RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ YAHIA FARÈS DE MÉDÉA



FACULTÉ DE TECHNOLOGIE

DÉPARTEMENT DE GÉNIE ÉLECTRIQUE

MÉMOIRE DE FIN D'ÉTUDES DE MASTER

FILIÈRE : TÉLÉCOMMUNICATIONS

SPÉCIALITÉ : SYSTÈME DES TÉLÉCOMMUNICATIONS

DÉTECTION D'ANOMALIES ET DIAGNOSTIC POUR L'OPTIMISATION DU TRAFIC ET LA GESTION DE LA CONGESTION DANS LES RÉSEAUX CELLULAIRES

PRÉSENTÉ PAR :

HAFSI MERIEM

ZEMIR RANIA

PROPOSÉS ET DIRIGÉS PAR :

MR. LACHENANI SID AHMED

PROMOTION DE

2024/2025

DÉDICACE

Avant tout, je tiens à remercier le bon Dieu, l'unique qui m'offre le courage et la volonté nécessaires pour affronter les différentes difficultés de la vie. Je dédie ce modeste travail :

À mes très chers parents, source de vie, d'amour et d'affection. Mon père **Hacene** et ma mère **Amina** , que je n'arriverai jamais à exprimer ma gratitude et ma reconnaissance, pour leurs amours, leurs soutiens tout au long de mes études, leurs prières et leurs encouragements.

Aucune dédicace ne pourra compenser leur sacrifice pour mon bien-être et mon bonheur. À la famille **Zemir** , surtout mes frères et sœurs : **Ilham et son marie zakaria** et **Walid et sa femme Rachida**,et **Karim** sources de joie et de bonheur.

A mes petits princes **Djawed** et **Nizar** À mes cousins :**Mohamed et sa femme Dalia** et ses enfatns surtout **Samy et Raouf** ,cousin **Amine** et cousine **ikram**.

À tous mes amis et collègues. En particulier à mes amis qui n'ont jamais manqué de me soutenir dans mon travail **Meriem, Dounia, Khaoula, Ibtihel, Rajaa, Hadia, Manel et sa soeur Assia , Feriel, Mahrez, et Remzi.**

À notre encadreur **Mr. Lachenani Sid Ahmed** pour son immense patience, sa grande disponibilité et ses conseils qui ont grandement contribué à la réalisation de ce travail. Qu'il trouve ici l'expression de ma profonde gratitude.

À vous, cher lecteur. Et à tous ceux qui nous ont soutenus pour l'achèvement de ce projet avec réussite.

- Rania

DÉDICACE

Je rends grâce à DIEU pour la force et la patience qu'Il m'a accordées tout au long de ce travail.

À moi-même

Merci pour ta force et ta persévérance, pour avoir surmonté la peur et la fatigue sans jamais abandonner. Ce travail est un hommage à ton courage et à ta détermination

À ma très chère mère

Tu es tout pour moi. Aucun mot ne peut exprimer ma reconnaissance. Chaque réussite est d'abord la tienne. Que Dieu te garde pour moi.

À mon très cher père

Pour ta dignité et ton appui constant. Tu es ma fierté, mon père bien-aimé.

À mes chers sœurs

À ceux qui ont toujours été mon soutien et partagé avec moi joies et peines, vous faites partie intégrante de ma vie. Merci de tout cœur pour chaque moment passé ensemble : **Sara**, **Nour El Houda**, **Hayat**.

À ma précieuse compagne de route

Ta présence a illuminé mon chemin, Ton soutien m'a portée dans les moments les plus durs. Tu es bien plus qu'une amie, tu es une part de mon âme **Lina**.

À mes amies et à tous ceux que j'aime et qui m'aiment

Du fond de mon cœur, je vous adresse un battement de gratitude, Vous êtes la joie du voyage et les compagnons d'âme irremplaçables : **Mahrez**,**Abd ellah**, **Imane**, **Rania**, **Sara**.

À notre encadreur

Mr. Lachenani Sid Ahmed pour son immense patience, pour ses conseils qui ont grandement contribué à la réalisation de ce travail. Qu'il trouve ici l'expression de ma profonde gratitude.

À vous, cher lecteur.

- Meriem

REMERCIEMENTS

Avant tout, nous remercions **DIEU**, le tout puissant, qui nous donne la volonté, la santé et la patience durant ces longues années d'étude afin que nous puissions Réaliser nos objectifs.

Tout d'abord, nous tenons à exprimer nos profonde gratitude à notre cher professeur et encadrant **Mr. LACHENANI Sid Ahmed** pour son suiviet pour son énorme soutien, qu'il n'a cessé de nous prodiguer tout au Lang de la période du projet.

nous remercions également la societé Djezzy pour nous accueillis et nous ont fourni diverses informations sur la plupart d'installations et services qu'ils offrent et spécialement Madame **BOUMELLAH Amina** .

Enfin, nos remerciements à tous ceux qui ont contribué de près ou de loin au bon déroulement de ce projet.

A tous mes amis de la section télécommunication, promos **2020-2025**.

Merci à tous

ملخص

يستعرض هذا البحث تطبيق تقنيات التعلم العميق في مهام كشف الاختلالات وتصنيف جودة الخدمة داخل شبكات الجيل الخامس. تم تطوير عدة نماذج، من بينها النماذج الكلاسيكية الخاضعة للإشراف، والشبكات العصبية الاصطناعية(ANN) ، وشبكات الأعصاب الالتفافية(CNN) ، وكذلك الشبكات العصبية المتكررة .(RNN/LSTM) وقد أُجريت مقارنة منهجية لأداء هذه النماذج.

تم تدريب واختبار هذه النماذج باستخدام قاعدة بيانات مخصصة تحتوي على تسجيلات حقيقية لحركة المرور في الشبكات الخلوية. وقد تم تدريب النماذج على التمييز بين السلوك العادي والسلوك الشاذ، وتصنيف مستويات جودة الخدمة. وقد أظهرت النتائج التجريبية أن الشبكات العصبية الاصطناعية (ANN) قدّمت أفضل أداء بدقة بلغت 98.18\%. كما سجلت نماذج LSTM أداءً جيدًا بدقة بلغت 96.55\%، رغم أنها أقل من .ANN أما نموذج CNN فقد كان الأقل كفاءة بدقة لم تتجاوز 90.32.%\

الكلمات المفتاحية: : 5G، جودة الخدمة(QoS) ، الذكاء الاصطناعي، كشف االاختلالات، DL، التعلم العميق، الشبكات العصبية.

ABSTRACT

This thesis studies the use of deep learning for anomaly detection and QoS classification in 5G networks. Several approaches were analyzed, including classical supervised models as well as advanced architectures such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (LSTM), with a systematic comparison of their performances.

The models were trained on real cellular traffic data to identify anomalies and classify QoS levels.

The results show that the Artificial Neural Network (ANN) achieved the highest overall accuracy of98.18, % followed by LSTM with 96.67 %. In contrast, the CNN approach yielded the lowest performance, with an overall accuracy of 90%

Keywords: 5G networks, Quality of Service, Anomaly detection, Artificial Intelligence, Deep Learning, ANN, CNN, LSTM

RÉSUMÉ

Ce mémoire étudie l'utilisation de l'apprentissage profond pour la détection d'anomalies et la classification de la qualité de service (QoS) dans les réseaux 5G. Plusieurs approches ont été analysées, incluant des modèles supervisés classiques, ainsi que des architectures avancées telles que les réseaux de neurones artificiels (ANN), convolutifs (CNN) et récur- rents (LSTM), avec une comparaison systématique de leurs performances. Les modèles ont été entraînés sur des données réelles de trafic cellulaire afin d'identifier les anomalies et de classifier les niveaux de QoS. Les résultats montrent que le réseau de neurones artificiels (ANN) offre la meilleure précision globale 98.18 %, suivi par le LSTM 96.67 %. En revanche, le CNN a obtenu les performances les plus faibles avec une précision globale de 90 %

Mots-clés : 5G, Réseaux mobiles, QoS, Supervision, Détection d'anomalies, IA, DL, Modèles supervisés, ANN, CNN, LSTM, Classification, Précision, Systèmes intelligents.

TABLE DES MATIÈRES

Dédicas **i**

Remerciements **i**

Abstract **iii**

Résumé **iii**

Table des matières **vi**

Liste des figures **viii**

Liste des tableaux **ix**

Liste des abréviations **x**

Introduction Générale **1**

I Évolution des réseaux mobiles **3**

I.1 Introduction 3

I.2 Les réseaux mobiles de première génération (1G) 3

I.3 Les réseaux mobiles de la deuxième génération (2G) 4

I.3.1 Global System for Mobile Communications 4

I.3.2 General Packet Radio Service (GPRS) 7

I.3.3 Enhaced Data Rates for GSM Evolution (EDGE) 8

I.3.4 Paramètres d'optimisation du réseau Deuxième génération de téléphonie mobile (2G) 8

I.4 Réseau mobile de troisième génération (3G) 10

I.4.1 L'architecture de l'UMTS 10

I.4.2 Techniques et innovations du réseau UMTS 12

I.4.3 Canaux logiques dans l'UMTS 12
I.4.4 Indicateur clé de performances (KPIs) en UMTS 13
I.5 Réseau mobile de quatrième génération (4G) 14
I.5.1 Long Term Evolution(LTE) 14
I.5.1.1 Objectif de Évolution à long terme (LTE) 14
I.5.1.2 Architecture du Réseau LTE 14
I.5.2 Interface LTE 18
I.5.3 Les caractéristiques Principales de la Quatrième génération de téléphonie mobile (4G) 18
I.5.4 Bandes de Fréquences 19
I.5.5 Les types de transmission 19
I.5.6 Gestion de la Mobilité et Qualité de Service (QoS) 19
I.5.7 Optimisation LTE et Analyse des KPI : 20
I.6 Réseau mobile de cinquième génération (5G) 20
I.6.1 Objectifs de la Cinquième génération de téléphonie mobile (5G) . . . 21
I.6.2 Applications visées pour le réseau de 5G 21
I.6.3 Les principales caractéristiques des domaines d'utilisation de la 5G . . 22
I.6.4 Architecture de la technologie 5G 23
I.6.5 L'évolution des bandes de fréquences 26
I.6.6 Les technologies émergentes de la 5G 27
I.6.7 Les interfaces 5G NR 33
I.6.8 Stratégies de déploiement de la 5G 33
I.6.9 la Qualité de Service (QoS) 34
I.6.10 D'optimisation du réseau 5G et analyse KPI 35
I.6.11 Conclusion 35

II Intelligence Artificielle et Deep Learning. 37
II.1 Introduction 37
II.2 Machine Learning 38
II.2.1 Fonctionement de Machine Learning 39
II.2.2 L'utilisation de Machine Learning 39
II.2.3 **Les deffirents types de Machine Learning** 40
II.3 Deep Learning 42
II.3.1 Porquoi le Deep Learning 42
II.3.2 Les Réseaux de neurones 44
II.3.2.1 Neurone biologique 44
II.3.2.2 Mode de fonctionnement du neurone biologique 45
II.4 **Les deffirents types de Deep Learning :** 45
II.4.1 Réseaux neurones artificiels 45
II.4.2 Fonctionnalité des réseaux de neurones artificiels 46
II.4.3 Fonctions D'activation 47

II.4.4 Types d'architectures de réseau de neurones 48
II.4.5 Réseau de neurones à convolution 50
II.4.5.1 Fonctionnement des réseaux de neurones convolutifs : 50
II.4.6 Réseaux de Neurones Récurrents : 51
II.4.7 Mémoire à long terme à court terme 51
II.4.7.1 Architecture du Mémoire à long court terme (LSTM) 52
II.4.7.2 Fonctionnement du LSTM 53
II.4.8 Évaluation des performances : 55
II.4.8.1 Matrice de confusion : 55
II.4.8.2 Métriques d'évaluation : 57
II.5 La Déffirence entre Deep Learning et Machine Learning 58
II.6 Conclusion 59

III Détection d'anomalies dans les réseaux 5G **60**
III.1 Introduction 60
III.2 Motivation 60
III.3 Objectives 61
III.4 Description de la base de données 61
III.4.1 Types de données utilisée 61
III.4.2 Prétraitement des données 62
III.5 Expérimentations et discutions 63
III.5.1 Matlab 63
III.5.2 Résultats de l'apprentissage et du processus d'entraînement du réseau de neurones Artificiel 64
III.5.2.1 Processus d'entraînement du réseau de neurones Artificiel . 64
III.5.2.2 Analyse de la progression de l'entraînement 65
III.5.2.3 Analyse de la matrice de confusion 67
III.5.2.4 Métriques d'évaluation du Réseau de neurones artificiels (ANN) 67
III.5.3 Processus d'entraînement du réseau de neurones convolutif 68
III.5.3.1 Analyse de la progression de l'entraînement 70
III.5.3.2 La matrice de confusion 71
III.5.3.3 Métriques d'évaluation du Réseau de neurones convolutifs (CNN) 72
III.5.4 Processus d'entraînement du réseau de neurones reseau neurons Long Short Term Memory 73
III.5.4.1 Analyse de la progression de l'entraînement 75
III.5.4.2 La matrice de confusion 77
III.5.4.3 Métriques d'évaluation du LSTM 77
III.6 Comparaison détaillée des performances des modèles CNN, ANN et LSTM . . 78
III.6.1 Discussion des resultats 80
III.6.2 Comparaison par nombre époques 80

III.6.3 Comparaison avec des travaux antérieurs 81
III.7 Conclusion 82

Conclusion Générale **83**

TABLE DES FIGURES

I.1 Structure générale d'un réseau GSM [3]. 5
I.2 Architecture des réseaux mobiles GSM et UMTS [14]. 10
I.3 architecture de réseau LTE [18]. 15
I.4 architecture de E-UTRAN. 15
I.5 architecture EPC. 16
I.6 Application de la 5G [25]. 22
I.7 L'architecture de réseau 5G. 23
I.8 Accès radio 5G (NG-RAN). 24
I.9 Architecteur d'un Réseau cœur 5G (5GC) [28]. 25
I.10 Système Massive MIMO [33]. 29
I.11 La technologie Beamforming. 31
I.12 La technologie Full duplex. 32
I.13 L'architecture NSA et SA [38]. 34

II.1 Les sous-catégories de L'IA [39]. 38
II.2 Schéma de DL avec plusieurs couches [44]. 42
II.3 neurone biologique [47]. 44
II.4 Réseaux à une couche (feed-forward). 49
II.5 Réseaux récurrents (feed-back) [50]. 50
II.6 Les réseaux de neurones convolutifs. 51
II.7 Schéma d'une cellule mémoire d'un LSTM. 53
II.8 déffirence entre Deep et Machine Learning. 59

III.1 Évolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau ANN : (a) Accuracy, (b) Loss. 66
III.2 matrice de confusion(ANN) 66
III.3 Mesures de Performance d'un Modèle de Classification ANN. 68

III.4 Evolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau CNN : (a) Accuracy, (b) Loss.. 70
III.5 matrice de confusion(CNN) 71
III.6 Mesures de Performance d'un Modèle de Classification cnn 73
III.7 Evolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau LSTM : (a) Accuracy, (b) Loss. 76
III.8 matrice de confusion (LSTM). 76
III.9 Mesures de Performance d'un Modèle de Classification LSTM. 78
III.10comparaison des Performances 79
III.11Comparaison des précisions moyennes entre notre travail et des travaux antérieurs. 82

LISTE DES TABLEAUX

I.1 Caractéristiques comparatives des types de cellules dans les réseaux mobiles 33

II.1 Les différentes fonctions d'activation 48
II.2 Matrice de confusion 56

III.1 Paramètres d'entraînement du modèle 65
III.2 Paramètres d'entraînement du modèle CNN 69
III.3 Paramètres d'entraînement du modèle LSTM avec descriptions 75
III.4 Précision de validation et nombre d'époques de convergence pour chaque modèle dans une tâche de classification binaire. 81

LISTE DES ABRÉVIATIONS

2G	Deuxième génération de téléphonie mobile
3G	Troisième génération de téléphonie mobile
3GPP	Projet de partenariat de troisième génération
4G	Quatrième génération de téléphonie mobile
5G	Cinquième génération de téléphonie mobile
AMPS	Système avancé de téléphonie mobile
ANN	Réseau de neurones artificiels
AUC	Aire sous la courbe ROC
BSC	Contrôleur de station de base
BSS	Sous-système de station de base
BTS	Station de base émettrice-réceptrice
CDMA	Accès multiple par répartition en code
CNN	Réseau de neurones convolutifs
CN	Réseau cœur
CS	Commutation de circuits
DL	Apprentissage profond
EDGE	Débit de données amélioré pour l'évolution du GSM
EIR	Registre d'identité des équipements
EPC	Cœur de réseau paquet évolué
EPS	Système de paquets évolué
E-UTRAN	Réseau d'accès radio terrestre UMTS évolué
FDMA	Accès multiple par répartition en fréquence
FDD	Répartition en fréquence duplex

GGSN Nœud passerelle de support GPRS

GMSC Centre de commutation mobile passerelle

GPRS Service général de radiocommunications par paquets

GSM Système mondial de communication mobile

gNB Nœud B de nouvelle génération

HLR Registre des abonnés locaux

HO Transfert d'appel ou de session (Handover)

HSPA Accès par paquets à haut débit

HSS Serveur d'abonnés domestiques

IA Intelligence artificielle

IMEI Identité internationale de l'équipement mobile

IMSI Identité internationale de l'abonné mobile

KPI Indicateur clé de performance

LTE Évolution à long terme

LSTM Mémoire à long court terme

MCC Code du pays mobile

MIMO Entrée multiple / sortie multiple

ML Apprentissage automatique

MME Entité de gestion de mobilité

MNC Code de réseau mobile

MSC Centre de commutation des services mobiles

MS Station mobile

MSIN Numéro d'identification de l'abonné mobile

NMT Téléphonie mobile nordique

NSA Mode non autonome

NSS Sous-système de commutation réseau

OFDM Multiplexage par répartition orthogonale de la fréquence

OFDMA Accès multiple par répartition orthogonale de la fréquence

OSS Système de support opérationnel

PCU Unité de contrôle de paquets

PCRF Fonction de règles de politique et de tarification

PDNGW Passerelle de réseau de données à paquets

PGW Passerelle de réseau de données (PDN Gateway)

PS Commutation par paquets

PSK Modulation par déplacement de phase

QoS Qualité de service

RNC Contrôleur de réseau radio

RRC Contrôle des ressources radio

RNN Réseau de neurones récurrents

SA Mode autonome

SC-FDMA Accès multiple par répartition en fréquence à porteuse unique

SGSN Nœud de support GPRS

SGW Passerelle de service

SIM Module d'identité de l'abonné

SMSS Sous-système de commutation et de gestion

TACS Système de communication à accès total

TDMA Accès multiple par répartition dans le temps

TDD Répartition dans le temps duplex

UMTS Système universel de télécommunications mobiles

VLR Registre des abonnés visiteurs

WCDMA Accès multiple par répartition en code à large bande

INTRODUCTION GÉNÉRALE

Depuis leur apparition, les réseaux mobiles ont connu une évolution rapide et profonde, bouleversant notre manière de communiquer et d'accéder à l'information. Cette progression s'est structurée en plusieurs générations technologiques, chacune apportant des avancées majeures en matière de capacité, de rapidité et de qualité des services.

