 : Intelligence Artificielle Explicable (XAI)

2.1	Introduction	31
2.2	L'intelligence artificielle	31
2.2.1	Concept de base	31
2.2.2	Machine learning	32
2.2.3	Deep learning.....	34
2.3	Fondements théoriques de l'Explainable AI.....	35
2.3.1	Définition et concepts clés.....	35
2.3.2	Typologie des méthodes XAI.....	36
2.3.2.1	Explications globales vs locales	37
2.3.2.2	Méthodes intrinsèques vs post-hoc	38
2.3.2.3	Méthodes dépendantes du modèle vs indépendantes.....	41
2.4	Principales techniques d'explicabilité.....	42
2.4.1	LIME (Local Interpretable Model-Agnostic Explanations)	42
2.4.1.1	Principe de construction d'un modèle local approximatif	42
2.4.1.2	Description mathématique du processus de perturbation et d'optimisation locale	43
2.4.1.3	Avantages et limites.....	44
2.4.2	SHAP (SHapley Additive exPlanations)	45
2.4.2.1	Fondements théoriques basés sur la théorie des jeux coopératifs et les valeurs de Shapley.....	45
2.4.2.2	Formulation mathématique des contributions marginales des variables	48
2.4.2.3	Interprétation des graphiques de type summary plot et force plot.....	49
2.4.2.4	Comparaison avec LIME	51
2.4.3	Grad-CAM et méthodes basées sur le gradient	51
2.4.3.1	Méthodes d'interprétabilité basées sur le gradient	53
2.4.3.2	Applications de Grad-CAM en imagerie médicale et en vision par ordinateur.....	54
2.5	Métriques d'évaluation des explications.....	54
2.5.1	Les critères d'évaluation.....	55
2.5.2	Protocoles expérimentaux	55
2.6	Conclusion	56
Chapitre 3 : RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION		
3.1	Introduction	57
3.2	Présentation du dataset RadioML2016.10a.....	57
3.3	Modèle CNN-TCN	62

3.3.1	Architecture du modèle	62
3.3.1.1	Branche d'extraction des caractéristiques du signal	62
3.3.1.2	Branche de traitement du SNR	62
3.3.1.3	Branche de fusion et de classification	62
3.3.2	Résultats et discussion.....	63
3.3.2.1	Analyse des courbes d'apprentissage.....	63
3.3.2.2	Rapport de classification	65
3.3.3	La matrice de confusion	66
3.3.3.1	Analyse de la matrice de confusion globale	66
3.3.3.2	Analyse des matrices de confusion par niveau de SNR	68
3.3.4	Les performances du modèle CNN-TCN	69
3.3.4.1	Performances en conditions de fort bruit (SNR faible)	69
3.3.4.2	Performances en conditions favorables (SNR élevé)	70
3.3.5	Les métriques d'évaluation en fonction du SNR	71
3.3.5.1	À faible SNR	71
3.3.5.2	À SNR élevé	71
3.3.6	Etude comparative	72
3.3.6.1	Robustesse en conditions dégradées (SNR faible : -20 dB à -8 dB)	73
3.3.6.2	Performances en régime intermédiaire (-6 dB à 4 dB)	73
3.3.6.3	Performances en conditions favorables (SNR élevé : 6 dB à 18 dB)	73
3.4	Interprétation du modèle par les méthodes XAI (gradients intégrés et SHAP)	73
3.4.1	Modèle hybride CNN-TCN avec fusion SNR et interprétabilité XAI	73
3.4.2	Analyse explicable des modulations AM-DSB, AM-SSB et 8PSK	74
3.4.2.1	Gradients intégrés et SHAP sous faible SNR (-20 dB)	74
3.4.2.2	Gradients intégrés et SHAP sous SNR intermédiaire (0 dB)	77
3.4.2.3	Gradients intégrés et SHAP sous SNR élevé (10 dB)	81
3.5	Conclusion.....	86
	CONCLUSION GENERALE.....	87

Liste des figures

Chapitre 1

Figure 1.1 : Chaîne de transmission numérique	4
Figure 1.2 : Principe de la numérisation d'une source analogique.	4
Figure 1.3 : Schéma d'une bande passante.	7
Figure 1.4 : Modulation d'amplitude AM.	11
Figure 1.5 : La représentation temporelle de la modulation AM Double Bande Sans Porteuse. .	13
Figure 1.6 : Spectre d'amplitude du signal modulé.	13
Figure 1.7 : Représentation temporelle du signal AM DBAP.	14
Figure 1.8 : Représentation spectrale du signal AM DBAP.	15
Figure 1.9 : Caractéristiques du signal BLU.	15
Figure 1.10 : Principe de modulation AM-BLR.	16
Figure 1.11 : Représentation de spectre d'un signal BLR.	16
Figure 1.12 : Représentation graphique de la modulation FM.	17
Figure 1.13 : Exemple d'un spectre FM.	18
Figure 1.14 : Représentation graphique des trois premières fonctions de Bessel.	18
Figure 1.15 : La représentation graphique de la modulation de phase.	19
Figure 1.16 : Chronogramme de modulation ASK.	21
Figure 1.17 : Modulation M-ASK symétriques.	21
Figure 1.18 : Représentation spectrale de signal modulé en FSK.	22
Figure 1.19 : Modulation par déplacement de phase PSK.	23
Figure 1.20 : Constellations M-PSK.	24
Figure 1.21 : Principe de modulation QAM.	24
Figure 1.22 : Constellation de la modulation 8-QAM et 16-QAM.	25

Chapitre 2

Figure 2.1 : La différence entre l'apprentissage humain et le Machine Learning.	34
Figure 2.2 : Le fonctionnement d'un réseau de Deep Learning.	35
Figure 2.3 : Deux objectifs principaux pour l'XAI : justification et validation.	36
Figure 2.4 : Régression linéaire. Ajustement affine.	39
Figure 2.5 : Une structure générale d'arbre de décision.	40
Figure 2.6 : Illustration du fonctionnement de la méthode SHAP et de l'attribution des valeurs de Shapley aux variables explicatives.	47
Figure 2.7 : Impact moyen des variables sur la prédiction du modèle selon les valeurs SHAP.	49

Figure 2.8 : Analyse de l'interprétabilité locale du modèle via les graphiques SHAP Force Plot.	50
Figure 2.9 : Illustration du pipeline Grad-CAM : de l'image d'entrée à la carte thermique d'activation	52
Figure 2.10 : Comparaison entre l'image originale et les heatmaps Grad-CAM associées aux classes « Cat » et « Dog ».	53

Chapitre 3

Figure 3.1 : Composantes I/Q dans le domaine temporel d'un signal à un SNR de 18 dB.	61
Figure 3.2 : Architecture du modèle CNN-TCN avec fusion SNR.	63
Figure 3.3: Training/Validation Accuracy vs. Epoch.	64
Figure 3.4: Training/Validation Loss vs. Epoch.	65
Figure 3.5 : Rapport de classification.....	66
Figure 3.6 : Matrice de confusion globale.....	67
Figure 3.7 : Matrices de confusion pour différentes valeurs SNR (-20, -10, 0, 10, 18 dB)	69
Figure 3.8 : Tracé des métriques d'évaluation en fonction du SNR.	72
Figure 3.9 : Architecture d'un Modèle Hybride CNN-TCN avec Fusion de Données SNR et Module d'Interprétabilité XAI (SHAP)	74
Figure 3.10 : Carte de saillance à gradients intégrés pour l'AM-DSB avec un rapport signal/bruit (SNR) de -20 dB.	75
Figure 3.11 : Carte de saillance à gradients intégrés pour l'AM-SSB avec un rapport signal/bruit (SNR) de -20 dB.	75
Figure 3.12: 20 principales features SHAP (I/Q) - AM-DSB (SNR = -20 dB)	76
Figure 3.13: 20 principales features SHAP (I/Q) - AM-SSB (SNR = -20 dB).	77
Figure 3.14 : Carte de saillance à gradients intégrés pour l'AM-DSB avec un rapport signal/bruit (SNR) de 0 dB.	78
Figure 3.15 : Carte de saillance à gradients intégrés pour l'AM-SSB avec un rapport signal/bruit (SNR) de 0 dB.	78
Figure 3.16 : Carte de saillance à gradients intégrés pour l'8PSK avec un rapport signal/bruit (SNR) de 0 dB.	79
Figure 3.17: 20 principales features SHAP (I/Q) – AM-DSB (SNR = 0 dB).	80
Figure 3.18: 20 principales features SHAP (I/Q) – AM-SSB (SNR = 0 dB).	80
Figure 3.19: 20 principales features SHAP (I/Q) – 8PSK (SNR = 0 dB).	81
Figure 3.20 : Carte de saillance à gradients intégrés pour l'AM-DSB avec un rapport signal/bruit (SNR) de 10 dB.	82

Figure 3.21 : Carte de saillance à gradients intégrés pour l'AM-SSB avec un rapport signal/bruit (SNR) de 10 dB.	82
Figure 3.22 : Carte de saillance à gradients intégrés pour l'8PSK avec un rapport signal/bruit (SNR) de 10 dB.	83
Figure 3.23: 20 principales features SHAP (I/Q) – AM-DSB (SNR = 10 dB).	84
Figure 3.24: 20 principales features SHAP (I/Q) – AM-SSB (SNR = 10 dB).	84
Figure 3.25: 20 principales features SHAP (I/Q) – 8PSK (SNR = 10 dB).	85

Liste des tableaux

Chapitre 1

Tableau 1.1 : Tableau des dix premières fonctions de Bessel en fonction de m	19
---	----

Chapitre 2

Tableau 2.1 : Comparaison synthétique entre SHAP et LIME.	51
Tableau 2.2 : Principaux critères d'évolution des méthodes d'explicabilité et leurs objectifs	55

Chapitre 3

Tableau 3.1 : Évaluation des performances du modèle de classification de modulation à différents niveaux de rapport signal/bruit	70
Tableau 3.2 : Comparaison des performances.	72

Liste des abréviations

AI : Artificial Intelligence
AMC : Automatic Modulation Classification
AM : Amplitude Modulation
AM-DSB : Amplitude Modulation Double Side Band
AM-SSB : Amplitude Modulation Single Side Band
AM-BLR : Amplitude Modulation à Bande Latérale Résiduelle
ASK : Amplitude Shift Keying
AWGN : Additive White Gaussian Noise
BER : Bit Error Rate
BLU : Bande Latérale Unique
BPSK : Binary Phase Shift Keying
CAN : Convertisseur Analogique-Numérique
CNN : Convolutional Neural Network
CPFSK : Continuous Phase Frequency Shift Keying
DL : Deep Learning
FFT : Fast Fourier Transform
FM : Frequency Modulation
FSK : Frequency Shift Keying
GFSK : Gaussian Frequency Shift Keying
Grad-CAM : Gradient-weighted Class Activation Mapping
IA : Intelligence Artificielle
I/Q : In-phase / Quadrature
IoT : Internet of Things
LIME : Local Interpretable Model-Agnostic Explanations
ML : Machine Learning
PAM4 : Pulse Amplitude Modulation à 4 niveaux
PM : Phase Modulation
PSK : Phase Shift Keying
QAM : Quadrature Amplitude Modulation
QPSK : Quadrature Phase Shift Keying
QoS : Quality of Service
RadioML : Radio Machine Learning Dataset
RFID : Radio Frequency Identification
SHAP : SHapley Additive exPlanations
SIGINT : Signals Intelligence
SNR : Signal-to-Noise Ratio
TCN : Temporal Convolutional Network
TU : Transferable Utility
WBFM : Wideband Frequency Modulation
XAI : Explainable Artificial Intelligence

INTRODUCTION GENERALE

L'émergence rapide des systèmes de télécommunication et des technologies radio a engendré une utilisation croissante des communications numériques dans des domaines aussi variés que les réseaux mobiles, les systèmes satellitaires, la radio cognitive et les applications de défense. Dans ce contexte, la reconnaissance automatique des modulations (AMC : Automatic Modulation Classification) constitue une problématique fondamentale pour garantir une transmission fiable, efficace et adaptative des signaux radio, dans des environnements fréquemment affectés par le bruit et les interférences.

Les approches traditionnelles de classification des modulations s'appuient essentiellement sur l'extraction manuelle de caractéristiques statistiques ou spectrales. Si ces méthodes ont longtemps prévalu, elles montrent aujourd'hui leurs limites face à la complexité croissante des signaux radio modernes et à la diversité des conditions réelles de propagation. L'avènement de l'intelligence artificielle, et en particulier des techniques de Machine Learning et de Deep Learning, a ouvert de nouvelles perspectives pour améliorer significativement les performances des systèmes de classification automatique.

Parallèlement, le recours aux modèles d'apprentissage profond soulève des interrogations légitimes quant à l'interprétabilité et à la transparence de leurs décisions. En effet, bien que ces architectures offrent des performances remarquables, leur fonctionnement interne demeure souvent opaque, ce qui a motivé l'émergence des approches d'intelligence artificielle explicable (XAI : eXplainable Artificial Intelligence), visant à rendre les décisions de ces modèles plus compréhensibles et justifiables.

C'est dans cette dynamique que s'inscrit le présent travail de fin d'études, dont l'objectif est de proposer une approche hybride basée sur les réseaux de neurones profonds pour la reconnaissance automatique des modulations radio, en y intégrant des méthodes d'interprétabilité permettant de mieux appréhender les mécanismes décisionnels du modèle.

Notre mémoire s'articule autour de trois chapitres :

Le premier chapitre pose les fondements théoriques des communications radio numériques, présente les principales techniques de modulation analogique et numérique, et passe en revue les approches classiques de classification des modulations.

INTRODUCTION GENERALE

Le deuxième chapitre développe les concepts fondamentaux de l'intelligence artificielle, du Machine Learning, du Deep Learning, ainsi que les principales méthodes d'intelligence artificielle explicable, notamment LIME, SHAP et Grad-CAM.

Le troisième chapitre expose les résultats de simulation obtenus sur le dataset RadioML 2016.10a, décrit l'architecture du modèle hybride CNN-TCN avec fusion SNR, analyse ses performances sous différentes conditions de bruit, et présente l'interprétation des décisions du modèle à l'aide des méthodes XAI.

Ce travail ambitionne ainsi de contribuer au développement de systèmes intelligents de reconnaissance des modulations radio, conjuguant à la fois performance, robustesse et interprétabilité.

CHAPITRE 1:
GENERALITES SUR LES
COMMUNICATIONS RADIO ET LA
CLASSIFICATION DES
MODULATIONS

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

1.1 Introduction

Un système de communication numérique garantit une transmission des informations sûre et performante, malgré les perturbations du canal, tout en maximisant l'utilisation des ressources à disposition. Les systèmes de télécommunication contemporains reposent sur une chaîne de communication numérique, qui assure la liaison entre émetteurs et récepteurs via différents canaux de transmission [1]. Elle se compose de plusieurs modules ou segments, chacun contribuant de manière significative à la fiabilité et à l'efficacité du transfert d'informations [2].

La communication radio numérique consiste à utiliser des signaux binaires pour transmettre la voix, les images et les données, par contre les communications analogiques modulent directement la voix ou le signal continu. Elle offre une robustesse face aux interférences et une compression efficace des données [1].

Aujourd'hui, les transmissions radio sont un élément fondamental des systèmes de télécommunication actuels, englobant diverses applications allant des réseaux mobiles et du Wi-Fi aux systèmes par satellite, militaires et de l'Internet des objets (IoT). Ces dispositifs s'appuient sur la diffusion d'informations par le biais d'un canal radio, intrinsèquement déficient, en faisant usage de signaux électromagnétiques modulés. Pour aborder la question de la reconnaissance automatique de modulation (AMR), il est primordial de comprendre les caractéristiques des signaux radio, les méthodes de modulation et le canal de propagation [3].

Ce chapitre présente les bases théoriques nécessaires à la compréhension des communications radio numériques et de la reconnaissance automatique des modulations. Il introduit les principes fondamentaux des systèmes radio, les principales techniques de modulation analogiques et numériques, ainsi que les approches classiques utilisées pour la classification des modulations.

1.2 Communication radio numérique

1.2.1 Chaîne de transmission numérique

Le schéma de principe d'une chaîne de transmission numérique est représenté dans la Figure 1.1.

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

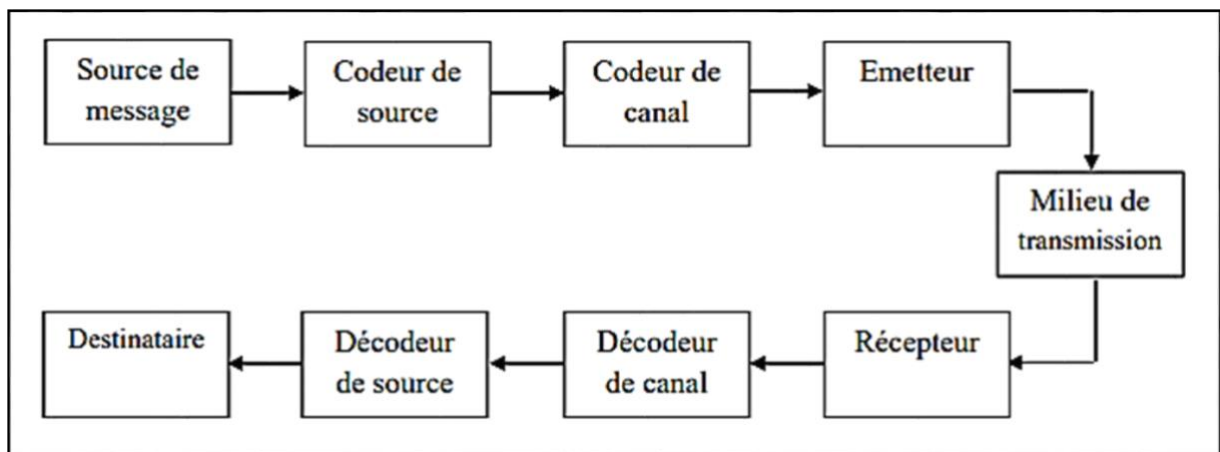


Figure 1.1 : Chaîne de transmission numérique.

La source de message, le milieu de transmission et le destinataire sont des données du problème, alors que le codage et le décodage de source, le codage et le décodage du canal, l'émetteur et le récepteur représentent des degrés de liberté du concepteur pour réaliser le système de transmission [4].

1.2.1.1 Source de message

Pour réaliser une transmission numérique et si la source délivre un message analogique tel que le signal de parole (sortie d'un microphone) ou le signal d'image (sortie d'une caméra), il faut le numériser. Nous rappelons que la numérisation se fait en échantillonnant le message analogique puis on quantifie les échantillons obtenus. Chaque échantillon quantifié est ensuite codé sur m éléments binaires (appelés bits) [4].

Les principales étapes de la numérisation d'un signal analogique sont résumées sur la Figure 1.2 :

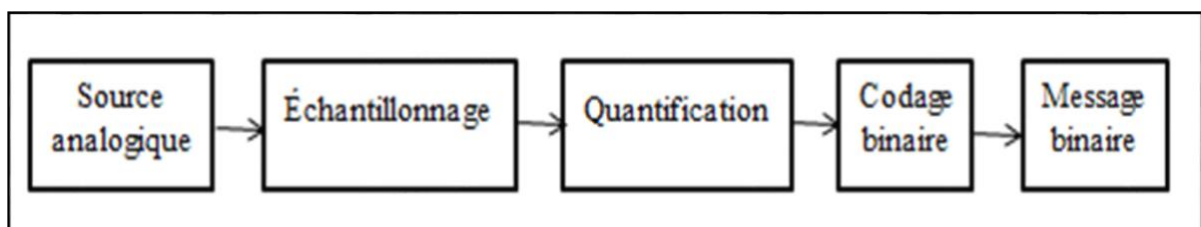


Figure 1.2 : Principe de la numérisation d'une source analogique.

1.2.1.2 Codage de source

Disons simplement qu'après codage de source, certains éléments binaires peu significatifs du message ont été supprimés. Ce message est alors sous forme concise. Après numérisation et codage, la source de message numérique est caractérisée par son débit binaire D défini comme étant le nombre d'éléments binaires qu'il émet par unité de temps. L'unité de débit binaire est bit par seconde. Si l'intervalle de temps séparant l'émission par la source de deux éléments binaires consécutifs est constant et égale à T_b , alors le débit binaire est :

$$D = \frac{1}{T_b} \text{ bit/s} \quad (1.1)$$

Une transmission est dite synchrone si l'émission d'éléments binaires par la source s'effectue à cadence constante, c'est-à-dire à raison d'un élément binaire tous les T_b secondes. Elle est dite asynchrone lorsque la cadence d'émission est variable dans le temps. Dans ce chapitre, nous nous placerons dans l'hypothèse de transmission synchrone et nous désignerons désormais par α_k l'élément binaire émis à l'instant kT_b [4].

1.2.1.3 Codage de canal

Le codage de canal, aussi appelé codage détecteur et/ou correcteur d'erreurs, est une fonction spécifique des transmissions numériques qui n'a pas son équivalent en transmission analogique. Le codage canal consiste à insérer dans le message des éléments binaires dits de redondance suivant une loi donnée. Cette opération conduit donc à une augmentation du débit binaire de la transmission.

Le décodeur de canal, qui connaît la loi de codage utilisée à l'émission, vient vérifier si cette loi est toujours respectée en réception. Si ce n'est pas le cas, il détecte la présence d'erreurs de transmission qu'il peut corriger sous certaines conditions [4].

1.2.1.4 Emetteur

Le message numérique en tant que suite d'éléments binaires est une grandeur abstraite. Pour transmettre ce message, il est donc nécessaire de lui associer une représentation physique sous forme d'un signal électrique (conversion bit/symbole) [4].

1.2.1.5 Milieu de transmission

Il représente le lien physique entre l'émetteur et le récepteur. Il est pratiquement constitué par l'un des supports suivants :

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

a) Un câble bifilaire

Dont la bande passante est faible et qui est en générale réservé pour les transmissions à bas débit (inférieures à 2 Mbit/s sur les réseaux téléphoniques).

b) Un câble coaxial

Ce dernier possède une bande passante plus importante que le câble bifilaire et qui permet de réaliser des transmissions avec un débit relativement élevé : plusieurs centaines de Mbit/s (Jusqu'à 565 Mbit/s sur le réseau téléphonique entre lequel transite un grand nombre de communications).

c) Une fibre optique

Qui apparait aujourd'hui grâce à sa bande passante très élevée et sa faible atténuation comme un support très intéressant. Les fibres optiques sont de plus en plus utilisées pour les réseaux terrestres à grande capacité (plusieurs Gbit/s voir dans le futur plusieurs dizaines de Gbit/s) pour les câbles sous-marins où elles ont supplanté le câble coaxial ainsi que pour les réseaux de distributions (c'est-à-dire sur les liaisons entre centraux téléphoniques et abonnés).

d) L'espace libre

Ce dernier utilise la propagation d'une onde électromagnétique dans l'atmosphère. Ce milieu est généralement réservé aux transmissions par satellite ou par faisceaux hertziens ainsi qu'aux radiocommunications avec les mobiles [4].

1.2.1.6 Le récepteur

Le récepteur a pour fonction la reconstitution du message émis par la source à partir du signal reçu.

Le bruit est une perturbation aléatoire dont les origines sont le milieu de transmission (bruit externe), ou les dispositifs électroniques utilisés dans le récepteur (bruit interne). Parmi les sources de bruit externe, on peut citer les rayonnements divers captés par l'antenne (cas des transmissions en espace libre), les interférences éventuelles entre les différents utilisateurs du milieu de transmission, ou encore les bruits d'origine industriels (ligne à haute tension). Le bruit interne a pour origine le mouvement brownien des électrons dans les composants passifs (résistance) et les composants actifs (amplificateur, filtre, mélangeur, etc.). Le bruit interne est généralement prépondérant dans les systèmes de transmission [4].

1.2.2 Notion de bande passante

En électronique et en télécommunications, la bande passante est la partie du spectre des valeurs de fréquence qui peut être assumée par un signal électrique, sans qu'il soit atténué par le

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

passage à travers un filtre. De manière informelle, on dit que ce sont les fréquences qui « passent » à travers le filtre [5].

La valeur de la fréquence, mesurée en Hertz, à partir de laquelle le signal ne passe plus à travers le filtre, est appelée fréquence de coupure. Ce sont les fréquences de coupures basses et hautes qui délimitent la bande passante [5].

Idéalement, les signaux dont les fréquences sont supérieures ou inférieures aux valeurs de coupure du filtre seraient atténués à zéro. En pratique, cependant, le critère de la demi-puissance est adopté : on considère comme fréquence de coupure celle dans laquelle la puissance du signal est atténuée à la moitié de l'original [5].

La Figure 1.3 montre la réponse fréquentielle d'un filtre passe-bande.

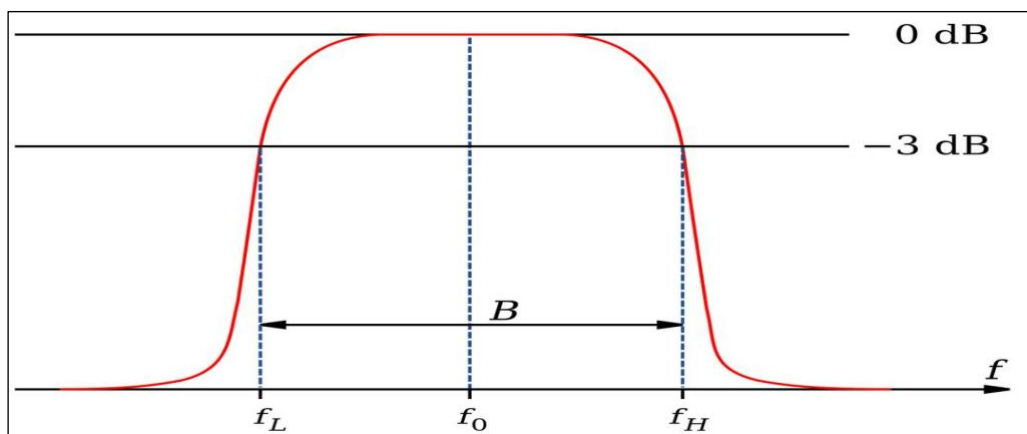


Figure 1.3 : Schéma d'une bande passante

La bande passante peut se définir comme la capacité maximale de débit sur une liaison donnée, déterminée par les technologies de transmission mises en œuvre à l'aide des équipements situés à chaque extrémité de cette liaison [5].

1.2.2.1 Applications pratiques de la bande passante

La bande passante joue un rôle crucial dans diverses applications électroniques :

- a) **Amplificateurs** : La bande passante d'un amplificateur détermine la plage de fréquences qu'il peut amplifier sans distorsion significative. Une bande passante plus large permet l'amplification fidèle de signaux complexes contenant une large gamme de fréquences, comme la musique [6].

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

- b) **Filtres** : Les filtres sont conçus pour laisser passer ou bloquer sélectivement certaines plages de fréquences. Leur bande passante définit les fréquences qu'ils autorisent à passer. Par exemple, un filtre passe-bas pourrait avoir une bande passante étroite ne permettant que le passage des basses fréquences, tandis qu'un filtre passe-haut laisserait passer les hautes fréquences [6].
- c) **Systèmes de communication** : Dans les systèmes de communication, la bande passante est essentielle pour transmettre et recevoir des données efficacement. Une bande passante plus large permet une transmission de données plus rapide, prenant en charge Internet haut débit, la visioconférence et d'autres applications exigeantes [6].

1.2.2.2 L'impact de la bande passante

- a) **Fidélité du signal** : Une bande passante plus large permet une représentation plus fidèle du signal d'origine, préservant les nuances et les détails [6].
- b) **Vitesse de transmission de données** : Une bande passante plus large permet des débits de transmission de données plus rapides, prenant en charge les applications haute performance [7].
- c) **Rapport signal sur bruit** : Une bande passante étroite peut réduire le bruit en filtrant les fréquences indésirables, améliorant le rapport signal sur bruit [6].

1.2.3 Bruit et canal de propagation

1.2.3.1 Le bruit

Le bruit est une perturbation aléatoire qui s'ajoute au signal lors de sa transmission. Il dégrade la qualité du signal reçu et peut provoquer des erreurs lors de la détection des symboles numériques. En communications radio, le bruit est souvent modélisé par un bruit blanc gaussien additif (AWGN), caractérisé par une distribution gaussienne et une densité spectrale constante [7].

Le niveau de bruit influence directement le rapport signal sur bruit (SNR), qui mesure la qualité de la transmission. Plus le SNR est élevé, meilleure est la fiabilité de la communication [7].

Le rapport signal sur bruit se calcule par la formule suivante :

$$SNR = \left(\frac{P_S}{P_n} \right) \quad (1.2)$$

En décibels :

$$SNR_{(dB)} = 10 \log_{10} \left(\frac{P_S}{P_n} \right) \quad (1.3)$$

Où :

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

- P_s = puissance du signal
- P_n = puissance du bruit

La capacité maximale d'un canal bruité est donnée par la formule suivante :

$$C = B \log_2(1 + \text{SNR}) \quad (1.4)$$

Où :

- C = capacité en bits/s
- B = bande passante
- SNR = rapport signal/bruit

Plus le bruit augmente, plus le SNR diminue, et plus la capacité diminue.

1.2.3.2 Types de bruit en communication radio

Dans un système de transmission radio, le bruit se réfère à toute interférence indésirable qui s'ajoute au signal principal et nuit à sa qualité. On identifie divers types de bruit en fonction de leur provenance et de leurs propriétés [7].

- a) **Bruit thermique (bruit blanc)** : L'agitation aléatoire des électrons dans les composants électroniques est à l'origine du bruit thermique. On le retrouve dans tous les systèmes électriques et son élimination totale n'est pas possible. On le représente habituellement par un bruit blanc gaussien additif (AWGN), qui se distingue par une densité spectrale constante sur une vaste gamme de fréquences [8].
- b) **Bruit atmosphérique** : Il provient des phénomènes naturels tels que les éclairs, les orages ou l'activité solaire. Il affecte particulièrement les communications radio à basses fréquences [8].
- c) **Bruit industriel** : Les équipements électriques, les moteurs, les lignes haute tension et les dispositifs électroniques produisent ce genre de bruit. On le retrouve souvent dans les milieux urbains [9].
- d) **Bruit de quantification** : Il apparaît lors de la conversion analogique-numérique (CAN). Il est dû à l'approximation des valeurs analogiques par un nombre fini de niveaux numériques [10].

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

- e) **Bruit impulsionnel** : Il se manifeste sous forme de pics de forte amplitude et de courte durée. Il peut être causé par des commutations électriques ou des décharges soudaines. Ce bruit est particulièrement perturbant pour les systèmes numériques [7].

1.2.3.3 Le canal de propagation

Le canal de propagation fait référence à l'environnement à travers lequel le signal radio se diffuse entre l'émetteur et le récepteur. À la différence d'un canal filaire, le canal radio est variable et soumis à divers phénomènes physiques tels que: la diminution du signal en fonction de la distance, la considération des obstacles, la diffusion et la diffraction, le multipath (multi-trajet) et la fluctuation (changement rapide de l'amplitude du signal). Ces phénomènes peuvent provoquer des déformations et des variations du signal capté, ce qui complique la transmission [11].

1.2.3.4 Influence du bruit dans le canal de propagation

Le bruit a une importance cruciale pour la qualité d'un canal de diffusion radio. Au cours de la transmission, le signal principal subit des perturbations aléatoires qui se superposent à lui et altèrent ses propriétés d'amplitude, de phase ou de fréquence. Dans la pratique, on modélise généralement le bruit par un bruit blanc gaussien additif (AWGN), qui symbolise une perturbation aléatoire suivant une distribution normale [11].

L'impact dominant du bruit se manifeste par une réduction du rapport signal-bruit (SNR). Quand le rapport signal-bruit est faible, le récepteur éprouve plus de difficultés à identifier correctement les symboles envoyés, ce qui provoque une hausse du taux d'erreur binaire (BER). Donc, plus le bruit est fort, plus la chance de commettre une erreur de détection s'accroît [12].

Dans un canal radio réel, le bruit agit en combinaison avec d'autres phénomènes tels que l'atténuation, le multi-trajet et le fading. Cette interaction peut provoquer des distorsions supplémentaires et rendre la transmission encore plus sensible aux erreurs, notamment dans les environnements urbains ou à longue distance.

Donc, le bruit dégrade la fiabilité et la performance globale du canal de propagation. Pour limiter son influence, les systèmes de communication numérique utilisent des techniques telles que le codage de canal, le filtrage, l'augmentation de la puissance d'émission ou l'adaptation de la modulation [1].

1.3 La modulation analogique et ses types de base

1.3.1 Principe de modulation analogique

La modulation analogique est un procédé de télécommunication qui consiste à transposer un signal d'information analogique, généralement de basse fréquence (voix, musique, signal continu), sur une onde porteuse analogique de haute fréquence, en faisant varier l'amplitude, la fréquence ou la phase de cette porteuse proportionnellement au signal d'information. Cette opération permet une transmission efficace du signal sur de longues distances, une meilleure adaptation au canal de transmission et une réduction des interférences [13].

1.3.2 La modulation d'amplitude (AM)

La modulation d'amplitude consiste à faire varier l'amplitude d'une onde porteuse de haute fréquence en fonction d'un signal de basse fréquence contenant l'information à transmettre, appelé signal modulant, est de fréquence relativement faible (par exemple inférieure à 20 kHz), tandis que la porteuse est de fréquence beaucoup plus élevée (souvent supérieure à 100 kHz) [19]. Elle constitue l'une des premières méthodes de modulation utilisées dans les systèmes de communication radio, dans les années 1920 [17].

La représentation temporelle de la modulation d'amplitude est donnée par la Figure 1.4.

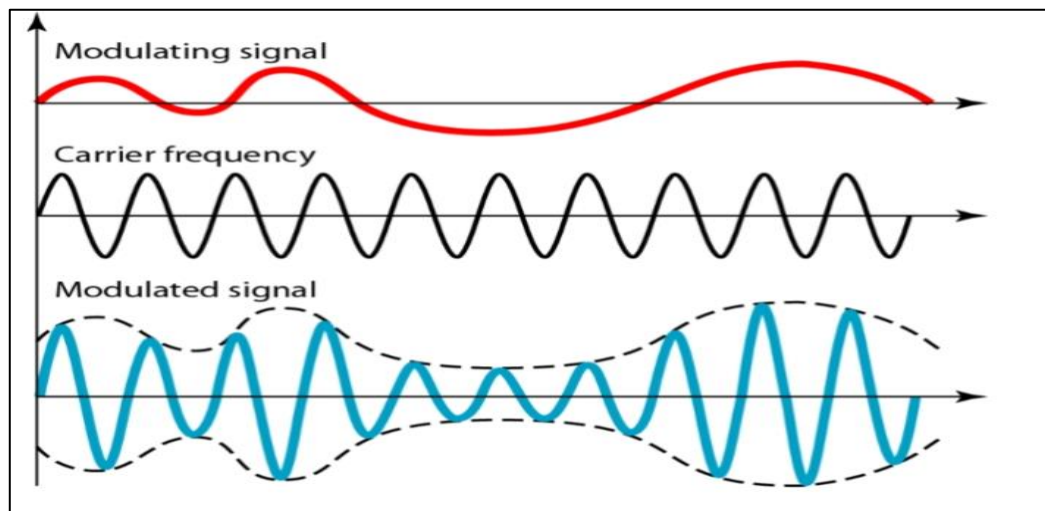


Figure 1.4 : Modulation d'amplitude.

Le signal modulé en amplitude s'écrit :

$$S(t) = A_c[1 + K_a m(t)] \cos(2\pi f_c t) \quad (1.5)$$

Si : $m(t) = A_m \cos(\omega_m t)$ (1.6)

Alors : $S(t) = A_c[1 + m \cos(\omega_m t)] \cos(2\pi f_c t)$ (1.7)

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

Où : $m = K_a \cdot A_m$

L'indice de modulation μ est défini par :

$$\mu = K_a \max |m(t)| \quad (1.8)$$

- $0 \leq \mu \leq 1$: modulation correcte.
- $(\mu > 1)$: surmodulation, entraînant une distorsion du signal.

Soit :

- $m(t)$: le signal d'information analogique,
- A_c : l'amplitude de la porteuse,
- f_c : la fréquence de la porteuse.
- $\omega_m = 2\pi f_m$
- K_a : coefficient de sensibilité en amplitude.

1.3.2.1 La modulation AM Double Bande Sans Porteuse (AM-DBSP)

Soit un signal sinusoïdal haute fréquence $p(t) = \cos(2\pi f_0 t)$, appelé porteuse. Le message $m(t)$ à transmettre est appelé "signal modulant".

Le signal AM modulé en amplitude Double Bande Sans Porteuse(DBSP) s'écrit :

$$S(t) = p(t) \cdot m(t) \quad (1.9)$$

On considère un signal modulant sinusoïdal $m(t) = A \cos(2\pi f_m t)$

Avec $f_m \ll f_0$.

Le signal AM s'écrit alors :

$$S(t) = \cos(2\pi f_0 t) \cdot A \cos(2\pi f_m t) \quad (1.10)$$

La représentation temporelle de la modulation AM Double Bande Sans Porteuse, est montrée par la Figure 1.5.

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

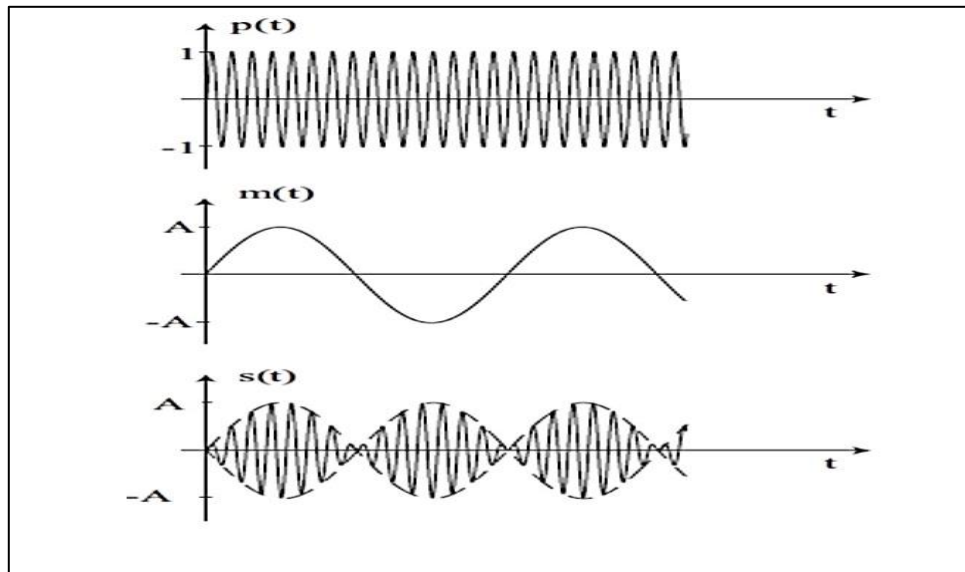


Figure 1.5 : La représentation temporelle de la modulation AM Double Bande Sans Porteuse.

Pour déterminer le spectre de $s(t)$, il faut décomposer $s(t)$ en une somme de signaux sinusoïdaux.

On a :

$$S(t) = \cos(2\pi f_0 t) \cdot A \cos(2\pi f_m t) = \frac{A}{2} (\cos 2\pi(f_0 + f_m)t + \cos 2\pi(f_0 - f_m)t) \quad (1.11)$$

Le spectre d'amplitude du signal modulé $s(t)$ est donc constitué de deux raies symétriques situées aux fréquences $f_0 - f_m$ et $f_0 + f_m$. De plus, il n'y a pas de composante spectrale à la fréquence f_0 de la porteuse. L'allure du spectre d'amplitude du signal modulé justifie l'appellation Double Bande Sans Porteuse (voir Figure 1.6).

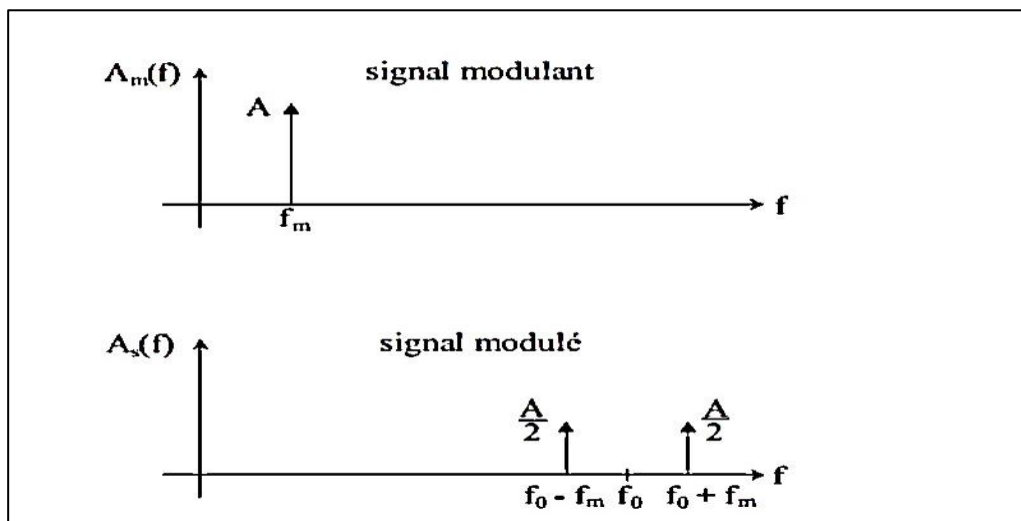


Figure 1.6 : Spectre d'amplitude du signal modulé.

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

Le signal modulé est un signal à bande étroite, centré autour de la fréquence f_0 de la porteuse. Le but de la modulation est atteint : le signal BF est transformé en un signal HF.

1.3.2.2 La modulation AM Double Bande Avec Porteuse

L'expression mathématique du signal s'écrit comme suit :

$$S(t) = A_c \cos(2\pi f_0 t) \quad (1.12)$$

La porteuse et $m(t)$ le message à transmettre. Le signal AM DBAP s'écrit :

$$S(t) = (A_c + m(t)) \cos(2\pi f_0 t) \quad (1.13)$$

Dans le cas d'un signal modulant sinusoïdal $s(t) = A_m \cos(2\pi f_m t)$, le signal AM DBAP devient :

$$s(t) = (A_c + A_m \cos(2\pi f_m t)) \cos(2\pi f_0 t) \quad (1.14)$$

$$s(t) = A_c \left(1 + \frac{A_m}{A_c} \cos(2\pi f_m t)\right) \cos(2\pi f_0 t) \quad (1.15)$$

$$s(t) = A_c (1 + k \cos(2\pi f_m t)) \cos(2\pi f_0 t) \quad (1.16)$$

Avec $K = \frac{A_m}{A_c}$: indice de modulation (ou taux de modulation) qui est le rapport entre l'amplitude du signal modulant et celle de la porteuse. Pour un signal modulant quelconque, l'indice de modulation est défini par :

$$k = \frac{|m(t)|_{\max}}{A} \quad (1.17)$$

La représentation temporelle du signal AM DBAP est illustrée dans la Figure 1.7 :

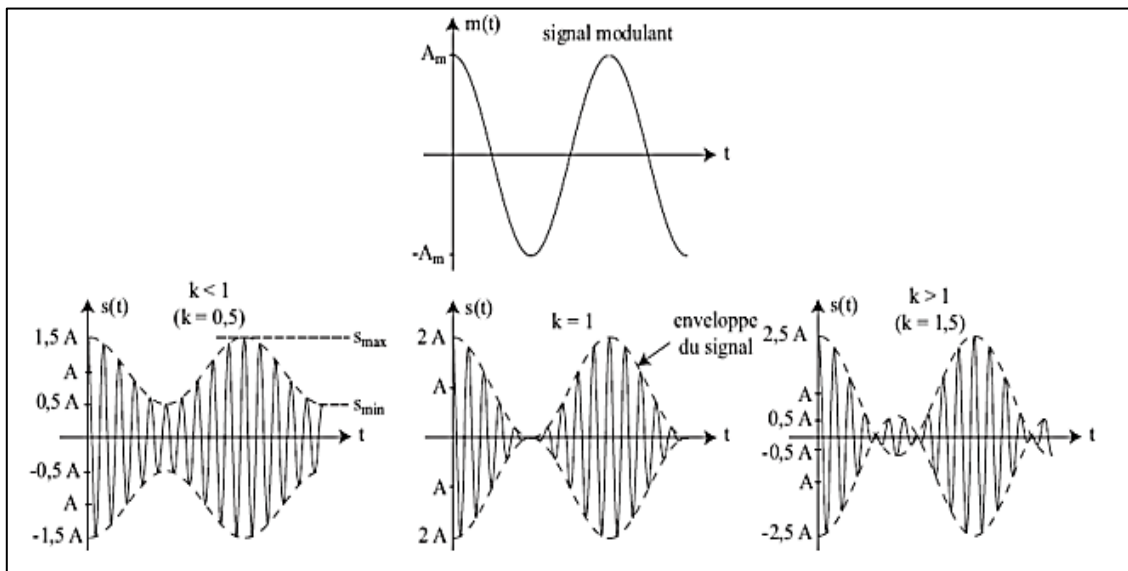


Figure 1.7 : Représentation temporelle du signal AM DBAP.

Lorsque $k < 1$, l'enveloppe reste positive (pas de distorsion), pour $k = 1$, on atteint la modulation critique (enveloppe tangente à zéro), et pour $k > 1$, il y a sur-modulation avec inversion d'enveloppe, ce qui explique les trois cas représentés.

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

Pour la représentation spectrale du signal AM DBAP (Figure 1.8), on a les expressions suivantes :

$$S(t) = (1 + k \cos(2\pi f_m t)) \cos(2\pi f_0 t) \quad (1.18)$$

$$S(t) = A_c(1 + k A \cos(2\pi f_m t)) \cos(2\pi f_0 t) \quad (1.19)$$

$$S(t) = A_c \cos(2\pi f_0 t) + \frac{kA}{2} \cos 2(2\pi f_0 - f_m)t + \frac{kA}{2} \cos 2(2\pi f_0 + f_m)t \quad (1.20)$$

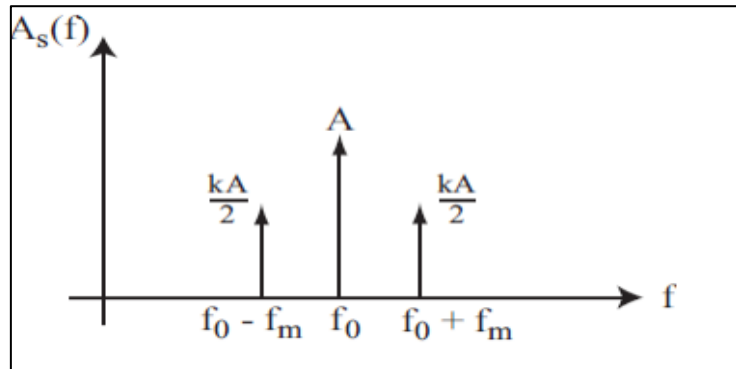


Figure 1.8 : Représentation spectrale du signal AM DBAP.

1.3.2.3 La modulation AM à Bande Latérale Unique

En modulation AM double bande, avec ou sans porteuse, le spectre du signal modulé présente deux bandes latérales symétriques autour de la fréquence de la porteuse. Ces deux bandes latérales se déduisent l'une de l'autre, donc elles contiennent chacune la même information. Leur occupation spectrale vaut le double de celle du signal en bande de base [14].

La Figure 1.9 montre les caractéristiques d'un signal BLU.

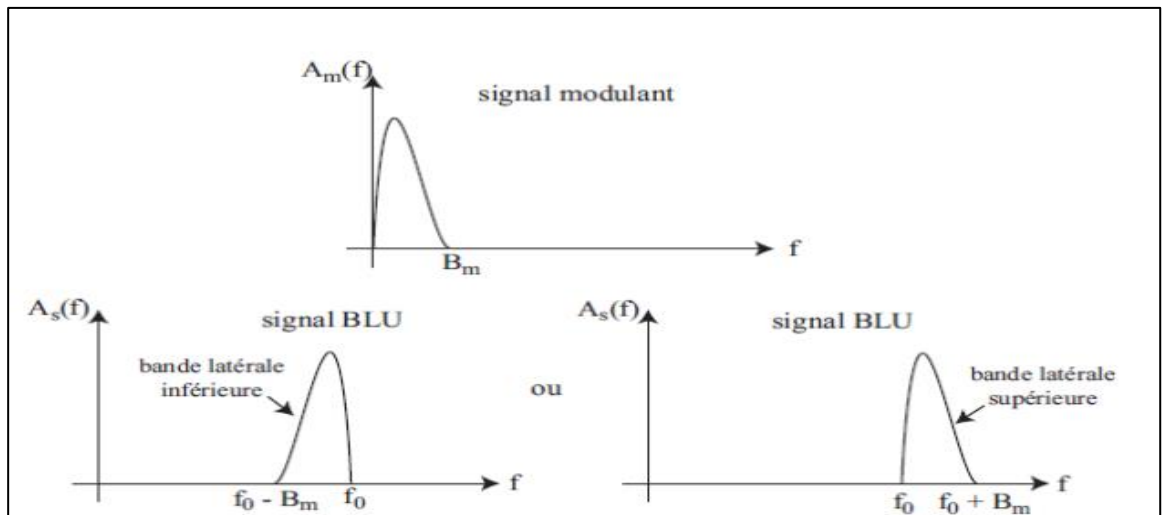


Figure 1.9 : Caractéristiques du signal BLU.

L'occupation spectrale du signal BLU est donnée par l'équation :

$$B_s = B_m \quad (1.21)$$

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

1.3.2.4 Modulation AM Bande Latérale Résiduelle (AM-BLR)

La modulation AM en bande latérale résiduelle emploie la discrimination de fréquence. Mais pour la modulation (BLR), le filtrage n'élimine pas complètement la bande latérale secondaire.

Aussi, la bande principale n'est pas complètement passée par le filtre [7].

Le principe de modulation AM-BLR est illustré par la Figure 1.10.

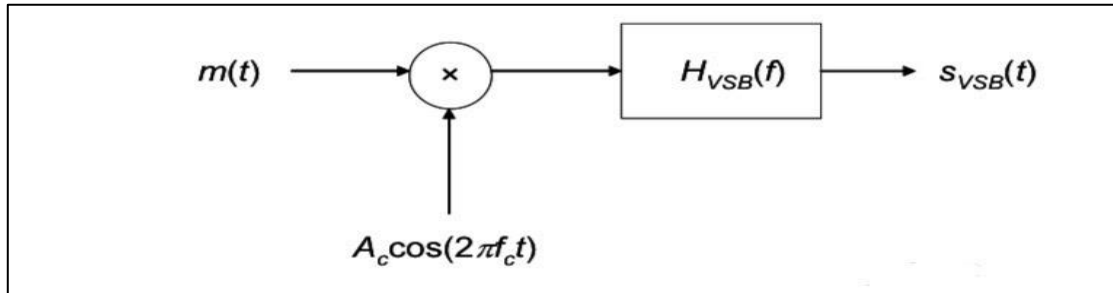


Figure 1.10 : Principe de modulation AM-BLR.

Un tel système qui utilise la bande latérale haute comme la bande principale et la bande latérale de basse comme la bande résiduelle. Mais c'est également possible d'employer la bande latérale basse comme la bande principale [16].

La représentation graphique du spectre est donnée par la Figure 1.11.

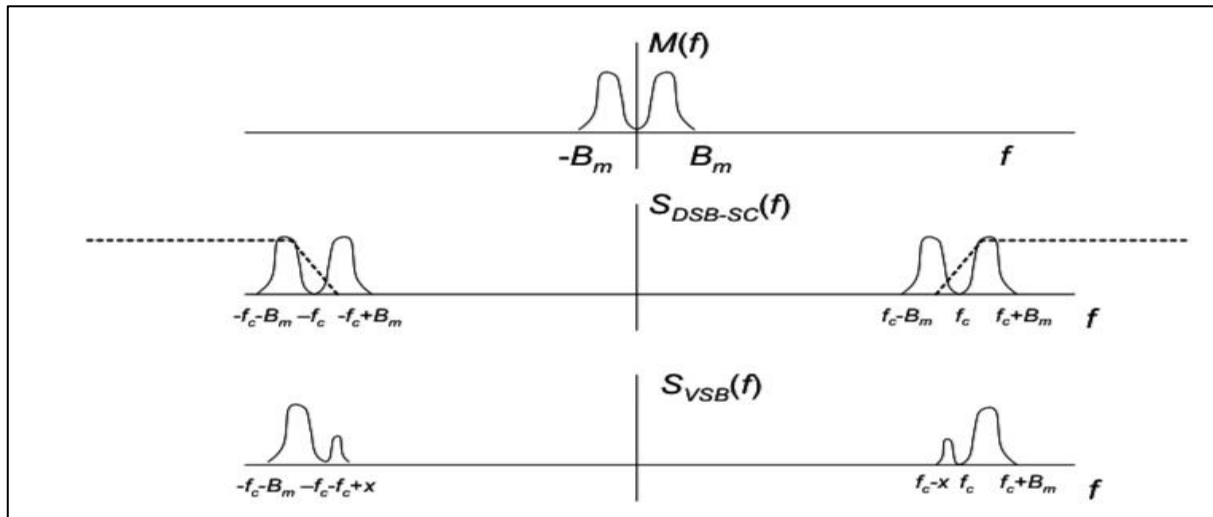


Figure 1.11 : Représentation du spectre d'un signal BLR.

1.3.3 La modulation de fréquence (FM)

La modulation de fréquence (FM) est une technique de modulation analogique dans laquelle la fréquence instantanée de la porteuse varie corrélativement à l'amplitude du signal d'information [15,17].

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

La modulation de fréquence (FM) consiste à transmettre une information en faisant varier la fréquence du signal porteur, tandis que son amplitude reste constante [18].

La représentation temporelle d'un signal FM est illustrée à la Figure 1.12 :

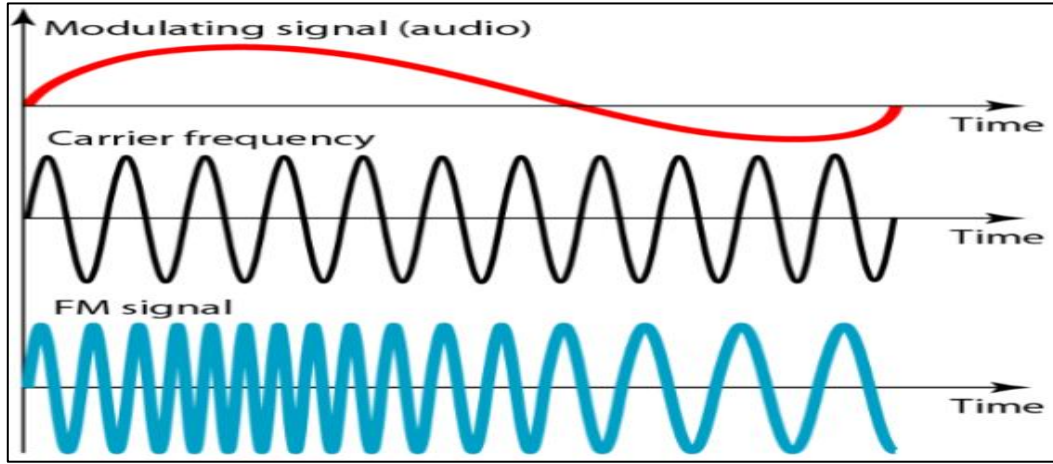


Figure 1.12 : Représentation temporelle de la modulation FM.

La modélisation mathématique du signal FM s'écrit :

$$S(t) = A_c \cos(2\pi f_c t + 2\pi k_f m(t)) \quad (1.22)$$

Soit :

- $m(t)$: signal modulant.
- A_c : amplitude de la porteuse.
- f_c : fréquence centrale de la porteuse.
- k_f : sensibilité en fréquence (Hz/V).
- A_m : Amplitude du signal d'information (signal modulant).
- ω_m : Pulsation du signal modulant

$$\text{Si :} \quad m(t) = A_m \cos(\omega_m t) \quad (1.23)$$

$$\text{Alors :} \quad \int_0^t m(t) dt = \frac{A_m}{\omega_m} \sin(\omega_m t) \quad (1.24)$$

$$\text{Donc :} \quad S(t) = A_c \cos(2\pi f_c t + \beta \sin(\omega_m t)) \quad (1.25)$$

$$\text{La phase instantanée :} \quad \theta(t) = 2\pi f_0 t + \varphi(t) + \varphi_0 \quad (1.26)$$

$$\text{La fréquence instantanée du signal est :} \quad f_i(t) = f_c + k_f m(t) \quad (1.27)$$

$$\text{La déviation de fréquence maximale :} \quad \Delta f = k_f \cdot \max |m(t)| \quad (1.28)$$

$$\text{L'indice de modulation FM est défini par :} \quad \beta = \frac{B_m}{\Delta f} = \frac{k_f A_m}{\omega_m} \quad (1.29)$$

Où (B_m) est la bande du signal modulant.

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

Afin de définir le type de modulation de fréquence selon β :

Si : ($\beta \ll 1$) : FM étroite .

($\beta \gg 1$) : FM large.

1.3.3.1 Spectre d'un signal FM

Le spectre de modulation de fréquence est infini théoriquement, mais la majorité de l'énergie est contenue dans une bande de largeur approximative donnée par la règle de Carson :

$$B_{FM} = 2(\Delta_f + B_m) = 2B_m(\beta + 1) \quad (1.30)$$

Le spectre comporte des composantes latérales multiples, avec des amplitudes données par les fonctions de Bessel $J_n(\beta)$ [21,3] :

$$S_{FM}(t) = A_{cn} = \sum_{-\infty}^{+\infty} J_n(\beta) \cos[2\pi(f_c + nf_m)t] \quad (1.31)$$

La représentation du spectre FM est donnée par la Figure 1.13.

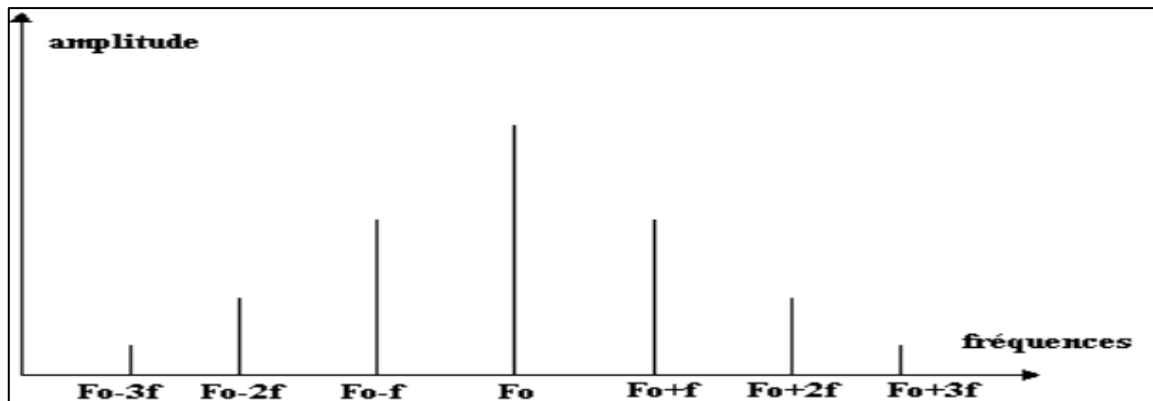


Figure 1.13 : Exemple d'un spectre FM.

La représentation graphique de la fonction de Bessel est montrée dans la Figure 1.14.

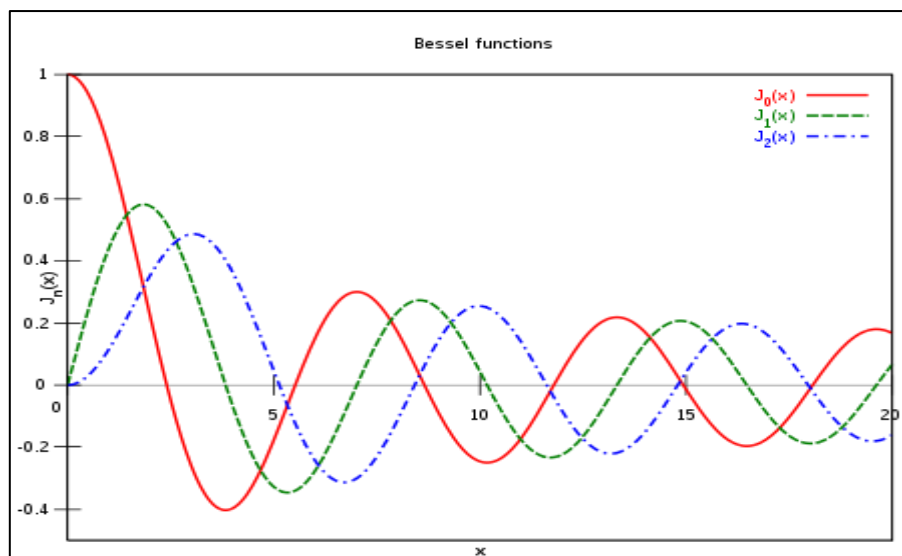


Figure 1.14 : Représentation graphique des trois premières fonctions de Bessel.

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

le Tableau 1.1 illustre les dix premières fonctions de Bessel en fonction de m

m	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}
0,00	1.00										
0,25	0.98	0.12									
0,5	0.94	0.24	0.03								
1,0	0.77	0.44	0.11	0.02							
1,5	0.51	0.56	0.23	0.06	0.01						
2,0	0.22	0.58	0.35	0.13	0.03						
2,5	-0.05	0.50	0.45	0.22	0.07	0.02					
3,0	-0.26	0.34	0.49	0.31	0.13	0.04	0.01				
4,0	-0.40	-0.07	0.36	0.43	0.28	0.13	0.05	0.02			
5,0	-0.18	-0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02		
6,0	0.15	-0.28	-0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02	
7,0	0.30	0.00	-0.30	-0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02
8,0	0.17	0.23	-0.11	-0.29	-0.10	0.19	0.34	0.32	0.22	0.13	0.6

Tableau 1.1 : Tableau des dix premières fonctions de Bessel en fonction de m .

1.3.4 La modulation de Phase (PM)

La modulation de phase (PM) est une technique de modulation analogique où la phase instantanée de la porteuse est proportionnelle au signal modulant ($m(t)$) [15].

La représentation graphique de la modulation de phase est donnée par la Figure 1.15.

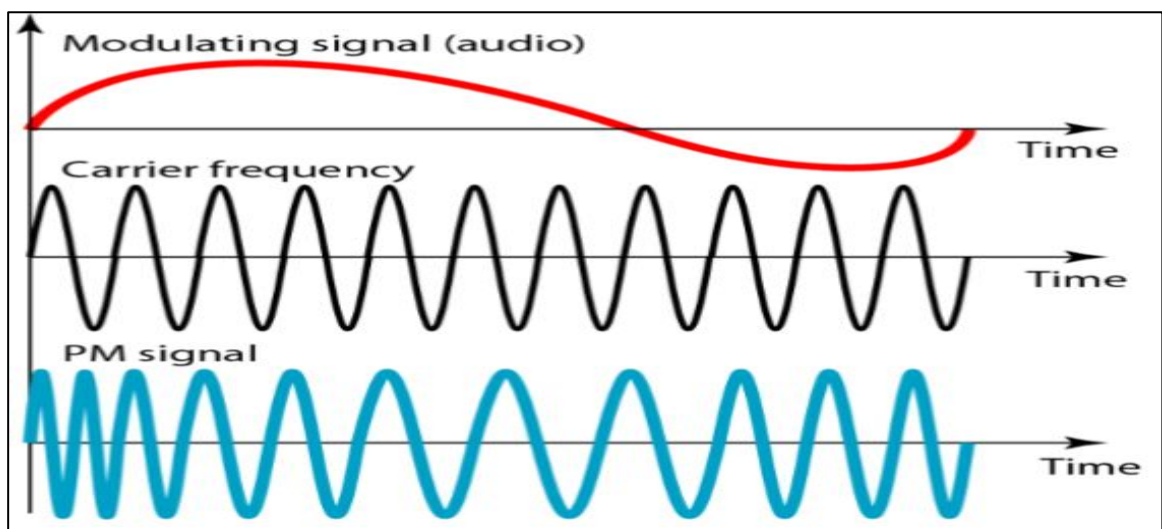


Figure 1.15 : Représentation graphique de la modulation de phase.

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

Le signal modulé en phase s'écrit :

$$S_{PM}(t) = A_c \cos(w_c t + k_p m(t)) \quad (1.32)$$

Contrairement à l'AM (modulation en amplitude) et FM (modulation en fréquence), l'information est codée dans la phase de la porteuse [14].

La fréquence instantanée varie indirectement avec ($m(t)$) :

$$f_i(t) = f_c + \frac{1}{2\pi} \frac{d\phi}{dt} \quad (1.33)$$

$$\text{Le signal PM s'écrit : } S_{PM}(t) = A_c \cos[2\pi f_c t + k_p m(t)] \quad (1.34)$$

Soit :

- A_c : amplitude de la porteuse.
- f_c : fréquence de la porteuse.
- k_p : sensibilité de phase (rad/V).

La phase instantanée est donnée par :

$$\phi(t) = 2\pi f_c t + k_p m(t) \quad (1.35)$$

$$\text{Pour PM, l'indice de modulation } \beta \text{ est : } \beta = k_p \cdot \max |m(t)| \quad (1.36)$$

Il définit l'amplitude des composantes latérales dans le spectre.

PM et FM larges (WBFM/WBPM) produisent de nombreuses bandes latérales [18,19].

Le spectre PM est similaire à FM : nombre infini de bandes latérales, amplitudes données par les fonctions de Bessel ($J_n(\beta)$) [21]:

$$S(t) = A_{cn} = -\sum_{n=-\infty}^{\infty} J_n(\beta) \cos[2\pi(f_c + n f_m)t] . \quad (1.37)$$

$$\text{Largeur de bande approximative (règle de Carson) : } B \approx 2(\Delta_f + B_m) \quad (1.38)$$

$$\text{Où } (\Delta_f) : \text{déviation de fréquence maximale } \Delta_f = \frac{k_p}{2\pi} \cdot \max \left| \frac{dm(t)}{dt} \right| \quad (1.39)$$

1.4 La modulation numérique et ses types de base

1.4.1 Principe de modulation numérique

En modulation numérique, la modulation est appliquée à une porteuse en modifiant l'amplitude, la fréquence ou la phase de cette dernière. Ce qui caractérise une modulation numérique est que le signal modulant (le message) est de nature numérique [20].

1.4.2 La modulation par déplacement d'amplitude (M-ASK -Amplitude Shift Keying)

La modulation ASK est une technique de modulation de signal qui varie l'amplitude de la porteuse en fonction du signal de données. En d'autres termes, elle manipule l'amplitude de la forme d'onde porteuse pour représenter les bits de données. Cette méthode offre une grande flexibilité et une efficacité accrue dans la transmission de données [21].

Le chronogramme de modulation ASK est illustré à la Figure 1.16.

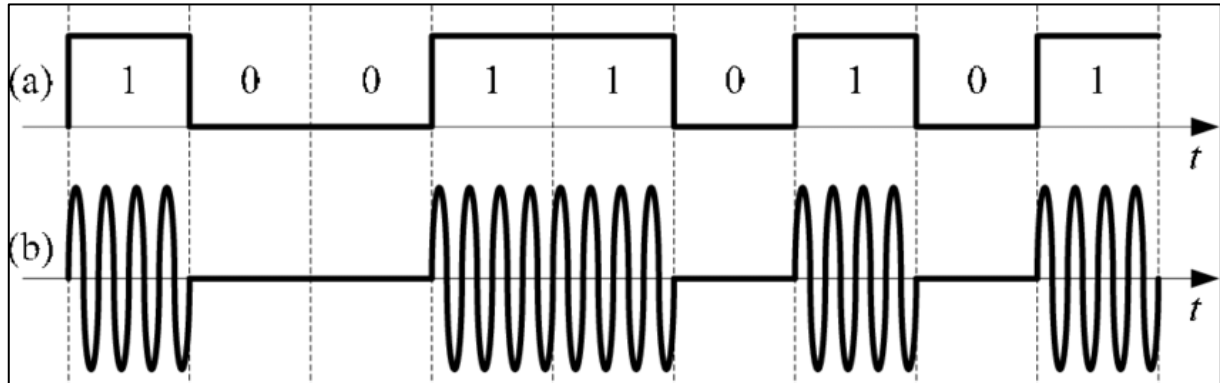


Figure 1.16 : Chronogramme de modulation ASK.