La 2G, introduite au début des années 1990, a marqué un tournant décisif en remplaçant les communications analogiques par des communications numériques. Représentée principalement par la norme GSM (Global System for Mobile Communications), elle a permis d'améliorer significativement la qualité des appels vocaux et la sécurité des transmissions.

Puis est arrivée la 3G, au début des années 2000, qui a introduit les services multimédias et l'accès Internet mobile à haut débit. Grâce à des technologies comme UMTS et HSPA, cette technologie a rendu possible la navigation web, les appels en visioconférence et le streaming, tout en garantissant une meilleure gestion du trafic et des priorités pour assurer une qualité de service adaptée aux besoins des utilisateurs.

Avec l'émergence de la 4G à partir de 2009, une nouvelle ère s'est ouverte pour les réseaux mobiles, marquée par une nette amélioration des débits, de la capacité et de la réactivité du réseau. Basée principalement sur la technologie LTE (Long Term Evolution), la 4G a permis d'atteindre des vitesses de téléchargement de plusieurs centaines de mégabits par seconde, rendant possibles des usages gourmands en bande passante comme la vidéo haute définition, les jeux en ligne en temps réel ou encore le cloud computing. En parallèle, la latence a été significativement réduite, améliorant l'expérience utilisateur dans les applications interactives. La 4G a aussi permis une meilleure prise en charge du nombre croissant d'appareils connectés, un élément clé pour l'émergence de l'Internet des objets (IoT). En adoptant un cœur de réseau entièrement basé sur le protocole IP, elle a favorisé l'intégration fluide des services voix et données, tout en renforçant la qualité de service (QoS) grâce à une gestion plus intelligente et flexible des ressources radio.

La cinquième génération de réseaux mobiles. la 5G, représente une avancée technologique majeure qui dépasse de loin les capacités de ses prédécesseurs. Lancée pour répondre aux exigences croissantes de connectivité, elle repose sur trois piliers fondamentaux : **les dé-**

bits ultra-élevés, **la latence extrêmement faible**, **et la connectivité massive**. Grâce à des technologies comme les ondes millimétriques, le Massive MIMO (Multiple Input Multiple Output).

La 5G offre des vitesses pouvant atteindre 10 Gbps avec une latence inférieure à une milliseconde, permettant des usages critiques comme la réalité augmentée, la chirurgie à distance ou les véhicules autonomes. Elle peut connecter massivement des objets, facilitant le déploiement de l'IoT. Grâce au network slicing, elle crée des sous-réseaux virtuels adaptés à chaque besoin (santé, industrie, télécoms). Bien plus qu'une simple amélioration, la 5G redéfinit l'architecture des réseaux mobiles vers une connectivité flexible, intelligente et ultra-fiable. Cependant, la complexité croissante des infrastructures 5G rend leur gestion plus difficile. Les opérateurs doivent faire face à des défis tels que l'augmentation du trafic, la variabilité des conditions réseau, et la prolifération des appareils connectés. Ces facteurs peuvent entraîner des dégradations de la QoS et des anomalies de fonctionnement difficiles à détecter avec les approches traditionnelles.

Dans ce contexte, l'intelligence artificielle (IA) et plus particulièrement le deep learning apparaît comme une solution prometteuse pour assurer une surveillance proactive du réseau et une gestion intelligente des ressources. En analysant en temps réel les données issues du réseau, les modèles d'apprentissage profond permettent de détecter automatiquement les anomalies, de prédire les défaillances potentielles et de classer les niveaux de qualité de service. Ces approches permettent une adaptation dynamique aux conditions réelles du réseau, tout en réduisant l'intervention humaine.

Dans ce projet, nous explorons l'utilisation de différentes approches d'apprentissage automatique et profond pour détecter les anomalies et évaluer la qualité de service dans les réseaux 5G. Nous proposons une architecture basée sur des données simulées issues de réseaux cellulaires, en mettant en œuvre des techniques avancées telles que les réseaux de neurones artificiels (ANN), les réseaux de neurones convolutifs (CNN), les réseaux de neurones récurrents (RNN/LSTM). Ces modèles sont évalués selon leur capacité à distinguer les comportements normaux des dysfonctionnements, et à classifier les niveaux de QoS,a la fin, une comparaison systématique de ces approches est également réalisée afin d'identifier les plus performantes selon différents critères d'efficacité et de précision.

Ce mémoire est structuré comme suit :

- Le **premier chapitre** présente l'évolution des réseaux mobiles avec ses différentes architectures, en soulignant les apports de la 5G en terme de qualité de service.
- Le **deuxième chapitre** décrit les principes de l'intelligence artificielle, de l'apprentissage automatique et profond , des techniques de deep learning utilisées dans la surveillance réseau.
- Le **troisième chapitre**, cœur de notre travail, détaille les modèles développés, les expérimentations menées et l'évaluation des performances des différentes approches pour la détection d'anomalies et l'analyse de la qualité de service dans les réseaux 5G, ainsi que la comparaison entre eux et avec d'autres travaux réalisés.

CHAPITRE I

ÉVOLUTION DES RÉSEAUX MOBILES

I.1 Introduction

L'évolution des réseaux mobiles a été classée en différentes générations, chacune représentant une avancée technologique majeure, permettant de différencier les capacités et les services offerts par chaque technologie. Depuis leur apparition, les réseaux mobiles ont connu une évolution rapide et spectaculaire, transformant profondément notre manière de communiquer, de travailler et d'accéder à l'information. Chaque génération technologique, de la 1G à la 5G, a apporté des avancées majeures en termes de performance, de couverture et de services.
Ce chapitre explore les architectures et les protocoles des réseaux mobiles 1G, 2G, 3G, 4G (LTE) et 5G. offrant ainsi une vue d'ensemble de leur conception et de leur optimisation. Comprendre cette progression permet non seulement d'apprécier les techniques relevées, mais aussi mieux anticiper les révolutions numériques à venir.

I.2 Les réseaux mobiles de première génération (1G)

La première génération de réseaux mobiles, connue sous le nom de **1G**, a marqué un tournant décisif dans l'histoire des communications. Apparue à la fin des années 1970 et déployée plus largement au début des années 1980, elle a permis aux utilisateurs de passer des appels en déplacement, une révolution pour l'époque.

Cette avancée technologique a été rendue possible grâce au développement rapide des semiconducteurs et des microprocesseurs, qui ont permis de concevoir des téléphones plus compacts, plus légers, et surtout plus accessibles au grand public.

Les réseaux 1G reposaient sur la transmission analogique de la voix, sans prise en charge des données. Plusieurs standards se sont imposés selon les régions du monde, notamment Système avancé de téléphonie mobile (AMPS) aux États-Unis, Téléphonie mobile nordique

(NMT) dans les pays nordiques et Système de communication à accès total (TACS) au Royaume-Uni.

L'arrivée de ces systèmes a rencontré un succès rapide : on comptait déjà près de 20 millions d'abonnés au début des années 1990. Toutefois, cette première génération présentait aussi ses limites. D'une part, le nombre de communications simultanées dans une même cellule restait très faible. D'autre part, chaque pays ayant adopté ses propres normes, les réseaux étaient incompatibles d'une région à l'autre, ce qui freinait l'interopérabilité [1].

I.3 Les réseaux mobiles de la deuxième génération (2G)

Le lancement des réseaux mobiles de **deuxième génération (2G)** a marqué une étape majeure dans l'évolution des communications sans fil. Apparue au début des années 1990, cette génération succède à la 1G en introduisant une innovation fondamentale : le passage de la transmission analogique à la transmission **numérique**. Cette transition a permis d'améliorer significativement la qualité des appels vocaux, la sécurité des communications, et d'optimiser l'utilisation du spectre radioélectrique.

La 2G a également permis une gestion plus efficace des ressources radio, offrant une plus grande capacité réseau et réduisant les coûts pour les opérateurs et les utilisateurs. Elle a ouvert la voie à de nouveaux services tels que les **SMS** (Short Message Service), devenus rapidement populaires à l'échelle mondiale.

Parmi les technologies emblématiques de la 2G et ses évolutions, on distingue [2] :

- Système mondial de communication mobile (GSM) : Standard 2G le plus répandu, il permet des appels vocaux fiables et introduit les SMS, utilisant le multiplexage temporel (TDMA).
- Service général de radiocommunications par paquets (GPRS) : Technologie 2.5G apportant la transmission de données en mode paquet, avec un accès mobile à Internet à faible débit (jusqu'à 114 kbit/s).
- Débit de données amélioré pour l'évolution du GSM (EDGE) : Amélioration du GPRS (2.75G), augmentant les débits jusqu'à 384 kbit/s, permettant des services multimédia plus riches tout en restant compatible avec le réseau GSM.

Grâce à ces évolutions, la 2G a rendu la téléphonie mobile plus accessible, plus performante et a amorcé l'ère de la connectivité mobile, ouvrant la voie aux générations futures.

I.3.1 Global System for Mobile Communications

Dans cette section, nous présentons l'architecture générale du réseau GSM.

a) Mobile Station (MS) : La station mobile est l'équipement que l'utilisateur emploie pour accéder au réseau GSM. Elle se compose de deux éléments :

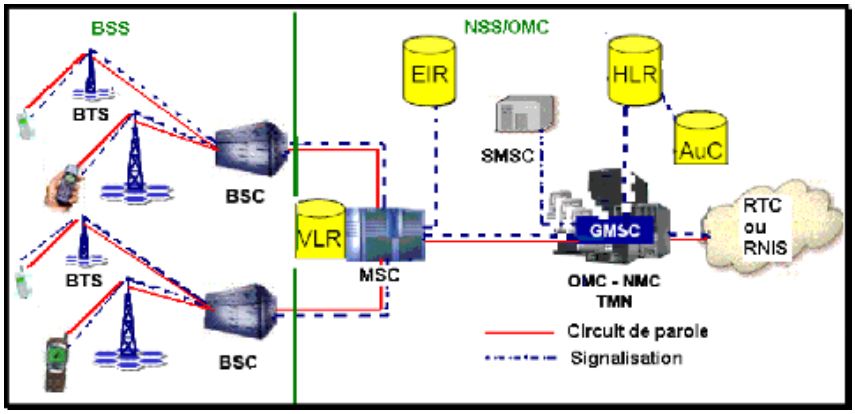


FIGURE I.1 – Structure générale d'un réseau GSM [3].

- **Le terminal mobile** : identifiée par un numéro unique appelé Identité internationale de l'équipement mobile (IMEI), cette partie correspond au téléphone lui-même.
- **La carte Module d'identité de l'abonné (SIM)** : cette carte sécurisée contient l'identité de l'abonné (Identité internationale de l'abonné mobile (IMSI)) et ses données d'authentification. L'IMSI, défini par la norme E212, se compose de :
  - **Code du pays mobile (MCC)** (Mobile Country Code) – 3 chiffres indiquant le pays (ex. **213** pour l'Algérie),
  - **Code de réseau mobile (MNC)** (Mobile Network Code) – identifiant l'opérateur (ex. **07** pour Djezzy),
  - **Numéro d'identification de l'abonné mobile (MSIN)** (Mobile Subscriber Identification Number) – numéro unique attribué à chaque abonné.

b) Le sous-système radio Sous-système de station de base (BSS)

Le sous-système radio (BSS) est une composante essentielle du réseau GSM, responsable de la gestion des communications radio entre les stations mobiles Station mobile (MS) et le réseau central. Il se compose principalement de deux éléments :

- **La station de base Station de base émettrice-réceptrice (BTS)** :
  Ce sont des émetteurs-récepteurs dotés d'un minimum de capacités intelligentes. Ils assurent la transmission et la réception des informations via le canal radio, en fournissant une interface physique entre la station mobile et le BSC.

  La BTS assure plusieurs fonctions essentielles , détaillées dans les points suivants [4] :
  - Mesures sur la liaison montante (up-link), servant à l'algorithme de décision du handover.
  - Détection des messages de *handover access* (HO ACCESS).
  - Détection des demandes d'accès des mobiles reçues sur le canal de contrôle commun (RACH).

- Contrôle de la puissance d'émission : La BTS ajuste dynamiquement la puissance d'émission des stations mobiles (MS) pour optimiser la qualité du signal, réduire les interférences et prolonger la durée de vie des batteries [5].

- **La station de contrôleContrôleur de station de base (BSC)**
  Le BSC gère une ou plusieurs stations de base et communique avec elles par le biais de l'interface A-bis (organe "intelligent" du BSS). Le BSC assure différentes fonctions tant au niveau de la communication qu'au niveau de l'exploitation.

  Pour les signaux en provenance des stations de base, le BSC agit comme un concentrateur, transférant les communications vers une sortie unique. Dans l'autre sens, il achemine les données vers la BTS appropriée [6].

  Fonctions principales :

  - Allocation des canaux de communication.
  - Traitement des mesures des niveaux d'émission des BTS et des MS.
  - Gestion des handovers intra-BSC.
  - Contrôle d'admission des appels.

c) Le sous-système réseau Sous-système de commutation réseau (NSS)
Le sous-système d'acheminement, appelé NSS ou Sous-système de commutation et de gestion (SMSS), comprend l'ensemble des fonctions nécessaires pour l'établissement des appels et la gestion de la mobilité [7]. Il se compose de plusieurs équipements :

- Centre de commutation des services mobiles (MSC)
  Le MSC est l'élément central du NSS. En s'appuyant sur les informations du HLR et du VLR, il gère l'établissement des appels, la commutation et le codage.

- Registre des abonnés visiteurs (VLR)

  Le VLR est une base de données temporaire contenant les informations des abonnés présents dans une zone donnée. Ces données sont obtenues du HLR. Souvent, le VLR est co-localisé avec le MSC, pour une gestion synchronisée des zones géographiques [8].

- Registre des abonnés locaux (HLR)

  Le HLR est une base de données principale contenant la localisation et les caractéristiques des abonnés. Contrairement au VLR, il stocke les informations de manière permanente. Le format de l'identifiant MSISDN est :

  $$\texttt{MSISDN = CC / NDC / SN}$$

  - CC : *Country Code* (ex. Algérie = 213)
  - NDC : *National Destination Code*, indicatif national sans le zéro initial
  - SN : *Subscriber Number*, numéro de l'utilisateur [5].

- Aire sous la courbe ROC (AUC)

  L'AUC est associé au HLR et contient une clé secrète unique pour chaque abonné mobile. Elle est utilisée pour :

  - Authentifier l'abonné dans le réseau GSM.
  - Générer une clé de chiffrement (Kc) pour sécuriser les appels.

- Registre d'identité des équipements (EIR)

  L'EIR est une base de données contenant les identifiants des équipements mobiles. Elle permet de restreindre l'accès au réseau aux appareils non autorisés ou volés. Trois listes y sont maintenues [9] :

  - Liste blanche : appareils autorisés.
  - Liste grise : appareils sous surveillance.
  - Liste noire : appareils volés ou non conformes.