Le signal modulé ASK s'écrit: $S(t) = A_k \cos(2 \pi f_c t)$ (1.40)

Avec : A_k Amplitude dépendant du symbole

f_c : Fréquence de la porteuse

Pour une modulation M-ASK symétrique, les M symboles sont uniformément répartis sur une droite comme le montre la Figure 1.17.

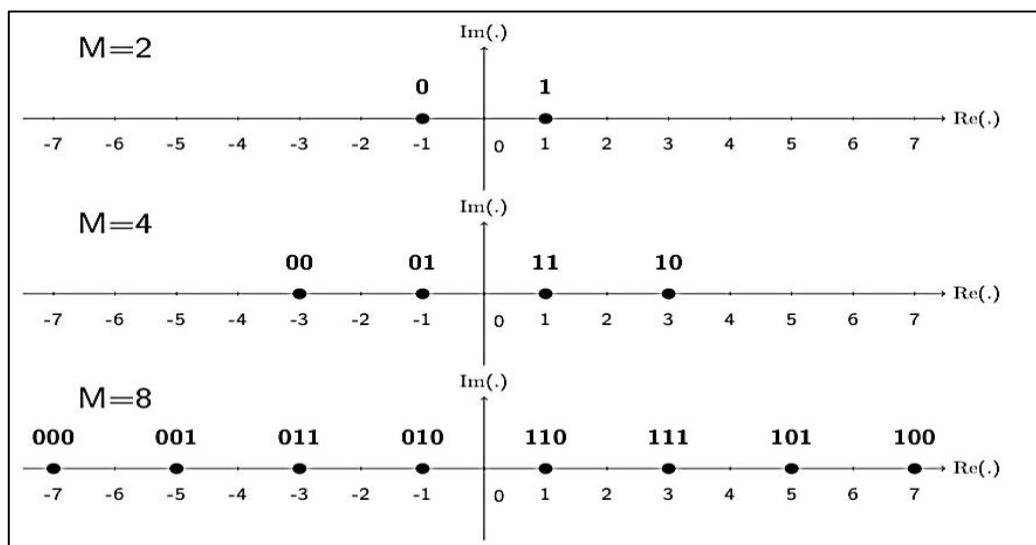


Figure 1.17 : Modulation M-ASK symétriques.

1.4.3 La modulation par déplacement de fréquence (M-FSK-Frequency Shift Keying)

La modulation FSK (Frequency Shift Keying) est une technique de modulation de fréquence utilisée en télécommunications pour transmettre des données numériques en variant la fréquence d'une onde porteuse. Cette méthode emploie deux fréquences distinctes pour représenter les états binaires 0 et 1, ce qui la rend robuste face au bruit et aux interférences. Le FSK est couramment utilisé dans les modems, les systèmes de communication radio et les technologies RFID [22].

Le signal FSK s'écrit :

$$S_i(t) = A_c \cos(2\pi f_i t + \phi_i) \quad , \quad i = 1, 2, \dots, M \quad (1.41)$$

La relation entre la déviation de la fréquence et le signal informatif est donnée comme suit :

$$\frac{1}{2\pi} \frac{d\phi(t)}{dt} = \frac{\delta f}{2} \sum_k a_k (g t - kT) . \quad (1.42)$$

L'intégrale de cette dernière équation, dans l'intervalle $[kT; (k+1)T]$, donne :

$$\phi(t) = \pi \Delta_f a_k (t - kT) + \theta_k \quad (1.43)$$

$$\theta_k : \text{est la constante d'intégration donnée par : } \theta_k = \phi kT \quad (1.44)$$

Cette équation montre que la fréquence varie d'une façon linéaire dans l'intervalle de temps $[kT; (k+1)T]$.

L'expression du signal modulé est donnée par :

$$S(t) = \cos[2\pi (f_c + \frac{\Delta_f}{2} a_k) t + \theta_k] . \quad (1.45)$$

La Figure 1.18 montre la représentation spectrale du signal modulé en FSK.

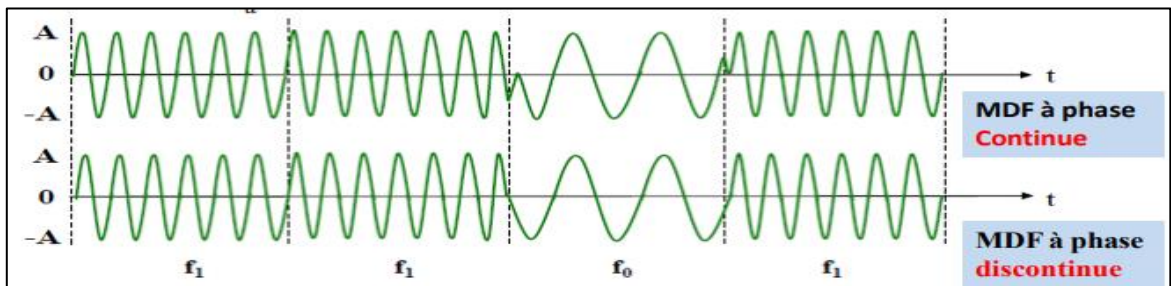


Figure 1.18 : Représentation spectrale du signal modulé en FSK.

f_0 : est la fréquence centrale.

$\frac{1}{2\pi} \frac{d\phi(t)}{dt}$: est la déviation de $2\pi dt$ la fréquence.

En fonction de la variation de phase, on distingue deux types de modulation MDF qui sont : la MDF à phase discontinue et la MDF à phase continue.

1.4.4 La modulation par déplacement de phase (M-PSK-Phase Shift Keying)

La modulation M-PSK (Modulation par déplacement de phase à M états) est une technique de modulation numérique dans laquelle l'information est transmise en faisant varier la phase du signal porteur, tout en gardant l'amplitude et la fréquence constantes.

Tels que :

- M représente le nombre de symboles possibles.
- Chaque symbole transporte $\log_2(M)$ bits [23].

La Figure 1.19 montre que le signal modulé possède deux fréquences différentes correspondant aux deux valeurs du symbole (0 et 1).

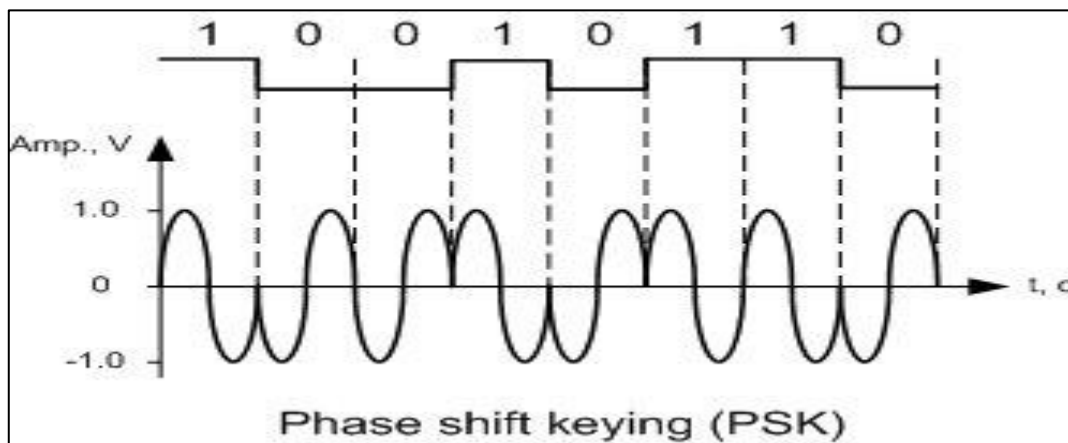


Figure 1.19 : Modulation par déplacement de phase PSK .

Le signal PSK se définit par l'expression suivante :

$$S(t) = A_c \cos (2\pi f_c t + \theta_k) \quad (1.45)$$

Où :

A_c : Amplitude de la porteuse

f_c : Fréquence de la porteuse

θ_k : Phase associée au symbole k

$$\theta_k = \frac{2\pi K}{M} \quad K = 0, 1, \dots, M-1 \quad (1.46)$$

On donne des exemples suivants : $M = 2 \rightarrow$ BPSK (1 bit/symbole), $M = 4 \rightarrow$ QPSK (2 bits/symbole) et $M = 8 \rightarrow$ 8-PSK (3 bits/symbole)

Les diagrammes de constellation de modulation M-PSK sont illustrés sur la Figure 1.20.

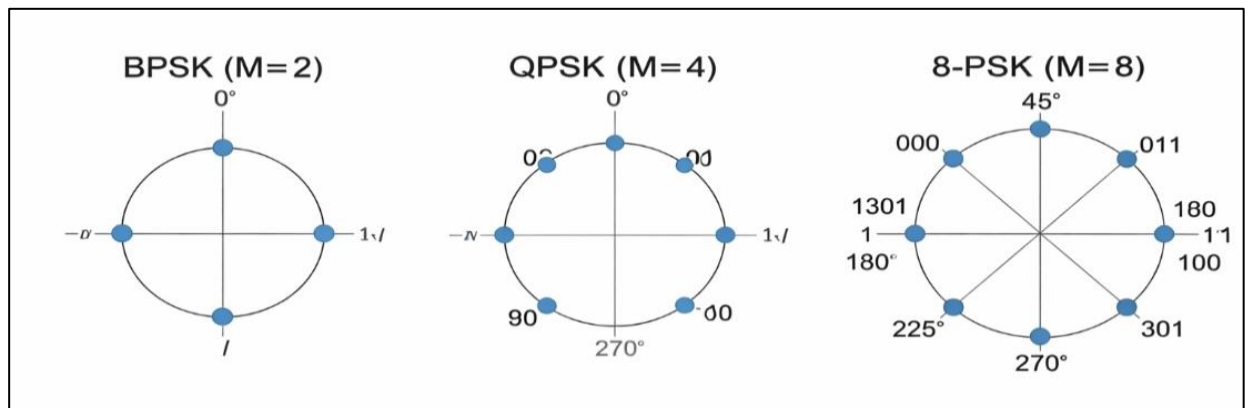


Figure 1.20 : Constellations M-PSK.

1.4.5 La modulation d'amplitude en quadrature QAM (Quadrature Amplitude Modulation)

La modulation d'amplitude en quadrature est une forme de modulation d'une porteuse par modification de l'amplitude de la porteuse elle-même et d'une onde en quadrature (une onde déphasée de 90° avec la porteuse) selon l'information transportée par deux signaux d'entrée.

Autrement dit, cela peut être considéré (utilisant une notation en nombre complexe) comme une simple modulation d'amplitude d'une onde, exprimée en complexe, par un signal, exprimé en complexe [24].

Le principe de la modulation d'amplitude en quadrature est illustré à la Figure 1.21.

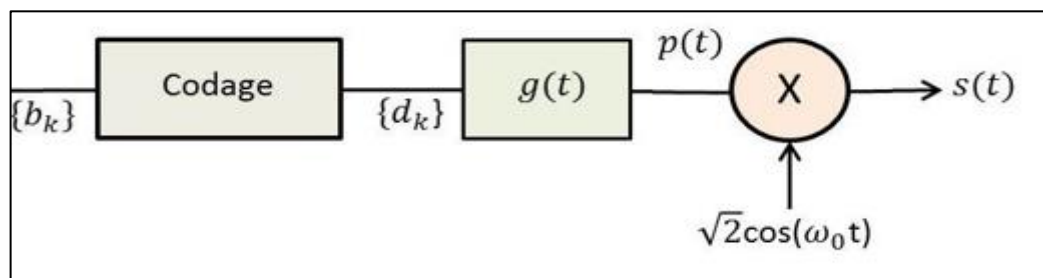


Figure 1.21 : Principe de modulation QAM.

Un signal QAM peut s'écrire sous la forme :

$$s(t) = I(t) \cos(2\pi f_c t) - Q(t) \sin(2\pi f_c t) \quad (1.47)$$

Où :

$I(t)$: Composante en phase (In-phase)

$Q(t)$: Composante en quadrature (Quadrature)

La Figure 1.22 montre deux exemples de la constellation 8 QAM et 16 QAM.

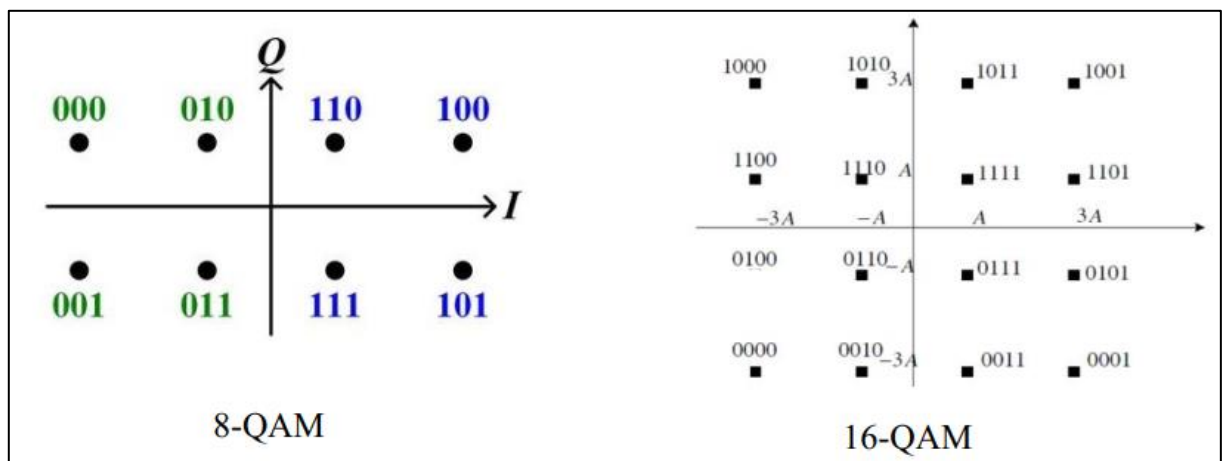


Figure 1.22 : Constellation de la modulation 8-QAM et 16-QAM.

1.5. Reconnaissance automatique des modulations

1.5.1. Définition de la reconnaissance automatique des modulations

La classification automatique des modulations revêt une importance significative dans le domaine des systèmes de communication sans fil. La prolifération des schémas de modulation, associée à la complexité croissante des environnements de canal, nécessite une précision accrue pour discerner les signaux modulés. Les méthodes traditionnelles peinent à classer efficacement cette variété de schémas de modulation. En réponse [25].

1.5.2. Domaines d'application

1.5.2.1. Communications militaires et guerre électronique

Dans le domaine militaire, la reconnaissance automatique des signaux permet d'identifier rapidement les émissions ennemies et de surveiller le spectre radio. Elle joue un rôle clé dans la détection des tentatives de brouillage et d'interception, et s'intègre notamment dans les systèmes de renseignement électronique (SIGINT) ainsi que dans l'adaptation des équipements de communication sécurisés face aux menaces [26].

1.5.2.2. Radios cognitives (Cognitive Radio)

Les radios cognitives tirent parti de la reconnaissance automatique des signaux pour analyser en permanence l'utilisation du spectre et identifier les types de transmissions actives sur une bande donnée. Sur la base de cette analyse, elles ajustent de manière autonome leurs paramètres de transmission fréquence, modulation, puissance afin de s'adapter aux conditions du canal. Cette intelligence spectrale leur permet d'optimiser l'utilisation des ressources disponibles tout en limitant les interférences entre utilisateurs [27].

1.5.2.3. Réseaux de télécommunications modernes

Dans les réseaux modernes tels que la 4G, la 5G et les futurs systèmes 6G, la classification des signaux joue un rôle central dans la gestion efficace des ressources radio. Elle facilite la surveillance et le diagnostic des transmissions dans des environnements de plus en plus complexes. Par ailleurs, elle contribue directement à l'amélioration de la qualité de service (QoS) et à la détection précoce des anomalies réseau. [28].

1.5.2.4. Communications satellitaires

Dans le domaine des communications satellitaires, la reconnaissance automatique de modulation permet d'identifier les signaux provenant de diverses sources orbitales ou terrestres. Elle assure une surveillance continue de la qualité des liaisons et facilite la détection des perturbations ou interférences susceptibles d'affecter les transmissions. Cette capacité renforce la fiabilité des échanges et simplifie la gestion des canaux dans des environnements spatiaux particulièrement contraignants. [26].

1.5.2.5. Recherche académique et intelligence artificielle

Sur le plan académique, la reconnaissance des modulations constitue un terrain de recherche fertile, notamment pour explorer de nouvelles approches fondées sur le Machine Learning et le Deep Learning. L'objectif est de concevoir des modèles capables d'apprendre automatiquement les caractéristiques distinctives des signaux radio sans intervention humaine. Des jeux de données spécialisés, tels que RadioML, servent de référence commune pour entraîner et évaluer les performances des algorithmes développés. [28].

1.5.2.6. Applications civiles et industrielles

Dans les environnements civils et industriels, la reconnaissance des signaux contribue à surveiller les interférences radio, à renforcer la sécurité des communications sans fil et à garantir la fiabilité des systèmes connectés. Elle s'applique à des domaines variés tels que les réseaux WiFi, l'Internet des objets (IoT), les systèmes de surveillance et les infrastructures intelligentes. Cette technologie permet également d'améliorer les performances des équipements déployés dans des environnements radio particulièrement chargés. [26].

1.5.3. Les enjeux de la Classification Automatique des Modulations (AMC) et contraintes

La classification automatique des modulations (AMC) joue un rôle stratégique dans les systèmes de communication modernes. Voici les principaux enjeux, appuyés par des références scientifiques reconnues [29].

CHAPITRE 1: GENERALITES SUR LES COMMUNICATIONS RADIO ET LA CLASSIFICATION DES MODULATIONS

1.5.3.1.Optimisation et gestion efficace du spectre radio

La classification automatique des modulations (AMC) offre la capacité d'identifier les signaux présents dans une bande de fréquence, contribuant ainsi à une exploitation plus rationnelle du spectre radio. Cette ressource, par nature limitée et très sollicitée, nécessite une gestion rigoureuse dans les systèmes de communication actuels. En identifiant précisément les signaux en présence, l'AMC permet de réduire les interférences, d'optimiser l'allocation des canaux et d'améliorer l'efficacité spectrale globale [27].

1.5.3.2.Sécurité et surveillance des communications

Sur le plan de la sécurité, l'AMC s'avère particulièrement utile pour détecter et identifier des signaux inconnus, suspects ou potentiellement malveillants, que ce soit dans les systèmes militaires, les dispositifs de surveillance ou les applications de renseignement radio. Elle permet d'anticiper les menaces avant qu'elles ne se concrétisent et de repérer rapidement toute tentative de brouillage ou d'intrusion électromagnétique. Cette capacité de détection précoce contribue à renforcer la protection des infrastructures de communication face à des risques toujours plus diversifiés [30].

1.5.3.3.Amélioration de la qualité de service (QoS) dans les réseaux modernes

Dans les réseaux 4G, 5G et les systèmes sans fil de demain, une identification précise du type de modulation contribue directement à l'amélioration de la qualité de service (QoS). Elle permet d'ajuster en temps réel les paramètres de transmission selon l'état du canal et la charge du réseau. Cette adaptabilité se traduit concrètement par un gain en débit, en fiabilité et en réduction de la latence [26].

1.5.3.4.Intégration de l'Intelligence Artificielle

L'un des défis les plus prometteurs de l'AMC aujourd'hui est l'intégration de l'intelligence artificielle, à travers le Machine Learning et le Deep Learning, pour automatiser l'extraction des caractéristiques et affiner la précision de classification même en conditions dégradées. Ces approches permettent de s'affranchir des limitations des méthodes traditionnelles en s'adaptant automatiquement à la variabilité des environnements radio. Elles ouvrent ainsi la voie à des systèmes de reconnaissance plus robustes, plus flexibles et capables de maintenir de bonnes performances face à des situations complexes et imprévisibles. [28].

1.6. Méthodes classiques de classification des modulations

Les méthodes classiques de classification automatique des modulations (AMC) reposent principalement sur l'extraction de caractéristiques statistiques et spectrales du signal reçu, suivie d'un classificateur (seuil, arbre de décision, maximum de vraisemblance, etc.) [26].

1.6.1. Approches basées sur des caractéristiques statistiques

Ces méthodes extraient des paramètres statistiques calculés à partir du signal reçu dans le domaine temporel ou complexe (I/Q), tels que : moments d'ordre supérieur, cumulant, kurtosis, variance instantanée et statistiques d'enveloppe.

Ces caractéristiques permettent de distinguer les modulations numériques (PSK, QAM, FSK, etc.) même sans connaissance complète du canal [30].

1.6.1.1.Approches basées sur des caractéristiques spectrales

Ces méthodes exploitent les propriétés fréquentielles du signal, tels que : densité spectrale de puissance (DSP), largeur de bande occupée, analyse cyclo stationnaire, spectre instantané et transformée de Fourier (FFT).

L'analyse cyclo-stationnaire est particulièrement efficace pour détecter les périodicités inhérentes aux signaux modulés, exploitables même en présence de bruit [31]

1.6.2. Limites en environnement bruité

En pratique, les performances des méthodes classiques se dégradent fortement lorsque le signal est affecté par le bruit, le fading ou les interférences.

1.6.2.1.Sensibilité au faible SNR

Quand le rapport signal sur bruit (SNR) chute à des niveaux trop bas, les caractéristiques extraites du signal se trouvent considérablement dégradées par le bruit ambiant. Les moments statistiques, les cumulant et les paramètres spectraux perdent alors leur fiabilité et deviennent difficiles à exploiter. Cette instabilité nuit directement à la précision des méthodes classiques et accroît le risque d'erreurs de classification, particulièrement dans les environnements radio soumis à de fortes perturbations [26].

1.6.2.2.Influence du fading et des canaux multi-trajets

Dans les conditions de propagation réelles, les phénomènes de fading et de multi-trajets perturbent en permanence l'amplitude, la phase et parfois même la structure fréquentielle du signal reçu. Ces effets introduisent une distorsion dans les caractéristiques exploitées pour la reconnaissance ou la classification des signaux. Il en résulte une dégradation notable des performances des approches classiques, surtout lorsque les conditions du canal évoluent rapidement et de manière imprévisible [30].

1.6.2.3.Dépendance à la synchronisation

Les méthodes classiques reposent en grande partie sur l'hypothèse d'une synchronisation temporelle et fréquentielle parfaite entre l'émetteur et le récepteur. Dès lors qu'un décalage temporel ou une erreur de fréquence porteuse survient, les caractéristiques extraites perdent leur exactitude et ne reflètent plus fidèlement le signal d'origine. Cette sensibilité aux imperfections de synchronisation peut rendre la décision finale peu fiable, voire totalement erronée [32].

1.6.2.4.Complexité computationnelle élevée (Méthode ML)

Bien que les approches par maximum de vraisemblance (ML) affichent de bonnes performances sur le plan théorique, elles impliquent une charge de calcul particulièrement lourde. L'estimation de multiples paramètres du canal, combinée à des procédures itératives coûteuses, mobilise d'importantes ressources de traitement. Cette exigence computationnelle constitue un frein majeur à leur déploiement dans des systèmes temps réel ou sur des plateformes embarquées aux capacités limitées [26].

1.6.2.5.Faible capacité de généralisation

Les méthodes classiques sont le plus souvent développées pour répondre à des scénarios bien définis, en s'appuyant sur des hypothèses strictes relatives au signal ou au canal de transmission. Confrontées à des environnements inédits ou à des conditions qui s'écartent de celles prévues lors de leur conception, leurs performances se détériorent rapidement. Cette rigidité les rend peu adaptées à la diversité et à la variabilité des situations rencontrées dans les environnements radio réels [33].

1.7. Conclusion

Ce chapitre constitue donc une base théorique solide pour aborder, dans la suite du travail, des méthodes plus avancées et plus performantes de classification automatique adaptées aux environnements radio réels.

Ce chapitre a présenté les fondements des communications radio numériques et de la reconnaissance automatique des modulations. Nous avons décrit la chaîne de transmission numérique, en expliquant le rôle de chaque bloc dans l'acheminement fiable de l'information.

Nous avons ensuite étudié la bande passante, le bruit et le canal de propagation, en montrant comment le bruit, souvent modélisé par un bruit blanc gaussien additif (AWGN), dégrade le rapport signal sur bruit (SNR) et augmente le taux d'erreur binaire (BER).

Les principales techniques de modulation, analogiques et numériques, ont été examinées afin de comprendre leur impact sur la robustesse, l'efficacité spectrale et la complexité du système.

Enfin, nous avons introduit la reconnaissance automatique des modulations et souligné son importance ainsi que les limites des approches classiques, notamment en conditions de faible SNR. Ce chapitre pose ainsi les bases théoriques nécessaires pour développer des méthodes de classification plus avancées adaptées aux environnements radio réel. Dans le chapitre suivant, nous allons aborder l'Intelligence Artificielle Explicable qui vise à rendre les décisions des modèles d'intelligence artificielle compréhensibles et interprétables par l'humain, afin de renforcer la confiance, la transparence et l'adoption des systèmes d'IA, en particulier dans les domaines critiques tels que la sécurité, la médecine, l'industrie, etc.

CHAPITRE 2:
INTELLIGENCE ARTIFICIELLE
EXPLICABLE (XAI)

2.1 Introduction

L'intelligence artificielle (IA) a récemment été réintroduite dans la plupart des domaines de recherche scientifique et d'innovation sociétale : les télécommunications fait partie de cette longue liste d'applications, avec d'innombrables cas d'usage étudiés et validés par les autorités publiques dans nombreux pays au monde.

Les modèles d'apprentissage automatique sont fréquemment sujets à une opacité intrinsèque, ce qui rend leurs décisions ardues à saisir. Cette « boîte noire » soulève d'importants défis concernant la transparence, la fiabilité et le respect des réglementations [34].

Pour répondre aux problèmes posés par l'utilisation croissante des modèles IA dans les applications à forts enjeux socio-économiques ou de sécurité, l'intelligence artificielle explicable (XAI) a connu un essor important durant les dernières années. Initialement dévolue à la recherche de solutions techniques permettant de produire automatiquement des explications, elle s'est heurtée à plusieurs difficultés, en particulier lorsque ces solutions ont été confrontées à des utilisateurs finaux non experts. L'XAI s'est alors attachée à s'inspirer des sciences sociales pour produire des explications plus faciles à comprendre. Malgré certains résultats encourageants, cette nouvelle approche n'a pas apporté autant qu'espéré [35].

Ce chapitre présente les concepts fondamentaux de l'Intelligence Artificielle Explicable (XAI). Il commence par introduire les bases de l'intelligence artificielle, ainsi que les notions de machine learning et de deep learning. Ensuite, les fondements théoriques du XAI sont abordés, en mettant l'accent sur les concepts d'interopérabilité, d'explicabilité et de transparence des modèles. Enfin, une typologie des principales méthodes d'explication est présentée, en distinguant les approches globales et locales, intrinsèques et post-hoc, ainsi que les méthodes dépendantes ou indépendantes du modèle.

2.2 L'intelligence artificielle

2.2.1 Concept de base

Un système d'intelligence artificielle (IA) est un ensemble de technologies et d'algorithmes conçus pour imiter ou reproduire certaines capacités humaines, telles que la perception, le raisonnement, l'apprentissage et la prise de décision.

Ces systèmes sont capables de traiter des données, d'apprendre de l'expérience, de s'adapter à des environnements changeants et de résoudre des problèmes de manière autonome ou semi-autonome.

L'IA fait appel à un certain nombre d'éléments dont les plus importants sont :

- Des algorithmes informatiques.
- De grandes bases de données.
- Des systèmes et matériels informatiques performants.

L'IA est apparue comme un domaine de recherche au milieu du XX^e siècle. Les premières bases théoriques ont été posées par Alan Turing en 1950 avec la proposition du test de Turing, destiné à évaluer la capacité d'une machine à imiter l'intelligence humaine. Le terme « intelligence artificielle » a ensuite été introduit en 1956 lors de la Dartmouth Conference par John McCarthy.

Dans les années 1960 et 1970, les chercheurs ont développé les premiers programmes capables de résoudre des problèmes logiques et les premiers systèmes experts. Cependant, les limitations technologiques ont entraîné une période de ralentissement appelée l'hiver de l'IA dans les années 1970 et 1980.

À partir des années 1990, l'IA connaît un renouveau grâce au développement du Machine Learning, qui permet aux machines d'apprendre à partir des données. Plus récemment, depuis les années 2010, les progrès du Deep Learning et l'augmentation de la puissance de calcul ont permis des avancées majeures dans plusieurs domaines comme la reconnaissance d'images, la traduction automatique et les systèmes intelligents [36].

2.2.2 Machine Learning

Le Machine Learning (apprentissage automatique) est un sous-domaine de l'Intelligence Artificielle qui permet aux systèmes informatiques d'apprendre à partir des données et d'améliorer leurs performances sans être explicitement programmés pour chaque tâche. Le concept repose sur la création d'algorithmes capables d'identifier des modèles et des relations dans les données afin de faire des prédictions ou de prendre des décisions.

Selon Tom M. Mitchell, le Machine Learning peut être défini comme suit : un programme informatique apprend à partir de l'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 .

Les méthodes de Machine Learning sont généralement classées en trois grandes catégories :

2.2.2.1 Apprentissage supervisé

L'apprentissage supervisé représente une méthode d'apprentissage automatique où le modèle est formé sur des données labellisées, autrement dit des paires entrée-sortie (x, y) . Le but est d'acquérir une fonction $f(x)$ qui peut correctement prédire les résultats pour de nouvelles données. On utilise principalement ce genre d'apprentissage pour des tâches de classification (associer une catégorie à une observation) et de régression (estimer une valeur continue). Parmi les algorithmes couramment utilisés, on trouve la régression linéaire, les machines à vecteurs de support et les réseaux neuronaux. En raison de sa précision, l'apprentissage supervisé est couramment employé en pratique. Toutefois, il requiert une quantité importante de données étiquetées, ce qui peut s'avérer coûteux à obtenir [37].

2.2.2.2 Apprentissage non supervisé

L'apprentissage non supervisé concerne les situations où les données ne sont pas étiquetées. L'algorithme cherche à découvrir des structures cachées ou des relations dans les données sans connaissance préalable des résultats attendus. Les principales tâches incluent le clustering (regroupement d'individus similaires), la réduction de dimensionnalité et la détection d'anomalies. Des méthodes telles que k-means, l'analyse en composantes principales (ACP) ou les modèles de mélange gaussien sont souvent utilisées. Ce type d'apprentissage est particulièrement utile pour explorer des données complexes et extraire des connaissances implicites [37].

2.2.2.3 Apprentissage par renforcement

Le paradigme de l'apprentissage par renforcement consiste à ce qu'un agent intelligent apprenne à faire des choix en interagissant avec un environnement en constante évolution. À chaque intervention, l'agent obtient un gain ou une sanction, ce qui lui donne la possibilité de perfectionner graduellement sa tactique (appelée politique) dans le but d'optimiser la récompense totale sur le long terme. On utilise ce genre d'apprentissage dans des secteurs comme la robotique, les jeux (comme les échecs ou les jeux vidéo) ainsi que les systèmes de recommandation. Il se démarque des autres méthodes par son aspect séquentiel et évolutif [37].

Le Machine Learning est aujourd'hui largement utilisé dans de nombreux domaines tels que la reconnaissance d'images, le traitement du langage naturel, les systèmes de recommandation et l'analyse prédictive, etc. [37].

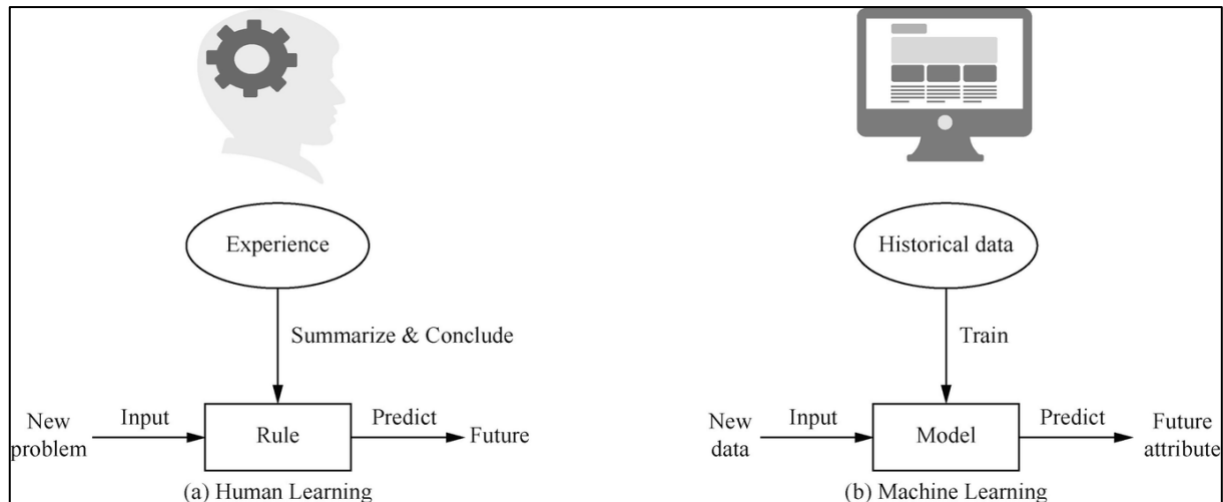


Figure 2.1 : Différence entre l'apprentissage humain et le Machine Learning.

La Figure 2.1 compare l'apprentissage humain et l'apprentissage automatique (Machine Learning). Dans l'apprentissage humain, l'individu utilise son expérience pour analyser les situations et en déduire des règles qui lui permettent de résoudre de nouveaux problèmes et de faire des prédictions. En revanche, dans le Machine Learning, la machine utilise des données historiques pour entraîner un modèle, qui sera ensuite capable de prédire des résultats à partir de nouvelles données.

2.2.3 Deep Learning

Le Deep Learning, ou apprentissage profond, est un sous-domaine du Machine Learning utilisant des réseaux de neurones artificiels multicouches inspirés du cerveau humain pour analyser de grandes quantités de données complexes, comme les images ou le texte, sans extraction manuelle de caractéristiques. Il excelle dans l'apprentissage hiérarchique où chaque couche traite des caractéristiques spécifiques et transmet les résultats aux suivantes, améliorant la précision grâce au big data et à la puissance de calcul accrue [38].

Les réseaux neuronaux profonds comportent des dizaines ou centaines de couches cachées : les données d'entrée passent par ces couches pour identifier des motifs progressifs, des bords simples aux objets complexes, via des ajustements de poids pendant l'entraînement. Les GPU accélèrent ce processus en gérant des calculs parallèles massifs, rendant possible l'entraînement sur des volumes énormes de données [38].

Le Deep Learning est particulièrement efficace dans des domaines tels que la reconnaissance d'images, le traitement du langage naturel, la reconnaissance vocale et les

systèmes de recommandation. Grâce à l'augmentation de la puissance de calcul et à la disponibilité de grandes bases de données, ces méthodes ont permis des avancées majeures dans plusieurs applications de l'intelligence artificielle [38].

La Figure 2.2 illustre le fonctionnement d'un réseau de Deep Learning qui transforme des données brutes en informations exploitables grâce à un réseau de neurones profonds capable d'apprendre automatiquement les caractéristiques pertinentes.

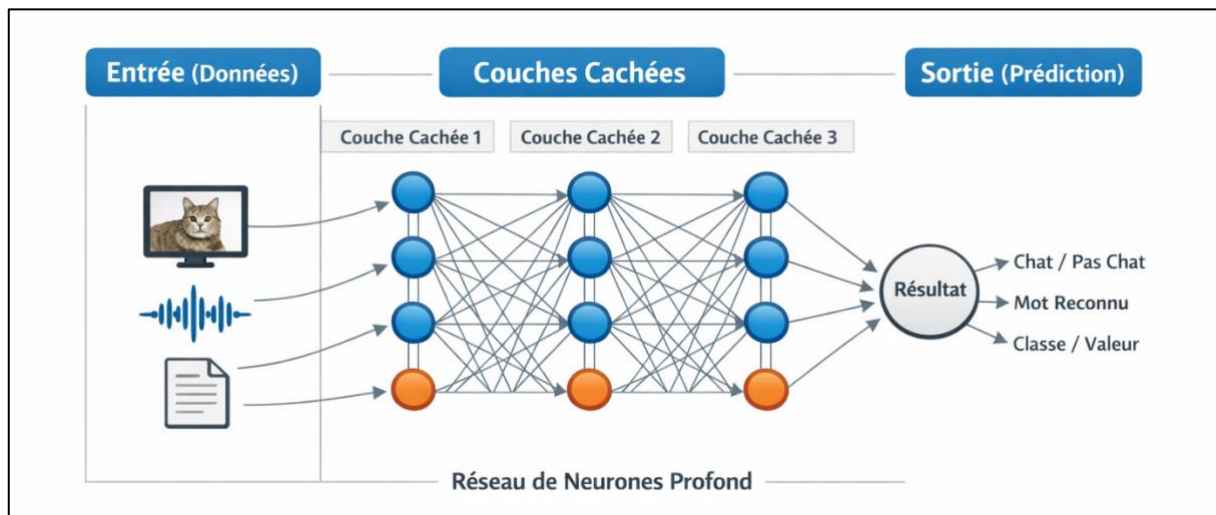


Figure 2.2 : Le fonctionnement d'un réseau de Deep Learning.

2.3 Fondements théoriques de l'IA explicable (eXplainable AI)

2.3.1 Définition et concepts clés

L'intelligence artificielle explicable (XIA) fournit à l'utilisateur des indices étayant l'hypothèse selon laquelle les corrélations à l'origine de la production d'un résultat à partir d'une entrée sont sous-tendues par l'existence de liens causaux [39].

Le besoin et la motivation pour l'XAI peuvent être résumés en deux objectifs principaux, comme illustré à la Figure 2.3. Les explications permettent de justifier les résultats d'un modèle d'apprentissage automatique en fournissant un motif du raisonnement suivi par le modèle. Une fois que les explications sont extraites, il est essentiel de vérifier la validité de ces explications pour s'assurer qu'elles correspondent aux intentions et aux attentes. Les explications qui révèlent des schémas différents de ceux qui étaient attendus par l'utilisateur peuvent également conduire à la découverte de nouvelles connaissances [40].

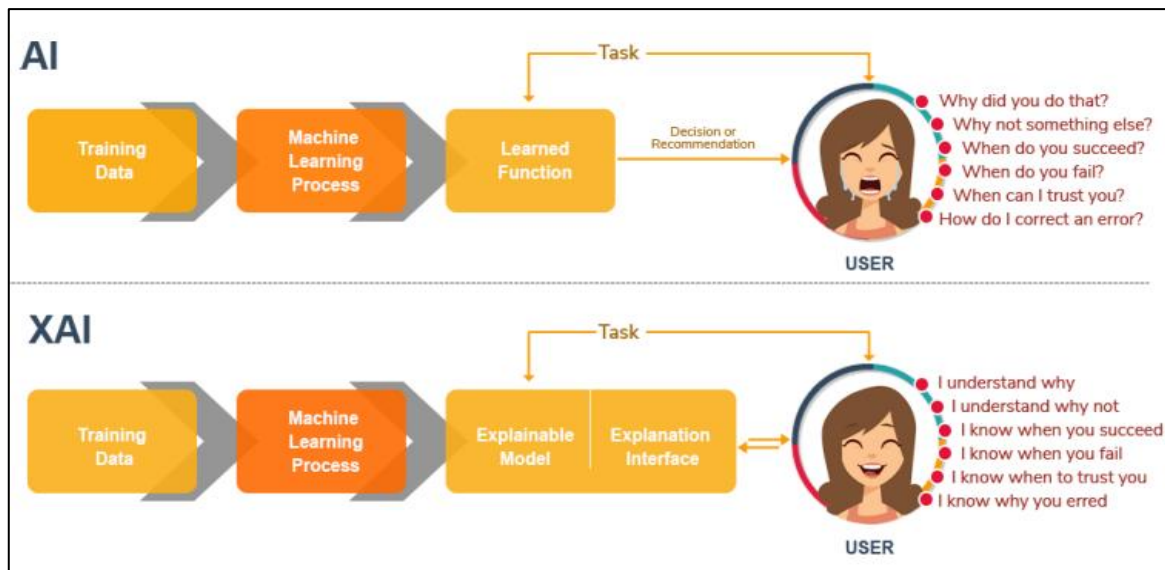


Figure 2.3 : Deux objectifs principaux pour l’XAI : justification et validation

2.3.1.1 Justification

La justification vise à expliquer pourquoi un modèle d’apprentissage automatique a pris une décision donnée. Concrètement, elle fournit un motif ou une raison du raisonnement suivi par le modèle pour arriver à un résultat. Par exemple, pour une classification, une explication pourrait décrire quelles caractéristiques (attributs) de l’entrée ont conduit à cette décision [40].

2.3.1.2 Validation

La validation consiste à vérifier la validité des explications produites par le modèle. Ici, il s’agit de s’assurer que ces explications correspondent réellement aux attentes et aux intentions de l’utilisateur du système d’IA, et qu’elles reflètent bien le comportement réel du modèle. Cela permet de contrôler si l’explication est fiable, cohérente et utile au-delà d’être seulement générée [40].

2.3.2 Typologie des méthodes XAI

Ces dernières années ont été marquées par une multiplication des méthodes d’explication. Ces méthodes génèrent des explications de différentes manières, aboutissant à divers types d’explications adaptées à différents types de données et de modèles d’apprentissage automatique. Nous pouvons identifier plusieurs types d’explications, chacun présentant un format spécifique. Outre le format des explications, un autre aspect crucial est la méthode de production de ces explications. En général, nous pouvons distinguer deux approches différentes pour expliquer un

modèle d'apprentissage automatique, qui permettent de générer des explications locales ou globale [40].

2.3.2.1 Explications globales vs locales

Dans le domaine de l'intelligence artificielle explicable (XAI), les méthodes d'interprétation peuvent être distinguées selon le niveau d'analyse des explications, à savoir les explications globales et les explications locales.

a) Explication globale du comportement du modèle

Les explications globales visent à comprendre le comportement général du modèle sur l'ensemble des données. Elles permettent d'identifier les relations entre les variables d'entrée et les prédictions du modèle, ainsi que l'importance relative de chaque caractéristique dans le processus de décision. Ce type d'explication est particulièrement utile pour analyser la logique globale du modèle, détecter d'éventuels biais et évaluer sa fiabilité. Par exemple, l'analyse de l'importance des variables ou les graphiques de dépendance partielle (Partial Dependence Plots) peuvent fournir une vision globale du fonctionnement du modèle [41].

b) Explication locale d'une prédiction individuelle

Les explications locales se concentrent sur une prédiction individuelle afin de comprendre pourquoi le modèle a produit un résultat spécifique pour une observation donnée. Elles analysent l'influence des différentes variables d'entrée dans la décision prise pour ce cas particulier. Des méthodes telles que LIME ou SHAP sont souvent utilisées pour générer ce type d'explication, car elles permettent d'identifier quelles caractéristiques ont le plus contribué à la prédiction.

Ainsi, les explications globales permettent d'obtenir une compréhension d'ensemble du modèle, tandis que les explications locales offrent une analyse détaillée d'une décision spécifique, ce qui contribue à améliorer la transparence et la confiance dans les systèmes d'apprentissage automatique [41].

Donc, Il est nécessaire d'établir une distinction en fonction des prédictions pour lesquelles les explications sont générées. D'un côté, les approches d'interopérabilité globale visent à fournir des explications pour aider l'utilisateur à acquérir une connaissance globale du modèle. Les informations extraites concernent donc le comportement général du modèle. D'un

autre côté, les approches locales visent à générer des explications pour une prédiction spécifique liée à une instance particulière en entrée [40].

2.3.2.2 Méthodes intrinsèques vs post-hoc

Les modèles d'apprentissage automatique peuvent être classés en deux catégories principales : les modèles intrinsèquement interprétables et les modèles boîte noire. Les modèles intrinsèquement interprétables, tels que la régression linéaire ou les arbres de décision, permettent de comprendre directement la relation entre les variables d'entrée et la prédiction grâce à une structure simple et transparente. À l'inverse, les modèles boîte noire, comme les réseaux de neurones profonds, offrent généralement de meilleures performances pour des problèmes complexes, mais leur processus de décision est difficile à interpréter en raison de leur structure interne complexe. Ainsi, le choix entre ces deux types de modèles repose souvent sur un compromis entre précision prédictive et interopérabilité, ce qui a conduit au développement de nombreuses méthodes d'explicabilité pour analyser les modèles complexes [41].

a) Méthodes intrinsèques

Les méthodes intrinsèques concernent des modèles naturellement interprétables, dont le fonctionnement peut être compris directement sans nécessiter d'outil supplémentaire. Ces modèles sont conçus de manière à rendre leurs décisions transparentes et faciles à analyser.

Parmi les exemples les plus courants :

- **Modèles linéaires**

Le modèle linéaire est l'un des modèles les plus simples et les plus interprétables en apprentissage automatique. Il repose sur l'hypothèse qu'il existe une relation linéaire entre les variables d'entrée (caractéristiques) et la variable de sortie. Dans ce type de modèle, la prédiction est obtenue en combinant les variables d'entrée avec des coefficients pondérés, auxquels s'ajoute une constante appelée biais. Chaque coefficient indique l'influence d'une variable sur la prédiction finale : un coefficient positif signifie que la variable augmente la valeur prédite, tandis qu'un coefficient négatif indique une influence inverse. Grâce à cette structure simple et transparente, les modèles linéaires sont considérés comme intrinsèquement interprétables, car il est possible de comprendre directement comment chaque variable contribue à la décision du modèle. Ils sont largement utilisés dans de nombreux domaines, notamment en statistique, en

économie et en apprentissage automatique, pour analyser les relations entre les variables et effectuer des prédictions [41].

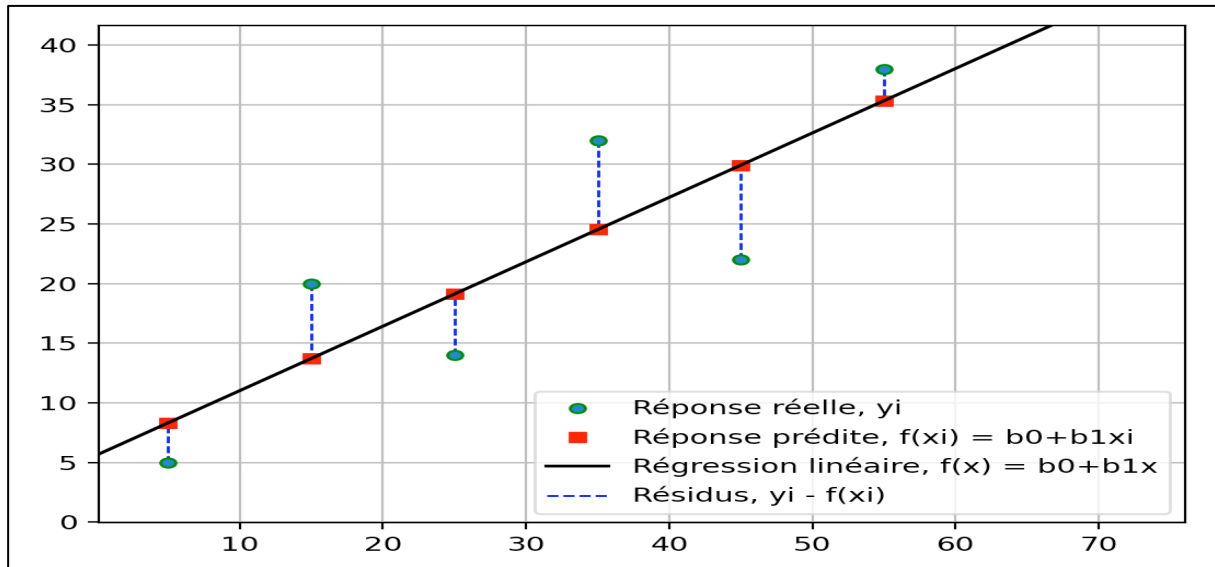


Figure 2.4 : Régression linéaire. Ajustement affine.

La Figure 2.4 illustre le principe de la régression linéaire, un type de modèle linéaire utilisé pour modéliser la relation entre une variable explicative x et une variable cible y .

La fonction estimée, qui est issue de la régression linéaire simple est tracée en ligne continue noire sur la même Figure. Cette fonction est sous la forme :

$$f(x) = b_0 + b_1 x_1. \quad (2.1)$$

Où :

- b_0 est l'ordonnée à l'origine (intercept),
- b_1 est le coefficient de pente, qui mesure l'influence de la variable x sur la variable y .

Lorsqu'on est face à une situation où les paramètres indépendants d'entrée sont au nombre de 2 ou plus, c'est la régression linéaire multiple qui permet de trouver un modèle mathématique approché.

S'il n'y a que deux variables indépendantes, lors la fonction de régression estimée est :

$$f(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2. \quad (2.2)$$

Le cas de plus de deux variables indépendantes est similaire, mais plus général. La fonction de régression estimée est :

$$f(x_1, \dots, x_n) = b_0 + b_1x_1 + \dots + b_nx_n, \quad (2.3)$$

- **Arbres de décision (decision trees)**

Les arbres de décision constituent l'un des modèles d'apprentissage automatique les plus populaires en apprentissage supervisé. Le processus de décision d'un arbre de décision est similaire à celui de l'être humain, prenant des décisions en suivant une série de règles (conjonction de règles). Cette logique simple dans le processus de décision rend l'arbre de décision interprétable. En général, un arbre de décision est composé d'un nœud racine (root node), de plusieurs nœuds de décision (decision nodes) et de nœuds feuilles (leaf nodes). Chaque nœud de décision utilise un attribut comme critère de test pour évaluer une instance. Les nœuds feuilles correspondent à des classes dans le cas d'un problème de classification ou à des valeurs réelles pour un problème de régression.

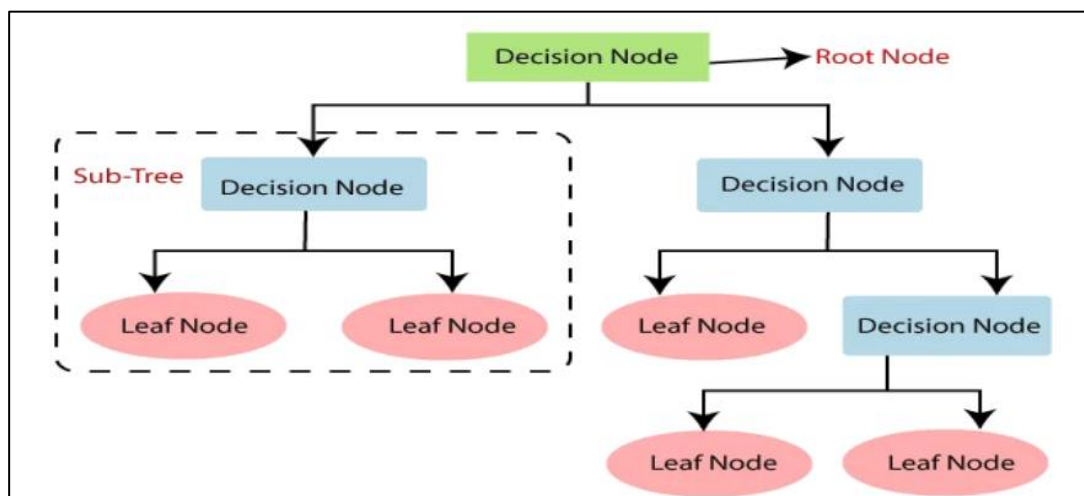


Figure 2.5 : Structure générale d'arbre de décision.

Un arbre de décision est présenté graphiquement à la Figure 2.5, Chaque sous-arbre représente un cas correspondant à un sous-ensemble de valeurs spécifiques de l'attribut sélectionné, selon un critère bien défini. Chaque chemin, de la racine à un nœud feuille, correspond à un terme qui regroupe les conditions rencontrées dans les nœuds de décision par lesquels le chemin passe. Pour prédire la sortie d'une nouvelle donnée inconnue, il faut trouver le chemin unique compatible avec cette donnée en parcourant l'arbre à partir du nœud racine jusqu'à un nœud feuille. La prédiction finale est la valeur associée au nœud feuille [40].

Ces modèles offrent une bonne transparence, mais ils peuvent être moins performants pour traiter des problèmes très complexes [41].

b) Méthodes post-hoc

Les méthodes post-hoc sont des techniques d'explicabilité utilisées pour interpréter les prédictions des modèles d'apprentissage automatique après leur phase d'entraînement, sans modifier leur structure interne. Elles sont principalement appliquées aux modèles boîte noire, tels que les réseaux de neurones profonds ou les modèles d'ensemble, afin de comprendre les facteurs qui influencent leurs décisions. Ces méthodes permettent de fournir des explications locales, qui analysent une prédiction spécifique, ou des explications globales, qui décrivent le comportement général du modèle. Parmi les approches post-hoc les plus connues figurent LIME et SHAP, qui évaluent l'influence des variables d'entrée sur la prédiction du modèle. Ces techniques contribuent à améliorer la transparence et la confiance dans les systèmes d'intelligence artificielle complexes [41].

2.3.2.3 Méthodes dépendantes du modèle vs indépendantes

Les méthodes d'explicabilité peuvent également être classées selon leur dépendance au modèle. On distingue principalement les méthodes spécifiques au modèle et les méthodes agnostiques.

Les méthodes spécifiques au modèle sont conçues pour un type particulier d'algorithme et exploitent sa structure interne afin de fournir des explications plus précises. Par exemple, certaines techniques sont développées spécialement pour analyser les réseaux de neurones profonds, comme les cartes de saillance (Saliency Maps) ou Grad-CAM, qui permettent d'identifier les caractéristiques de l'entrée ayant le plus d'influence sur la prédiction du modèle [41].

En revanche, les méthodes agnostiques ne dépendent pas du type de modèle utilisé et peuvent être appliquées à n'importe quel système d'apprentissage automatique, qu'il s'agisse d'un réseau de neurones, d'un arbre de décision ou d'un modèle d'ensemble. Des approches telles que LIME ou SHAP analysent le comportement du modèle uniquement à partir des entrées et des sorties, sans nécessiter d'accès à son fonctionnement interne. Cette distinction est essentielle dans l'XAI, car elle permet de choisir la méthode d'interprétation la plus adaptée au type de modèle utilisé [41].

Cette distinction est importante dans l'analyse de l'explicabilité, car les méthodes dépendantes du modèle offrent souvent des explications plus précises pour un type d'algorithme spécifique, tandis que les méthodes agnostiques présentent l'avantage d'être générales et applicables à différents modèles.

2.4 Principales technique d'explicabilité

2.4.1 LIME (Local Interpretable Model-Agnostic Explanations)

Proposée en 2016 par Ribeiro, Singh et Guestrin, la méthode LIME part d'un constat simple : il n'est pas toujours nécessaire de comprendre le fonctionnement interne d'un modèle pour en expliquer les décisions. Son principe repose sur la construction d'un modèle local simplifié généralement linéaire qui approxime le comportement du modèle complexe au voisinage d'une observation précise. L'un de ses atouts majeurs réside dans son caractère model-agnostic : LIME s'applique indifféremment à tout type de modèle, qu'il s'agisse d'un réseau de neurones, d'un arbre de décision ou de tout autre algorithme, sans nécessiter l'accès à sa structure interne [42].

2.4.1.1 Principe de construction d'un modèle local approximatif

LIME est un modèle complexe souvent incompréhensible dans sa globalité, mais son comportement autour d'une observation précise peut être bien plus facile à saisir.

La méthode fonctionne en trois temps.

D'abord, on crée un "voisinage" autour de l'observation qu'on veut expliquer. Concrètement, on génère des versions légèrement modifiées de cette observation, en perturbant un peu ses caractéristiques, comme si on testait des scénarios proches du cas réel.

Ensuite, on soumet toutes ces variantes au modèle complexe pour récupérer ses prédictions. Cela permet de voir comment le modèle réagit quand les variables changent légèrement, et donc de cartographier son comportement local.

Enfin, on ajuste un modèle simple généralement une régression linéaire sur ces données, en accordant plus d'importance aux variantes les plus proches de l'observation originale. Les coefficients de ce modèle simple deviennent alors l'explication : ils révèlent quelles variables ont joué en faveur ou en défaveur de la prédiction.

Ce qui rend LIME particulièrement puissant, c'est qu'il est agnostique au modèle : peu importe la complexité de la boîte noire utilisée, la méthode s'applique de la même façon. On renonce à comprendre le modèle en entier, mais on gagne une explication claire et locale, centrée sur ce qui compte vraiment pour une décision donnée [42].

2.4.1.2 Description mathématique du processus de perturbation et d'optimisation locale

a) Processus de perturbation

La description mathématique de LIME repose sur la formulation d'un problème d'optimisation visant à approximer localement un modèle complexe autour d'une observation donnée.

Soit :

- $f : \mathbb{R}^d \rightarrow \mathbb{R}$ le modèle complexe (boite noire).
- $x_0 \in \mathbb{R}^d$ l'observation à expliquer.
- $g \in G$ une famille de modèles interprétables.
- $\pi_{x_0}(x)$ une fonction de pondération mesurant la proximité entre x et x_0 .

LIME introduit une fonction de proximité $\pi_{x_0}(x)$ afin de privilégier les points proches de x_0 .

Chaque point est pondéré par une fonction noyau, généralement exponentielle :

$$\pi_{x_0}(x_i) = \exp\left(-\frac{D(x_i, x_0)}{\sigma^2}\right)^2 \quad (2.4)$$

où $D(\cdot)$ est une mesure de distance et σ contrôle l'étendue du voisinage.

Plus x_i est proche de x_0 , plus son poids est élevé.

b) Optimisation locale

On cherche un modèle interprétable $g \in G$ qui approxime localement f .

LIME cherche à résoudre le problème d'optimisation suivant :

$$\underset{g \in G}{\operatorname{argmin}} L(f, g, \pi_{x_0}) + \Omega(g) \quad (2.5)$$

où :

- $L(f, g, \pi_{x_0})$ mesure la fidélité locale de g par rapport à f ,
- $\Omega(g)$ pénalise la complexité du modèle interprétable (pour garantir sa simplicité).

Dans le cas d'une approximation linéaire locale, le modèle interprétable s'écrit :

$$g(x) = \beta_0 + \sum_{j=1}^d \beta_j x_j \quad (2.6)$$

Les paramètres β sont estimés en minimisant une erreur quadratique pondérée :

$$\min_{\beta} \sum_{i=1}^N \pi_{x_0}(x_i) (f(x_i) - g(x_i))^2 + \lambda \|\beta\|_1 \quad (2.7)$$

Le terme $\lambda \|\beta\|_1$ favorise la parcimonie afin de conserver un nombre limité de variables explicatives.

Ainsi, les coefficients estimés constituent l'explication locale de la prédiction $f(x_0)$, valable uniquement dans le voisinage immédiat de l'observation étudiée [42].

2.4.1.3 Avantages et limites

a) Avantages de LIME

LIME présente plusieurs atouts qui expliquent son adoption large dans le domaine de l'explicabilité des modèles.

Indépendance vis-à-vis du modèle (Model-Agnostic) : LIME n'a besoin que des prédictions du modèle pour fonctionner, sans accès à sa structure interne, ce qui lui permet de s'appliquer universellement à tout type de modèle [42].

Interopérabilité locale claire : LIME produit des explications centrées sur un cas précis, où chaque coefficient de l'approximation linéaire indique directement l'influence d'une variable sur la prédiction, qu'elle soit positive ou négative [42].

Applicabilité à différents types de données : LIME fonctionne indifféremment sur des données tabulaires, textuelles ou images, ce qui lui confère une flexibilité rare parmi les méthodes d'explicabilité [42].

Simplicité conceptuelle : LIME repose sur une régression locale pondérée, un mécanisme intuitif qui le rend accessible à des non-spécialistes tout en conservant la rigueur de ses explications [41].

b) Limites de LIME

Malgré ses atouts, plusieurs limites ont été identifiées dans la littérature.

Validité strictement locale : Les explications de LIME sont cantonnées au voisinage immédiat de l'observation et ne peuvent en aucun cas être généralisées à l'ensemble du modèle [42].

Instabilité des explications : La génération aléatoire des perturbations rend les résultats variables d'une exécution à l'autre, ce qui nuit à leur fiabilité et leur reproductibilité [43].

Sensibilité au choix du voisinage : Le paramètre de taille du voisinage est difficile à calibrer : trop large, il dilue l'explication locale ; trop étroit, il manque d'information. Dans les deux cas, la qualité de l'explication en souffre [41].

Vulnérabilité aux manipulations adversaires : LIME peut être délibérément trompé par des modèles conçus pour produire des explications fallacieuses, ce qui remet en cause sa fiabilité dans des contextes sensibles [44].

2.4.2 SHAP (SHapley Additive exPlanations)

SHAP (SHapley Additive exPlanations) est une méthode d'interprétabilité des modèles d'apprentissage automatique fondée sur la théorie des jeux coopératifs, introduite par le mathématicien Lloyd Shapley. Cette approche permet d'expliquer la prédiction d'un modèle en attribuant à chaque variable d'entrée une contribution quantitative équitable [45].

2.4.2.1 Fondements théoriques basés sur la théorie des jeux coopératifs et les valeurs de Shapley

La méthode SHAP puise ses origines dans la théorie des jeux coopératifs, une branche des mathématiques et de l'économie qui s'intéresse à la manière dont des agents peuvent collaborer pour maximiser un gain collectif. À la différence des jeux non coopératifs où chaque joueur agit pour son propre compte, les jeux coopératifs mettent l'accent sur les coalitions — ces groupes de joueurs qui s'associent pour obtenir ensemble un meilleur résultat. La question fondamentale qui se pose alors est celle de l'équité : comment répartir le gain collectif entre les membres en fonction de la contribution réelle de chacun ?

Cette logique se transpose naturellement au domaine du Machine Learning. Le jeu devient la prédiction du modèle, les joueurs sont les variables d'entrée, et le gain à répartir correspond à l'écart entre la prédiction obtenue pour une instance donnée et la moyenne globale des prédictions. L'enjeu est donc de déterminer, de façon rigoureuse et équitable, dans quelle mesure chaque variable a contribué à la prédiction finale.

Ce cadre repose formellement sur la notion de jeu coopératif à utilité transférable, où le gain peut librement circuler entre les joueurs, et qui se définit par un couple associant l'ensemble des joueurs à une fonction caractéristique mesurant ce que chaque coalition peut obtenir par elle-même [46].

Un jeu coopératif à utilité transférable (TU – Transferable Utility) est généralement défini par un couple :

$$(N, v)$$

Où :

- $N = \{1, 2, 3, \dots, n\}$ est l'ensemble des joueurs.
- v est la fonction caractéristique :

$$v: 2^N \rightarrow \mathbb{R}$$

Cette fonction associe à chaque coalition de joueurs un gain réel.

Une coalition est tout sous-ensemble de joueurs :

$$S \subseteq N$$

- S représente un groupe de joueurs qui coopèrent.
- $v(S)$ représente le gain total obtenu par cette coalition.

Dans l'analyse coopérative, on mesure l'apport d'un joueur i à une coalition S par sa contribution marginale :

$$\Delta_i(S) = v(S \cup \{i\}) - v(S) \quad (2.8)$$

Elle représente l'augmentation du gain lorsque le joueur i rejoint la coalition S [45]. Les valeurs de Shapley fournissent une solution mathématiquement fondée pour répartir équitablement le gain entre les joueurs.

Pour une variable i , la valeur de Shapley est définie par :

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (2.9)$$

Où :

N : ensemble total des variables.

S : Sous-ensemble de variables n'incluant pas i .

$f(S)$: Prédiction du modèle en utilisant uniquement les variables de S .

ϕ_i : Contribution moyenne marginale de la variable i .

Cette formule calcule la contribution moyenne de la variable i sur toutes les combinaisons possibles des autres variables [45].

La Figure 2.6 illustre le fonctionnement global de la méthode SHAP, en montrant comment les valeurs de Shapley sont calculées et attribuées à chaque variable d'une observation pour expliquer la prédiction finale du modèle.

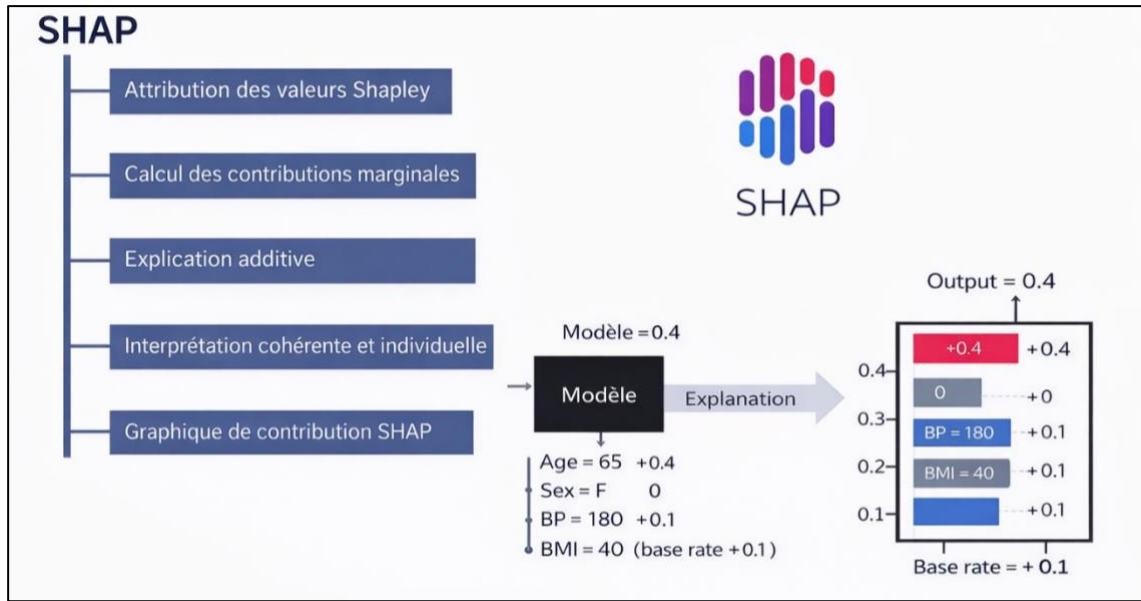


Figure 2.6 : Illustration du fonctionnement de la méthode SHAP et de l'attribution des valeurs de Shapley aux variables explicatives.

Emprunté à la théorie des jeux coopératifs, SHAP fournit des explications individuelles pour chaque prédiction d'un modèle d'apprentissage automatique en attribuant des valeurs Shapley à chaque caractéristique (feature) d'une observation.

a) Les propriétés fondamentales

Les valeurs de Shapley satisfont quatre propriétés essentielles :

Efficacité : La somme des contributions est égale à la différence entre la prédiction et la valeur moyenne :

$$\sum_i \phi_i = f(x) - E[f(x)] \quad (2.10)$$

Symétrie : Deux variables ayant la même contribution obtiennent la même valeur.

Nullité (Dummy) : Une variable sans influence reçoit une contribution nulle.

Additivité : Pour deux modèles combinés f et g :

$$\phi_i(f + g) = \phi_i(f) + \phi_i(g) \quad (2.11)$$

Ces propriétés garantissent une interprétation cohérente et équitable.

b) SHAP en apprentissage automatique (modèle explicatif additif)

L'extension des valeurs de Shapley aux modèles de Machine Learning a été formalisée par Scott M. Lundberg et Su-In Lee dans leur article fondateur sur SHAP [47].

Ils montrent que toute méthode d'explication additive locale cohérente doit être équivalente aux valeurs de Shapley. La prédiction d'un modèle peut alors être exprimée sous forme additive :

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i \quad (2.12)$$

où :

- $\phi_0 = E[f(x)]$: représente la prédiction moyenne du modèle,
- ϕ_i : correspond à la contribution de la variable i ,
- M : est le nombre total de variables.

SHAP unifie ainsi plusieurs méthodes d'interprétation existantes (LIME, Deep LIFT, etc.) sous un cadre théorique unique et cohérent [47].

c) Intérêt en intelligence artificielle explicable (XAI)

Grâce à ses garanties théoriques, SHAP est aujourd'hui l'une des méthodes les plus utilisées en intelligence artificielle explicable (XAI). Elle permet :

- L'interprétation locale des prédictions.
- L'analyse globale de l'importance des variables.
- L'étude de la robustesse des modèles.

Dans des applications comme la reconnaissance automatique de modulation (AMR), SHAP permet d'identifier les caractéristiques spectrales et statistiques les plus discriminantes.

2.4.2.2 Formulation mathématique des contributions marginales des variables

La contribution marginale d'une variable, dans le cadre de SHAP, représente l'apport supplémentaire que cette variable apporte à la prédiction du modèle lorsqu'elle vient s'ajouter à un groupe de variables déjà présentes. Autrement dit, elle mesure à quel point la prédiction évolue du fait de l'inclusion de cette variable dans la coalition [47].