- Centre de commutation mobile passerelle (GMSC)

  Le GMSC est le premier point de passage pour les appels entrants vers un mobile, même si sa position exacte n'est pas connue. Il récupère le MSRN (Mobile Station Roaming Number) via le HLR, à partir du MSISDN, pour acheminer l'appel vers le MSC approprié.

d) Sous-système d'exploitation et de support Système de support opérationnel (OSS) : Le OSS est essentiel pour la gestion des réseaux de télécommunications. Il supervise l'ensemble du réseau GSM, contrôle la charge de trafic du BSS et prend en charge les opérations matérielles ainsi que la facturation. L'OSS est connecté au NSS et au BSS.

I.3.2 General Packet Radio Service (GPRS)

Le GPRS est une évolution du réseau 2G existant, amélioré pour prendre en charge les services basés sur la transmission par paquets tout en augmentant les débits de données utilisés par ces réseaux. Le terme 2.5G désigne les systèmes 2G qui ont intégré un domaine à commutation par paquets en complément du domaine à commutation de circuits.

En utilisant la commutation par paquets, le GPRS permet une transmission plus efficace des signaux et des données à haut débit par rapport aux systèmes GSM traditionnels.

Le GPRS offre des débits de données allant de **56 Kbps à 384 Kbps** et intègre des bases de données essentielles telles que le HLR, VLR, EIR et AuC [10].

Un réseau GPRS est un réseau IP, constitué principalement de routeurs IP.
L'introduction de la mobilité nécessite l'ajout de nouvelles entités fonctionnelles, notamment :

- **Nœud de support GPRS (SGSN)** : Ce nœud de service offre une gamme de services liés aux parties IP du système, notamment le routage, le transfert de paquets, la gestion de la mobilité, l'authentification et la gestion des données de charge.

- **Nœud passerelle de support GPRS (GGSN)** : Ce nœud passerelle joue le rôle de jonction entre le réseau mobile et les réseaux de données externes. Il agit comme un routeur, une passerelle, et un pare-feu. Avant de transférer les données, il vérifie la disponibilité de l'utilisateur ciblé.
- **Unité de contrôle de paquets (PCU)** : Connecté au BSC, ce routeur physique convertit les données entre le réseau GSM basé sur la commutation de circuits, et le réseau GPRS basé sur la commutation de paquets.

I.3.3 Enhaced Data Rates for GSM Evolution (EDGE)

L'introduction de la technologie EDGE a permis de pallier certaines limitations du réseau GPRS. Compatible avec les réseaux Accès multiple par répartition dans le temps (TDMA) et GSM, EDGE peut être intégré à toute infrastructure supportant le GPRS, étant considéré comme une extension de ce dernier. Avec un débit théorique allant jusqu'à 500 kbps, EDGE utilise la modulation par décalage de phase Modulation par déplacement de phase (PSK) pour atteindre cette performance. Les améliorations apportées par EDGE incluent une augmentation significative du débit par abonné, passant d'environ 8,8 kbps à un maximum théorique de 384 kbps, grâce à l'introduction de la modulation 8-PSK en remplacement de la GMSK. Ces évolutions permettent également une réduction de la sensibilité au signal, ainsi qu'une amélioration de la capacité et de la couverture du réseau.

En dehors de quelques ajustements matériels mineurs au niveau des stations de base BTS et des mises à jour logicielles du réseau, la mise en œuvre d'EDGE ne nécessite pas de modifications matérielles importantes. Cela en fait une solution économique et efficace pour améliorer les performances des réseaux existants [11].

I.3.4 Paramètres d'optimisation du réseau 2G

Afin d'optimiser les performances des réseaux 2G, des indicateurs appelés **KPI** (Key Performance Indicators) sont utilisés. Ils permettent d'évaluer la qualité du service offert et de repérer les éventuelles défaillances. La section suivante présente les principaux KPI spécifiques au réseau GSM.

a)Indicateurs clés de performance dans le GSM

Les indicateurs clés de performance (KPI) mesurent la performance et la qualité d'un réseau. Ils peuvent être utilisés pour les fonctions suivantes :

- Fournir des informations détaillées aux planificateurs RF.
- Améliorer les performances du réseau radio afin d'offrir un meilleur service aux abonnés.
- Détecter immédiatement les problèmes de performances inacceptables sur un réseau cellulaire.

Généralement, les indicateurs de performance clés (KPI) peuvent être classés comme suit :

- **Retenabilité** : mesure la capacité à maintenir la connexion existante.
- **Accessibilité** : Mesure la capacité à établir un appel.
- **Mobilité** : mesure la capacité à maintenir la connexion de l'utilisateur lors de déplacements sur le réseau.
- **Intégrité** : mesure le niveau de service après que l'utilisateur a accédé au service.

b) Principaux indicateurs clés de performance (KPI) du GSM
Inclure tous les paramètres suivants :

- **Taux de réussite d'établissement d'appel (CCSR)** : mesure du nombre de tentatives d'appel réussies par rapport au nombre total de tentatives. Un appel est considéré comme réussi si le SDCCH est accédé, occupé, puis que le TCH est accédé avec succès. Le CCSR est généralement affecté par un blocage (SDCCH/TCH) ou par des problèmes opérationnels.
- **Le taux d'appels abandonnés (DCR)** : compare le nombre d'appels abandonnés au nombre d'appels établis avec succès. Un appel est confirmé comme abandonné si le TCH est libéré de façon anormale. Du point de vue RF, le DCR dépend du RLT, un compteur basé sur la perte ou la réception des trames SACCH, influencé par l'environnement radio (interférences, faible couverture). L'augmentation des échecs de transfert accroît le DCR, causé par des problèmes RF ou opérationnels. Pour le réduire, le plan de voisinage doit être optimal, sans conflits de fréquences, et les paramètres de transfert et de contiguïté doivent être à leur valeur maximale.
- **Le taux de réussite des transferts intercellulaires (HSR)** : compare le nombre de transferts réussis au total des tentatives. Il existe deux types : HSR sortant et HSR entrant. Le HSR sortant dépend des paramètres de contiguïté et de contrôle de transfert, et peut être affecté par le blocage TCH dans les cellules adjacentes. Le HSR entrant dépend des mêmes paramètres et peut être impacté par le blocage TCH sur la cellule source.
- **Minutes par interruption (MPD)** : nombre moyen de minutes après lesquelles un appel est interrompu sur le réseau. Par exemple, si une cellule a un MPD de 60 minutes, cela implique qu'un appel y reste en moyenne 60 minutes avant d'être interrompu. Le MPD est calculé en divisant les minutes de trafic vocal par le total des interruptions TCH.
- **Blocage brut TCH** : mesure du blocage TCH lors de l'établissement de l'appel. Il n'inclut pas le blocage TCH dû aux tentatives de transfert entrantes, ni le blocage TCH dû aux tentatives de prise TCH par nouvelle tentative dirigée entrante.
- **Blocage SDCCH** : mesure des tentatives de prise SDCCH qui n'ont pas pu être facilitées en raison de l'indisponibilité du SDCCH [12].

I.4 Réseau mobile de troisième génération (3G)

La Troisième génération de téléphonie mobile (3G) est une technologie de téléphonie mobile numérique de troisième génération, succédant aux réseaux 2G comme le GSM en Europe, le PDC et le PHS au Japon, ainsi que le Accès multiple par répartition en code (CDMA) One aux États-Unis et en Corée. Cette technologie utilise des bandes de fréquences plus larges et un protocole de transfert de données par paquets. Considérée comme une avancée majeure dans le domaine de la téléphonie mobile, l'UMTS a été développé au sein du Projet de partenariat de troisième génération (3GPP) avec la participation de nombreux organismes de régulation régionaux et nationaux, garantissant ainsi une interopérabilité et un roaming optimisés. Fondé en décembre 1998, le 3GPP résulte d'une collaboration entre l'Europe, les États-Unis et le Japon pour définir les standards des technologies mobiles de troisième génération sous l'égide de l'Union internationale des Télécommunications (UIT) [13].

I.4.1 L'architecture de l'UMTS

L'ajout des données en mode paquet dans le GPRS a nécessité de nouvelles entités réseau. La conception de réseau Système universel de télécommunications mobiles (UMTS) a été construite sur la fusion de ces deux pièces de réseau. Figure II.3 montre l'architecture UMTS totale avec le réseau GPRS.

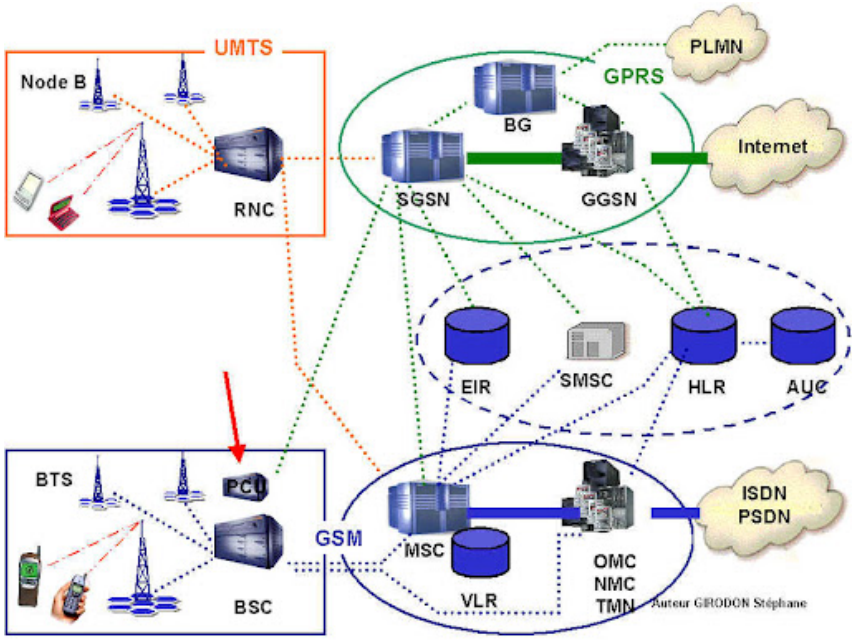


FIGURE I.2 – Architecture des réseaux mobiles GSM et UMTS [14].

La conception du réseau UMTS se compose de trois sous-systèmes principaux :

a) Réseau d'accès radio (UTRAN - UMTS Terrestrial Radio Access Network)

Le sous-système radio se compose de deux éléments distincts, à savoir le Node B et le contrôleur de réseau radio RNC

- **Node B (équivalent BTS en 2G)** : Son rôle principal est d'assurer la réception et la transmission des signaux radio entre l'équipement utilisateur et une ou plusieurs cellules du réseau d'accès UMTS. Le Node B opère au niveau de la couche physique du modèle OSI, où il est responsable des processus de codage et de décodage des données.
- **Contrôleur de réseau radio (RNC)** : Son rôle principal est d'assurer le routage des communications entre le Node B et le cœur de réseau UMTS. Il opère aux couches 2 et 3 du modèle OSI et prend en charge plusieurs fonctions essentielles, notamment :
  - Le contrôle de puissance en boucle externe.
  - La gestion du *handover* (transfert d'une communication d'une cellule à une autre).
  - Le contrôle de l'admission des mobiles au réseau et la gestion de la charge.
  - L'allocation des codes CDMA.
  - Le séquencement de la transmission des données en mode paquet.
  - La combinaison et la distribution des signaux en situation de macro-diversité permettant la gestion des communications entre plusieurs Node B [15].

b) Réseau central (Core Network - CN)

Le cœur du réseau UMTS est composé de deux domaines :

- **Commutation de circuits (CS)** :) est une méthode de transmission dédiée aux services en temps réel tels que la téléphonie, la visiophonie, les jeux en ligne et certaines applications multimédias, où la latence doit être minimale pour garantir une communication fluide. Lors de l'introduction de l'UMTS, ce mode permettait un débit de $384\,\mathrm{kbit\,s^{-1}}$. Il repose sur l'architecture du GSM (MSC/VLR, GMSC), assurant une interconnexion efficace avec les réseaux extérieurs.
- **Commutation de paquets (PS)** : Le mode paquet est utilisé pour les services qui ne nécessitent pas une transmission instantanée. Il est principalement destiné à la navigation sur Internet, à la gestion des jeux en réseau et à d'autres services similaires [16].

Éléments clés du cœur de réseau :

- **MSC** : Gestion des appels et de la mobilité.
- **HLR/VLR** : Gestion des abonnés.
- **SGSN** : Transfert de données mobiles.
- **GGSN** : Connexion du réseau mobile à Internet.

c) Réseau de gestion et exploitation OSS

Ce sous-système surveille et optimise le fonctionnement global du réseau.

I.4.2 Techniques et innovations du réseau UMTS

a) WCDMA (Wideband Code Division Multiple Access)

Le WCDMA est un mode d'accès numérique permettant le partage de la bande de fréquence. Il repose sur la technique d'étalement de spectre, qui attribue des codes uniques aux utilisateurs pour les différencier au sein d'une même bande. Cette technologie offre plusieurs avantages :

- **Capacité élevée** : grâce au grand nombre de codes disponibles, elle permet d'accueillir un grand nombre d'utilisateurs simultanément.
- **Protection contre le brouillage** : l'élargissement du spectre réduit les interférences, améliorant ainsi la qualité du signal.

Cependant, le Accès multiple par répartition en code à large bande (WCDMA) requiert un contrôle de puissance extrêmement précis pour assurer une transmission efficace et éviter les interférences entre utilisateurs [17].

b)Gestion de la mobilité et des handovers

- Handover **intra-système** : entre cellules 3G.
- Handover **inter-système** : passage entre 3G et 2G si nécessaire.

c) Sécurité et authentification

- Chiffrement des communications.
- Authentification basée sur la carte SIM et l'IMSI.

I.4.3 Canaux logiques dans l'UMTS

En fonction des spécificités du service offert, tout service fourni à l'UE est associé à un support radio qui définit la configuration et le réglage du contrôle de liaison radio (RLC), du contrôle d'accès moyen (MAC) et de la couche physique. À différents niveaux de protocole, le flux d'information associé à un signal d'accès radio (RAB) est cartographié dans différents types de canaux. La couche RLC crée des canaux logiques, qui sont ensuite cartographiés pour transporter des canaux dans la couche MAC. Deux types de canaux logiques fournissent les porteurs de plan de commande et de plan d'utilisateur : les canaux de contrôle pour la signalisation et les canaux de circulation pour le porteur. Les canaux utilisés dans le réseau UMTS sont classés en différentes catégories en fonction de leur rôle et de leur direction.

a) Canaux de contrôle

- **Canal de contrôle de radiodiffusion (BCCH)** : transmet des informations système, telles que les séquences de code, l'identifiant de la cellule et les chronomètres.
- **Canal de contrôle des pages (PCCH)** : annonce les appels entrants ou d'autres messages devant être envoyés aux utilisateurs dans la zone de localisation.
- **Canal de contrôle dédié (DCCH)** : canaux de signalisation bidirectionnels utilisés pour les rapports de mesure et les messages de contrôle RRC entre l'UE et la RNC.

- **Canaux de contrôle communs (CCCH)** : utilisés pour la configuration de la connexion, l'affectation du canal et la réélection des cellules. Ils peuvent être utilisés en DL ou UL.

b) Canaux de trafic

- **Chaîne de trafic dédiée (DTCH)** : utilisée pour transférer des informations pour le service spécifique d'un seul utilisateur. Plusieurs services peuvent être fournis à un seul utilisateur sur plusieurs DTCH simultanément, en DL et en UL.
- **Canaux de trafic communs (CTCH)** : comme les messages de diffusion cellulaire SMS, par exemple, c'est un canal point à multipoint qui transporte des informations pour un groupe d'utilisateurs.

c) Canaux de transport

- **Chaîne de diffusion (BCH)** : assure le transport du canal logique BCCH.
- **Canaux de pagination (PCH)** : utilisés pour transporter le canal logique PCCH.
- **Canal d'accès aléatoire (RACH)** : avant qu'un canal de trafic puisse être attribué, des informations de signalement courtes doivent être fournies lors de l'accès initial au système.
- **Envoyer le canal d'accès (FACH)** : un canal logique dédié à une UE spécifique pour la réponse au RACH.
- **Canaux de paquets communs (CPCH)** : le canal RACH a été élargi pour permettre la transmission de plus grands paquets de données [12].
- **Télécharger le canal partagé (DSCH)** : alloue un pool de ressources physiques pour la transmission de paquets, selon une politique de planification spécifique, similaire au GPRS.

I.4.4 KPIs en UMTS

L'évaluation de la performance du réseau UMTS repose largement sur les principaux indicateurs de performance et leurs mesures associées. Les KPI du réseau UMTS sont classés dans les sous-catégories suivantes :

- **Accessibilité** : Ceci est l'efficacité dans laquelle l'utilisateur réussit à se connecter au réseau UMTS et à utiliser les services disponibles.
- **Retenabilité** : Cela se définit comme la capacité d'un utilisateur de garder un service souhaité pour le délai précisé.
- **Mobilité** : Cette caractéristique concerne la capacité d'un utilisateur à se déplacer entre les cellules UMTS et GSM voisines tout en maintenant le service requis, qu'il s'agisse de services de paquets (Commutation par paquets (PS)) ou de services de circuit (Commutation de circuits (CS)).

- **Intégrité** : Représentent la perception de la qualité par l'utilisateur.

Les principaux KPIs du réseau UMTS sont la baisse des appels CS ou PS, l'échec de la configuration de la connexion Contrôle des ressources radio (RRC), la défaillance de la Transfert d'appel ou de session (Handover) (HO) (3G-3G), l'échec de l'IRAT-HO (3G-2G) et le problème de la congestion d'alimentation.

I.5 Réseau mobile de quatrième génération (4G)

À la différence des générations précédentes, la 4G se caractérise par la nécessité d'élaborer de nouvelles normes et standards. Conçus pour répondre aux besoins des utilisateurs finaux, les systèmes 4G offrent une connectivité haut débit continue et une large gamme de services. Cependant, la transition des infrastructures actuelles vers cette quatrième génération représente un défi majeur. De nombreux travaux de recherche ont abordé cette problématique sous différents angles.

Dans cette partie, nous mettrons en lumière le passage de la 3G à la 4G en explorant les technologies WIMAX et LTE, ainsi que les spécificités des techniques qu'elles exploitent.

I.5.1 Long Term Evolution(LTE)

La technologie LTE ou la 4G est une norme de communication mobile développée par le 3GPP . Elle constitue une évolution de la 3G et du Accès par paquets à haut débit (HSPA) .

I.5.1.1 Objectif de LTE

LTE vise à améliorer l'efficacité spectrale et à augmenter la capacité de gestion du nombre de mobiles dans une même cellule. Elle tente aussi d'offrir des débits plus élevés, une réduction de la latence. LTE a été conçu pour améliorer les performances des réseaux mobiles et répondre à la demande croissante de données mobiles.

I.5.1.2 Architecture du Réseau LTE

Les réseaux LTE sont des réseaux cellulaires constitués de milliers de cellules radio qui utilisent les mêmes fréquences hertziennes grâce aux codages radio Accès multiple par répartition orthogonale de la fréquence (OFDMA) dans le sens descendant « Downlink », et le Accès multiple par répartition en fréquence à porteuse unique (SC-FDMA) dans le sens montant « Uplink ».

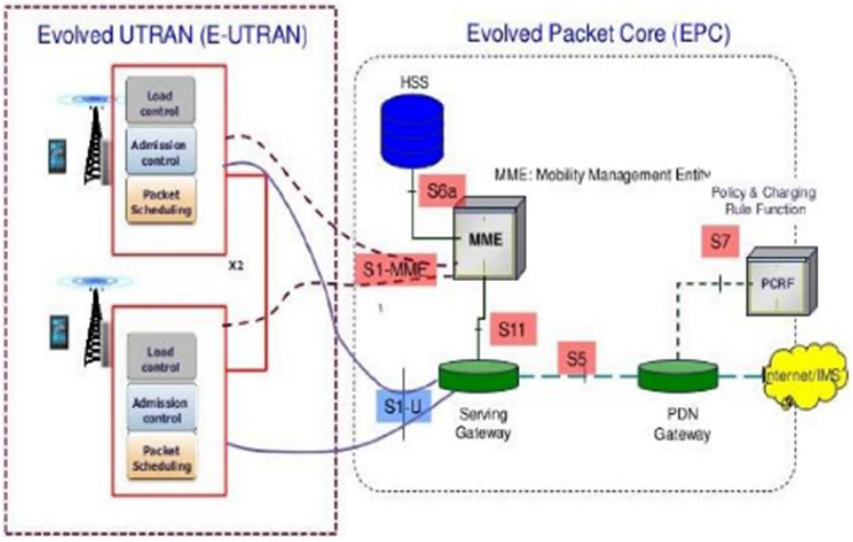


FIGURE I.3 – architecture de réseau LTE [18].

La technologie LTE a apporté une efficacité spectrale, une amélioration de débit, une augmentation de couverture et du nombre d'appels supporté par la cellule. De même que ces précédentes, elle est caractérisée par son architecture connue sous le nom d'Système de paquets évolué (EPS) , qui comporte :

- Un réseau d'accès : l'Réseau d'accès radio terrestre UMTS évolué (E-UTRAN) .
- Un réseau cœur : l'Cœur de réseau paquet évolué (EPC) [19].

- **Le Réseau d'accès (e-UTRAN)** : La partie G, par l'intégration des fonctions de contrôle dans les stations de base « eNodeB », qui étaient auparavant implémentées dans les RNC des réseaux 3G UMTS. La partie radio d'un réseau LTE (voir Figure II.5), se compose donc des eNodeBs, d'antennes locales ou distantes, de liaisons en fibres optiques vers les antennes distantes et des liens IP reliant les eNodeBs entre eux (liens X2) et avec le cœur de réseau.

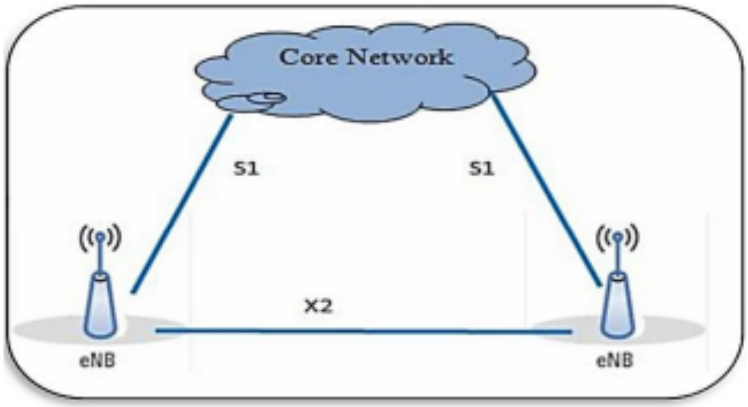


FIGURE I.4 – architecture de E-UTRAN.

- **eNodeB (Evolved Node B)** : L'eNodeB est l'équivalent de la BTS dans le réseau GSM et le NodeB dans l'UMTS. La fonctionnalité de Handover est plus robuste dans le LTE. Ce sont des antennes qui relient les UE (User Equipment) avec le réseau cœur du LTE via les RF (Radio Frequency) air interface. La fonctionnalité du contrôleur radio réside dans eNodeB, le résultat est plus efficace, et le réseau est moins latent, par exemple la mobilité est déterminée par eNodeB à la place du BSC ou RNC des réseaux GSM et l'UMTS.
- **EPC (Evolved Packet Core)** : Le cœur de réseau appelé EPC, utilise des technologies Tout-IP, c'est-à-dire basées sur les protocoles Internet pour la signalisation, le transport de la voix et des données. Ce cœur de réseau permet l'interconnexion via des routeurs avec les autres eNodeBs, les réseaux des autres opérateurs mobiles, les réseaux de téléphonie fixe et le réseau Internet. Il assure la gestion des utilisateurs, la gestion de la mobilité, la gestion de la qualité de service QoS et la gestion de la sécurité, au moyen des équipements tels que le MME, le SGW, le Passerelle de réseau de données à paquets (PDNGW) et le Fonction de règles de politique et de tarification (PCRF) [20]. L'utilisation du protocole IP de bout en bout dans le cœur du réseau, permet des temps de latence réduits pour l'accès internet et les appels vocaux LTE.

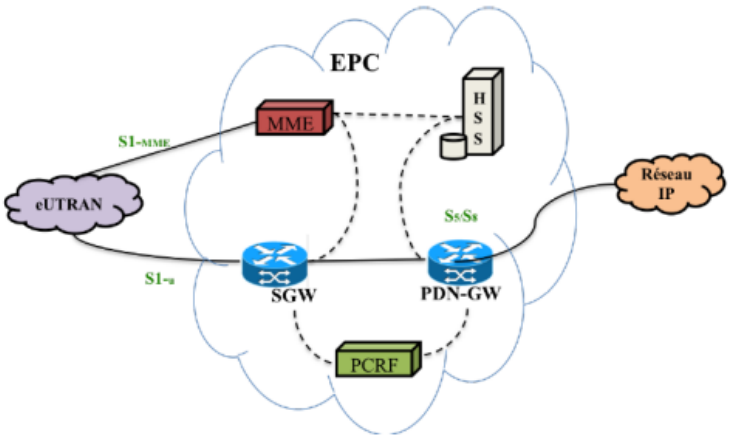


FIGURE I.5 – architecture EPC.

● **Fonctionnement du Cœur de Réseau EPC**

- **MME (Mobility Management Entity)** : C'est une entité de gestion de mobilité, le Entité de gestion de mobilité (MME) c'est le nœud principal de contrôle du réseau d'accès LTE/SAE, elle gère toutes les procédures (authentification, chiffrement, mobilité.) des UE.

  Les fonctions de l'entité MME incluent :

  – **Signalisation « EMM » et « ESM » avec l'UE** : Les terminaux LTE disposent des protocoles EMM et ESM qui leur permettent de gérer leur mobilité (atta-

chement, détachement, mise à jour de localisation) et leur session (établissement/libération de session de données) respectivement. Ces protocoles sont échangés entre l'UE et le MME.

- **Authentification** : Le MME est responsable de l'authentification des UEs à partir des informations recueillies du HSS.
- **Gestion de la liste de « Tracking Area »** : L'UE est informé des zones de localisation prises en charge par le MME, appelées Tracking Area. L'UE met à jour sa localisation lorsqu'il se retrouve dans une Tracking Area qui n'est pas prise en charge par son MME.
- **Sélection du « SGW » et du « PDN GW »** : C'est au MME de sélectionner le Serving GW et le PDN GW qui serviront à mettre en œuvre le Default Bearer au moment du rattachement de l'UE au réseau.
- **Sélection du MME lors du Handover avec changement de MME** : Lorsque l'usager est dans l'état ACTIF et qu'il se déplace d'une zone prise en charge par un MME à une autre zone qui est sous le contrôle d'un autre MME, alors il est nécessaire que le Handover implique l'ancien et le nouveau MME.
- **Roaming avec interaction avec le HSS nominal** : Lorsque l'usager se rattache au réseau, le MME s'interface au HSS nominal afin de mettre à jour la localisation du mobile et obtenir le profil de l'usager [21].

- **Passerelle de service (SGW)** : C'est la jonction principale entre le réseau radio accès et le réseau cœur SGW . Elle achemine les paquets de données, maintient la connexion de l'inter-eNodeB Handover, puis inter-système Handover entre LTE et GSM/UMTS. L'échange des paquets est acheminé par le SGW au PDN-GW par l'interface S5. Le SGW est connecté à l'E-UTRAN via l'interface S1-U qui sert de relai entre l'utilisateur et l'EPC.
- **Passerelle de réseau de données (PDN Gateway) (PGW)** : Le P-GW est le nœud qui relie l'utilisateur mobile aux autres réseaux PDN, tels que les réseaux IP, PSTN et NON-3GPP. L'accès aux réseaux IP et PSTN se fait par l'intermédiaire de l'IMS. Le PDN Gateway agit comme un routeur par défaut par lequel transitent les requêtes de l'utilisateur. Il effectue l'allocation d'adresses IP pour chaque Terminal Mobile, le filtrage des paquets pour chaque usager et comptabilise les octets échangés dans la session de ce dernier à des fins de facturation.
- **Serveur d'abonnés domestiques (HSS)** : Le HSS se présente comme une version évoluée du HLR, il permet de stocker des informations d'abonnement pouvant servir au contrôle des appels et à la gestion de session des utilisateurs réalisé par le MME. Il entrepose, pour l'identification des utilisateurs, la numérotation et le profil des services auxquels ils sont abonnés. En plus des données d'authentification des utilisateurs, il contient les informations de souscription pour les autres réseaux, comme le GSM, le GPRS, 3G, LTE et IMS.

- **PCRF** : Le PCRF est une entité qui exécute principalement deux grandes tâches :
  - La première est de gérer la qualité de service Qualité de service (QoS) que requiert le réseau et alloue en conséquence les porteuses Bearer appropriées.
  - La deuxième tâche se rapporte principalement à la tarification. En effet, le PCRF gère les politiques de facturation qui doivent être prises en compte par le PDN-GW et applicables en fonction des actions de l'utilisateur.

I.5.2 Interface LTE

Dans l'architecture LTE, plusieurs interfaces assurent la communication entre les différentes entités du réseau. Chacune de ces interfaces remplit des fonctions spécifiques, qu'il s'agisse du transport de données, de la signalisation de contrôle, ou encore de la gestion de la mobilité. Les principales interfaces sont décrites ci-dessous :

- LTE-Uu : Assurant la liaison entre les équipements utilisateur (UE) et les stations de base (eNodeBs), le LTE-Uu est chargé de gérer tous les échanges de signalisation entre le MME et le eNodeB, ainsi que du transfert de données entre le S-GW et l'UE.

- S1 : L'interface S1 relie l'E-UTRAN et l'EPC pour les besoins de contrôle et d'utilisateur. Elle se compose de deux parties distinctes : la S1-AP (GTP-AP) pour le contrôle, et la S1U (GTP-U) pour les données utilisateur.

- X2 : Facilitant la connexion entre deux ou plusieurs eNodeBs, l'interface X2 se divise en deux parties : X2-C pour le contrôle inter-eNodeBs, et X2-U pour les données utilisateur. La structure des interfaces X2-C et X2-U est similaire à celle de l'interface S1.

- S5 : Connecte S-GW et P-Gw, il fournit la gestion de tunnels et de tunnel à l'utilisateur. Il est basé sur le protocole GTP et est utilisé pour connecter un P-GW non collocé pour la connectivité PDN, et le déplacement de S-GW en raison de la mobilité.

- SGi : Il est responsable de la connexion du P-GW au réseau de données de paquets. En tant que fourniture pour les services IMS, un réseau de données de paquets peut être le réseau public ou privé externe d'un opérateur ou un réseau intra-opérateur.

I.5.3 Les caractéristiques Principales de la 4G

La quatrième génération de réseaux mobiles, principalement basée sur la technologie LTE, introduit plusieurs innovations majeures par rapport aux générations précédentes. Ces caractéristiques permettent d'offrir une connectivité plus rapide, plus stable et plus adaptée aux nouveaux usages numériques. Les principaux avantages techniques de la 4G sont les suivants :

- **Haute vitesse de transmission** : LTE permet des débits allant jusqu'à 100 Mbps en downlink et 50 Mbps en uplink.
- **Faible latence** : Temps de réponse réduit (10 ms), améliorant l'expérience utilisateur.
- **Efficacité spectrale** : Utilisation optimale des bandes de fréquences.

- **Support du MIMO (Multiple Input Multiple Output)** : Amélioration du débit et de la robustesse de la connexion.
- **Interopérabilité** avec les réseaux précédents (2G, 3G, HSPA) pour une transition fluide.

I.5.4 Bandes de Fréquences

LTE fonctionne en Répartition en fréquence duplex (FDD) et Répartition dans le temps duplex (TDD) selon les besoins du réseau. Les principales bandes de fréquences utilisées incluent 700 MHz, 1800 MHz et 2600 MHz.

I.5.5 Les types de transmission

- **Multiplexage par répartition orthogonale de la fréquence (OFDM)** : Multiplexage par division de fréquence orthogonale, augmentant la robustesse au brouillage.
- **SC-FDMA** : Utilisé en uplink pour réduire la consommation d'énergie.
- **Entrée multiple / sortie multiple (MIMO)** : Utilisation de plusieurs antennes pour améliorer le débit et la couverture.

I.5.6 Gestion de la Mobilité et Qualité de Service (QoS)

La gestion efficace de la mobilité et la garantie de la qualité de service constituent deux piliers essentiels du réseau LTE. D'une part, la mobilité permet aux utilisateurs de se déplacer entre différentes cellules sans perte de connexion ; d'autre part, la qualité de service vise à offrir une expérience utilisateur optimale en fonction du type de service utilisé (voix, vidéo, données, etc.). Ces deux mécanismes sont détaillés ci-dessous :

a) Handover

Processus permettant à une UE de passer d'une cellule LTE à une autre sans interruption :

- **Handover intra-LTE** : Changement de cellule au sein du réseau LTE.
- **Handover inter-RAT** : Transition entre LTE et d'autres technologies (3G, Wi-Fi).

b) Qualité de Service (QoS)

LTE utilise des "EPS Bearers" pour allouer des priorités différentes aux services (voix, vidéo, données), garantissant une expérience fluide. La qualité de service (QDS) ou Quality of Service (QoS), est la capacité de transmission d'un certain nombre de paquets, dans de bonnes conditions, entre un émetteur et un récepteur, et cela peut être présenté sous plusieurs termes tels que :

- La disponibilité.
- Le débit.
- Les délais de transmission.

- La gigue.
- Le taux de perte de paquets.

Elle regroupe un ensemble de technologies mises en œuvre pour assurer des débits suffisants et constants sur tous les types de réseaux.

I.5.7 Optimisation LTE et Analyse des KPI :

Les indicateurs de performance clés (KPI) en LTE peuvent être regroupés en plusieurs catégories distinctes :

- Accessibilité : Cet indicateur permet de vérifier si les services demandés par les utilisateurs sont accessibles dans des conditions spécifiques. Il mesure également la qualité de la disponibilité des services lorsque les utilisateurs en ont besoin. Par exemple, cela peut concerner l'accès au réseau, la capacité à passer un appel vocal ou à échanger des données.