Mathématiquement, la contribution marginale de la variable i est donnée par :

$$\Delta_i(S) = f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \quad (2.13)$$

Soit:

- $S \subseteq N \setminus \{i\}$ un sous-ensemble de variables.
- $f_S(x_S)$ la prédiction du modèle utilisant uniquement les variables de S .
- $f_{S \cup \{i\}}(x)$: Prédiction lorsque la variable i est incluse.
- $f_S(x_S)$: Prédiction sans la variable i .

Donc:

Contribution marginale = Prédiction avec i - Prédiction sans i [47].

2.4.2.3 Interprétation des graphiques de type summary plot et force plot

Les visualisations SHAP, développées par Lundberg et Lee, offrent une lecture des prédictions d'un modèle à deux niveaux complémentaires : l'un global, pour comprendre le comportement général du modèle, l'autre local, pour expliquer une décision précise.

a) L'analyse globale : le summary plot

Le SHAP summary plot permet d'identifier les variables qui pèsent le plus sur les prédictions du modèle dans son ensemble. Pour établir ce classement, on calcule pour chaque variable la moyenne de ses valeurs SHAP en valeur absolue sur l'ensemble des observations. Plus cette moyenne est élevée, plus la variable exerce une influence déterminante sur les décisions du modèle [47].

Dans la Figure 2.7 analysée, c'est la variable sex qui ressort en tête du classement, avec la contribution moyenne la plus forte. Elle s'impose donc comme le facteur le plus déterminant pour la prédiction. Viennent ensuite pclass, age et fare, qui exercent une influence appréciable sur le modèle, bien qu'en retrait par rapport à sex. À l'inverse, sibsp et parch ont un poids bien plus limité, ce qui suggère qu'elles n'interviennent que marginalement dans le processus de prédiction.

Les barres du graphique sont colorées selon les classes du modèle (Class 0 et Class 1), ce qui permet de distinguer comment chaque variable oriente la probabilité d'appartenance à l'une ou l'autre classe.

En somme, ce graphique offre une lecture synthétique et intuitive du comportement global du modèle, en mettant en lumière les caractéristiques qui pèsent réellement dans la balance lors de chaque prédiction.

La Figure 2.7 présente le graphique récapitulatif (summary plot) des valeurs SHAP, permettant d'identifier l'importance globale des variables dans les prédictions du modèle.

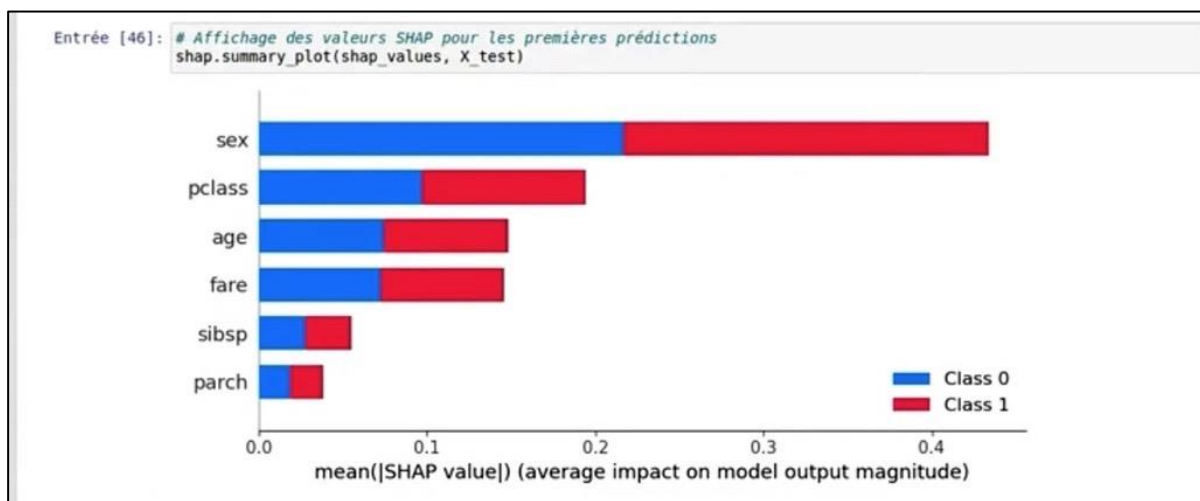


Figure 2.7 : Impact moyen des variables sur la prédiction du modèle selon les valeurs SHAP.

b) L'analyse locale : le force plot

Le force plot se concentre sur une observation précise et explique pourquoi le modèle a produit telle prédiction pour ce cas particulier. Il représente visuellement comment chaque variable pousse la prédiction vers le haut ou vers le bas, à partir d'une valeur de référence généralement la moyenne des prédictions jusqu'à aboutir à la valeur finale retenue par le modèle [47].

La Figure 2.8 présente un SHAP Force Plot appliqué à une observation individuelle. Ce graphique illustre comment chaque variable contribue à faire évoluer la prédiction depuis la valeur de base (*base value*) jusqu'à la valeur finale prédite $f(x) = 1.17$.

Trois variables sont visibles dans cette représentation :

MedInc = 4.152 apparaît en rouge, ce qui signifie qu'elle exerce une influence positive sur la prédiction elle pousse $f(x)$ vers le haut par rapport à la moyenne.



Figure 2.8 : Analyse de l'interprétabilité locale du modèle via les graphiques SHAP Force Plot.

Longitude = -117 et AveOccup = 4.181 apparaissent en bleu, indiquant qu'elles exercent une influence négative elles tirent la prédiction vers le bas, en dessous de la valeur de base. La largeur de chaque segment est proportionnelle à l'intensité de la contribution : plus la bande est large, plus l'impact de la variable est important. Ici, c'est Longitude qui présente la contribution négative la plus marquée, tandis que MedInc constitue la principale force positive. La valeur de base correspond à l'espérance des prédictions :

$$\phi_0 = E[f(x)] \quad (2.14)$$

La prédiction expliquée s'écrit :

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i \quad (2.15)$$

Les clés d'interprétation sont :

- Point de départ : valeur moyenne du modèle.
- Segments rouges : variables qui augmentent la prédiction.

- Segments bleus : variables qui la diminuent.
- Point final : prédiction spécifique de l'observation étudiée [48].

Ce graphique permet donc de comprendre précisément pourquoi une instance donnée reçoit une prédiction élevée ou faible.

2.4.2.4 Comparaison avec LIME

Afin de mettre en évidence les différences fondamentales entre les deux approches d'explicabilité, une comparaison synthétique entre SHAP et LIME est présentée dans le Tableau 2.1, en s'appuyant sur leurs fondements théoriques et leurs caractéristiques méthodologiques [47].

Critères	SHAP	LIME
Fondement théorique	Théorie des jeux coopératifs (valeurs de Shapley)	Approximation locale par régression linéaire pondérée
Type d'explication	Locale et globale	Principalement locale
Forme d'explication	Modèle additif exact où additive, nullité)	Modèle linéaire local approximatif
Garantie mathématique	Oui (efficacité, symétrie, additivité, nullité)	Non garantie formellement
Stabilité des résultats	Plus stable	Peut varier selon l'échantillonnage
Complexité computationnelle	Elevée (optimisations disponibles comme TreeSHAP)	Plus faible
Interprétation globale	Oui (summary plot, dependance plot)	Non native
Dépendance au voisinage	Non (basé sur toutes les coalitions)	Oui (dépend des perturbation locales)
Cohérence des contributions	Oui	Non garantie

Tableau 2.1 : Comparaison synthétique entre SHAP et LIME [42,47].

2.4.3 Grad-CAM et les méthodes basées sur le gradient

Grad-CAM, pour Gradient-weighted Class Activation Mapping, est une technique permettant de produire des "cartes de chaleur" (heatmaps). Cet algorithme a été initialement créé pour les modèles CNN en utilisant les informations de gradient spécifiques à la classe qui circulent dans la dernière couche convolutionnelle du modèle [49].

Grad-CAM consiste à chercher quelles parties de l'image ont conduit un réseau neuronal convolutif à sa décision finale. Son principe repose sur la génération des cartes thermiques représentant les classes d'activation sur les images reçues en entrée. Une classe d'activation est associée à une classe de sortie spécifique.

Concrètement, chaque pixel se voit attribuer un niveau d'importance selon sa contribution à la classification : plus son influence est élevée, plus son intensité sur la carte thermique sera marquée. Ainsi, les zones chaudes (rouge/jaune) correspondent aux régions déterminantes pour la décision, tandis que les zones froides (bleu) signalent des parties peu pertinentes.

Pour illustrer ce mécanisme, prenons l'exemple d'un CNN entraîné à distinguer des chiens et des chats. Grad-CAM produira une heatmap distincte pour chaque classe : l'une mettra en évidence les zones de l'image évoquant un chat, l'autre celles associées à un chien. Cela permet de comprendre sur quels indices visuels le modèle s'est appuyé pour trancher.

D'un point de vue technique, la méthode calcule, au niveau de la dernière couche convolutive, une combinaison linéaire des cartes d'activation pondérées par les gradients propres à la classe cible. La carte obtenue est ensuite redimensionnée à la taille de l'image d'entrée et superposée à celle-ci, offrant une visualisation intuitive et directement interprétable des zones d'intérêt du modèle. Ceci est illustré par la carte thermique (Figure 2.9) ci-dessous.

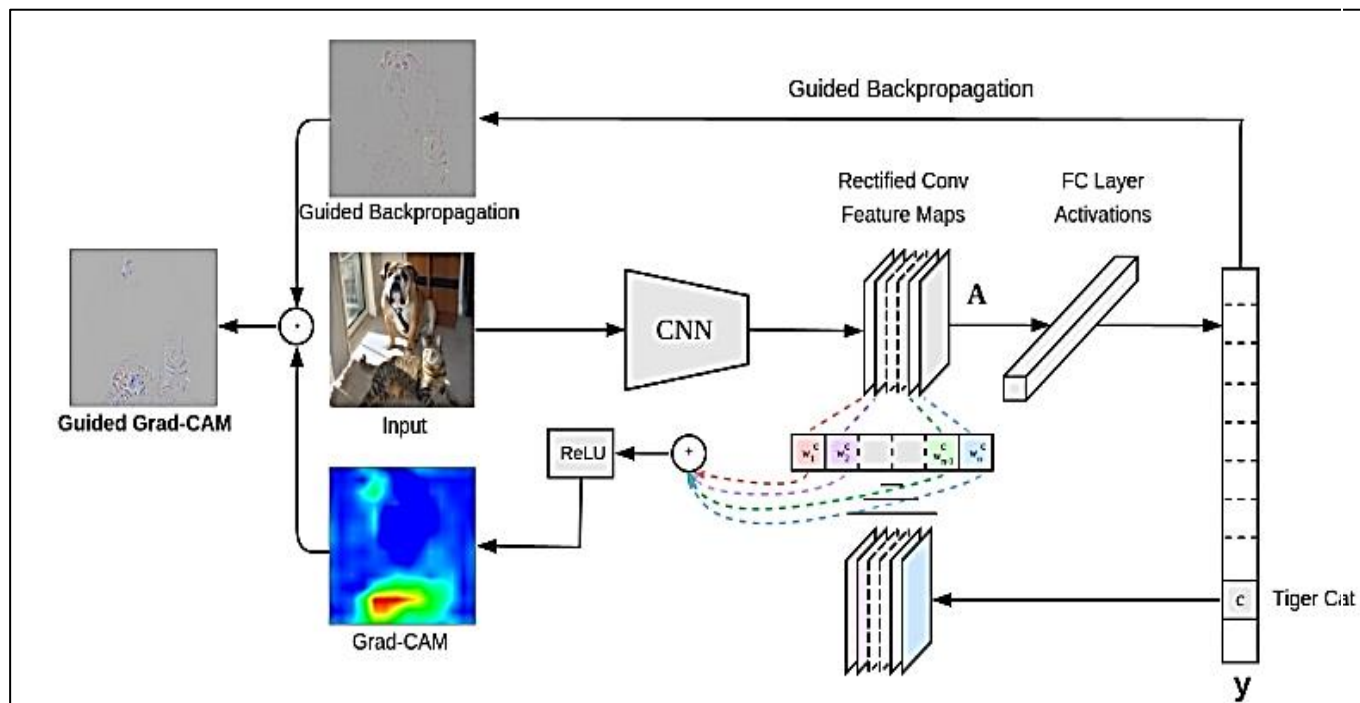


Figure 2.9 : Illustration du pipeline Grad-CAM : de l'image d'entrée à la carte thermique d'activation.

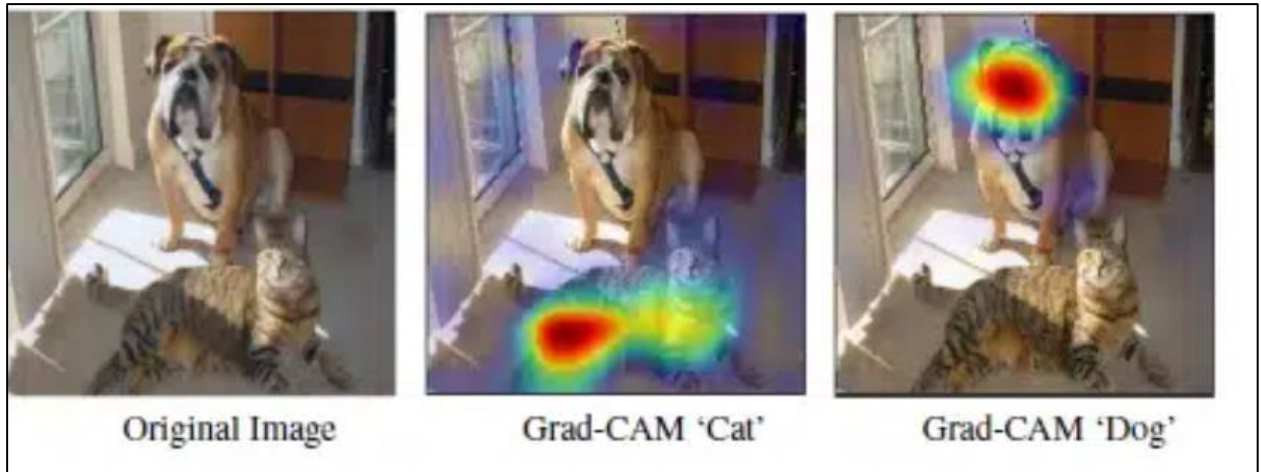


Figure 2.10 : Comparaison entre l'image originale et les heatmaps Grad-CAM associées aux classes « Cat » et « Dog ».

Reprenons maintenant l'exemple de la classification de chiens et de chats. La méthode Grad-CAM va donc permettre de rendre thermique les zones utilisées pour la classification de chacun des objets de notre image et cela produira le résultat suivant dans la Figure 2.10 :

Cette section a présenté une technique récente permettant d'interpréter les réseaux de neurones convolutifs, qui représentent aujourd'hui une architecture avancée particulièrement utilisée pour les tâches liées au traitement et à l'analyse d'images. La recherche dans le domaine de l'apprentissage automatique interprétable connaît une évolution rapide et joue un rôle essentiel pour renforcer la confiance des utilisateurs et favoriser l'amélioration des performances et de la transparence des modèles [50].

2.4.3.1 Méthodes d'interprétabilité basées sur le gradient

Les méthodes basées sur le gradient exploitent les dérivées partielles de la sortie d'un réseau neuronal par rapport à ses entrées ou à ses cartes de caractéristiques internes.

En notant :

$$y^c = f_c(x) \quad (2.16)$$

la sortie du réseau pour la classe, ces approches calculent le gradient :

$$\frac{\partial y^c}{\partial x} \quad (2.17)$$

Qui mesure la sensibilité locale de la prédiction aux variations infinitésimales des pixels d'entrée.

Ces approches permettent d'identifier les régions de l'image ayant le plus contribué à la décision du modèle.

a) Principe mathématique

Soient A^k : la k -ième carte de caractéristiques d'une couche convolutionnelle, y^c : le score avant softmax pour la classe c .

On calcule d'abord les poids d'importance :

$$a_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (2.18)$$

Où : Z est le nombre total de pixels dans la carte d'activation, i, j indexent les positions spatiales [50].

2.4.3.2 Applications de Grad-CAM en imagerie médicale et en vision par ordinateur

La méthode Grad-CAM est largement utilisée pour interpréter les décisions des réseaux neuronaux convolutionnels (CNN) dans des domaines où la compréhension des régions influentes est essentielle.

En imagerie médicale, elle intervient notamment dans la détection de tumeurs sur des images IRM, l'analyse de radiographies pulmonaires et le diagnostic assisté par intelligence artificielle. Elle permet de vérifier que le modèle se concentre effectivement sur des zones anatomiquement pertinentes, ce qui constitue un élément fondamental pour garantir la fiabilité clinique des systèmes d'aide au diagnostic [51].

En vision par ordinateur, Grad-CAM est appliqué à la classification d'objets, à la reconnaissance faciale, à la détection d'anomalies industrielles ainsi qu'aux systèmes autonomes. Dans ces contextes, la visualisation des régions importantes améliore la transparence des modèles et renforce la confiance dans les systèmes basés sur des réseaux convolutionnels [50].

2.5 Métriques d'évaluation des explications

Évaluer la qualité d'une explication produite par une méthode XAI ne se résume pas à une simple mesure chiffrée. Plusieurs critères complémentaires entrent en jeu : la fidélité garantit que l'explication reflète réellement le comportement du modèle, tandis que la stabilité et la robustesse s'assurent que des observations similaires produisent des explications cohérentes, même en présence de légères perturbations. La complétude vérifie que la somme des contributions individuelles des variables correspond bien à la prédiction finale, ce qui est notamment garanti par la méthode SHAP. Enfin, la compréhensibilité et la cohérence s'intéressent à la dimension humaine de l'évaluation, en s'assurant que les explications sont intelligibles pour l'utilisateur et uniformes entre des modèles au comportement comparable. L'ensemble de ces critères forme un cadre d'évaluation rigoureux, indispensable pour juger de la fiabilité et de l'utilité des explications générées [53].

2.5.1 Critères d'évaluation des méthodes d'explicabilité

Pour évaluer la qualité des explications d'un modèle d'apprentissage automatique, il est important de définir des critères clairs et mesurables, tels que la fidélité, la stabilité, la robustesse, la cohérence et la compréhensibilité humaine. Ces métriques permettent de comparer et valider les méthodes d'explicabilité, notamment dans des domaines critiques comme la santé ou la finance.

Les détails de ces critères sont présentés dans le Tableau 2.2.

Critère	Définition	Objectif
Fidélité	Mesure dans quelle mesure l'explication reflète réellement le comportement du modèle sous-jacent	S'assurer que l'explication correspond à la prédiction réelle du modèle
Stabilité	Capacité de la méthode à produire des explications similaires pour des observations proches	Vérifier la cohérence locale de la méthode
Robustesse	Résistance de l'explication aux perturbations mineures de l'entrée ou du modèle	Garantir que les explications ne changent pas de façon erratique
Cohérence	Comparaison des explications pour différents modèles ou configurations sur la même tâche	Vérifier que des modèles similaires produisent des explications similaires
Compréhensibilité humaine	Facilité avec laquelle un expert ou un utilisateur peut comprendre et interpréter l'explication	Évaluer l'utilité réelle pour la prise de décision humaine

Tableau 2.2 : Principaux critères d'évaluation des méthodes d'explicabilité et leurs objectifs [42, 50,53].

2.5.2 Protocoles expérimentaux

Pour évaluer ces métriques, plusieurs approches sont utilisées :

- Tests basés sur les données : masquer ou perturber les variables identifiées comme importantes et observer la variation de la prédiction pour mesurer la fidélité.
- Tests de cohérence et stabilité : comparer les explications pour des observations proches ou pour des modèles légèrement modifiés.

- Évaluation humaine : demander à des experts d'évaluer la pertinence et la compréhension des explications [54].

2.6 Conclusion

À travers ce chapitre, nous avons établi le cadre théorique qui sous-tend notre démarche d'explicabilité des modèles d'apprentissage automatique. Pour ce faire, nous sommes partis des concepts fondateurs de l'intelligence artificielle, du Machine Learning et du Deep Learning trois domaines étroitement liés qui constituent aujourd'hui le cœur des systèmes intelligents.

Nous avons ensuite mis en lumière les enjeux de l'Intelligence Artificielle Explicable (XAI), dont l'objectif principal est de rendre les modèles complexes plus lisibles et dignes de confiance. Les différentes familles de méthodes d'explicabilité ont été passées en revue, qu'il s'agisse des explications globales ou locales, des approches intrinsèques ou post-hoc, ou encore des techniques liées ou indépendantes du modèle.

Enfin, nous avons détaillé les méthodes les plus couramment mobilisées dans ce domaine, notamment LIME, SHAP et Grad-CAM, chacune apportant un éclairage différent sur le comportement interne des modèles.

Les éléments théoriques développés tout au long de ce chapitre ne constituent pas une fin en soi : ils serviront de grille de lecture pour la partie pratique de notre travail, dans laquelle ces mêmes méthodes seront mises en œuvre pour analyser et interpréter concrètement le modèle étudié. Dans le chapitre suivant, nous allons discuter et analyser en détail les résultats expérimentaux obtenus, afin d'interpréter les performances du modèle CNN-TCN dans différents scénarios de bruit et d'évaluer plus précisément son comportement en fonction du SNR.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

3.1 Introduction

Ce chapitre est consacré à l'étude expérimentale menée pour la reconnaissance automatique des modulations radio, en s'appuyant sur un modèle hybride CNN-TCN appliqué au dataset RadioML 2016.10a. Cette base de données constitue une référence incontournable dans le domaine de la classification automatique des modulations (AMC), regroupant une large variété de modulations numériques et analogiques générées sous différentes conditions de bruit, caractérisées par le rapport signal sur bruit (SNR).

L'objectif central de cette étude est d'évaluer la capacité du modèle proposé à extraire efficacement les caractéristiques discriminantes des signaux radio et à maintenir de bonnes performances de classification dans des conditions de transmission variées. À cet effet, une architecture hybride associant les réseaux de neurones convolutifs (CNN) et les réseaux convolutifs temporels (TCN) est développée, permettant d'exploiter conjointement les structures locales et les dépendances temporelles des signaux. Une branche dédiée au traitement de l'information SNR est par ailleurs intégrée à l'architecture, afin d'enrichir la représentation fournie au modèle et de renforcer sa robustesse face aux dégradations du canal. La démarche adoptée dans ce chapitre s'articule en plusieurs étapes progressives. Une présentation détaillée du dataset RadioML 2016.10a et des différentes classes de modulation considérées est d'abord proposée. L'architecture complète du modèle CNN-TCN est ensuite décrite, en mettant en évidence les rôles respectifs des branches d'extraction, de fusion et de classification. Les résultats expérimentaux sont par la suite analysés à travers les courbes d'apprentissage, les rapports de classification, les matrices de confusion et l'évolution des métriques de performance en fonction du SNR.

Une étude comparative avec plusieurs architectures récentes de la littérature est également conduite afin d'évaluer le comportement du modèle dans des conditions de bruit faible, intermédiaire et élevé. Le chapitre s'achève sur une analyse d'interprétabilité fondée sur les méthodes d'intelligence artificielle explicable (XAI), notamment les gradients intégrés et l'analyse SHAP, permettant d'éclairer les mécanismes décisionnels du modèle pour différentes modulations et différents niveaux de SNR.

Les simulations ont été effectuées sous Google Colab et Kaggle, ce qui permet un traitement efficace de grands volumes de données.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

3.2 Présentation du dataset RadioML2016.10a

La base de données RadioML 2016.10a [54], abrégée RML2016a, est une base de référence utilisée pour l'évaluation des algorithmes de classification de modulation automatique (AMC). Elle a été développée pour simuler un environnement de communication réaliste, incluant des effets de bruit, de décalage de fréquence, et de gigue temporelle. Elle contient des signaux I/Q complexes générés pour 11 types de modulations (analogiques et numériques) suivantes :

- **BPSK (Binary Phase Shift Keying)**

Modulation numérique où chaque symbole représente 1 bit via deux phases opposées (0° et 180°). Elle est très robuste au bruit et aux interférences, ce qui la rend adaptée aux environnements à faible SNR, mais son efficacité spectrale reste limitée.

- **QPSK (Quadrature Phase Shift Keying)**

Utilise quatre états de phase pour coder 2 bits par symbole. Elle offre un bon compromis entre robustesse et efficacité spectrale. Dans RadioML, elle est souvent utilisée comme référence pour comparer les performances des modèles.

- **8PSK (8-Phase Shift Keying)**

Extension de la PSK avec 8 phases, permettant de coder 3 bits par symbole. Elle améliore le débit mais devient plus sensible au bruit, ce qui la rend plus difficile à classifier à faible SNR.

- **QAM16 (16-Quadrature Amplitude Modulation)**

Combine amplitude et phase pour transmettre 4 bits par symbole. Les points de constellation sont répartis sur un plan I/Q. Elle offre un bon compromis entre débit et robustesse, mais reste sensible aux perturbations du canal.

- **QAM64 (64-Quadrature Amplitude Modulation)**

Version plus dense de la QAM avec 64 symboles (6 bits/symbole). Elle permet un débit élevé mais nécessite un SNR élevé, car les points de constellation sont très rapprochés, ce qui augmente les erreurs en présence de bruit.

- **GFSK (Gaussian Frequency Shift Keying)**

Modulation de fréquence avec un filtrage gaussien appliqué avant la modulation. Cela réduit la largeur de bande et les interférences intersymboles. Utilisée dans des systèmes réels comme Bluetooth, elle est relativement robuste.

- **CPFSK (Continuous Phase Frequency Shift Keying)**

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

Variante de FSK où la phase est continue dans le temps, évitant les discontinuités brutales. Cela améliore l'efficacité spectrale et réduit les émissions hors bande, ce qui est utile dans les systèmes contraints en bande passante.

- **AM-DSB (Amplitude Modulation - Double Side Band)**

Modulation analogique où l'information est portée par l'amplitude, avec deux bandes latérales symétriques. Elle est simple à implémenter mais inefficace en termes de bande passante et sensible au bruit.

- **AM-SSB (Amplitude Modulation - Single Side Band)**

Version optimisée de l'AM où une seule bande latérale est transmise. Elle réduit la bande passante et améliore l'efficacité énergétique, mais nécessite une démodulation plus complexe.

- **FM (Frequency Modulation)**

L'information est transmise via les variations de fréquence de la porteuse. Elle est très résistante au bruit d'amplitude et largement utilisée dans les communications analogiques.

- **WB-FM (Wideband Frequency Modulation)**

Version à large bande de la FM avec une déviation de fréquence importante. Elle offre une meilleure qualité (notamment audio) et une meilleure résistance au bruit, mais au prix d'une bande passante plus large.

Pour chaque type de modulation le signal reçu est évalué à vingt valeurs du rapport signal sur bruit (SNR) allant de -20 dB à +18 dB avec un pas de 2.

Pour chaque type de modulation et pour chaque valeur du rapport signal sur bruit la base offre 1000 exemples de signal. Chaque exemple de signal est représenté par une séquence complexe de 128 échantillons I/Q, accompagnée de son label de modulation et de la valeur SNR correspondante. Cette base est particulièrement bien adaptée aux techniques d'apprentissage profond, car elle offre un grand volume de données annotées. Elle permet de tester la robustesse des modèles face à des conditions de canal variées, simulant un environnement réaliste pour les systèmes de radio cognitive et les applications de guerre électronique.

La Figure 3.1 illustre les composantes I/Q dans le domaine temporel pour onze types de modulations du dataset RadioML 2016.10a, évaluées à un SNR de 18 dB. À ce niveau favorable, les structures caractéristiques de chaque modulation sont clairement distinguables. Parmi les modulations numériques, la BPSK et la QPSK se caractérisent par des oscillations régulières reflétant leurs transitions de phase respectives, tandis que la CPFSK et la GFSK se

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

distinguent par des évolutions temporelles lisses et continues. La PAM4 présente des variations d'amplitude multi-niveaux, et les QAM16 et QAM64 affichent des formes d'onde de plus en plus complexes et denses, traduisant la richesse de leurs constellations respectives.

Du côté des modulations analogiques, l'AM-DSB se reconnaît par son enveloppe symétrique et sa composante Quasi nulle, tandis que l'AM-SSB présente des amplitudes plus faibles liées à la suppression d'une bande latérale. La WBFM se distingue par des amplitudes très faibles et une apparente régularité des formes d'onde.

L'ensemble de ces signatures I/Q confirme la diversité du dataset RadioML 2016.10a, qui constitue un support riche et bien adapté à l'entraînement et à l'évaluation des modèles de classification automatique des modulations.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

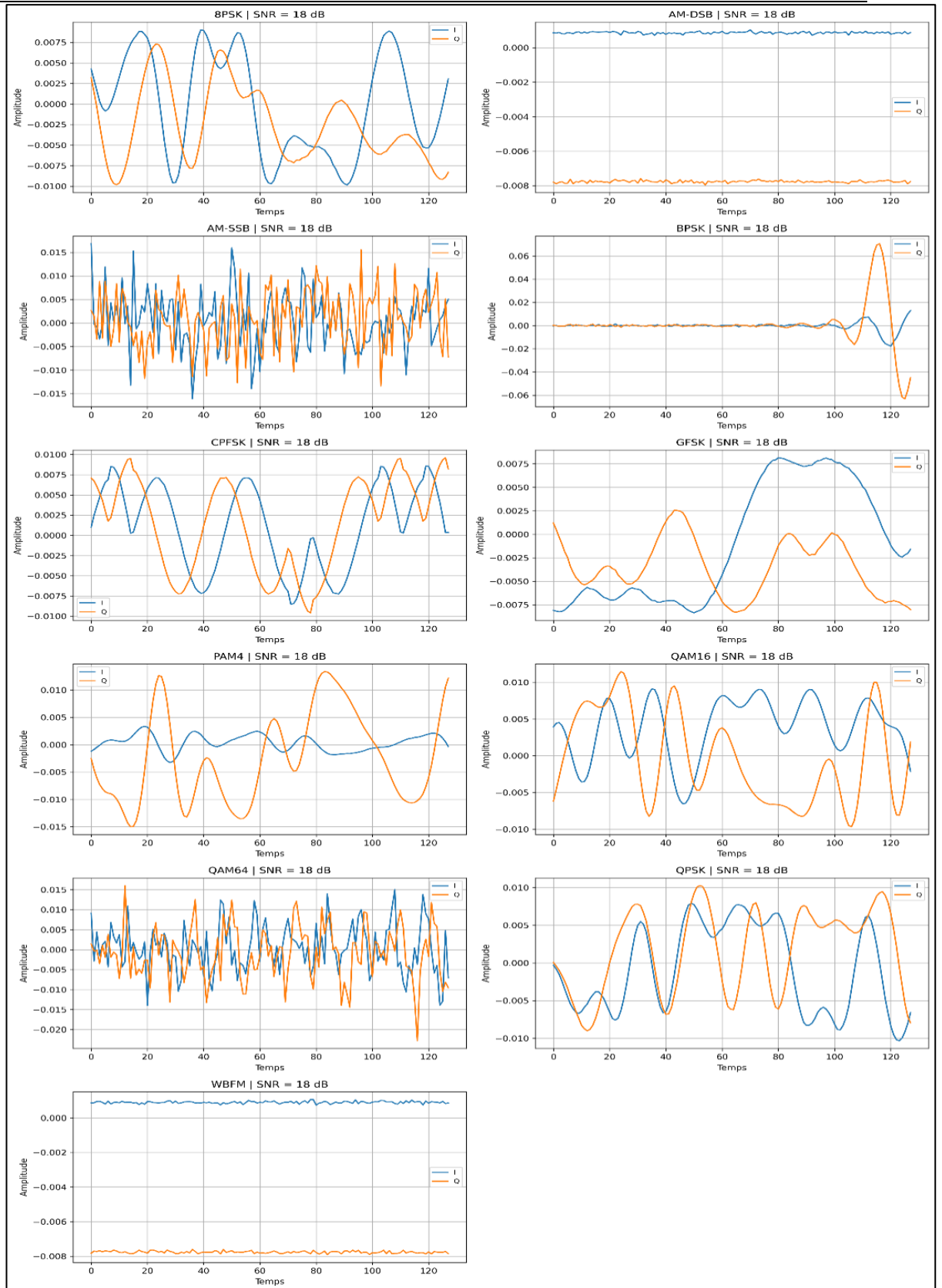


Figure 3.1 : Composantes I/Q dans le domaine temporel d'un signal à un SNR de 18 dB.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

3.3 Modèle CNN-TCN

3.3.1 Architecture du modèle

Le modèle que nous proposons s'appuie sur une architecture hybride qui associe deux types de réseaux de neurones complémentaires : les réseaux convolutifs classiques (CNN) et les réseaux convolutifs temporels (TCN). Cette combinaison vise à reconnaître automatiquement le type de modulation d'un signal radio à partir de ses composantes en phase et en quadrature (I/Q), en tirant parti à la fois des structures locales et des dynamiques temporelles du signal. L'architecture se décompose en trois branches fonctionnelles qui travaillent en parallèle avant de se rejoindre pour la décision finale.

3.3.1.1 Branche d'extraction des caractéristiques du signal

La première branche reçoit directement les échantillons I/Q du signal, représentés sous forme de deux canaux. Elle commence par deux blocs convolutifs successifs, chacun comprenant une convolution 1D, une normalisation par batch, une activation ReLU et une opération de max-pooling. Ces blocs jouent le rôle d'un extracteur de primitives locales, capturant les motifs caractéristiques du signal à différentes échelles. Les représentations ainsi obtenues sont ensuite transmises à un réseau TCN composé de plusieurs blocs à convolutions dilatées, avec des facteurs de dilatation croissants (1, 2, 4 et 8). Cette conception permet au modèle de percevoir aussi bien les corrélations à court terme qu'à long terme dans le signal, tout en maintenant une complexité de calcul raisonnable.

3.3.1.2 Branche de traitement du SNR

La deuxième branche traite une information complémentaire : la valeur du rapport signal sur bruit (SNR), fournie sous forme scalaire. Une couche entièrement connectée, suivie d'une activation ReLU et d'un dropout de régularisation, transforme cette valeur en une représentation vectorielle utile. L'idée sous-jacente est d'équiper le modèle d'une conscience du niveau de bruit ambiant, ce qui l'aide à adapter sa décision selon les conditions de transmission.

3.3.1.3 Branche de fusion et de classification

Les sorties des deux branches précédentes sont concaténées pour former un vecteur de caractéristiques enrichi, associant information spectro-temporelle et contexte bruitage. Ce

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

vecteur est ensuite traité par plusieurs couches "fully connected" avec activation ReLU et dropout, avant de produire en sortie la classe de modulation prédite.