- Retenabilité : Cet aspect évalue la capacité du réseau à maintenir la connexion d'un utilisateur ou à lui fournir de manière continue les services nécessaires.

- Mobilité : L'objectif ici est de mesurer la performance du réseau lorsqu'un utilisateur se déplace, en s'assurant qu'il bénéficie d'une expérience continue et sans coupure.

- Intégrité : Cet indicateur évalue la fiabilité du réseau envers ses utilisateurs, en particulier en ce qui concerne des critères comme le débit, la latence, et la capacité à supporter un nombre important d'utilisateurs.

- Disponibilité : Il permet de vérifier si le réseau est prêt, disponible, et capable de répondre aux besoins des utilisateurs pour l'utilisation des services.

- Utilisation : Ce KPI permet d'analyser si la capacité du réseau est pleinement exploitée et si les ressources disponibles sont proches de leur limite.

I.6 Réseau mobile de cinquième génération (5G)

La technologie «5G » représente une nouvelle génération des standards de la téléphonie mobile. Cette technologie de communication sans fil promet de transformer la façon dont le monde interagit. Absolument tout, des véhicules autonomes aux robots chirurgicaux, en passant par les dispositifs de réalité virtuelle et augmentée, les drones, « Internet des Objets» (IoT) et, plus généralement, l'ensemble des communications mobiles, sera pris en charge par une nouvelle tranche de bande passante située entre 6 GHz et 300 GHz. Une partie de cette tranche sera soumise à licence, tandis qu'une grande partie restera libre.

La 5G se distingue des générations précédentes. Pour la première fois, l'infrastructure physique sera dissociée de l'infrastructure logique ou « virtuelle ». Un réseau de type « SDN » (Software Defined Networks - réseaux définis par logiciel) pourra être déployé, supprimé, étendu ou réduit selon les besoins. La gestion complexe du réseau reposera sur l'apprentissage automatique (Machine Learning ou ML) et l'Intelligence Artificielle (IA). Toutefois, pour garantir son bon fonctionnement, les principaux opérateurs devront adopter les normes

internationales et ouvrir leurs interfaces à d'autres acteurs, y compris la concurrence. C'est sur cet aspect que se jouera la bataille réglementaire et juridique [22].

I.6.1 Objectifs de la 5G

L'application de la technologie 5G est considérée comme la réalisation d'un rêve, car elle est dotée de fonctionnalités avancées qui dépassent les limites des technologies précédentes. Comparée aux technologies radio précédentes, la 5G offre les avancées suivantes :

- **Débit ultra-rapide** : Jusqu'à 10 Gbit/s, soit environ 10 fois plus rapide que la 4G.
- **Latence ultra-faible** : Réduction du temps de réponse à moins de 1 milliseconde, permettant des interactions quasi-instantanées.
- **Connectivité massive** : Capacité à supporter jusqu'à 1 million d'appareils par km^2 (IoT, villes intelligentes, objets connectés).
- **Fiabilité et stabilité accrues** : Amélioration de la couverture et de la qualité du signal, même en mouvement.
- **Efficacité énergétique** : Réduction de la consommation d'énergie par bit transmis, prolongeant l'autonomie des appareils.
- **Réseaux intelligents et flexibles** : Introduction du "network slicing" permettant d'adapter les performances aux besoins des utilisateurs et des industries.
- **Bande passante par unité de surface** : augmentée d'un facteur 1000 [23].

I.6.2 Applications visées pour le réseau de 5G

La 5G offrira une connectivité sans fil avancée pour de nombreuses nouvelles applications, comme illustré dans la FIGURE II.6. Elle s'intégrera profondément dans tous les aspects de notre société future, établissant un écosystème d'information multidimensionnel centré sur l'utilisateur. En supprimant les barrières de temps et d'espace, la 5G permettra une expérience utilisateur immersive et interactive. Elle réduira également la distance entre les individus et les objets, facilitant une intégration transparente pour une interconnexion intelligente et fluide entre les personnes et leur environnement.

Grâce à la 5G, la vision « L'information au bout des doigts, tout reste connecté » deviendra une réalité.

De nombreuses organisations ont proposé divers cas d'utilisation pour cette technologie. L'essor de l'Internet mobile et de l'Internet des objets (IoT) constitue les principaux moteurs du développement futur des communications mobiles, ouvrant ainsi la voie à un large éventail d'applications innovantes [24].

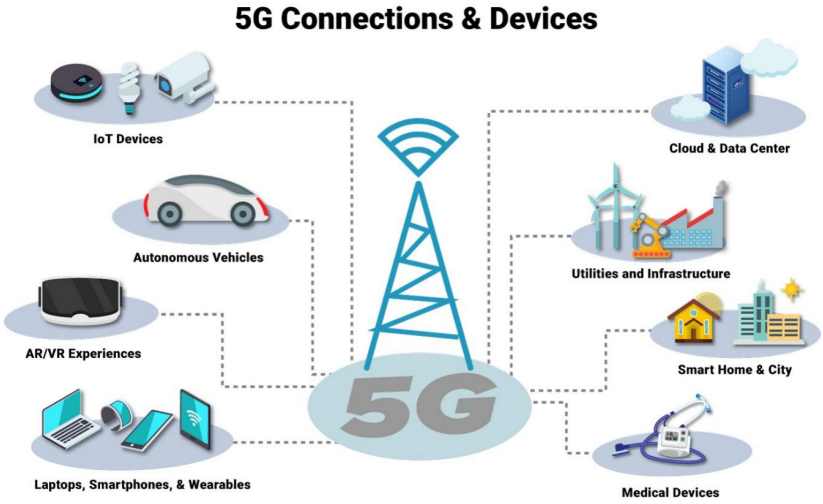


FIGURE I.6 – Application de la 5G [25].

I.6.3 Les principales caractéristiques des domaines d'utilisation de la 5G

L'Union Internationale des Télécommunications (UIT) identifie trois principaux cas d'usage pour la 5G : l'amélioration du haut débit mobile (eMBB – Enhanced Mobile Broadband), les communications massives entre machines (mMTC – Massive Machine-Type Communication), et les communications ultra-fiables à faible latence (URLLC – Ultra-Reliable Low Latency Communication).

- **Enhanced Mobile Broadband (eMBB)** :
  - Débits de données élevés et variables pouvant atteindre 20 Gbit/s.
  - Prise en charge d'un grand nombre d'utilisateurs et de scénarios de mobilité élevée (ex. : trains à grande vitesse).
  - Une transition fluide entre les réseaux fixes et mobiles, garantissant un accès mobile à haut débit optimisé.
- **Massive Machine-Type Communication (mMTC)** :
  - Connexion d'un très grand nombre d'appareils, notamment pour l'Internet des objets (IoT), avec une transmission de petits volumes de données non sensibles au temps (ex. : compteurs d'eau intelligents).
  - Nécessité de dispositifs à faible coût et à très faible consommation énergétique, avec une autonomie pouvant atteindre 10 ans.
- **Ultra-Reliable and Low Latency Communication (URLLC)** :
  - Connexions ultra-fiables et à très faible latence, adaptées aux applications critiques nécessitant une transmission garantie des données.
  - Temps de latence de quelques millisecondes et taux de défaillance extrêmement faible, répondant aux exigences des systèmes sensibles au temps [26].

I.6.4 Architecture de la technologie 5G

Un réseau 5G est composé d'un réseau d'accès 5G (AN) et d'un réseau central 5G (5GC). Le réseau d'accès lui-même est constitué d'un réseau d'accès radio de nouvelle génération (NG-RAN), qui utilise l'interface radio nouvelle génération 5G (NR), et/ou d'un réseau d'accès non-3GPP (par exemple, WiFi, xDSL, etc.) se connectant à un réseau central 5G.

Les différentes entités du réseau sont connectées par un réseau de transport sous-jacent basé sur TCP/IP, qui prend en charge la qualité de service (QoS) différenciée (Quality Of Service) [27].

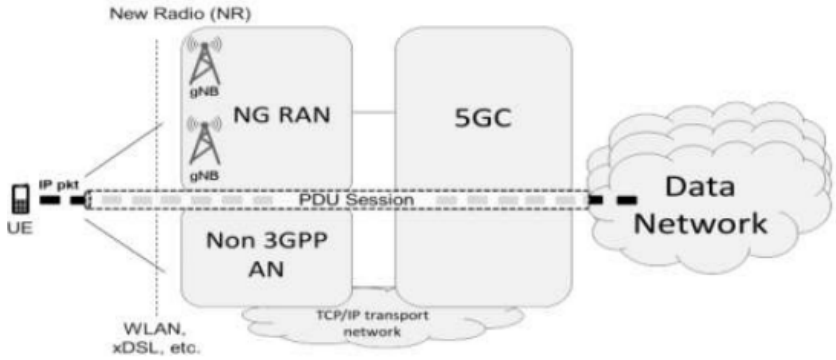


FIGURE I.7 – L'architecture de réseau 5G.

- **Le mobile UE (User Equipement)**

Le dispositif mobile UE (Équipement Utilisateur) dans un réseau 5G conserve le même principe que dans les réseaux 3G et 4G. Il est équipé d'une USIM, qui stocke l'identité de l'abonné, les paramètres de sécurité et les informations d'abonnement.

L'identification de l'abonné se fait via le SUPI (SUbscription Permanent Identifier), une version améliorée de l'IMSI utilisé en 4G. Il peut également être identifié par un identifiant temporaire appelé GUTI (Globally Unique Temporary User Equipment Identity), attribué localement par le réseau visité, qui correspond au TMSI dans la 3G.

L'UE ne peut communiquer avec le 5GC que s'il prend en charge le protocole NAS (« Non-Access Stratum »), qui désigne la « couche non accès » ou le mode « Non-Standalone Access ». De plus, l'UE peut établir simultanément deux connexions radio avec deux stations de base, ce qui correspond au mode 5G Dual, permettant ainsi une compatibilité avec la 5G SA (Standalone) et la 5G NSA (Non-Standalone).

- **L'accès radio 5G (NG-RAN)**

Cette composante est constituée de stations de base de nouvelle génération. Les équipements utilisateurs (UE) peuvent communiquer avec la station de base soit via un lien radio 5G, soit par un lien radio 4G.

Lorsque la communication s'effectue en 5G, la station de base est appelée Nœud B de nouvelle génération (gNB) . En revanche, si la communication passe par la 4G, la station de base est désignée sous le nom de eNB, mais elle doit être adaptée pour s'interfacer avec le cœur de réseau 5G, prenant ainsi l'appellation ng-eNB (Next Generation eNB).

Les fonctions assurées par la gNB sont comparables à celles d'un eNB en 4G. Toutefois, les différences résident principalement dans la gestion de QoS, qui est basée sur des flux plutôt que sur des supports, ainsi que dans la gestion des tranches de réseau (Network Slices) sur l'interface radio. Un slice correspond à une instance logique du réseau mobile, permettant d'assurer un niveau de sécurité, de performance et de QoS adapté à chaque cas d'usage .

L'architecture de la gNB est structurée en plusieurs niveaux fonctionnels. Elle comprend une unité centrale (gNB-CU, Central Unit), qui est associée à une ou plusieurs unités distribuées (gNB-DU, Distributed Units) et reliée à une tête radio distante (RRH, Remote Radio Head). L'unité centrale (gNB-CU) est elle-même subdivisée en deux parties : l'une dédiée aux fonctions du plan de contrôle (gNB-CU-CP, Control Plane Functions) et l'autre consacrée aux fonctions du plan utilisateur (gNB-CU-UP, User Plane Functions) [27], comme illustré dans la figure ci-dessous :

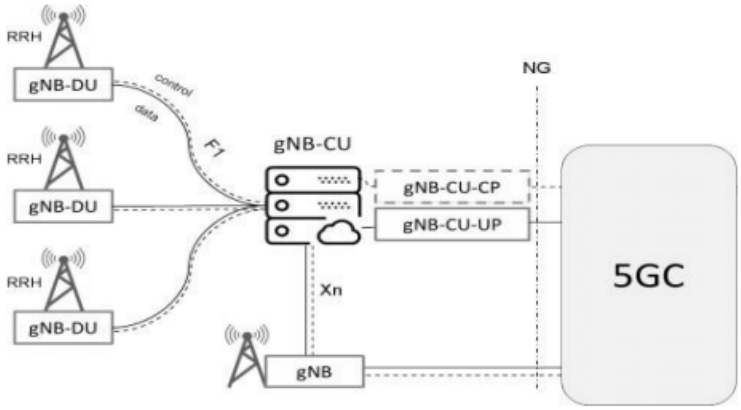


FIGURE I.8 – Accès radio 5G (NG-RAN).

- **Réseau cœur 5G (5GC)**

La principale mission du 5G Core Network est d'assurer une connectivité externe au réseau de données par paquets (notamment Internet) pour les utilisateurs, tout en garantissant leur authentification et en veillant à ce que la qualité de service perçue corresponde aux conditions de leur abonnement. Un élément clé du 5G Core Network est sa capacité à gérer la mobilité des abonnés en suivant leur dernière position avec une précision allant jusqu'à la station de base qui les dessert.

Le 5G Core Network, désigné par le 3GPP sous le nom de *NG-Core*, repose sur une architecture de type « microservices ». En effet, bien que la norme 3GPP définisse un certain degré de modularité, elle se limite à spécifier un ensemble de blocs fonctionnels sans imposer une méthode de mise en œuvre spécifique. Il est important de distinguer ces blocs

fonctionnels des choix techniques nécessaires à la conception d'un système basé sur des microservices. Cela étant dit, considérer l'ensemble des composants représentés dans la FIGURE II.9 comme une architecture de microservices constitue un modèle pertinent pour appréhender le fonctionnement du réseau.

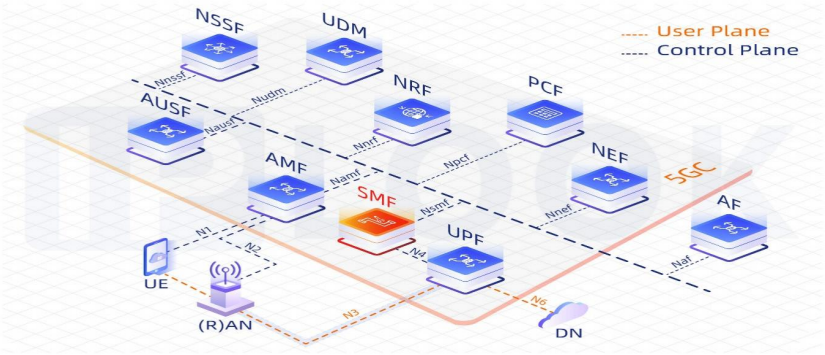


FIGURE I.9 – Architecteur d'un Réseau cœur 5G (5GC) [28].

- **Blocs fonctionnels du réseau cœur 5G (5GC)**

L'ensemble des blocs fonctionnels présentés ci-dessus peut être regroupé en trois catégories principales :

Première catégorie : Fonctions opérant dans le plan de contrôle (CP) et disposant d'un équivalent dans l'EPC (Evolved Packet Core).

• AMF (Fonction de gestion de l'accès et de la mobilité - Core Access and Mobility Management Function)

Responsable de la gestion des connexions et de l'accessibilité, de la gestion de la mobilité, ainsi que de l'authentification et de l'autorisation des accès et des services de localisation. Elle prend en charge les aspects liés à la mobilité du MME (Mobile Management Entity) dans l'EPC.

• SMF (Fonction de gestion des sessions - Session Management Function)

Assure la gestion de chaque session utilisateur (UE), notamment l'attribution des adresses IP, la sélection de la fonction UP correspondante, le contrôle de la qualité de service (QoS) et le routage des paquets. Elle correspond en partie au MME de l'EPC et aux fonctionnalités de contrôle du PGW (Packet Network Data Gateway) dans l'EPC.

• PCF (Fonction de contrôle des politiques - Policy Control Function)

Gère les règles de politique appliquées par les autres fonctions du plan de contrôle (CP). Elle remplit un rôle similaire au PCRF (Policy and Charging Rules Function) dans l'EPC.

• UDM (Gestion unifiée des données - Unified Data Management)

S'occupe de l'identité des utilisateurs, y compris la génération des informations d'authentification. Elle intègre certaines fonctionnalités du HSS (Home Subscriber Server) de l'EPC.

• AUSF (Fonction de serveur d'authentification - Authentication Server Function)

Agit principalement comme un serveur d'authentification et englobe une partie des fonctionnalités du HSS dans l'EPC.

Deuxième catégorie : Fonctions opérant également dans le plan de contrôle (CP) mais sans équivalent direct dans l'EPC.

• SDSF (Fonction de stockage des données structurées - Structured Data Storage Network Function)
Service d'assistance destiné au stockage des données structurées. Dans un environnement basé sur les microservices, il pourrait être implémenté via une base de données SQL.
• UDSF (Fonction de stockage des données non structurées - Unstructured Data Storage Network Function)
Service d'assistance dédié au stockage des données non structurées. Il pourrait être mis en œuvre à l'aide d'un « Key/Value Store » dans une architecture basée sur les microservices.
• NEF (Fonction d'exposition du réseau - Network Exposure Function)
Permet d'exposer certaines fonctionnalités du réseau à des services tiers, y compris la traduction entre les formats de données internes et externes. Elle pourrait être réalisée via un serveur KPI dans un environnement basé sur les microservices.
• NRF (Fonction de répertoire des fonctions réseau - NF Repository Function)
Permet la découverte des services disponibles. Elle pourrait être mise en place sous la forme d'un « Service Discovery » dans une architecture de microservices.
• NSSF (Fonction de sélection de tranche réseau - Network Slicing Selector Function)
Permet de choisir une tranche de réseau spécifique pour desservir un utilisateur donné. Le découpage en tranches permet une gestion différenciée des ressources afin d'adapter la qualité de service aux besoins spécifiques des utilisateurs.

Troisième catégorie : Fonction opérant dans le plan utilisateur (UP).

• UPF (Fonction du plan utilisateur - User Plane Function)
Assure le transfert des données entre le réseau d'accès radio (RAN) et Internet. Elle correspond à la combinaison des fonctions S/PGW dans l'EPC. Outre le transport des paquets, elle est chargée de l'application des politiques, de l'interception légale, de la gestion des rapports d'utilisation du trafic et du contrôle de la qualité de service QoS [29].

I.6.5 L'évolution des bandes de fréquences

La 5G ne se limite pas à une simple amélioration des générations précédentes de la téléphonie mobile. Bien plus qu'une évolution de la 4G, elle représente une nouvelle architecture réseau, exploitant des bandes de fréquences jusqu'alors inaccessibles au grand public. Ces fréquences sont réparties en trois catégories :

- **Les basses fréquences** (< 1 GHz), offrant une large couverture mais avec un débit plus limité.

- **Les fréquences moyennes** (entre 1 GHz et 6 GHz), qui constituent un compromis entre couverture et vitesse.
- **Les hautes fréquences** (> 20 GHz), permettant des débits très élevés, mais avec une portée plus restreinte.

Réutilisation des fréquences 4G et intégration progressive de nouvelles bandes

Dans un premier temps, la 5G s'appuiera sur les mêmes bandes de fréquences que la 4G grâce à la technologie DSS (Dynamic Spectrum Sharing). Cette innovation permettra aux antennes de basculer dynamiquement entre la 4G et la 5G en fonction des besoins, permettant ainsi aux opérateurs de proposer la 5G sans déployer immédiatement de nouvelles infrastructures.

Ensuite, la 5G intégrera deux nouvelles bandes de fréquences :

- **Les fréquences Sub-6 GHz** : Elles englobent les fréquences inférieures à 6 GHz, notamment entre 3,4 GHz et 3,8 GHz pour la 5G. Ces fréquences garantissent une bonne couverture et un meilleur équilibre entre portée et capacité, bien qu'elles offrent un débit inférieur à celui des fréquences plus élevées. Ce groupe inclut aussi bien des bandes moyennes nouvellement attribuées que des basses fréquences déjà exploitées en 4G, comme les 700 MHz.
- **Les ondes millimétriques (mmWave)** : Ce sont des ondes radioélectriques qui couvrent des fréquences allant jusqu'à 300 GHz, mais pour la 5G, elles seront situées entre 24 GHz et 30 GHz. Leur longueur d'onde étant de l'ordre du millimètre, elles permettent des vitesses de connexion ultra-rapides et une capacité réseau accrue. Cependant, ces ondes sont plus sensibles aux obstacles et présentent une portée limitée à moins d'1 km, en particulier dans les environnements urbains denses.

I.6.6 Les technologies émergentes de la 5G

La norme 5G s'appuie sur plusieurs avancées technologiques, notamment l'exploitation de fréquences ultra-hautes pour un transfert de données à grande échelle, le déploiement massif de petites cellules et l'intégration de la technologie Massive MIMO. Ces innovations permettent d'optimiser la vitesse et l'efficacité des transmissions de données, constituant ainsi l'un des principaux atouts de la 5G.

Les principes de ces avancées seront détaillés ci-dessous :

a) Les ondes millimétriques

Généralement, une fréquence inférieure à 6 GHz est utilisée pour la communication cellulaire, tandis que les fréquences supérieures sont principalement employées pour d'autres services tels que l'imagerie médicale, la télédétection par micro-ondes, la radio amateur, l'informatique térahertz et l'astronomie radio. L'augmentation massive du trafic de données a entraîné une congestion du spectre des fréquences radio. En conséquence, la bande passante disponible pour un utilisateur est limitée, ce qui provoque une connexion plus lente et moins fiable.

Une solution à ce problème consiste à utiliser des fréquences supérieures à 6 GHz pour les communications sans fil. Jusqu'à présent, ces fréquences n'ont jamais été utilisées dans ce domaine, et de nombreuses recherches portent sur la transmission des ondes millimétriques. Ces ondes correspondent à des fréquences comprises entre 30 GHz et 300 GHz et sont appelées ainsi en raison de leur longueur d'onde variant entre 1 et 10 mm, comparée aux ondes radio utilisées dans les systèmes de communication mobile actuels, dont la longueur atteint plusieurs dizaines de centimètres.

De nombreux aspects des ondes millimétriques ont été étudiés ces dernières années. Les auteurs de analysent le potentiel et les défis liés à cette technologie, tandis que présente l'avenir du réseau 5G avec l'intégration des ondes millimétriques. Ces dernières permettent d'offrir une bande passante jusqu'à dix fois supérieure à celle de l'ensemble du spectre 4G.

Bien que ces ondes à haute fréquence soient déjà utilisées dans certaines applications satellitaires, elles n'ont jamais été exploitées pour le haut débit mobile. En raison de leur faible longueur d'onde, elles ne conviennent pas aux applications longue portée. De plus, elles présentent certaines limitations, notamment leur incapacité à traverser les bâtiments et les obstacles, ainsi que leur tendance à être absorbées par la pluie [30]. **b) Les petites cellules** Les petites cellules sont des stations de base de faible puissance et de petite taille, pouvant être déployées tous les 100 mètres pour couvrir des zones géographiques restreintes. Ces stations permettent d'éviter la perte de signal dans les zones densément peuplées et d'améliorer la qualité de la couverture, ce qui les rend idéales pour les environnements urbains et les espaces intérieurs. Grâce à leur légèreté et leur compacité, elles peuvent être installées presque partout, facilitant ainsi leur déploiement à grande échelle.

L'utilisation des ondes millimétriques au lieu du spectre traditionnel inférieur à 6 GHz permet de réduire encore davantage la taille des petites cellules, leur permettant ainsi d'être placées dans des espaces extrêmement restreints. De plus, la réduction de la taille des cellules à moins de 100 mètres contribue à répartir la capacité du réseau entre un nombre plus limité d'utilisateurs, réduisant ainsi la congestion et améliorant la qualité des connexions.

Les petites cellules jouent un rôle clé dans la fourniture d'un haut débit mobile ultrarapide et d'une latence extrêmement faible pour les réseaux 5G. Elles sont classées en microcellules, femtocellules et picocellules, en fonction de la zone de couverture et du nombre d'utilisateurs qu'elles peuvent prendre en charge. De nombreuses études ont été menées sur l'impact des petites cellules dans l'amélioration des performances et de l'efficacité des réseaux 5G

c) Massive MIMO

La technologie Massive MIMO (entrées multiples, sorties multiples massives) est une technologie émergente qui étend le MIMO à une échelle potentiellement plusieurs fois supérieure par rapport à l'état de l'art actuel [31]. Lorsque la technologie MIMO est utilisée pour communiquer simultanément avec plusieurs terminaux, on parle de MIMO multi-utilisateurs. Par simplification, on utilise simplement l'abréviation MU-MIMO.

Le MU-MIMO dans les réseaux cellulaires offre plusieurs avantages majeurs :

- Débit amélioré grâce à l'émission simultanée de plusieurs flux de données.
- Fiabilité accrue en exploitant plusieurs chemins de propagation du signal.
- Efficacité énergétique optimisée en dirigeant le signal vers les terminaux ciblés.
- Réduction des interférences en évitant les émissions nuisibles.

Toutefois, ces avantages dépendent des conditions de propagation et ne peuvent pas être maximisés en même temps.

La technologie MU-MIMO est désormais bien établie et intégrée aux normes récentes comme 4G LTE et LTE-Advanced (LTE-A). Malgré son potentiel, le nombre d'antennes utilisées reste limité, avec LTE-A permettant jusqu'à 8 ports d'antenne, bien que les équipements actuels en utilisent moins. Les systèmes MIMO massifs visent à former des faisceaux avec précision et à assurer un décodage efficace à l'aide d'algorithmes simplifiés, souvent linéaires. Cependant, l'utilisation de techniques avancées de traitement du signal aux deux extrémités est essentielle pour réduire la complexité liée à la surcharge de signalisation. Ces systèmes reposent sur les informations de l'état du canal (CSI) disponibles à la station de base afin de garantir un traitement cohérent du signal, permettant ainsi une précodage et une détection optimisée. Lorsque le nombre d'antennes d'émission est largement supérieur à celui des antennes de réception, des algorithmes linéaires simples suffisent pour obtenir des performances quasi optimales. En revanche, si le nombre d'antennes d'émission est égal ou inférieur à celui des antennes de réception, la conception des récepteurs, du précodage et des détecteurs devient plus complexe, rendant ainsi l'optimisation du système plus difficile [32],FIGURE II.10 illustre la structure de ce type de système.

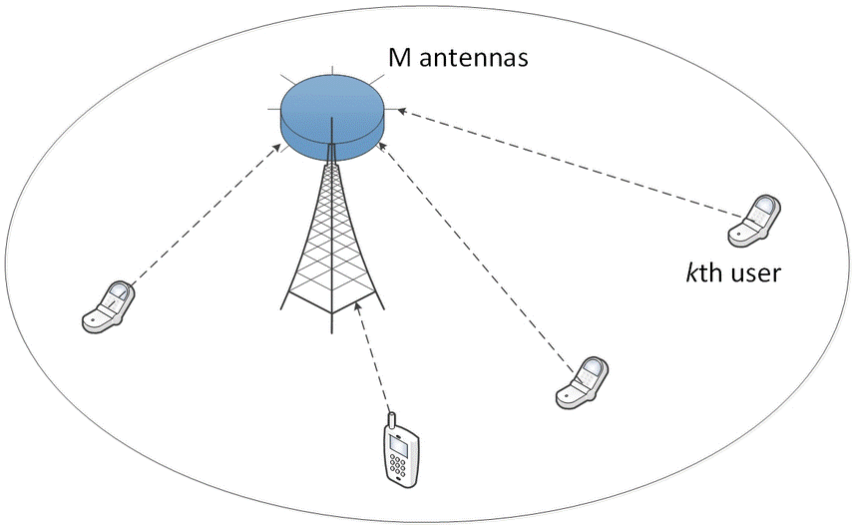


FIGURE I.10 – Système Massive MIMO [33].

d) Le Beamforming

Le beamforming est une technique avancée utilisée pour concentrer le signal sans fil émis par plusieurs antennes en un faisceau puissant, réduisant ainsi l'énergie dispersée dans les directions latérales au niveau de l'émetteur. Du côté du récepteur, le beamforming est une

forme de multiplexage spatial (Spatial Multiplexing), où les signaux reçus sont combinés de manière cohérente dans une direction spécifique, tandis que ceux provenant d'autres directions sont atténués et considérés comme des interférences.

Le contrôle de la direction du signal est assuré en ajustant les poids et les phases du signal à travers plusieurs antennes, ce qui permet d'optimiser les performances et d'orienter les faisceaux vers les utilisateurs de manière efficace. Les techniques de beamforming se divisent en deux catégories principales :

- **Beamforming fixe (Fixed Beamforming)** : utilise des poids et des phases prédéfinis pour combiner les signaux, sans prendre en compte les variations dynamiques du signal reçu.
- **Beamforming adaptatif (Adaptive Beamforming)** : plus avancé, il permet d'orienter dynamiquement le faisceau principal vers la direction souhaitée, en réduisant activement les interférences pour améliorer la qualité de transmission.

Dans les systèmes Massive MIMO, cette technique joue un rôle crucial en permettant à la station de base d'envoyer des données aux utilisateurs via plusieurs chemins. Le beamforming synchronise également l'échange de paquets et le timing de leur arrivée, facilitant ainsi la transmission simultanée de données à plusieurs utilisateurs, ce qui améliore l'efficacité du réseau.

De plus, comme les ondes millimétriques utilisées dans les réseaux 5G ne traversent pas facilement les obstacles et ne couvrent pas de longues distances en raison de leur faible longueur d'onde, le beamforming représente une solution idéale. Il permet d'orienter précisément les faisceaux concentrés vers les utilisateurs, optimisant ainsi la couverture et la qualité de la connexion [34].

Grâce à ces avantages, le beamforming est une technologie clé qui contribue à améliorer les performances globales des réseaux sans fil modernes, en faisant un élément essentiel dans le développement des communications de nouvelle génération.

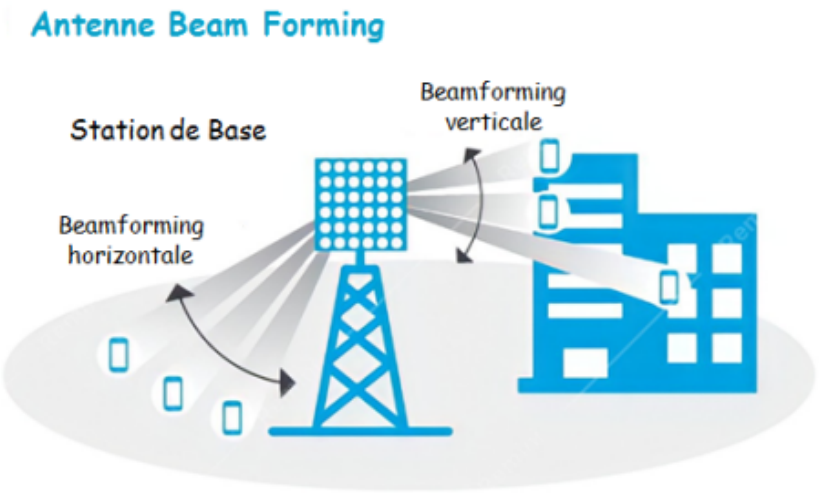


FIGURE I.11 – La technologie Beamforming.

e) Full Duplex Dans les communications mobiles, il est important d'Assurer une transmission bidirectionnelle en temps réel c.-à-d. de l'émetteur vers le récepteur et l'inverse à la fois. Pour les systèmes de téléphonie mobile précédents ils ont utilisé des systèmes FDD ou TDD, nommée semi-duplex nécessitant l'utilisation de deux canaux différents, un pour la liaison montante (UL) et un autre pour la liaison descendante (DL) [35].

Le full duplex ambitionne de permettre l'émission et la réception simultanée d'information, sur les mêmes fréquences, au même moment et au même endroit .

Les Avantages du Full Duplex est :

- Communication en temps réel .
- Filtrage.
- Nouvelles solutions de relais .
- Efficacité spectrale.

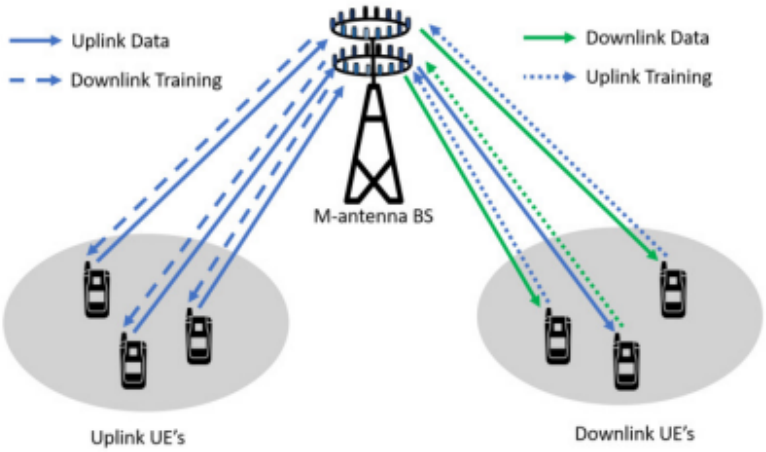


FIGURE I.12 – La technologie Full duplex.

f) Small cells

La 5G a pour objectif de garantir une connexion haut débit aussi bien à l'intérieur qu'à l'extérieur des bâtiments. Pour atteindre ce niveau de couverture tout en maintenant une bonne qualité de service, les chercheurs ont dû explorer de nouvelles approches. L'une des solutions les plus prometteuses consiste à rapprocher les points d'accès des utilisateurs, afin de réduire la distance que doivent parcourir les signaux. Ce concept, connu sous le nom d'approximation du réseau, est actuellement considéré comme indispensable pour répondre aux exigences de la 5G. En effet, dans les macrocelles classiques, la capacité par utilisateur est souvent limitée à cause du grand nombre d'usagers connectés. Pour pallier ce problème, il devient nécessaire d'introduire de petites cellules, appelées 'small cells', à l'intérieur des grandes cellules existantes. Ces petites cellules permettent de déployer les équipements 5G plus près des utilisateurs, améliorant ainsi à la fois la couverture et les performances du réseau [36].