L'architecture du modèle CNN-TCN avec fusion SNR est montrée dans la Figure 3.2 :

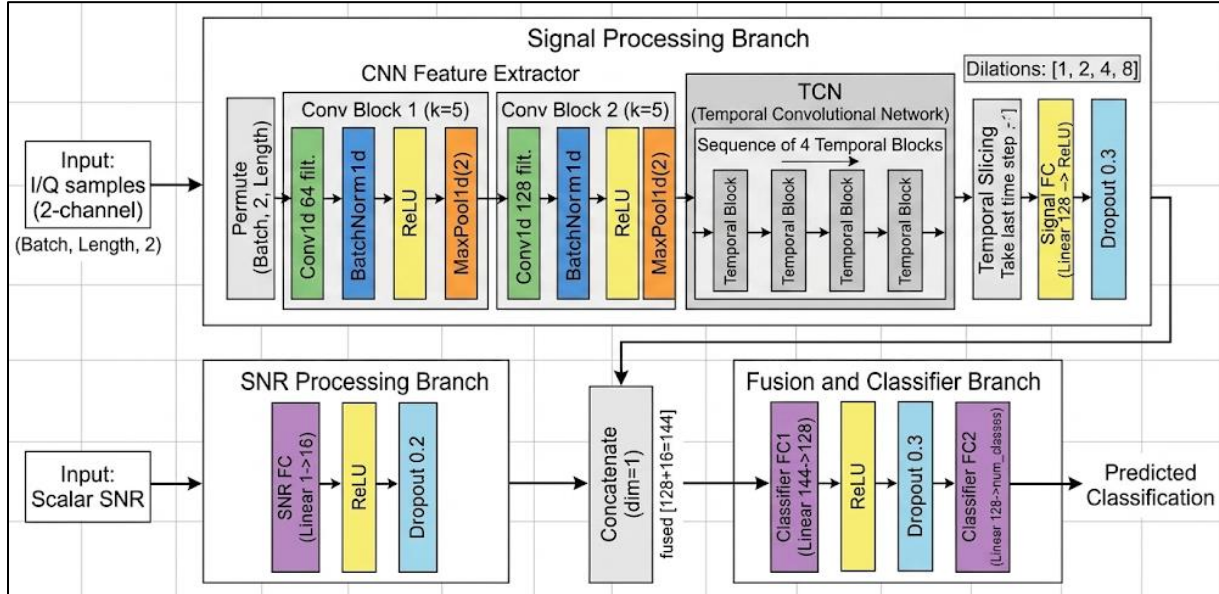


Figure 3.2 : Architecture du modèle CNN-TCN avec fusion SNR.

3.3.2 Résultats et discussion

3.3.2.1 Analyse des courbes d'apprentissage

Afin d'évaluer le comportement du modèle durant la phase d'entraînement, deux courbes ont été tracées : l'une suivant l'évolution de la précision, l'autre celle de la perte, toutes deux mesurées simultanément sur les ensembles d'entraînement et de validation au fil des époques.

a) Exactitude au fil des époques

En observant la Figure 3.3, on constate que le modèle démarre avec une exactitude relativement modeste, aux alentours de 44 %, puis progresse de façon continue et régulière pour atteindre près de 68 % à la cinquantième époque. Cette montée progressive est encourageante car elle montre que le modèle assimile réellement les structures des signaux radio au fur et à mesure de l'apprentissage, sans stagner ni diverger. La courbe de validation suit globalement la même trajectoire et se stabilise autour de 65 %. On y observe certes quelques irrégularités passagères, mais celles-ci sont tout à fait normales dans un contexte de classification de signaux radio, où le bruit et la variabilité des données peuvent perturber momentanément l'évaluation. Ce qui importe davantage, c'est que l'écart entre les deux courbes reste limité et constant, signe que le modèle ne mémorise pas

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

simplement les données d'entraînement mais apprend à généraliser.

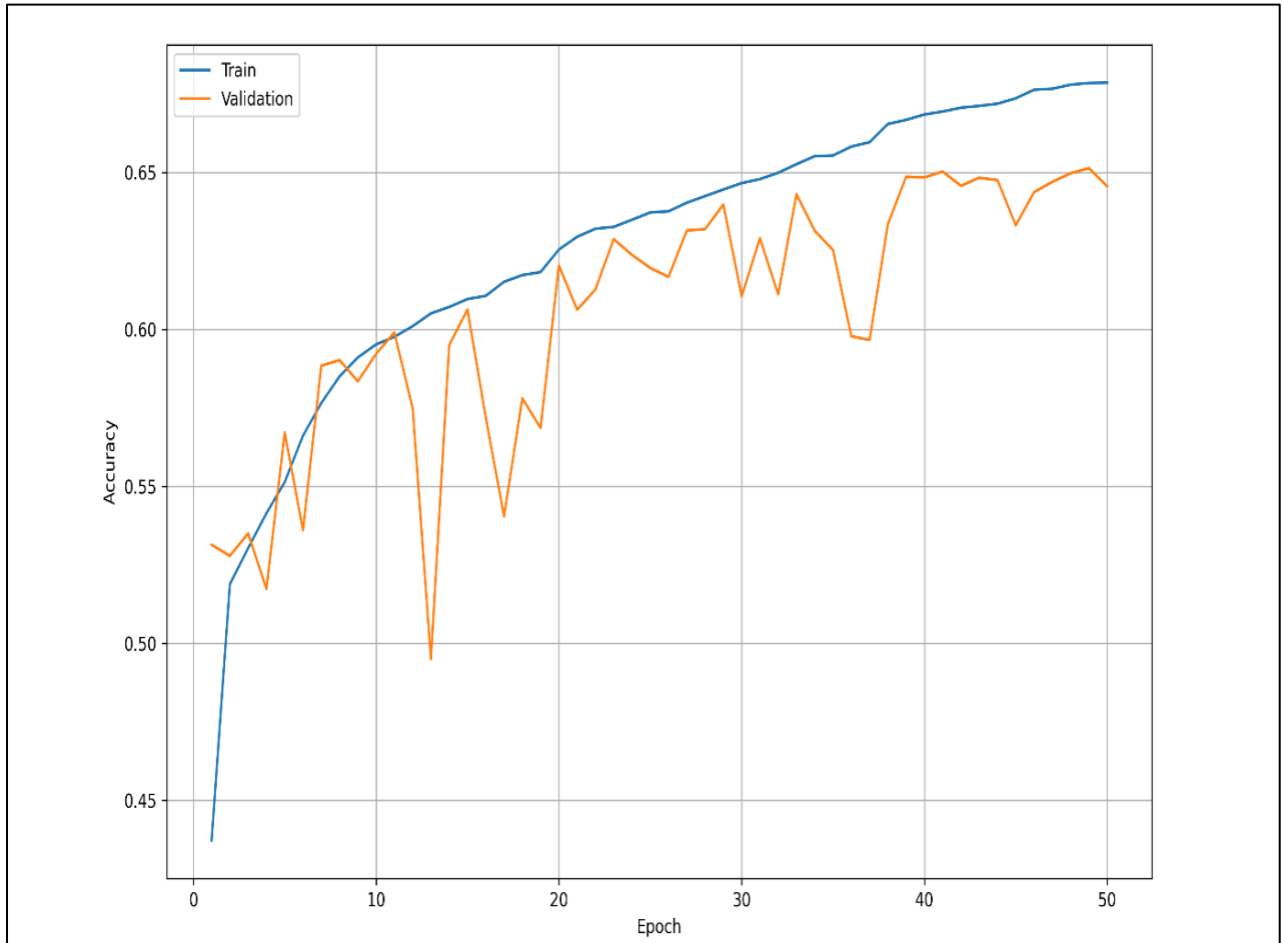


Figure 3.3: Training/Validation Accuracy vs. Epoch.

a) Perte au fil des époques

La Figure 3.4 vient confirmer et compléter cette lecture. La perte d'entraînement diminue de manière régulière et continue, passant d'environ 1.4 en début d'apprentissage à près de 0.85 à la dernière époque, ce qui reflète une optimisation saine et progressive du modèle.

La courbe de validation, en revanche, présente plusieurs pics prononcés, notamment autour des époques 10, 20 et 40, avant de se stabiliser aux alentours de 0.9. Ces soubresauts s'expliquent par la complexité inhérente des données radio, en particulier dans les conditions de faible rapport signal sur bruit où certaines modulations deviennent difficiles à distinguer. Malgré cela, la perte de validation finit par converger vers une valeur proche de celle de l'entraînement, ce qui confirme l'absence d'un surapprentissage problématique.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

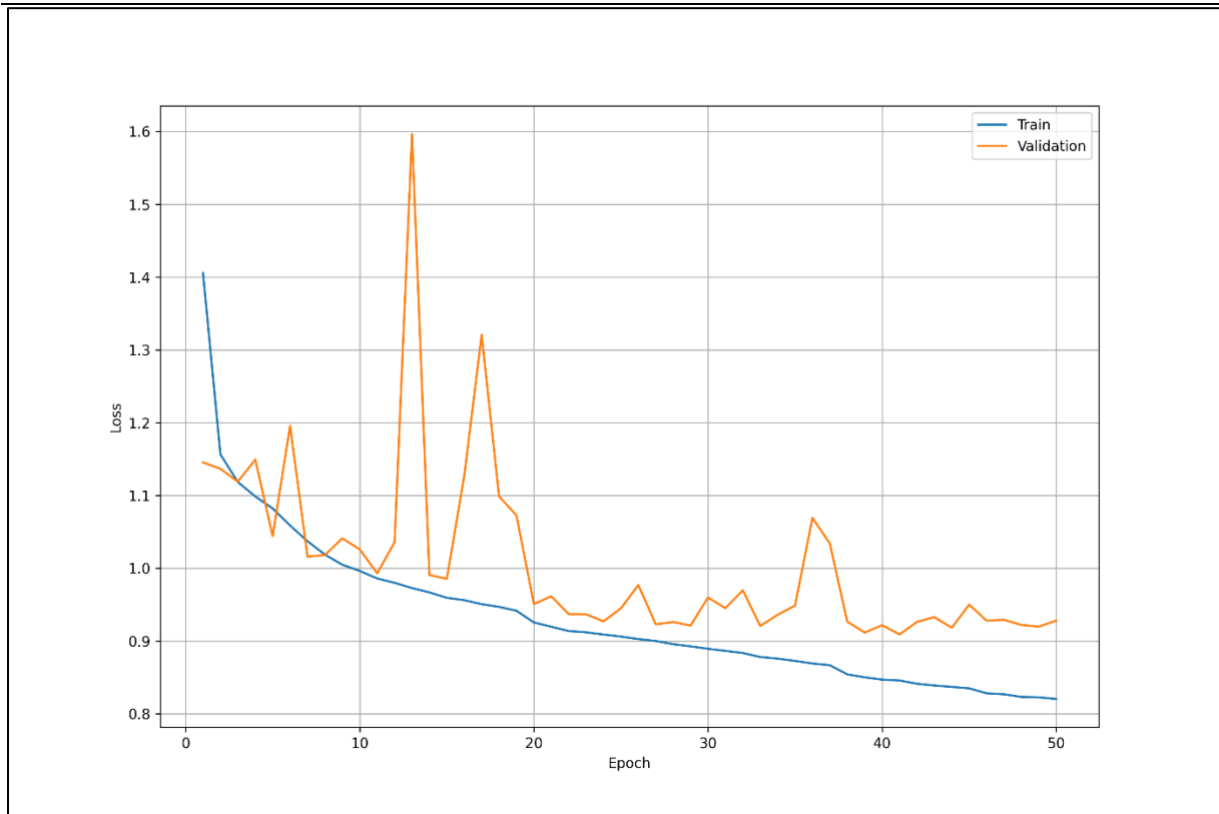


Figure 3.4 : Training/Validation Loss vs. Epoch.

En combinant l'analyse de ces deux Figures, on peut affirmer que le modèle CNN-TCN avec fusion SNR présente un comportement d'apprentissage stable et maîtrisé. La précision finale de test de 65.84 %, cohérente avec les valeurs observées sur la courbe de validation, confirme que les performances obtenues ne sont pas le fruit du hasard mais bien le reflet d'une architecture bien conçue et capable de s'adapter à des données variées et complexes.

3.3.2.2 Rapport de classification

La Figure 3.5 synthétise les résultats obtenus par le modèle CNN-TCN avec fusion SNR sur l'ensemble des types de modulation du dataset RadioML 2016, en s'appuyant sur quatre métriques d'évaluation : l'accuracy, la précision, le rappel, le F1-score. Le modèle affiche une accuracy globale de 65,84 %, témoignant d'une capacité satisfaisante à identifier correctement les signaux radio sur l'ensemble du jeu de test.

L'analyse par classe révèle des disparités notables entre les modulations. Parmi les mieux reconnues, PAM4 se distingue avec un F1-score de 0,7923, suivi de QAM64 (0,7601) et CPFSK (0,7233), des performances qui témoignent d'une extraction efficace des caractéristiques propres à ces modulations.

À l'inverse, certaines classes demeurent difficiles à classifier. WBFM enregistre un F1-

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

score de seulement 0,3650, et des modulations analogiques telles qu'AM-DSB obtiennent également des résultats plus modestes. Ces difficultés s'expliquent vraisemblablement par la proximité entre certaines signatures spectrales et par la sensibilité accrue de ces signaux aux effets du bruit.

Sur le plan global, les moyennes macros et pondérée affichent des valeurs cohérentes précision : 0,7064 ; rappel : 0,6584 ; F1-score : 0,6560 ce qui traduit un comportement relativement équilibré du modèle à travers l'ensemble des classes.

En définitive, ces résultats confirment la pertinence de l'architecture CNN-TCN avec fusion SNR pour la reconnaissance automatique des modulations radio, tout en soulignant des marges de progression pour les classes les plus sensibles aux conditions de bruit.

	precision	recall	f1-score	support
8PSK	0.6948	0.6135	0.6516	2000
AM-DSB	0.4159	0.8205	0.5520	2000
AM-SSB	0.5675	0.7585	0.6493	2000
BPSK	0.6905	0.6915	0.6910	2000
CPFSK	0.8455	0.6320	0.7233	2000
GFSK	0.5846	0.7290	0.6489	2000
PAM4	0.8512	0.7410	0.7923	2000
QAM16	0.6880	0.7200	0.7036	2000
QAM64	0.8080	0.7175	0.7601	2000
QPSK	0.8142	0.5830	0.6795	2000
WBFM	0.8107	0.2355	0.3650	2000
accuracy			0.6584	22000
macro avg	0.7064	0.6584	0.6560	22000
weighted avg	0.7064	0.6584	0.6560	22000

Figure 3.5 : Rapport de classification.

3.3.3 La matrice de confusion

3.3.3.1 Analyse de la matrice de confusion globale

La matrice de confusion donnée par la Figure 3.6 offre une lecture détaillée des performances du modèle CNN-TCN avec fusion SNR, modulation par modulation, sur le jeu de données RadioML 2016. Contrairement à un simple chiffre de précision globale, elle permet de voir exactement où le modèle réussit et où il se trompe.

Plusieurs types de modulation sont identifiés avec une bonne fiabilité, notamment la PAM4, la QAM64, la GFSK et la BPSK. Ces résultats montrent que l'architecture parvient à extraire des caractéristiques suffisamment distinctives pour ces signaux, même dans des conditions de transmission difficiles. Cela valide l'intérêt de combiner les blocs CNN et TCN pour capturer à la fois les structures locales et les dynamiques temporelles des signaux.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

Le modèle rencontre en revanche certaines difficultés lorsqu'il s'agit de distinguer des modulations dont les signatures se ressemblent fortement. C'est notamment le cas entre la QAM16 et la QAM64, entre la QPSK et la 8PSK, ou encore entre l'AM-DSB et la WBFM. Ces confusions ne sont pas surprenantes car ces paires de modulations partagent des structures temporelles proches, et leur séparation devient d'autant plus délicate lorsque le niveau de bruit est élevé. Il s'agit d'un défi bien connu dans le domaine de la classification automatique des modulations.

Malgré ces quelques zones de confusion, le modèle affiche une accuracy globale de 65.84 % sur l'ensemble du jeu de test, un résultat cohérent avec les courbes d'apprentissage présentées précédemment. Ce score, obtenu sur un dataset réputé pour sa complexité et la diversité de ses conditions de canal, témoigne de la robustesse de l'approche proposée et confirme que l'intégration de l'information SNR contribue positivement à la prise de décision du modèle.

	8PSK	AM-DSB	AM-SSB	BPSK	CPFSK	GFSK	PAM4	QAM16	QAM64	QPSK	WBFM
True 8PSK	1227	118	178	93	75	138	21	38	13	92	7
True AM-DSB	25	1641	84	48	9	86	14	20	20	8	45
True AM-SSB	45	153	1517	58	16	112	25	33	9	25	7
True BPSK	59	126	163	1383	23	109	32	36	25	39	5
True CPFSK	93	134	165	72	1264	146	23	30	27	39	7
True GFSK	26	169	151	45	37	1458	22	30	20	14	28
True PAM4	23	130	81	77	6	100	1482	58	24	17	2
True QAM16	29	103	67	48	8	72	51	1440	164	14	4
True QAM64	11	75	27	46	0	38	22	337	1435	8	1
True QPSK	205	137	151	91	54	102	34	32	24	1166	4
True WBFM	23	1160	89	42	3	133	15	39	15	10	471
	Predicted 8PSK	Predicted AM-DSB	Predicted AM-SSB	Predicted BPSK	Predicted CPFSK	Predicted GFSK	Predicted PAM4	Predicted QAM16	Predicted QAM64	Predicted QPSK	Predicted WBFM

Figure 3.6 : Matrice de confusion globale.

3.3.3.2 Analyse des matrices de confusion par niveau de SNR

Afin d'évaluer la robustesse du modèle face à différentes conditions de transmission, les matrices de confusion présentées dans la Figure 3.7 ont été calculées séparément pour cinq niveaux de SNR représentatifs : -20 dB, -10 dB, 0 dB, 10 dB et 18 dB. Cette décomposition permet de mieux comprendre dans quelle mesure la qualité du signal influence les décisions du modèle.

a) Condition très bruitées (-20 dB et -10 dB)

À ces niveaux de bruit extrêmes, le modèle se retrouve dans une situation particulièrement défavorable. Le bruit masque en grande partie les caractéristiques propres à chaque modulation, rendant leur discrimination très difficile, voire aléatoire dans certains cas. Les matrices de confusion le reflètent clairement : les prédictions sont largement dispersées hors de la diagonale principale, et les erreurs de classification touchent la quasi-totalité des classes. Il serait d'ailleurs surprenant qu'un quelconque modèle parvienne à des résultats satisfaisants dans de telles conditions, tant l'information utile est noyée dans le bruit.

b) Conditions intermédiaire (0 dB)

Dès que le SNR atteint 0 dB, une amélioration sensible se manifeste. La diagonale principale commence à se densifier, témoignant d'un nombre croissant de classifications correctes. Le modèle commence à tirer profit des structures caractéristiques des signaux, même si des confusions persistent entre certaines modulations aux formes similaires. Ce palier marque clairement une transition vers un régime de fonctionnement plus fiable.

c) Conditions favorables (10 dB et 18 dB)

C'est dans ces conditions que le modèle exprime pleinement son potentiel. Les matrices de confusion affichent une diagonale principale très prononcée, signe que la grande majorité des modulations sont correctement identifiées. Les confusions inter-classes se raréfient considérablement, et le taux d'erreur chute de manière significative. La clarté accrue des signaux à ces niveaux de SNR permet au modèle d'exploiter pleinement les représentations apprises durant l'entraînement.

L'ensemble de ces résultats met en évidence une relation directe et cohérente entre la qualité du signal reçu et les performances de classification. Plus le SNR est élevé, plus les caractéristiques des signaux sont nettes et exploitables, et plus le modèle est en mesure de

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

prendre des décisions fiables. Ce comportement, parfaitement aligné avec les attentes physiques des systèmes de communication, confirme également la pertinence de l'intégration du SNR comme entrée explicite dans l'architecture du modèle, lui permettant de contextualiser ses prédictions en fonction des conditions de canal rencontrées.

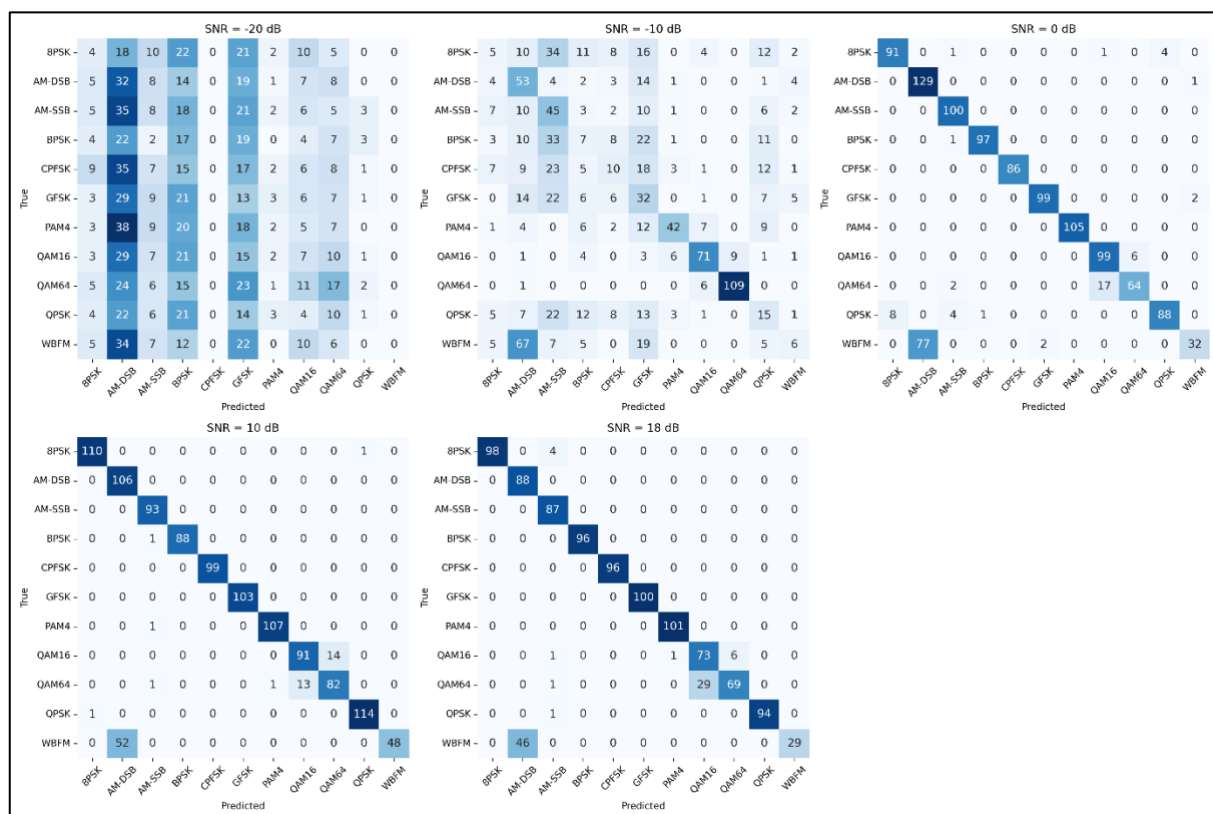


Figure 3.7 : Matrices de confusion pour différentes valeurs SNR (-20, -10, 0, 10, 18 dB).

3.3.4 Les performances du modèle CNN-TCN

Le Tableau 3.1 présente l'évolution des performances du modèle CNN-TCN avec fusion SNR pour des valeurs de SNR comprises entre -20 dB et 18 dB, évaluées selon quatre métriques : l'accuracy, la précision, le rappel et le F1-score. Les résultats mettent en évidence une dépendance marquée entre la qualité du signal reçu et les capacités de classification du modèle.

3.3.4.1 Performances en conditions de fort bruit (SNR faible)

Dans la plage allant de -20 dB à -10 dB, les performances demeurent très limitées. À -20 dB, l'accuracy ne dépasse pas 9,7 % et le F1-score plafonne à 0,0747. À ce niveau de bruit, les caractéristiques discriminantes des signaux sont largement masquées, rendant toute classification fiable quasi impossible.

Une amélioration sensible s'amorce à partir de -8 dB, avec une accuracy qui s'élève à environ 50,98 %, puis franchit le seuil des 64 % à -6 dB. Cette progression rapide traduit la

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

capacité du modèle à mieux exploiter les structures des signaux à mesure que le rapport signal sur bruit s'améliore.

3.3.4.2 Performances en conditions favorables (SNR élevé)

Au-delà de 0 dB, les résultats deviennent nettement satisfaisants, avec une accuracy dépassant 88 % et des valeurs de précision, de rappel et de F1-score avoisinant 0,90. Le modèle atteint ses meilleures performances dans la plage 10 dB – 18 dB, où l'accuracy excède 92 %, la précision dépasse 93 % et le F1-score se maintient au-dessus de 0,90.

Ces résultats confirment la robustesse de l'architecture CNN-TCN avec fusion SNR et sa capacité à tirer pleinement parti de l'information liée au rapport signal sur bruit pour améliorer la reconnaissance automatique des modulations radio.

SNR (dB)	Accuracy	Preccision	Recall	F1-Score
-20	0.0970	0.0833	0.0970	0.0747
-18	0.1337	0.1189	0.1337	0.1034
-16	0.1502	0.1074	0.1502	0.1180
-14	0.2042	0.1706	0.2042	0.1702
-12	0.2842	0.2772	0.2842	0.2580
-10	0.3773	0.3802	0.3773	0.3540
-8	0.5098	0.5036	0.5098	0.4959
-6	0.6443	0.6391	0.6443	0.6241
-4	0.7806	0.7728	0.7806	0.7568
-2	0.8528	0.8716	0.8528	0.8407
0	0.8863	0.9069	0.8863	0.8746
2	0.9070	0.9294	0.9070	0.9010
4	0.9168	0.9387	0.9168	0.9120
6	0.9032	0.9250	0.9032	0.8990
8	0.9045	0.9338	0.9045	0.8974
10	0.9245	0.9396	0.9245	0.9208
12	0.9102	0.9294	0.9102	0.9044
14	0.9017	0.9231	0.9017	0.8954
16	0.9206	0.9378	0.9206	0.9140
18	0.9127	0.9327	0.9127	0.9072

Tableau 3.1 : Évaluation des performances du modèle de classification de modulation à différents niveaux de rapport signal/bruit.

3.3.5 Les métriques d'évaluation en fonction du SNR

La Figure 3.8 retrace l'évolution de quatre métriques clés l'exactitude (accuracy), la précision, le rappel et le F1-score en fonction du rapport signal sur bruit. Cette représentation offre une vision synthétique et continue du comportement du modèle sur l'ensemble de la plage de SNR testée.

3.3.5.1 À faible SNR

Entre -20 dB et -10 dB, les quatre métriques affichent des valeurs relativement basses. Dans ces conditions, le bruit domine largement le signal utile, privant le modèle des informations nécessaires pour distinguer les différentes modulations. Les performances obtenues dans cette zone reflètent davantage les limites physiques du canal de transmission que celles de l'architecture elle-même.

C'est entre -8 dB et 0 dB que l'on observe la progression la plus spectaculaire. Les courbes de toutes les métriques grimpent rapidement, traduisant une amélioration marquée des capacités de classification du modèle. À mesure que le bruit recule, le modèle parvient à exploiter des caractéristiques de plus en plus fiables et discriminantes, ce qui se répercute directement sur ses performances.

3.3.5.2 À SNR élevé

Au-delà de 0 dB, les métriques se stabilisent à des niveaux élevés et cohérents entre elles. L'accuracy franchit la barre des 90% dès 2 dB, tandis que la précision, le rappel et le F1-score pondérés se maintiennent tous aux alentours de 0.9 . Les meilleurs résultats sont atteints entre 10 dB et 18 dB, où le modèle démontre une excellente capacité à identifier correctement les modulations radio, avec très peu de confusions inter-classes.

Au-delà des chiffres, la Figure 3.8 illustre de manière convaincante la complémentarité entre la qualité du signal et l'intelligence du modèle. L'architecture CNN-TCN avec fusion SNR tire pleinement parti des conditions favorables, tout en maintenant un comportement cohérent même dans des environnements dégradés. La progression régulière et homogène de toutes les métriques confirme que le modèle ne sacrifie aucune dimension de la classification au profit d'une autre, ce qui est un gage de fiabilité pour une utilisation dans des scénarios réels de communication radio.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

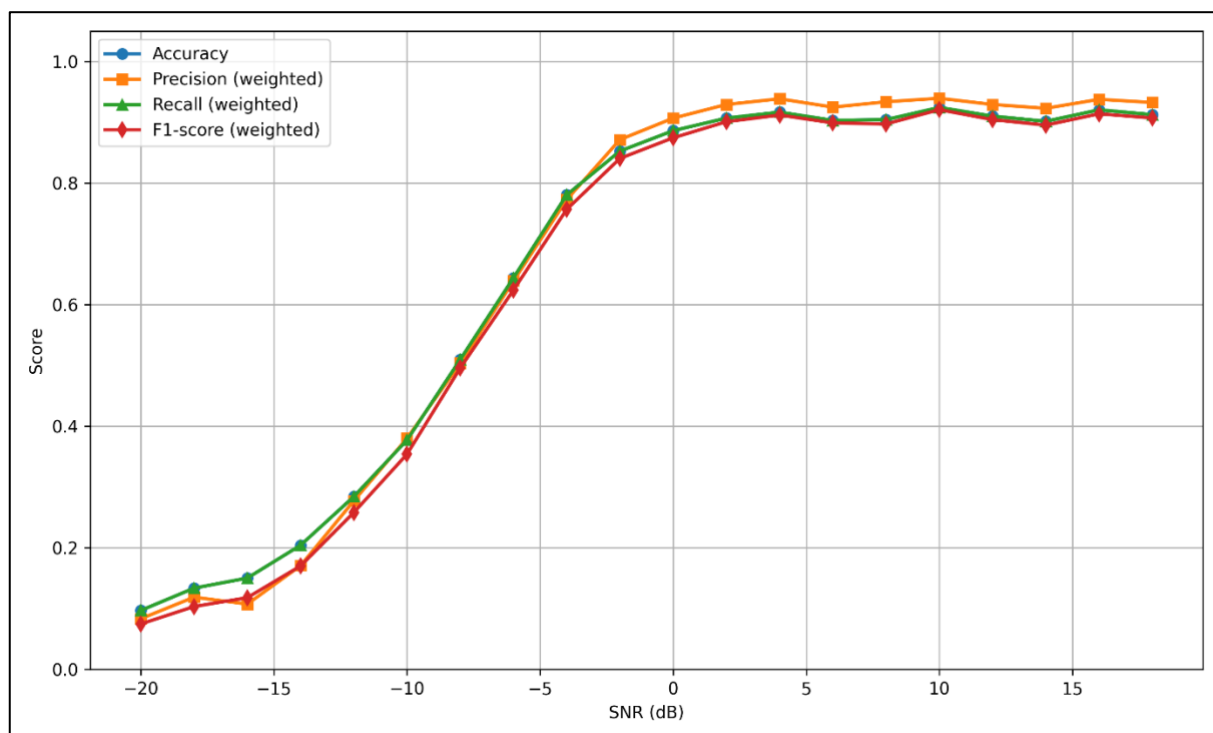


Figure 3.8 : Tracé des métriques d'évaluation en fonction du SNR.

3.3.6 Etude comparative

Le Tableau 3.2 propose une comparaison entre le modèle CNN-TCN avec fusion SNR et plusieurs architectures récentes dédiées à la reconnaissance automatique des modulations (AMR), toutes évaluées sur le dataset RadioML 2016. L'analyse s'appuie sur quatre indicateurs : l'accuracy moyenne par plage de SNR, l'accuracy globale, ainsi que les valeurs minimale et maximale atteintes.

Modèle	Average Accuracy(%)			Overall Accuracy %	Min Accuracy %	Max Accuracy %
	-20~-8dB	-6dB~4dB	6dB~18dB			
SMTrans [55]	17.95	81.47	92.99	63.27	9.52	93.34
MCLDNN [R4]	17.09	79.29	92.30	62.08	9.25	92.71
PET-CGDNN [59]	17.35	76.20	90.78	60.71	9.40	91.07
AMC-Net [60]	17.33	78.70	90.24	61.26	9.32	90.64
FEA-T [61]	17.14	76.09	88.68	59.86	9.42	89.08
ICRNNA [56]	--	--	--	63.24	--	93.40
CBADNN [57]	--	--	--	64.02	9.94	93.66
CNNTCNWithSNR (Ours)	25.09	83.13	91.11	65.84	9.70	92.45

Tableau 3.2 : Comparaison des performances.

3.3.6.1 Robustesse en conditions dégradées (SNR faible : -20 dB à -8 dB)

Dans cette plage où le bruit est prépondérant, le modèle proposé se distingue avec une accuracy moyenne de 25,09 %, la meilleure parmi l'ensemble des architectures comparées. Ce résultat témoigne d'une robustesse supérieure face aux conditions de transmission difficiles.

3.3.6.2 Performances en régime intermédiaire (-6 dB à 4 dB)

Le modèle affiche une accuracy moyenne de 83,13 % dans cette plage, surpassant plusieurs approches récentes telles que SMTTrans, PET-CGDNN, AMC-Net et FEA-T. Cette supériorité illustre la capacité du modèle à exploiter efficacement les caractéristiques des signaux même dans des conditions de bruit modérées.

3.3.6.3 Performances en conditions favorables (SNR élevé : 6 dB à 18 dB)

Lorsque le rapport signal sur bruit est favorable, le modèle maintient une accuracy moyenne de 91,11 %, confirmant sa fiabilité pour la classification des signaux dans des environnements peu bruités.

Sur l'ensemble du jeu de test, le modèle atteint une accuracy globale de 65,84 % le meilleur résultat parmi tous les modèles du tableau, ainsi qu'une accuracy maximale de 92,45 %. Ces performances s'expliquent par la complémentarité des composantes de l'architecture :

L'extraction des caractéristiques spatiales par le CNN, la modélisation des dépendances temporelles par le TCN, et l'intégration de l'information SNR qui confère au modèle une meilleure résistance aux effets du bruit.

Au final, cette comparaison confirme la pertinence et la compétitivité de l'architecture CNN-TCN avec fusion SNR pour la classification automatique des modulations radio dans des conditions variées de transmission.