Il existe principalement trois types de small cells, classés selon leur puissance d'émission et la taille de la zone qu'elles peuvent couvrir.

i) **Femtocellules** : ce type de cellule permet une puissance d'émission maximale de **24 dBm**, offrant ainsi une couverture jusqu'à **100 mètres**. Elles sont principalement déployées dans les environnements résidentiels pour couvrir un petit nombre d'utilisateurs.

ii) **Picocellules** : la puissance de transmission varie entre **24 et 30 dBm**, avec un rayon de couverture allant de **200 à 300 mètres**. Elles sont utilisées pour améliorer la couverture à l'intérieur comme à l'extérieur des bâtiments, par exemple dans les hôtels, les entreprises ou les centres commerciaux.

iii) **Microcellules** : elles offrent une couverture plus étendue pouvant atteindre un rayon de **2 kilomètres**, avec une puissance d'émission maximale de **40 dBm**. Ce type de

cellule convient aussi bien pour une couverture intérieure qu'extérieure, et peut supporter jusqu'à **2000 utilisateurs simultanés**. Le tableau I.1 présente un résume les différentes cellules selon leurs propriétés utilisées dans les systèmes de communication radio.

TABLE I.1 – Caractéristiques comparatives des types de cellules dans les réseaux mobiles

Type de cellule	Puissance (W)	Rayon de couverture (km)	Nombre d'utilisateurs	Zone de déploiement
Femtocellule	0,001 à 0,25	0,01 à 0,1	1 à 30	Intérieur
Picocellule	0,25 à 1	0,1 à 0,2	30 à 100	Intérieur / Extérieur
Microcellule	1 à 10	0,2 à 2	100 à 2000	Intérieur / Extérieur
Macrocellule	10 à >50	8 à 30	>2000	Extérieur

I.6.7 Les interfaces 5G NR

Il existe cinq interfaces principales :
-UU : une interface radio entre un UE et le réseau d'accès 5G.
-X2 : interface entre 4G et 5G.
-NG : L'interface entre le gNodeB et le CN est similaire à l'interface S1 du LTE.
-S1-U : interface du plan utilisateur entre le réseau LTE et le réseau central.
-Xn : interface entre les gNodeB.

I.6.8 Stratégies de déploiement de la 5G

Le déploiement de la 5G s'effectue selon deux principales approches techniques, qui diffèrent par leur niveau d'indépendance vis-à-vis des infrastructures 4G existantes. Ces stratégies, dites non autonome (NSA) et autonome (SA), représentent des étapes clés dans l'évolution vers un réseau 5G complet et performant. Elles sont présentées comme suit :
a) Non autonome (NSA)

La 5G NR Mode non autonome (NSA) est considérée comme la première version du mode SA 5G NR, dans lequel les réseaux sont soutenus par l'infrastructure LTE existante. Elle se concentre fondamentalement sur les fréquences eMBB pour une augmentation de la capacité des données. Mais continue d'utiliser les systèmes de communication 4G tels que le service vocal pour les téléphones cellulaires. De plus, la NSA offre également une voie de migration vers l'AS par le biais de la double connectivité NR et LTE. Cela contribuera à un déploiement plus rapide des services commerciaux.

b) Autonome (SA)

La 5G NR Mode autonome (SA) utilise une nouvelle architecture de bout en bout basée sur des ondes millimétriques (mm-Wave). Ce mode ne dépend pas de l'infrastructure 4G LTE, mais fait appel à l'Enhanced Mobile Broadband (eMBB), aux Ultra Reliable And Low Latency Communications (URLLC) et à l'Huge Machine Type Communication ou mMTC pour des débits de données plus élevés avec une efficacité accrue et des coûts moindres. La 5G SA permet d'utiliser le Network Slicing de manière plus avancée [37].

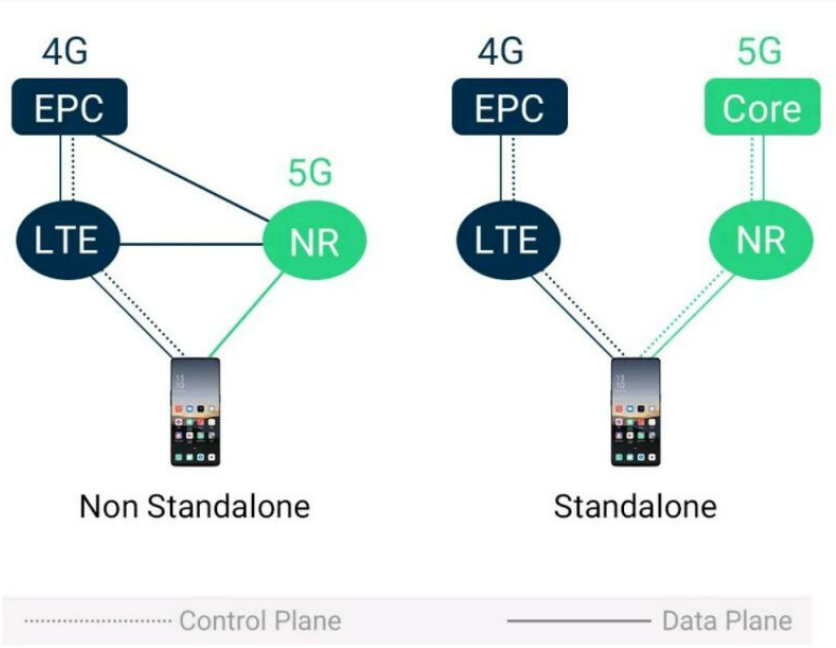


FIGURE I.13 – L'architecture NSA et SA [38].

I.6.9 la Qualité de Service (QoS)

La qualité de service (QoS) est essentielle pour gérer la qualité et déterminer la priorité de la transmission du trafic de données sur un réseau. Elle garantit que les différents types de trafic de données, tels que la voix, la vidéo et les données, reçoivent des niveaux de service appropriés. Ses principaux objectifs sont de permettre aux réseaux et aux entreprises de donner la priorité au trafic, notamment en fournissant une bande passante dédiée, en contrôlant la gigue et en réduisant le temps de latence. Il est essentiel pour améliorer les performances des applications d'entreprise, des réseaux étendus (WAN) et des réseaux de fournisseurs de services. L Les types de services qui nécessitent l'application de la QoS sont présentés ci-après :

-Les services de voix sur IP (VoIP) : Les applications de communication vocale exigent une faible latence et une gestion efficace de la gigue afin d'assurer une qualité d'appel sans coupures ni retards. La QoS permet de maintenir la clarté et la continuité de la communication.

- Les visioconférences et les diffusions en direct (streaming) : Ces services requièrent une bande passante suffisante et une connexion stable. La QoS contribue à minimiser les interruptions, les délais de transmission et les pertes de qualité audio/vidéo.

- Les applications en temps réel : Comme les jeux en ligne ou les systèmes de contrôle à distance, qui sont très sensibles à la latence. Une bonne gestion de la QoS permet d'améliorer le temps de réponse et d'offrir une expérience interactive fluide.

-La transmission de données sensibles : Dans des secteurs critiques tels que la finance ou la santé, certaines applications nécessitent une fiabilité et une sécurité élevées. La QoS aide à garantir la livraison des données dans le respect de contraintes strictes.

- Les applications de Big Data et de Cloud computing : Ces solutions requièrent des performances constantes et prévisibles pour l'accès aux données ou aux services distants. La QoS permet une gestion efficace des ressources réseau pour assurer un service de qualité.

I.6.10 D'optimisation du réseau 5G et analyse KPI

Les KPI en 5G peuvent être divisés en plusieurs sous-catégories suivantes :
- **Accessibilité** : Une requête provenant d'un réseau incite probablement l'utilisateur d'un service à poursuivre dans les conditions spécifiées. En 5G, l'enregistrement réussi des tranches de réseau indique la migration vers l'UE. Si les utilisateurs ou les clients ne peuvent pas s'enregistrer dans l'instance de tranche de réseau, ils ne peuvent utiliser aucun service réseau de cette instance.
- **Retenabilité** : Il est utilisé pour mesurer la manière dont le réseau conserve la possession de l'utilisateur ou est capable de détenir et de fournir les services aux utilisateurs.
- **Intégrité** : Il s'agit de la propriété selon laquelle les données ne sont pas modifiées de manière non autorisée et l'intégrité du service est le degré auquel un service est rendu sans perte excessive.
- **Utilisation** : Il décrit le nombre moyen de sessions KPI PDU qui sont établies avec succès dans l'instance de tranche de réseau.
- **Disponibilité** : Il mesure la disponibilité du réseau, son adéquation ou sa disponibilité pour les utilisateurs.
- **Mobilité** : Il mesure la performance du réseau, sa capacité à gérer les déplacements des utilisateurs et à maintenir le service pour eux.

I.6.11 Conclusion

L'analyse présentée ici porte sur l'évolution des réseaux mobiles, que nous avons suivie pas à pas,depuis la toute première génération, la 1G, jusqu'à la 5G qui est aujourd'hui la plus avancée. On a vu comment chaque génération a apporté son lot de nouveautés : que ce soit au niveau des architectures réseau, des interfaces, des fréquences utilisées, ou encore des innovations techniques propres à chaque époque.

Ce qui ressort clairement, c'est que chaque nouvelle génération répond à des besoins toujours plus exigeants, que ce soit en termes de débit, de mobilité, de sécurité, mais surtout de qualité de service (QoS). Du GSM à l'UMTS, puis du LTE à la 5G, les améliorations ont été constantes : meilleure performance réseau, réduction des délais de communication, augmentation de la capacité du réseau et suivi précis des indicateurs clés de performance.

La 5G représente une étape majeure, avec des technologies innovantes comme le Massive MIMO, le beamforming, les petites cellules, ou encore les ondes millimétriques. Ces avancées permettent de répondre aux nouveaux usages, notamment dans l'industrie, les véhicules autonomes, l'Internet des objets ou la réalité augmentée, grâce à une connexion plus rapide et plus réactive.

En somme, l'évolution des réseaux mobiles ne se limite pas à offrir plus de vitesse. Elle ouvre la voie à une transformation numérique profonde et rapide, essentielle pour construire la société connectée de demain, capable de relever les défis technologiques et d'accompagner les innovations futures.

CHAPITRE II

INTELLIGENCE ARTIFICIELLE ET DEEP LEARNING.

II.1 Introduction

La transformation numérique accélérée a bouleversé notre quotidien, notamment grâce à des technologies comme Intelligence artificielle (IA), le machine learning ML, le deep learning (DL) et le transfer learning. Ces domaines, bien que liés, possèdent chacun des caractéristiques et usages spécifiques.

L'intelligence artificielle englobe l'ensemble des techniques permettant à une machine de simuler l'intelligence humaine pour résoudre des tâches complexes. Le machine learning et le deep learning en sont des sous-domaines : le Apprentissage automatique (ML) permet aux machines d'apprendre à partir de données, tandis que le Apprentissage profond (DL), basé sur des réseaux de neurones profonds, améliore ces capacités. Le transfer learning, quant à lui, consiste à réutiliser un modèle entraîné sur une tâche pour en accélérer l'apprentissage sur une nouvelle tâche.

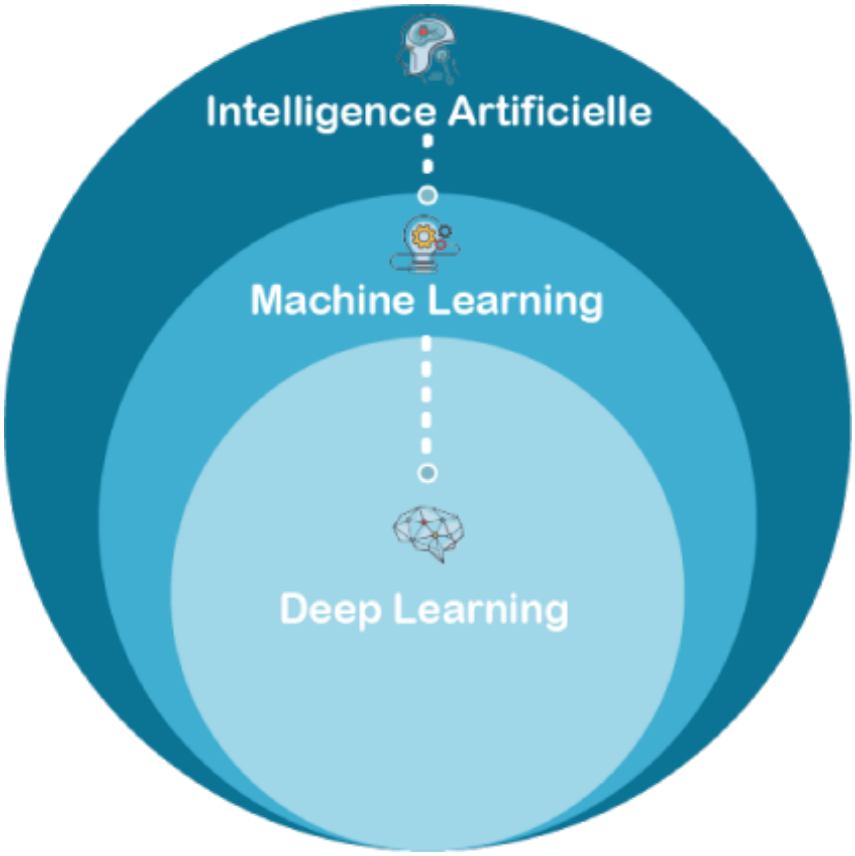


FIGURE II.1 – Les sous-catégories de L'IA [39].

II.2 Machine Learning

ML consiste à utiliser des exemples de données ou des connaissances préalables pour programmer les ordinateurs de manière à optimiser un critère de performance. Un modèle est défini avec quelques paramètres, puis un programme logiciel est exécuté pour optimiser progressivement ces paramètres à l'aide des données d'entraînement ou des connaissances antérieures, tout en apprenant.

Les modèles de ML peuvent être soit descriptifs, pour apprendre à partir des données, soit prédictifs, pour aider à prendre des décisions futures. Pour tirer des conclusions à partir d'un échantillon, le ML s'appuie sur la science des statistiques afin de développer des modèles mathématiques.

L'informatique joue deux rôles distincts dans le ML :

1)D'une part, la grande quantité de données utilisées pour l'apprentissage doit être stockée et traitée afin de créer des algorithmes efficaces pour optimiser le modèle.

2)D'autre part, une fois le modèle appris, la représentation et la solution algorithmique pour l'inférence doivent également être efficaces.

Dans certains cas, l'efficacité de l'algorithme d'apprentissage ou d'inférence en particulier sa complexité en espace et en temps peut être tout aussi importante que la précision de ses

prédictions

II.2.1 Fonctionement de Machine Learning

Le développement d'un modèle de machine learning se structure généralement autour de quatre grandes étapes, souvent dirigées et supervisées par un data scientist.

La première étape consiste à sélectionner et à préparer un ensemble de données d'apprentissage. Ces données serviront à entraîner le modèle afin qu'il puisse apprendre à résoudre le problème ciblé.

Selon les cas, les données peuvent être étiquetées (c'est-à-dire annotées avec des informations spécifiques que le modèle doit reconnaître) ou non étiquetées, auquel cas le modèle devra lui-même identifier les motifs et régularités présents dans les données.

Quelle que soit la nature des données, il est essentiel qu'elles soient nettoyées structurées et préparées avec soin. Un ensemble de données mal préparé peut introduire des biais durant l'entraînement et compromettre la fiabilité des prédictions futures du modèle.

La deuxième étape concerne le choix de l'algorithme à appliquer à ces données. Le choix dépend du type et du volume des données disponibles, ainsi que de la nature du problème à résoudre (classification, régression, clustering, etc.).

Vient ensuite l'étape d'entraînement de l'algorithme, qui est un processus itératif. On fait passer les données à travers l'algorithme, puis on compare les résultats produits avec les résultats attendus. Les paramètres internes du modèle (tels que les poids et les biais)sont alors ajustés pour améliorer la précision.

Ce cycle est répété jusqu'à ce que l'algorithme fournisse des résultats satisfaisants de manière cohérente. À ce stade, l'algorithme devient un modèle de ML entraîné.

Enfin, la quatrième étape est l'exploitation du modèle sur de nouvelles données, avec pour objectif de résoudre un problème concret. Par exemple, un modèle destiné à détecter les spams analysera des courriels, tandis qu'un modèle intégré à un aspirateur robot traitera des données issues de son environnement (comme des déplacements d'objets ou l'apparition d'obstacles).

Au fil du temps, ces modèles peuvent s'améliorer et gagner en précision, grâce à l'apprentissage continu et à l'exposition à de nouvelles données [40].

II.2.2 L'utilisation de Machine Learning

De manière générale, l'apprentissage automatique (machine learning) est utilisé lorsqu'on dispose d'un accès plus restreint ou structuré aux données. La plupart des algorithmes de ML sont conçus pour entraîner des modèles à partir de données structurées. Si les données sont non structurées, il est tout de même possible d'appliquer le Machine learning , mais cela nécessite des techniques de traitement et de préparation des données.