3.4 Interprétation du modèle par les méthodes XAI (gradients intégrés et SHAP)

3.4.1 Modèle hybride CNN-TCN avec fusion SNR et interprétabilité XAI

Ce travail propose une architecture hybride CNN-TCN avec fusion SNR pour la classification automatique des modulations radio, évaluée sur le dataset RadioML 2016.10a. Le modèle atteint une accuracy globale de 65,84 %, avec des performances dépassant 92 % en conditions de SNR élevé, surpassant plusieurs architectures récentes de la littérature. L'interprétabilité des décisions est assurée par des méthodes XAI (gradients intégrés et SHAP), confirmant la robustesse et la fiabilité de l'approche proposée.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

La Figure 3.9 présente l'architecture d'un Modèle Hybride CNN-TCN avec Fusion de Données SNR et Module d'Interprétabilité XAI (SHAP) :

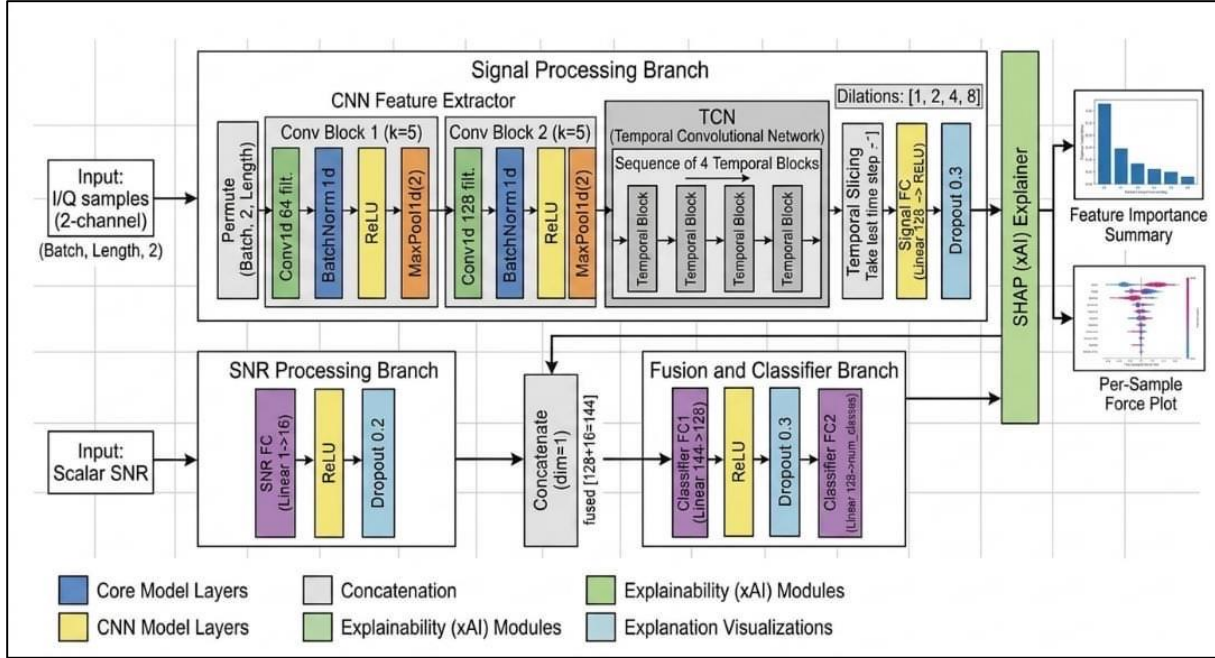


Figure 3.9 : Architecture d'un Modèle Hybride CNN-TCN avec Fusion de Données SNR et Module d'Interprétabilité XAI (SHAP).

3.4.2 Analyse explicable des modulations AM-DSB, AM-SSB et 8PSK

Cette section s'intéresse à l'interprétabilité du modèle CNN-TCN proposé à travers le prisme de l'intelligence artificielle explicable (XAI). Afin de mieux comprendre les mécanismes sous-jacents aux décisions du modèle, deux méthodes d'interprétation sont mises en œuvre : les cartes de saillance par gradients intégrés et l'analyse SHAP, toutes deux appliquées directement aux signaux I/Q.

3.4.2.1 Gradients intégrés et SHAP sous faible SNR (−20 dB)

a) Gradients intégrés

Les Figures 3.10 et 3.11 montrent les cartes de saillance à gradients intégrés pour deux types de modulations analogiques AM-DSB et AM-SSB prises à un rapport signal/bruits égal à 0 dB.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

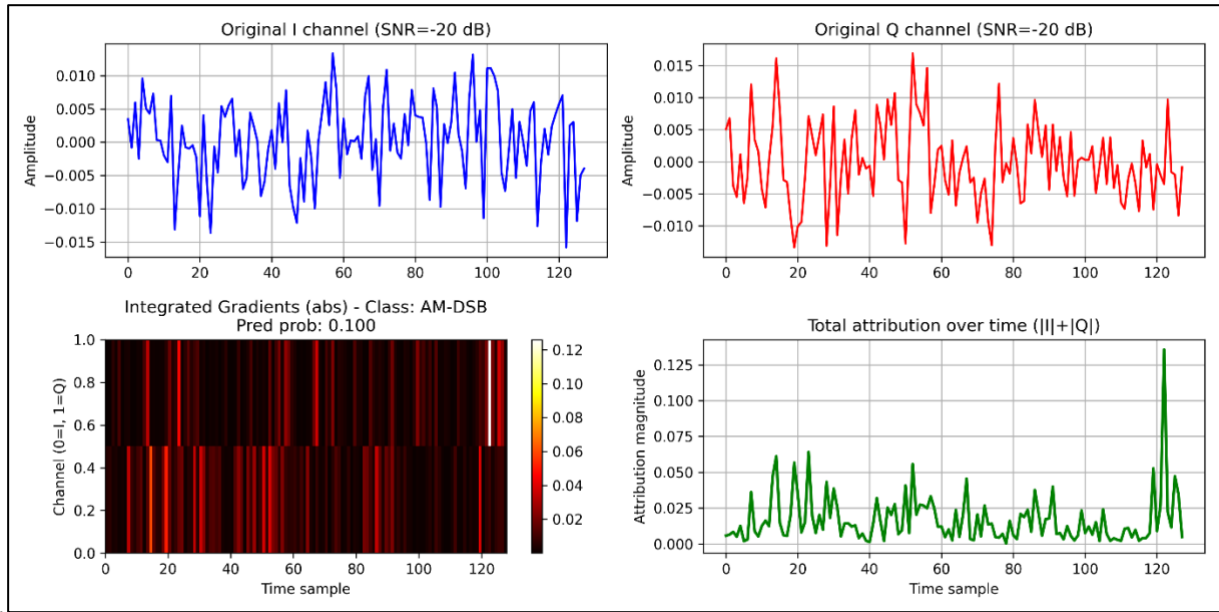


Figure 3.10 : Carte de saillance à gradients intégrés pour l'AM-DSB avec un rapport signal/bruit (SNR) de -20 dB.

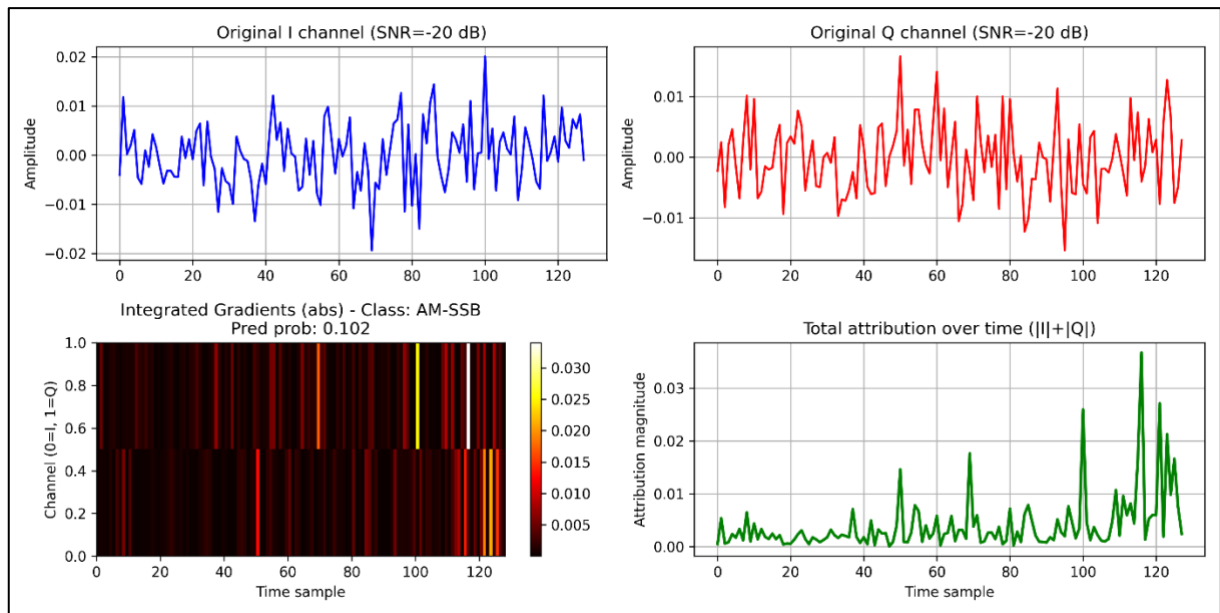


Figure 3.11 : Carte de saillance à gradients intégrés pour l'AM-SSB avec un rapport signal/bruit (SNR) de -20 dB.

A partir des deux Figures (AM-DSB et AM-SSB) on peut remarquer :

- les signaux I/Q sont fortement dégradés par le bruit (SNR = -20 dB), ce qui rend leur structure difficile à distinguer. Les cartes d'attribution (Integrated Gradients) révèlent que le modèle s'appuie sur quelques instants spécifiques avec des pics localisés, mais ces contributions restent globalement faibles et dispersées. Les probabilités de prédiction sont

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

très basses (≈ 0.10), indiquant que le modèle manque de confiance et a du mal à différencier correctement les deux types de modulation dans ces conditions très bruitées.

b) SHAP

Les Figures 3.12 et 3.13 montrent les 20 principales features SHAP des signaux I/Q pour deux types de modulations analogiques AM-DSB et AM-SSB prises à un rapport signal/bruits égal à -20 dB.

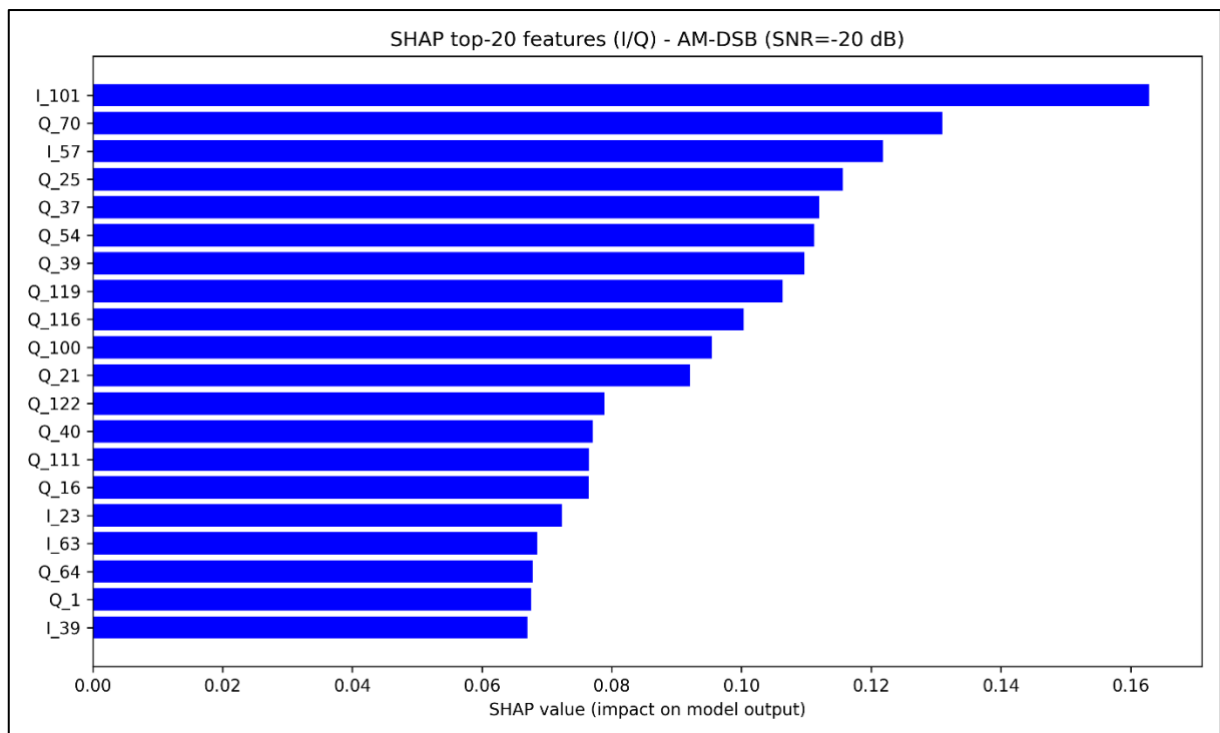


Figure 3.12: 20 principales features SHAP (I/Q) - AM-DSB (SNR = -20 dB).

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

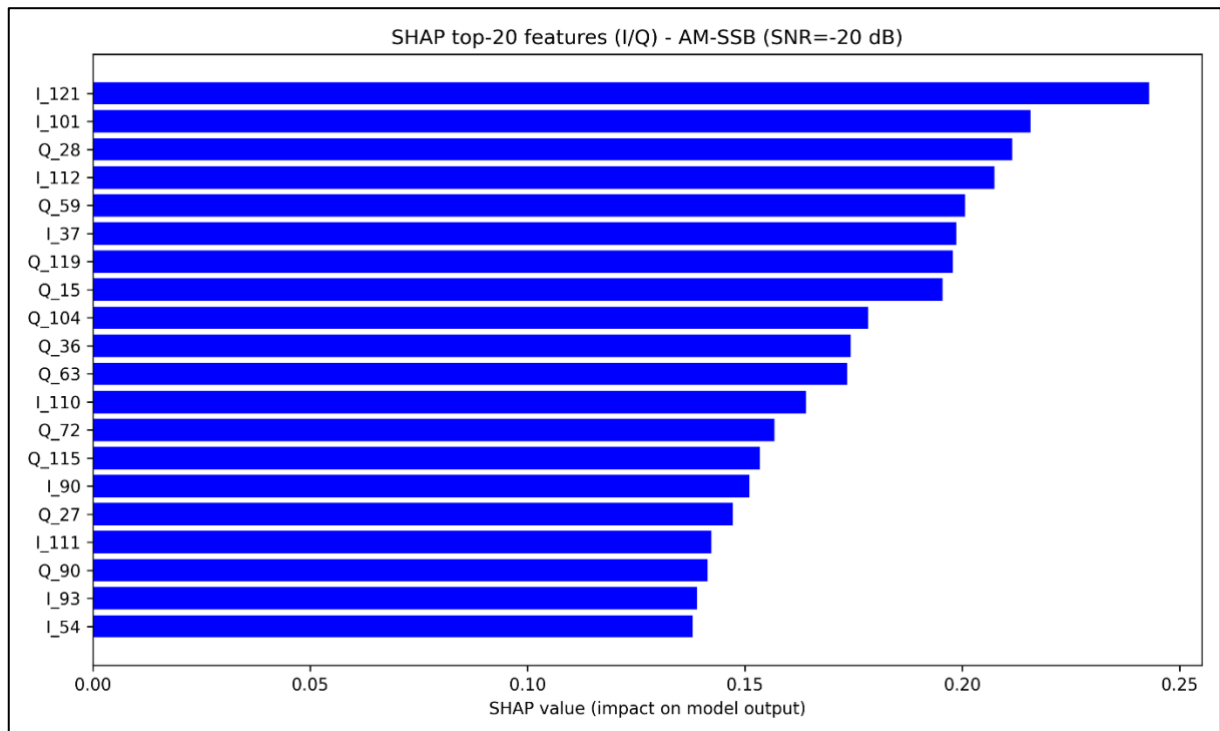


Figure 3.13: 20 principales features SHAP (I/Q) - AM-SSB (SNR = -20 dB).

A partir des deux Figures précédentes on peut tirer les observations suivantes :

- Pour AM-DSB, les contributions SHAP sont globalement plus faibles et réparties entre les canaux I et Q, avec une légère dominance de certains points isolés. Cela traduit une prise de décision basée sur des indices locaux et peu stables, cohérente avec la faible confiance du modèle dans ce cas.
- AM-SSB, les valeurs SHAP sont nettement plus élevées et concentrées sur certains échantillons clés, indiquant que le modèle identifie des caractéristiques plus discriminantes malgré le bruit. L'impact des variables est plus marqué et structuré, ce qui explique une meilleure capacité de reconnaissance par rapport à AM-DSB.
- Globalement, le modèle est plus performant pour AM-SSB que pour AM-DSB, car il exploite des caractéristiques plus claires et structurées, malgré le bruit.

3.4.2.2 Gradients intégrés et SHAP sous SNR intermédiaire (0 dB)

a) Gradients intégrés

Les Figures 3.14, 3.15 et 3.16 montrent respectivement les cartes de saillance à gradients intégrés pour deux types de modulations analogiques AM-DSB et AM-SSB et un type de modulation numérique 8PSK prises à un rapport signal/bruits égal à 0 dB. Malgré un bruit

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

encore présent, le modèle parvient à mieux exploiter les signaux I/Q, avec des niveaux de confiance variables selon la modulation.

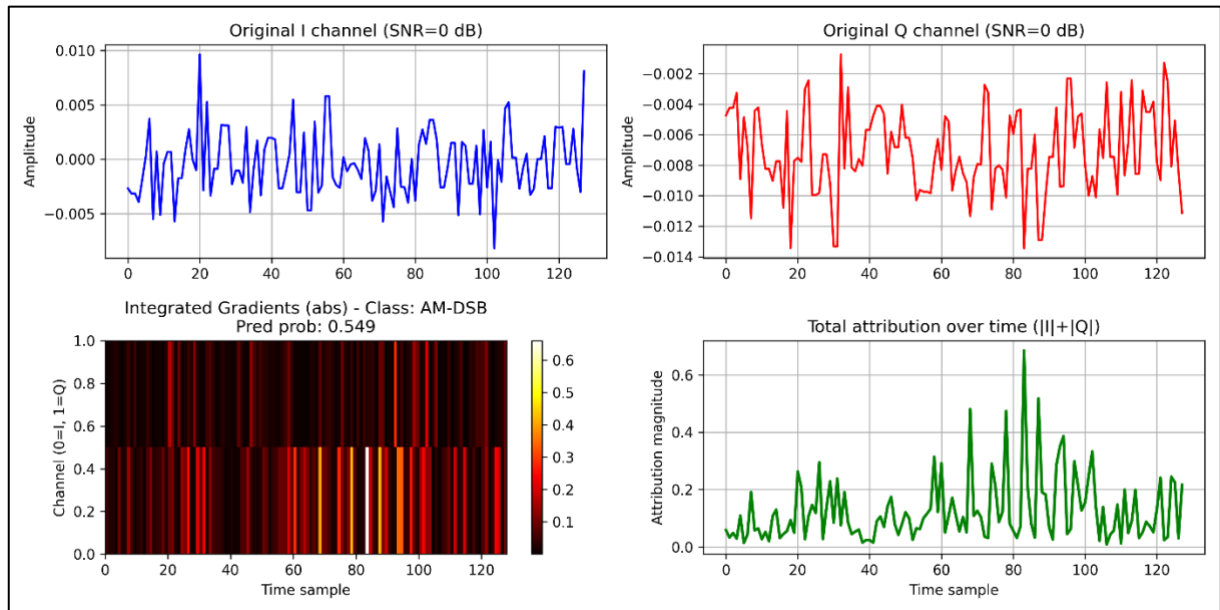


Figure 3.14 : Carte de saillance à gradients intégrés pour l'AM-DSB avec un rapport signal/bruit (SNR) de 0 dB.

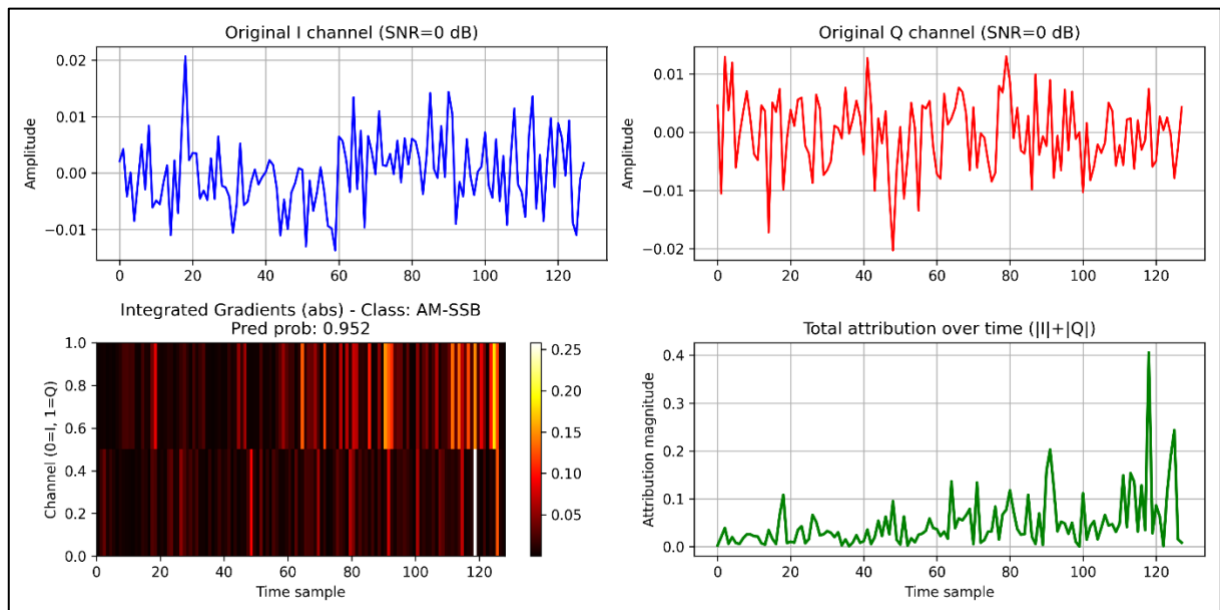


Figure 3.15 : Carte de saillance à gradients intégrés pour l'AM-SSB avec un rapport signal/bruit (SNR) de 0 dB.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

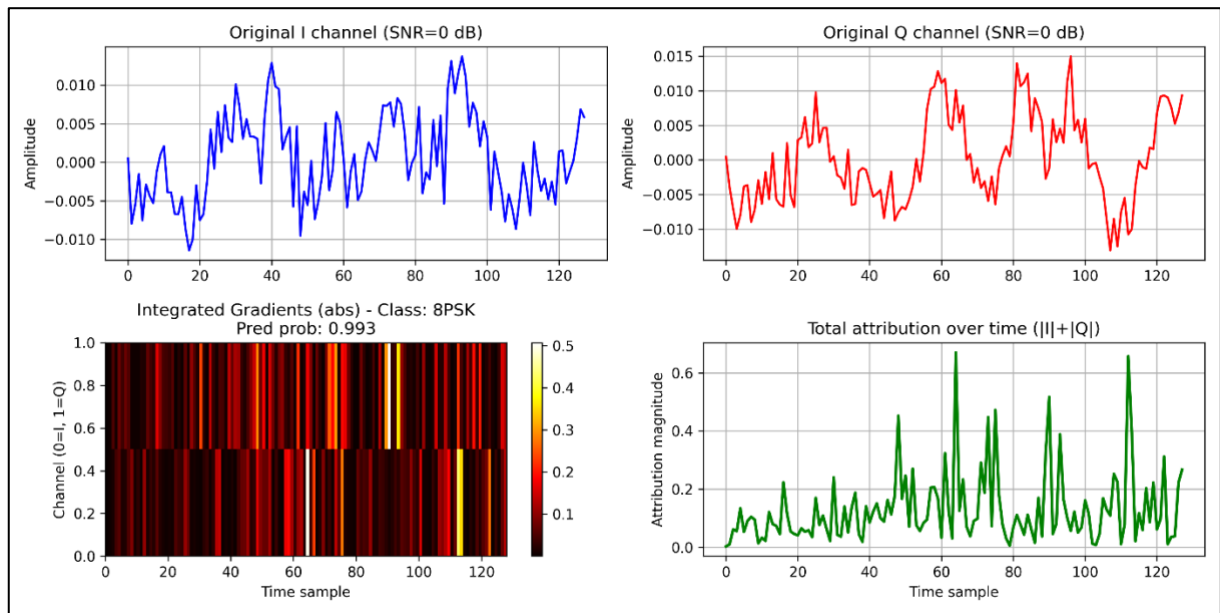


Figure 3.16 : Carte de saillance à gradients intégrés pour l'8PSK avec un rapport signal/bruit (SNR) de 0 dB.

À partir des Figures 3.14, 3.15 et 3.16, on peut dégager les observations suivantes :

- Pour l'AM-DSP, la prédiction est plus incertaine (≈ 0.55), et les attributions sont plus diffuses, montrant que le modèle a plus de difficulté à extraire des caractéristiques distinctives.
- AM-SSB, la confiance est élevée (≈ 0.95), avec des contributions concentrées sur certaines zones temporelles pertinentes.
- 8PSK, la prédiction est très fiable (prob ≈ 0.99) avec des attributions marquées et bien réparties, indiquant que le modèle identifie clairement des motifs discriminants.
- Globalement, même à SNR moyen, certaines modulations (comme 8PSK) sont nettement mieux reconnues que d'autres.

b) SHAP

Les Figures 3.17, 3.18 et 3.19 montrent les 20 principales features SHAP des signaux I/Q pour deux types de modulations analogiques AM-DSB et AM-SSB et un type de modulation numérique 8PSK prises à un rapport signal/bruits égal à 0 dB.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

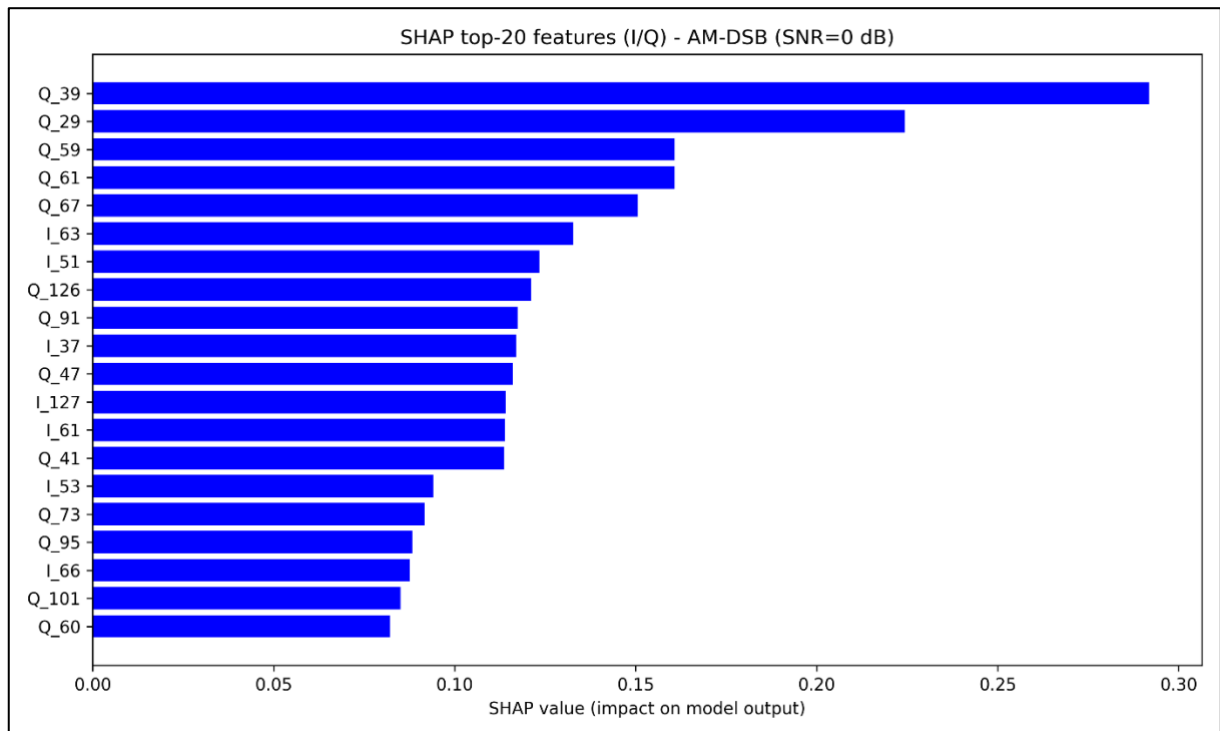


Figure 3.17: 20 principales features SHAP (I/Q) – AM-DSB (SNR = 0 dB).

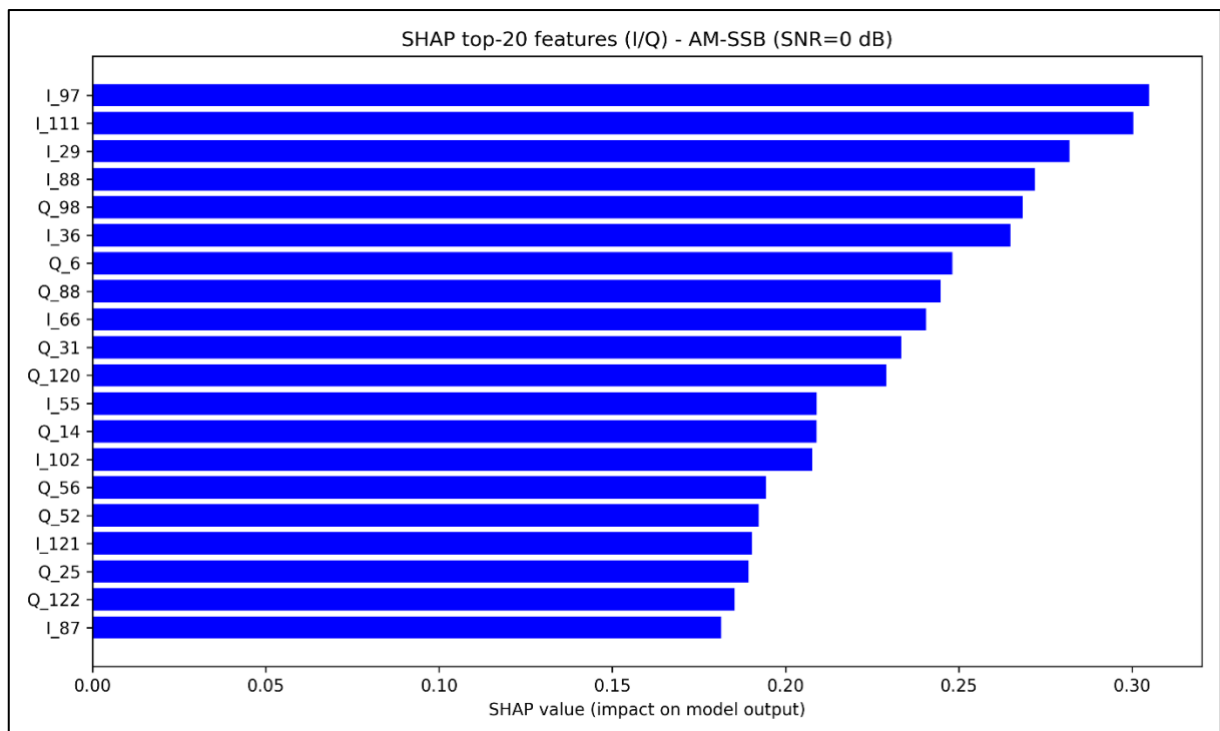


Figure 3.18: 20 principales features SHAP (I/Q) – AM-SSB (SNR = 0 dB).

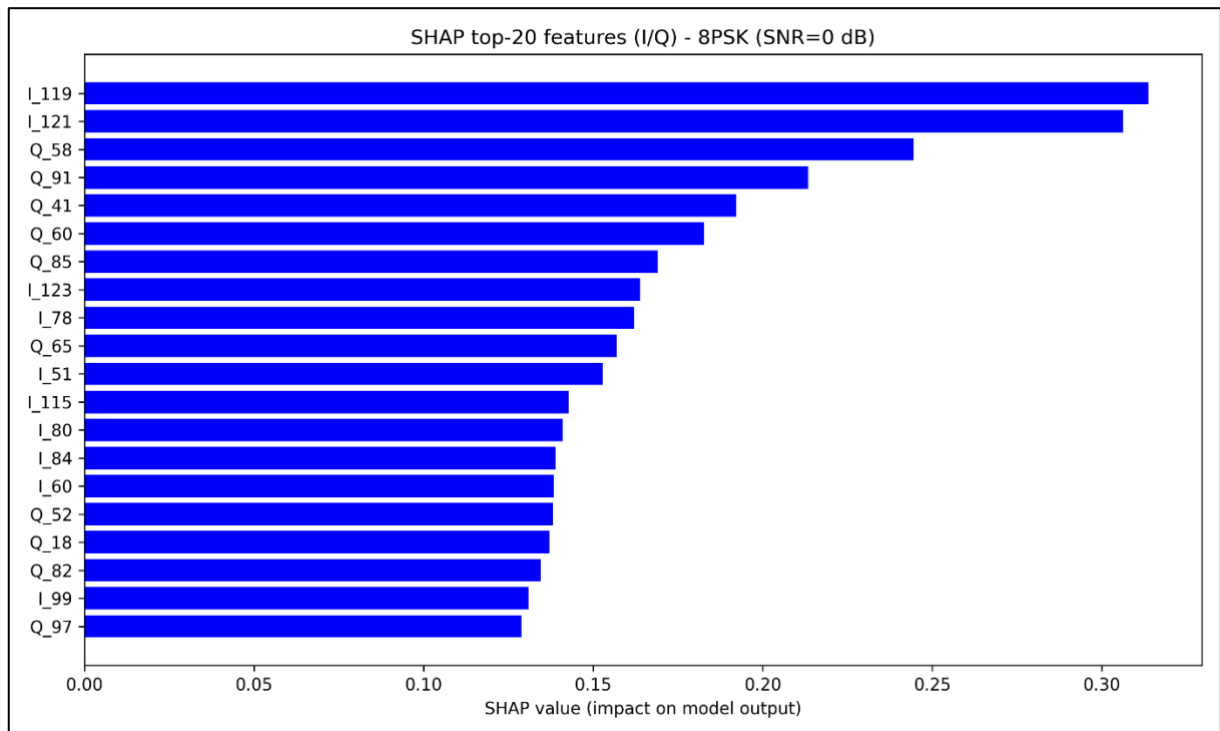


Figure 3.19: 20 principales features SHAP (I/Q) – 8PSK (SNR = 0 dB).

Les trois Figures à SNR = 0 dB (AM-DSB, AM-SSB et 8PSK) montrent :

- Une augmentation globale de l'importance des variables avec des valeurs SHAP plus élevées et mieux structurées
- Pour l'AM-DSB, les contributions restent dominées par quelques échantillons, souvent du canal Q, ce qui traduit une exploitation encore partielle et locale du signal.
- l'AM-SSB, les variables importantes sont plus nombreuses et mieux réparties entre I et Q, indiquant une extraction de caractéristiques plus robuste et globale.
- 8PSK, les contributions sont à la fois élevées et largement distribuées sur plusieurs échantillons des deux canaux, ce qui montre que le modèle capte efficacement la structure du signal.
- Globalement, plus la modulation est discriminante, plus les attributions SHAP sont fortes, réparties et cohérentes, ce qui se traduit par une meilleure performance du modèle.

3.4.2.3 Gradients intégrés et SHAP sous SNR élevé (10 dB)

a) Gradients intégrés

Les Figures 3.20 , 3.21 et 3.22 montrent respectivement les cartes de saillance à gradients intégrés pour deux types de modulations analogiques AM-DSB et AM-SSB et un type de modulation numérique 8PSK prises à un rapport signal/bruits égal à 10 dB.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

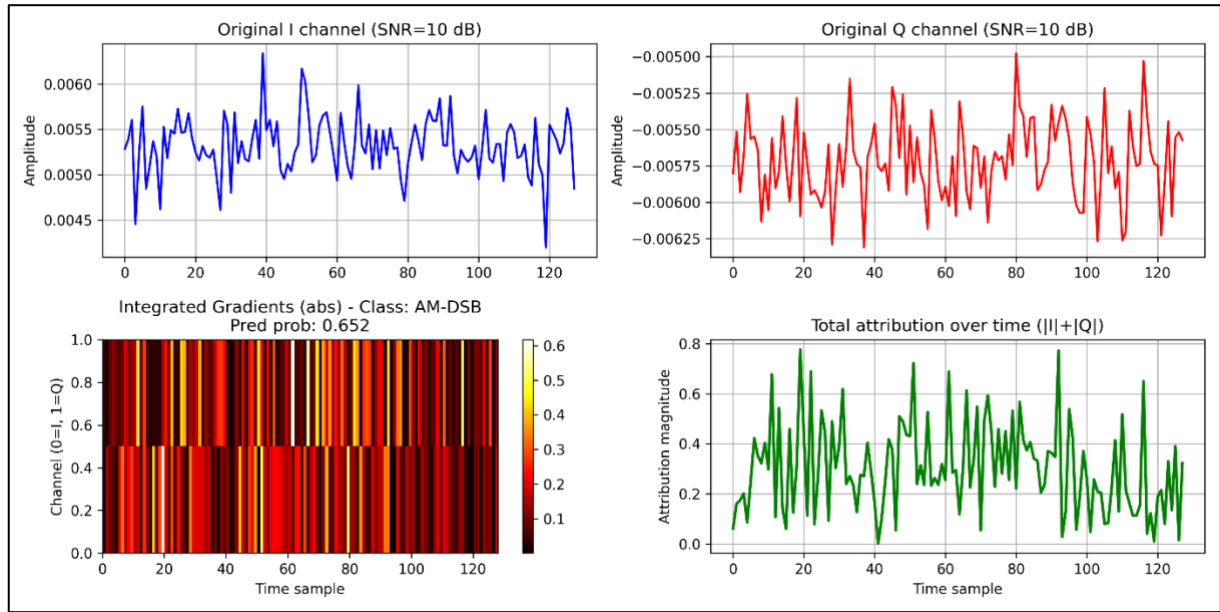


Figure 3.20 : Carte de saillance à gradients intégrés pour l'AM-DSB avec un rapport signal/bruit (SNR) de 10 dB.

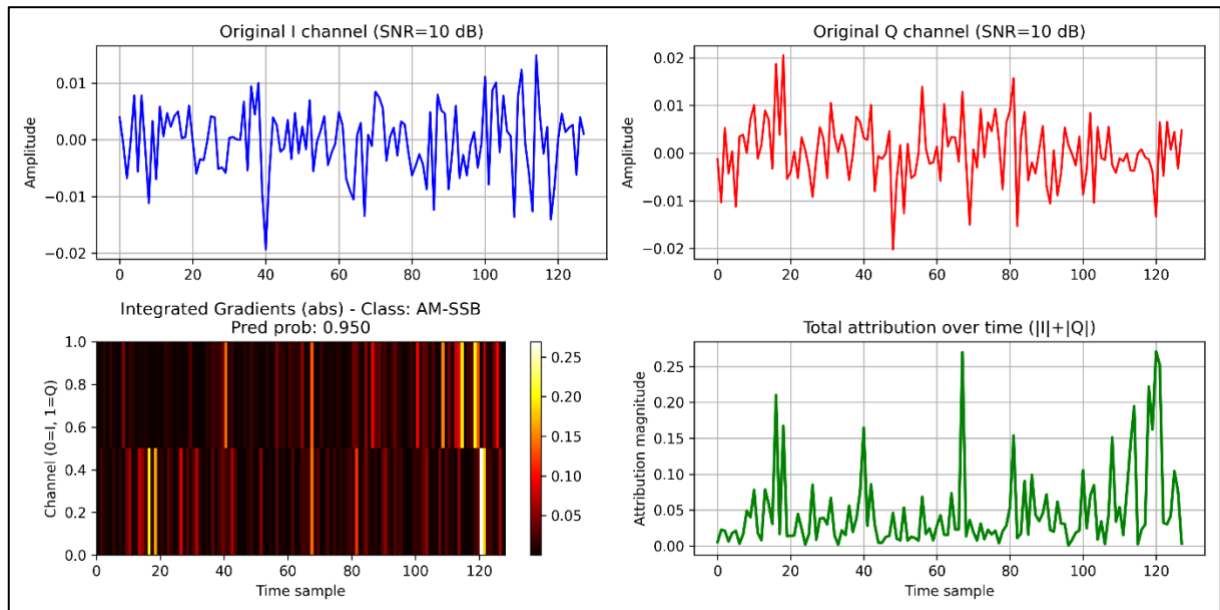


Figure 3.21 : Carte de saillance à gradients intégrés pour l'AM-SSB avec un rapport signal/bruit (SNR) de 10 dB.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

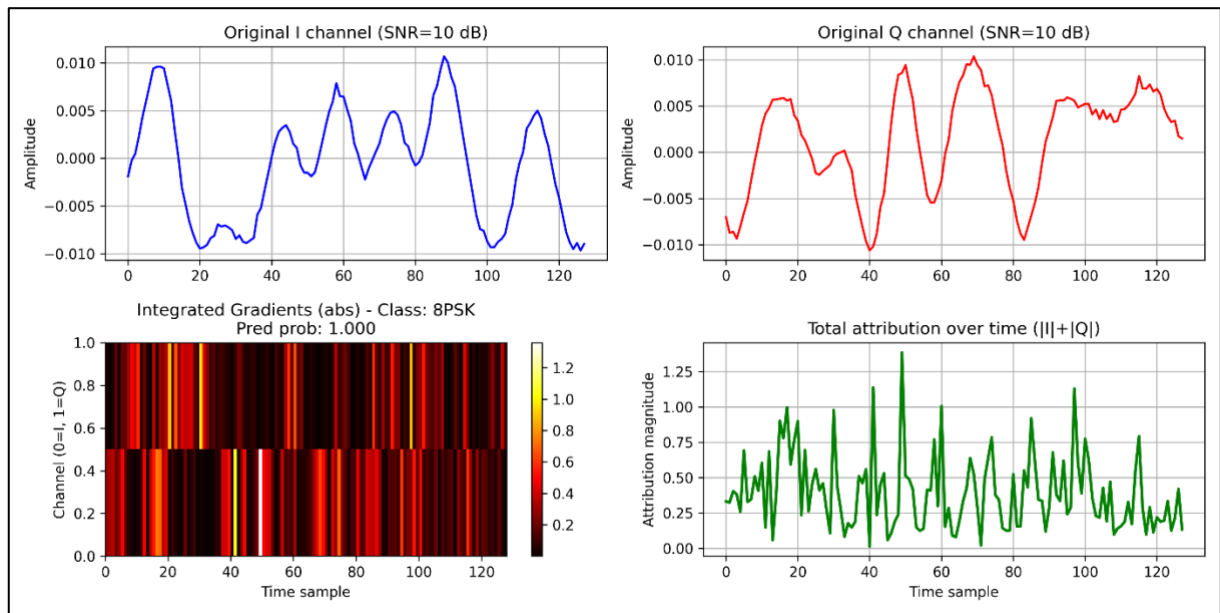


Figure 3.22 : Carte de saillance à gradients intégrés pour l'8PSK avec un rapport signal/bruit (SNR) de 10 dB.

A partir des Figures à SNR = 10 dB, on peut dégager les observations suivantes :

- une nette amélioration de la qualité des signaux et des performances du modèle. Le bruit devient moins dominant, rendant les formes des signaux I/Q plus structurées et interprétables. Les attributions sont plus intenses et mieux réparties dans le temps, indiquant que le modèle exploite davantage d'informations sur l'ensemble du signal.
- La reconnaissance est très fiable pour 8PSK (prob ≈ 1.00) et reste élevée pour AM-SSB (≈ 0.95), avec des contributions bien localisées sur des parties pertinents. En revanche, AM-DSB, bien qu'amélioré (≈ 0.65), reste moins bien distingué, avec des attributions plus diffuses.
- Globalement, l'augmentation du SNR permet une meilleure discrimination, surtout pour les modulations plus facilement séparables comme 8PSK.

b) SHAP

Les Figures 3.23 ,3.24 et 3.25 montrent les 20 principales features SHAP des signaux I/Q pour deux types de modulations analogiques AM-DSB et AM-SSB et un type de modulation numérique 8PSK prises à un rapport signal/bruits égal à 10 dB.

CHAPITRE 3: RESULTATS DE SIMULATION : PRESENTATION ET INTERPRETATION

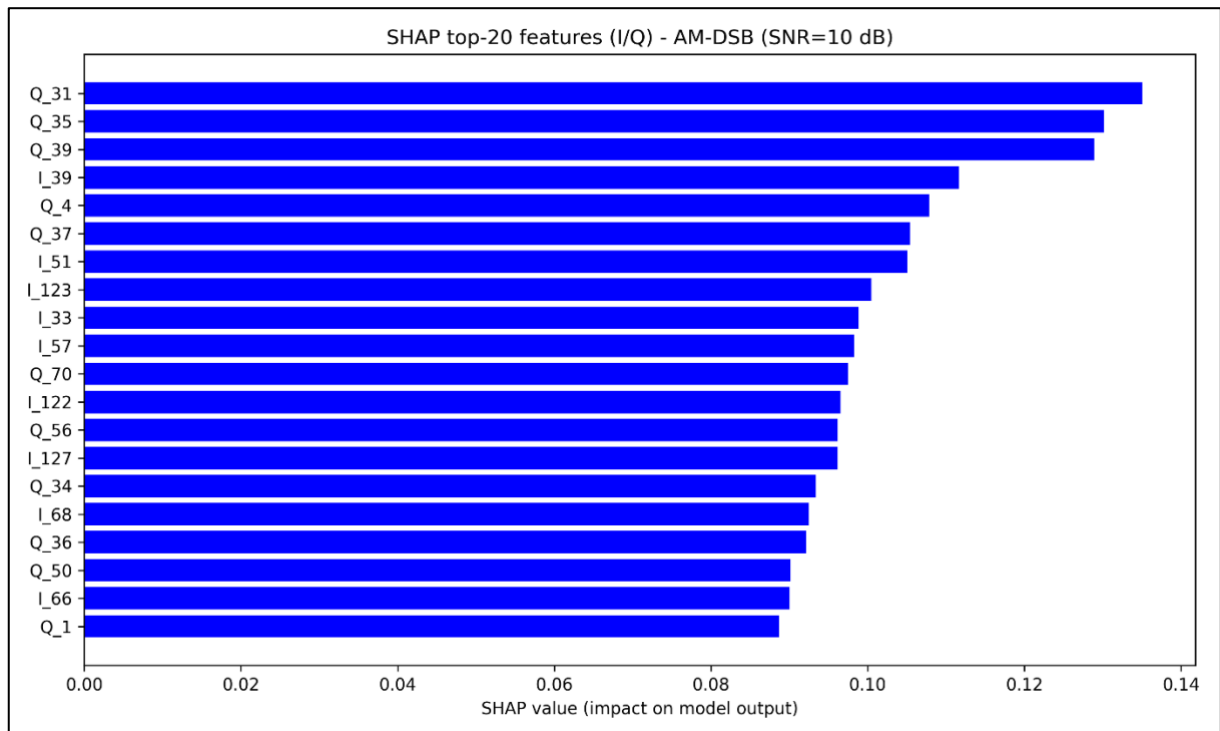


Figure 3.23: 20 principales features SHAP (I/Q) – AM-DSB (SNR = 10 dB).

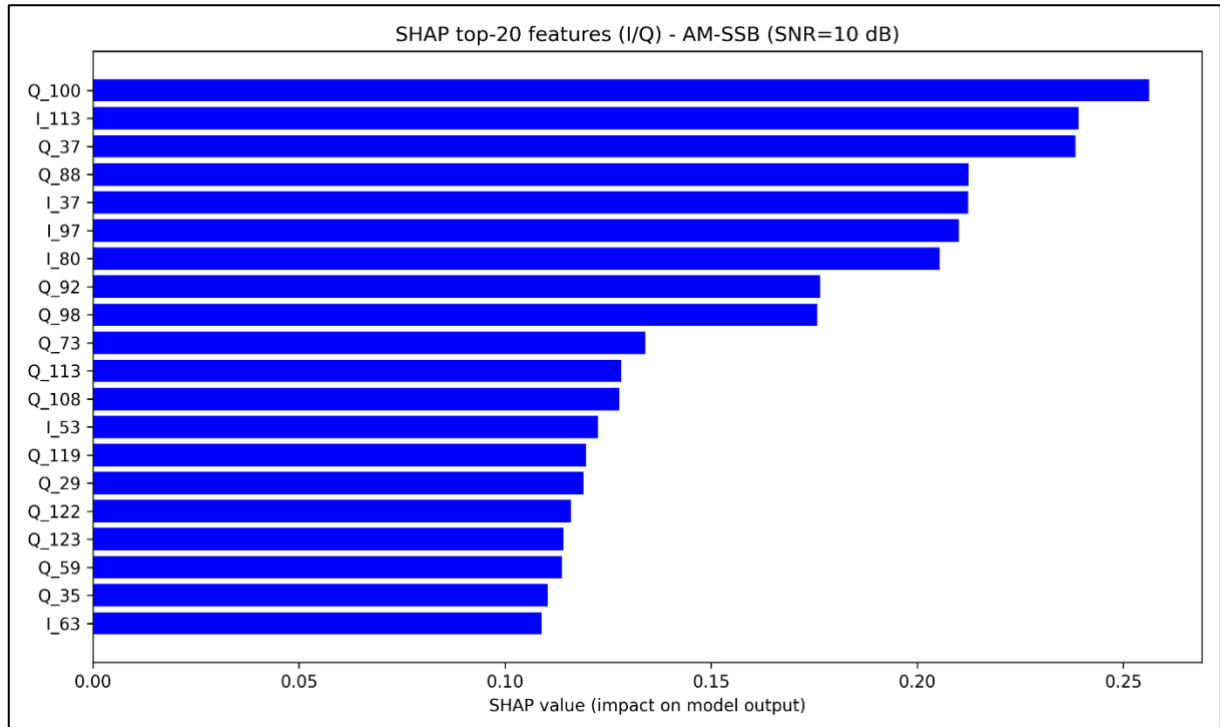


Figure 3.24: 20 principales features SHAP (I/Q) – AM-SSB (SNR = 10 dB).

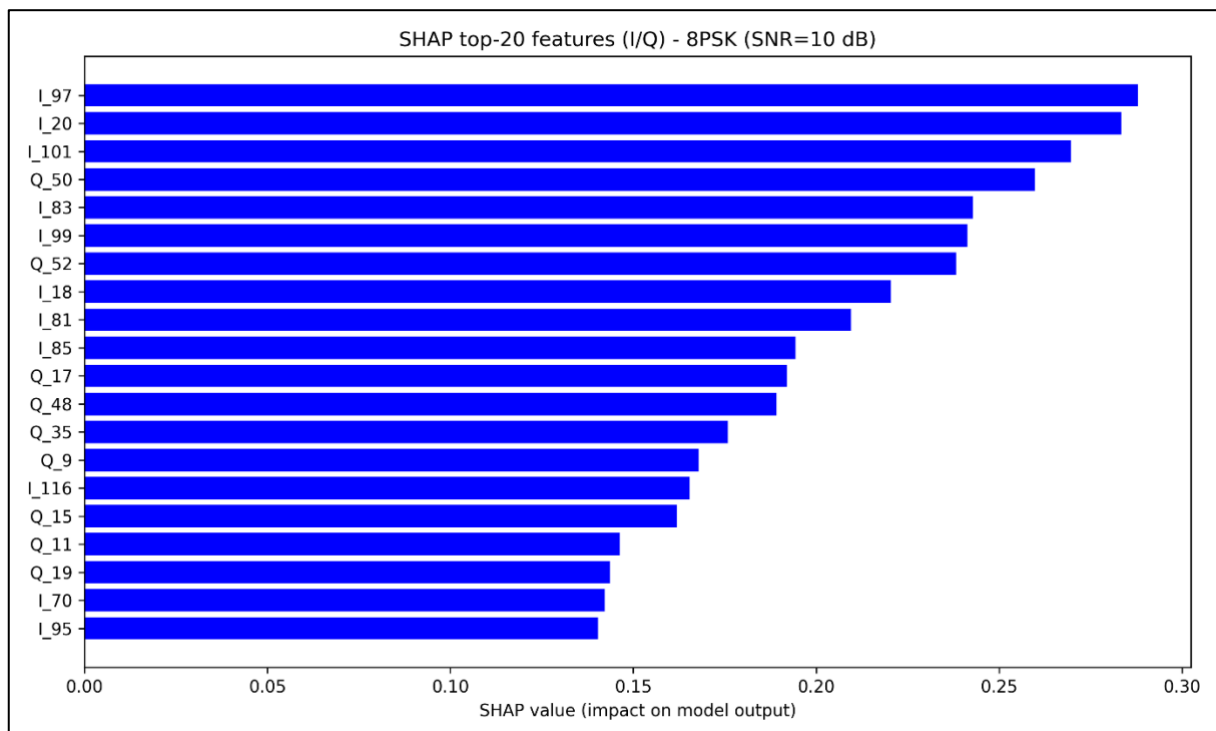


Figure 3.25: 20 principales features SHAP (I/Q) – 8PSK (SNR = 10 dB).

A partir des Figures précédentes , on peut dégager les observations suivantes

- Pour la modulation AM-DSB, les valeurs SHAP sont globalement modestes et distribuées de façon relativement uniforme sur l'ensemble des composantes I/Q, ce qui indique que la décision du modèle repose sur la contribution conjointe de plusieurs caractéristiques sans qu'aucune ne se démarque véritablement.
- La modulation AM-SSB présente un profil différent : certaines caractéristiques affichent des valeurs SHAP nettement plus élevées, témoignant d'une contribution plus marquée de quelques échantillons I/Q ciblés dans le processus de reconnaissance. Le modèle exploite ainsi de manière plus sélective les informations discriminantes propres à cette modulation.
- La modulation 8PSK se distingue quant à elle par les valeurs SHAP les plus élevées et les contributions les plus concentrées parmi les trois modulations étudiées.
- Les caractéristiques discriminantes y sont nettement mieux définies, reflétant une capacité remarquable du modèle à identifier les motifs structurels propres à ce type de signal numérique.

Ces résultats font de la 8PSK la modulation la mieux reconnue dans les conditions analysées, confirmant que l'architecture CNN-TCN tire pleinement parti des structures caractéristiques de cette modulation lorsque le rapport signal sur bruit est favorable.

3.5 Conclusion

Ce chapitre a été consacré à la présentation et à l'analyse approfondie des résultats expérimentaux obtenus par le modèle CNN-TCN avec fusion SNR pour la classification automatique des modulations radio sur le dataset RadioML 2016.10a.

L'étude des courbes d'apprentissage a mis en évidence un comportement d'entraînement stable et maîtrisé, sans surapprentissage notable, confirmant la bonne capacité de généralisation de l'architecture proposée. L'analyse des matrices de confusion, aussi bien globales que par niveau de SNR, a permis de caractériser précisément les forces et les limites du modèle : d'excellentes performances pour des modulations telles que AM-DSB, AM-SSB ou 8PSK, et des difficultés persistantes pour les modulations analogiques dont les signatures spectrales sont proches, notamment dans les conditions fortement bruitées.

L'évaluation des métriques en fonction du SNR a mis en évidence une progression cohérente et régulière des performances à mesure que la qualité du signal s'améliore, avec une accuracy dépassant 92 % dans la plage favorable de 10 dB à 18 dB. L'étude comparative avec plusieurs architectures récentes de la littérature a confirmé la compétitivité du modèle proposé, qui se distingue notamment par sa robustesse dans les conditions de faible SNR, avec la meilleure accuracy moyenne de 25,09 % dans la plage -20 dB à -8 dB, ainsi que par la meilleure accuracy globale de 65,84 % parmi tous les modèles comparés.

Enfin, l'application des méthodes d'intelligence artificielle explicable, à travers les cartes de saillance par gradients intégrés et l'analyse SHAP, a apporté un éclairage précieux sur les mécanismes internes du modèle. Ces outils ont permis d'identifier les régions et les caractéristiques I/Q les plus déterminantes dans le processus de décision, renforçant ainsi la transparence et la fiabilité de l'approche proposée.

En définitive, les résultats présentés dans ce chapitre valident l'efficacité et la pertinence de l'architecture hybride CNN-TCN avec fusion SNR pour la reconnaissance automatique des modulations radio dans des conditions de transmission variées, et ouvrent des perspectives prometteuses pour son déploiement dans des systèmes de radio cognitive et de guerre électronique.

CONCLUSION GENERALE

Ce travail de fin d'études a porté sur la problématique de la reconnaissance automatique des modulations radio dans des environnements de transmission bruités, en mobilisant les techniques modernes d'intelligence artificielle et d'apprentissage profond. L'ambition centrale était de concevoir un modèle capable d'identifier efficacement différents types de modulations tout en maintenant une robustesse satisfaisante face aux variations du rapport signal sur bruit.

Dans un premier temps, les fondements théoriques des communications radio numériques ont été établis, couvrant les principales techniques de modulation analogique et numérique ainsi que les méthodes classiques de classification. Cette étape a permis de mettre en lumière les difficultés inhérentes à la reconnaissance des signaux dans des conditions réelles de propagation et de souligner les limites des approches conventionnelles.

Le deuxième volet a été consacré aux concepts fondamentaux de l'intelligence artificielle, du Machine Learning, du Deep Learning et de l'intelligence artificielle explicable. Une attention particulière a été accordée aux méthodes d'explicabilité, notamment SHAP et les gradients intégrés, dont le rôle s'avère déterminant pour interpréter et valider les décisions des modèles profonds.

La partie expérimentale a conduit à la proposition et à l'évaluation d'une architecture hybride CNN-TCN avec fusion de l'information SNR, appliquée au dataset RadioML 2016.10a. Les résultats démontrent que le modèle atteint des performances satisfaisantes en classification des modulations, en particulier dans les conditions de SNR moyen et élevé. L'intégration explicite de l'information SNR s'est révélée bénéfique pour renforcer la robustesse du système face au bruit. Par ailleurs, l'application des méthodes XAI a offert une lecture plus transparente des caractéristiques exploitées par le modèle lors de la prise de décision, contribuant ainsi à accroître la confiance dans le système proposé.

Ce travail ouvre par ailleurs plusieurs perspectives prometteuses. Les performances du modèle pourraient être améliorées en explorant des architectures plus avancées, telles que les Transformers ou les mécanismes d'attention, et en s'appuyant sur des jeux de données plus complexes et plus représentatifs des conditions réelles. L'intégration de méthodes d'explicabilité complémentaires permettrait également d'approfondir l'analyse du comportement des réseaux de neurones appliqués aux communications radio intelligentes.

En définitive, cette étude apporte une contribution au domaine de la reconnaissance automatique des modulations radio basée sur l'intelligence artificielle, en proposant une approche qui conjugue performance, robustesse et interprétabilité.

Bibliographie

Bibliographie

- [1] Mhamdi, M. (2017). *Méthodes de transmission d'images optimisées utilisant des techniques de communication numériques avancées pour les systèmes multi-antennes* (Thèse de doctorat, Université de Poitiers & École nationale d'ingénieurs de Tunis).
- [2] Bauer, S. (2024). *Contribution à la modélisation de réseaux de spécifications sémantiques et tissage des liens entre les données « produit » tout au long de son cycle de vie* (Thèse de doctorat, Université de Bordeaux).
- [3] Lathi, B. P., & Ding, Z. (1998). *Modern digital and analog communication systems* (3rd ed.). Oxford University Press.
- [4] Fortaki, T. (s.d.). *Communication numérique – Partie 1 : Transmission numérique en bande de base*. Université Batna 2. https://staff.univ-batna2.dz/sites/default/files/fortaki_tarek/files/L3_TELECOM_S6_Communication_numerique.pdf
- [5] Futura-Sciences. (2020, 20 octobre). *Bande passante : qu'est-ce que c'est ?* <https://www.futura-sciences.com/tech/definitions/internet-bande-passante-412/>
- [6] Tidjma. (s.d.). *Notion de bande passante — définition et explication*. <https://www.tidjma.tn/fr/electrical/bandwidth-1/>
- [7] Ben Sik Ali, I. (2015). *Traitement du bruit impulsionnel dans des postes électriques pour les systèmes de communication sans fil* (Thèse de doctorat, École de technologie supérieure).
- [8] Gallin, L.-J. (2014). *Caractérisation acoustique des éclairs d'orage* (Thèse de doctorat, Université Pierre et Marie Curie).
- [9] Sacuto, F. (2010). *Caractérisation des milieux hautes tensions en vue d'adapter un réseau sans fil de capteurs* (Thèse de doctorat, École de technologie supérieure).
- [10] Mechouk, N. (2010). *Étude et conception d'une structure BIST pour convertisseur analogique-numérique* (Thèse de doctorat, Université Bordeaux 1).
- [11] Guillet, J. (2004). *Caractérisation et modélisation spatio-temporelles du canal de propagation radioélectrique dans le contexte MIMO* (Thèse de doctorat, INSA de Rennes).

Bibliographie

- [12] La Villette, P. (s.d.). *De la perception à la compréhension : Applications audioprothétiques – Résumés des communications*.
- [13] Proakis, J. G. (2001). *Digital communications* (4th ed.). McGraw-Hill.
- [14] Haykin, S. (2009). *Communication systems* (5th ed.). Wiley.
- [15] Lathi, B. P., & Ding, Z. (2010). *Modern digital and analog communication systems* (4th ed.). Oxford University Press.
- [16] *Unique lateral band AM modulation*. (s.d.).
https://shms-prod.s3.amazonaws.com/media/editor/143832/Unique_Lateral_Band_AM_modulation.pdf
- [17] Rappaport, T. S. (2002). *Wireless communications: Principles and practice* (2nd ed.). Prentice Hall.
- [18] Sarita. (s.d.). *Cours modulation FM*. Scribd.
<https://fr.scribd.com/document/3994017240/Cours-Modulation-FM>
- [19] Goldsmith, A. (2005). *Wireless communications*. Cambridge University Press.
- [20] Abrika, L., & Chabni, A. (s.d.). *La modulation numérique* [Document PDF]. Scribd.
<https://fr.scribd.com/document/544390875/La-Modulation-Numerique>
- [21] Bouaouina, A. (2018). *Étude et simulation des performances de la modélisation OFDM dans les liaisons optiques*.
- [22] De Beaupré, V. C. (2008). *Dispositifs à faible coût appliqués à la synthèse de fréquences et à la modulation FSK pour les systèmes de radiocommunication* (Thèse de doctorat, Université de Provence-Aix-Marseille I).
- [23] Proakis, J. G. (2008). *Digital communications* (5th ed.). McGraw-Hill.
- [24] Muller, J. P. (2000). *Les modulations numériques dans les systèmes de communication*.
- [25] Wang, Y., Liu, M., Yang, J., & Gui, G. (2019). Data-driven deep learning for automatic modulation recognition in cognitive radios. *IEEE Transactions on Vehicular Technology*, 68(4), 4074–4077. <https://doi.org/10.1109/TVT.2019.2900460>

Bibliographie

- [26] Dobre, O. A., Abdi, A., Bar-Ness, Y., & Su, W. (2007). Survey of automatic modulation classification techniques: Classical approaches and new trends. *IET Communications*, 1(2), 137–156.
- [27] Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications*, 23(2), 201–220.
- [28] O’Shea, T. J., & Corgan, J. (2016). Convolutional radio modulation recognition networks. In *Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN)*.
- [29] Lamnabhi-Lagarrigue, F. (2011). *Prospective en automatique HORIZON 2020* (Thèse de doctorat, L2S, CNRS, Supélec, Université Paris-Sud).
- [30] Azzouz, E. E., & Nandi, A. K. (1996). *Automatic modulation recognition of communication signals*. Springer.
- [31] Durnerin, M. (1999). *Une stratégie pour l’interprétation en analyse spectrale : Détection et caractérisation des composantes d’un spectre* (Thèse de doctorat, Institut National Polytechnique de Grenoble).
- [32] Haddad, M. T. (2014). *Anticipation des crises d’épilepsie temporale combinant des méthodes statistiques et non-linéaires d’analyse d’électroencéphalographie* (Thèse de doctorat, Université du Québec en Outaouais).
- [33] Jeanpierre, L. (2002). *Apprentissage et adaptation pour la modélisation stochastique de systèmes dynamiques réels* (Thèse de doctorat, Université Henri Poincaré-Nancy I).
- [34] Harispe, S. (2024). *Exploitation de connaissances formelles dans des processus décisionnels automatisés et apprentissage de modèles prédictifs* (Thèse de doctorat). Université de Montpellier.
- [35] Bollaert, M., & Coppin, G. (2024). Vers l’XAI sémantique : la troisième vague de l’intelligence artificielle explicable. *OpenScience.fr*, 12(3), 48-65.
- [36] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4e éd.). Pearson.
- [37] Mitchell, T. M. (1997). Artificial neural networks. *Machine Learning*, 45(81), 127.

Bibliographie

- [38] EDC Paris Business School. (2024, 12 mai). *Deep learning : définition, applications et avantages*. <https://www.edcparis.edu/fr/blog/deep-learning-definition-applications-et-avantages> (consulté mars 2026).
- [39] Bollaert, M., & Coppin, G. (2024). Vers l'XAI sémantique : la troisième vague de l'intelligence artificielle explicable. *OpenScience.fr*, 12(3), 48-65.
- [40] Bounia, L. (2023). *Modèles formels pour l'IA explicable : des explications pour les arbres de décision* (Thèse de doctorat). Université d'Artois.
- [41] Molnar, C. (2022). *Interpretable Machine Learning*. Lulu.com.[consulté mars 2026].
- [42] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.
- [43] Alvarez-Melis, D., & Jaakkola, T. (2018). On the robustness of interpretability methods.
- [44] Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post-hoc explanation methods.
- [45] Shapley, L. (1953). A value for n-person games. In *Contributions to the Theory of Games* (Vol. II). Princeton University Press.
- [46] Von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.
- [47] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- [48] Lundberg, S. M., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, 252-262.
- [49] Aryayi, M., Benzinou, A., Nasreddine, K., Foulon, V., Borremans, C., & Zeppilli, D. Développement d'un modèle d'apprentissage profond guidé par la connaissance pour l'identification de nématodes marins.
- [50] Selvaraju, R. R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618-626. <https://litora.io/grad-cam> (consulté mars 2026).

Bibliographie

- [51] Marchadour, W. (2024). *De la boîte noire à la boîte grise : évaluation de l'interprétabilité des réseaux de neurones profonds en classification d'images médicales* (Thèse de doctorat). Université de Bretagne Occidentale.
- [52] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
- [53] Ross, A. S., Hughes, M. C., & Doshi-Velez, F. (2017). Right for the right reasons: Training differentiable models by constraining their explanations. *arXiv preprint arXiv:1703.03717*.
- [54] O'Shea, T. J., & West, N. (2016). *Radio machine learning dataset generation with GNU radio. Proceedings of the GNU Radio Conference, 1*(1), 1.
<https://pubs.gnuradio.org/index.php/grcon/article/view/11> [consulté mai 2026]
- [55] Wu, S., Wang, C., Liang, J., et al. (2025). A transformer-based framework with complex-valued convolution and enhanced Bi-LSTM for automatic modulation recognition. *Physical Communication*, 102824.
- [56] Ke, Y., et al. (2025). GIGNet: A graph-in-graph neural network for automatic modulation recognition. *IEEE Transactions on Vehicular Technology*, 74, 10058–10062.
- [57] Zhang, L., et al. (2025). Automatic modulation recognition of radio frequency proximity sensor signals based on adaptive relational graph attention network. *IEEE Transactions on Instrumentation and Measurement*, 74, 1–13.
- [58] Qin, X., et al. (2025). Multilevel adaptive wavelet decomposition network-based automatic modulation recognition: Exploiting time-frequency multiscale correlations. *IEEE Transactions on Cognitive Communications and Networking*, 11, 3218–3231.
- [59] Feng, Y., et al. (2025). A dual-branch network with feature assistance for automatic modulation recognition. *IEEE Signal Processing Letters*, 32, 701–705.
- [60] Li, Y., Shi, X., Tan, H., Zhang, Z., Yang, X., & Zhou, F. (2025). Multi-representation domain attentive contrastive learning based unsupervised automatic modulation recognition. *Nature Communications*, 16.
- [61] El-Haryqy, N., Kharbouche, A., Ouamna, H., et al. (2025). Improved automatic modulation recognition using deep learning with additive attention. *Results in Engineering*, 26, 104783.