L'apprentissage profond (deep learning), quant à lui, exige généralement une grande quantité de données pour garantir que l'architecture (souvent composée de nombreuses va-

riables)ne surajuste pas le modèle durant l'entraînement. Les réseaux de neurones convolutifs (CNN) sont conçus pour fonctionner sur des jeux de données d'imagerie, même s'ils peuvent aussi être utilisés sur des données de capteurs en appliquant une transformation temps-fréquence, comme un spectrogramme. Les réseaux de neurones récurrents, tels que les réseaux LSTM (Long Short-Term Memory), sont conçus pour traiter des données séquentielles comme les séries temporelles ou les textes.

Le machine learning est couramment utilisé dans des projets visant à prédire un résultat ou à révéler des motifs cachés. Dans ces modèles, un ensemble limité de données est utilisé pour permettre aux machines d'apprendre des schémas qu'elles pourront ensuite appliquer à de nouvelles données. Les algorithmes standards utilisés en ML incluent la régression linéaire, les arbres de décision, les machines à vecteurs de support (SVM), le classificateur naïf bayésien et l'analyse discriminante linéaire.

Les modèles de deep learning nécessitent généralement plus de temps pour l'entraînement. L'utilisation de réseaux pré-entraînés et de jeux de données en accès libre peut réduire le temps d'apprentissage, mais leur exécution reste souvent longue. De plus, cela demande une capacité de calcul importante. Les informaticiens qui appliquent le DL doivent consacrer la majeure partie de leur temps à entraîner les modèles et à ajuster l'architecture du réseau de neurones.

Les algorithmes de machine learning peuvent être préférés lorsqu'on souhaite obtenir des résultats plus rapidement. Ils sont plus rapides à entraîner et nécessitent moins de puissance de calcul. Le nombre de caractéristiques (features) et d'observations sont les facteurs clés qui influencent le temps d'entraînement. Les informaticiens utilisant le ML devraient consacrer la majeure partie de leur temps à la création et à l'évaluation des caractéristiques pour améliorer la précision du modèle [41].

II.2.3 Les deffirents types de Machine Learning

1. Apprentissage supervisé : L'apprentissage supervisé consiste à entraîner un modèle sur un ensemble de données **étiquetées** (où la variable cible ou de sortie est connue).

Par exemple, pour un modèle de prévision des tornades, les variables d'entrée peuvent inclure la date, la localisation, la température, les modèles de flux de vent, etc., et la sortie serait l'activité tornadique réellement observée ces jours-là.

Il est couramment utilisé pour l'évaluation des risques, la reconnaissance d'images, l'analyse prédictive ou la détection de fraudes, en utilisant plusieurs types d'algorithmes :

· **Algorithmes de régression :** prédisent des valeurs continues (ex : température, salaire) en identifiant des relations linéaires. Exemples : régression linéaire, forêts aléatoires, gradient boosting.

· **Algorithmes de classification** : prédisent des catégories (ex : "spam" ou "non spam"). Exemples : régression logistique, k plus proches voisins, SVM.

· **Classificateurs de Bayes naïf** : efficaces pour des jeux de données volumineux, ils mo-

délisent la distribution des entrées par classe. Les arbres de décision en font partie, pouvant gérer régression et classification.

· **Réseaux neuronaux** : simulent le fonctionnement du cerveau humain avec des milliers de nœuds interconnectés. Utilisés pour la traduction, la reconnaissance vocale et d'image, ou encore la génération d'images.

· **Forêts aléatoires** : combinent plusieurs arbres de décision pour prédire une valeur ou une catégorie.

2. Apprentissage non supervisé

Les algorithmes non supervisés comme :

—Apriori les modèles de mélange gaussiens (GMM) ou l'analyse en composantes principales (ACP).

—tirent des conclusions à partir de données non étiquetées.

Ils sont utilisés pour l'analyse exploratoire, la reconnaissance de motifs et la modélisation prédictive.

Les méthodes les plus courantes sont :

- **Clustering (regroupement)** : catégorise les données par similarité (ex : segmentation clients, détection d'anomalies).
- **K-means** : groupe les données en K clusters selon la proximité avec un centroïde. Utilisé pour segmentation marketing, compression d'image, etc.
- **Clustering hiérarchique** : regroupe ou divise les données par similitude (ex : agglomératif ou divisif).
- **Clustering probabiliste** : regroupe les données selon la probabilité d'appartenir à une distribution donnée.
- **Algorithmes d'association** : identifient des liens entre objets de données (ex : les clients ayant acheté ceci ont aussi acheté cela).

4. Apprentissage par renforcement L'apprentissage par renforcement (RLHF (Reinforcement Learning from Human Feedback)) est une technique basée sur un système de récompenses et de punitions.

Un agent agit dans un environnement pour atteindre un objectif et reçoit un score en fonction de ses décisions.

À force d'expériences, l'agent apprend la stratégie optimale.

Utilisé notamment dans les jeux vidéo et la robotique pour simuler des comportements humains.

5. Apprentissage semi-supervisé

Ce type d'apprentissage combine apprentissage supervisé et non supervisé.

Le modèle est entraîné sur un petit ensemble de données étiquetées, et un grand ensemble non étiqueté. Les données étiquetées servent de guide.

Exemple :

Les réseaux antagonistes génératifs (GAN) sont des modèles de deep learning semi-supervisés composés de deux réseaux neuronaux qui s'affrontent pour générer des données réalistes [42].

II.3 Deep Learning

Le Deep Learning est une branche de l'apprentissage automatique (machine learning), C'est un algorithme qui tente d'exploiter l'abstraction de haut niveau des données à l'aide de plusieurs couches de traitement composées de structures complexes ou de multiples transformations non linéaires.

Dans le domaine de l'apprentissage automatique, l'apprentissage profond est un algorithme basé sur la caractérisation des données d'apprentissage. Le concept d'apprentissage profond s'oppose à celui de l'apprentissage superficiel (shallow learning). Les modèles d'apprentissage superficiel, tels que les machines à vecteurs de support (SVM) et la régression logistique, ont été introduits dans les années 1990. Ces modèles ne comportent qu'une seule couche, ou aucune couche cachée.

L'apprentissage profond, en revanche, repose sur plusieurs couches cachées. L'essence même de l'apprentissage profond est le réseau de neurones à couches multiples. Il utilise la sortie de la couche précédente comme entrée de la couche suivante afin d'apprendre des caractéristiques de données hautement abstraites [43].

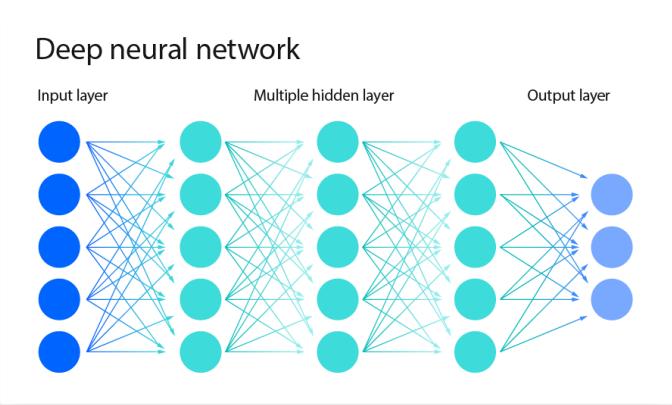


FIGURE II.2 – Schéma de DL avec plusieurs couches [44].

II.3.1 Porquoi le Deep Learning

L'apprentissage profond (ou deep learning) transforme de nombreux secteurs grâce à ses performances impressionnantes. Voici pourquoi il est si avantageux :
1. Il apprend tout seul les bonnes caractéristiques

Avec les méthodes traditionnelles, il faut souvent passer beaucoup de temps à choisir et préparer les données qu'on va donner à un modèle. L'apprentissage profond, lui, peut détecter automatiquement les informations importantes dans les données brutes, comme des images ou du texte, sans intervention humaine. C'est un énorme gain de temps.

2. Il est parfait pour les données non structurées

La majorité des données aujourd'hui ne sont pas bien organisées : images, vidéos, sons, textes.Les réseaux de neurones sont justement très bons pour ce type d'informations. Par exemple, on les utilise pour reconnaître des visages, comprendre le langage naturel ou encore transformer la voix en texte.

3. Il gère très bien les tâches complexes

Dès qu'on entre dans des domaines complexes, comme les voitures autonomes, les assistants vocaux ou le diagnostic médical, l'apprentissage profond fait souvent mieux que les autres techniques. Il est capable de comprendre des relations complexes dans les données, ce qui le rend très efficace là où d'autres modèles atteignent leurs limites.

4. Il s'adapte à grande échelle

Plus il y a de données, mieux l'apprentissage profond fonctionne. Grâce aux services cloud, il est désormais plus simple et plus abordable de traiter d'immenses volumes de données, sans avoir à investir dans du matériel coûteux. Cela le rend accessible à plus d'entreprises.

5. Il devient plus performant avec le temps

Ces modèles s'améliorent à mesure qu'ils voient plus de données. Ils peuvent apprendre en continu et s'adapter à de nouveaux contextes. C'est très utile, par exemple, pour détecter de nouvelles formes de fraude ou de cyberattaques.

6. Il peut apprendre sans supervision

Alors que beaucoup de modèles ont besoin de données étiquetées (c'est-à-dire déjà analysées par des humains), le deep learning peut aussi fonctionner sans. Il est capable de repérer des tendances ou des groupes dans des données brutes, ce qui est très pratique quand on n'a pas le temps ou les ressources pour tout classer à la main.

7. Il peut apprendre en temps réel

Certains modèles sont capables d'évoluer en temps réel, sans avoir besoin d'être complètement réentraînés. C'est un gros avantage pour les systèmes qui doivent réagir vite, comme les véhicules autonomes ou les plateformes de recommandation.

8. Moins besoin d'expertise métier

Comme le modèle découvre tout seul les éléments importants dans les données, on a moins besoin d'avoir des experts du domaine pour préparer ces données. Cela permet de se concentrer sur les décisions stratégiques plutôt que sur la technique.

9. Il peut gérer plusieurs types de données en même temps

L'apprentissage profond peut facilement traiter différents formats ensemble (texte, images, chiffres). Cela en fait un outil très flexible pour toutes sortes d'applications.

10. Il est à la pointe de l'intelligence artificielle

C'est grâce au deep learning qu'on a vu apparaître des innovations majeures en intelligence artificielle : les modèles de langage comme GPT, la reconnaissance d'images ultra-précise, les robots intelligents, ou encore la création d'œuvres d'art ou de musique générées par IA. C'est vraiment le cœur des avancées actuelles en IA [45].

II.3.2 Les Réseaux de neurones

L'étude des réseaux de neurones en intelligence artificielle s'inspire directement du fonctionnement du cerveau humain et des neurones biologiques. Avant d'aborder les modèles artificiels, il est essentiel de comprendre le comportement et la structure des neurones naturels, car ils constituent la base conceptuelle des réseaux de neurones artificiels. Ce qui suit présente le neurone biologique et son mode de fonctionnement.

II.3.2.1 Neurone biologique

Le cerveau humain est une structure extraordinairement complexe, composée d'environ 100 milliards de neurones. Ces cellules nerveuses sont fortement interconnectées, avec près de 10^{15} connexions, formant un vaste réseau dédié à la communication et au traitement de l'information. Les neurones jouent un rôle essentiel dans le fonctionnement du système nerveux, en transmettant les signaux électriques et chimiques qui permettent au corps de percevoir, de penser, d'agir et de ressentir. Le terme neurone a été introduit pour la première fois par Waldeyer en 1881, marquant une étape importante dans la compréhension du système nerveux [46].

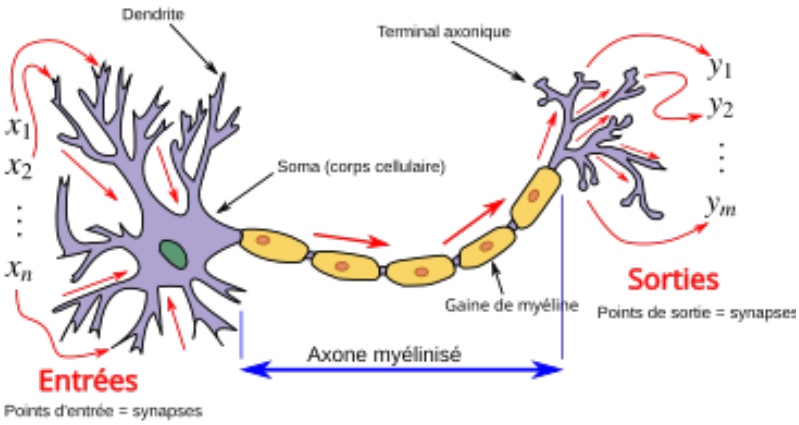


FIGURE II.3 – neurone biologique [47].

- **Corps Cellulaire :** Souvent désigné " Le Soma" effectue la somme des informations qui lui parviennent, puis transmet le résultat informations sous forme de signaux électriques aux autres neurones à travers l'axone.

- **Axone :** C'est un long prolongement fibreux du neurone, il fait la liaison entre les neurones et leur conduit l'influx nerveux.
- **Dendrite :**Joue le rôle de capteur, car il reçoit les informations venant de l'extérieur vers le corps cellulaire. Certaines dendrites peuvent avoir un effet moteur favorisant la transmission d'une information dans l'axone, d'autres au contraire ont un effet inhibiteur qui bloque la transmission de l'influx dans l'axone.
- **Synapses :** c'est une zone située entre deux neurones (cellules nerveuses) et assurant la transmission des informations de l'une à l'autre.
- **Influx Nerveux :** c'est une activité électrique qui parcourt les axones sous la forme d'une séquence de potentiel d'action, provocant la libération des neurotransmetteurs au niveau des synapses. Ces potentiels d'action sont générés par des échanges d'ions entre l'intérieur et l'extérieur des neurones. La vitesse de propagation de l'influx nerveux est de 100m/s.

II.3.2.2 Mode de fonctionnement du neurone biologique

Les dendrites reçoivent des informations d'autres neurones sous forme de signaux électriques (influx nerveux), Ces signaux sont ensuite transmis au corps cellulaire (Soma),plus précisément au noyau, ce qui fera que la somme des impulsions atteignant la dendrite affectera la sou- appelés poids en même temps" Poids Synaptiques".

Le résultat se propage le long de l'axone jusqu'au terminal synaptique lorsque le signal arrive, la vésicule synaptique fusionne avec la membrane cellulaire libérer des neurotransmetteurs.

II.4 Les deffirents types de Deep Learning :

On va présenter dans cette section les modèles de deep learning utilisés dans notre proposition (à savoir les ANN,CNN et les LSTM)

II.4.1 Réseaux neurones artificiels

C'est une nouvelle technologie informatique qui repose sur l'utilisation de nombreux petits processeurs fonctionnant en parallèle, tous connectés entre eux par un réseau intégré. Ensemble, fortement intégrés par un réseau de connexions qui créent un modèle de calcul distribué.

Leur architecture était donc très innovante dans les années 50, à leur naissance, s'inspirant clairement de la structure du cerveau : de nombreux neurones 10^{10} fortement connectés par des synapses, à travers lesquelles les calculs se propagent en parallèle dans le cortex cérébral. Cela permet d'atteindre de nouvelles performances : trouver des solutions en temps

réel à des problèmes complexes, apprendre par eux-mêmes, résister aux pannes et aux erreurs, etc [48].

II.4.2 Fonctionnalité des réseaux de neurones artificiels

La force des réseaux de neurones artificiels réside dans leur capacité à apprendre à partir des données. Pendant la phase d'apprentissage (ou d'entraînement), le réseau est exposé à un ensemble de données étiquetées. Il ajuste alors ses paramètres internes, appelés poids et biais, afin de minimiser la différence entre ses prédictions et les sorties réelles. Ce processus, souvent appelé rétropropagation backpropagation, consiste à mettre à jour de manière itérative les paramètres du réseau à l'aide d'algorithmes d'optimisation comme la descente de gradient (gradient descent).

Le processus d'apprentissage d'un réseau de neurones est un processus itératif, dans lequel les calculs sont effectués en avant et en arrière à travers chaque couche du réseau jusqu'à ce que la fonction de perte soit minimisée.

L'ensemble du processus d'apprentissage peut être divisé en trois étapes principales :
-Propagation avant (*Forward propagation*).
-Calcul de la fonction de perte.
-Rétropropagation (*Backward propagation ou backpropagation*).

1. **Propagation avant (Forward Pass)**

   Lors de la propagation avant, également appelée passage en avant, les données d'entrée sont introduites dans le réseau de neurones, et des calculs sont effectués couche par couche jusqu'à ce que la sortie soit générée. Chaque neurone du réseau reçoit des signaux d'entrée, applique une fonction d'activation spécifique, puis transmet le résultat aux neurones de la couche suivante. Ce processus se poursuit jusqu'à la production de la sortie finale.

   La propagation avant est responsable de la prédiction en fonction des paramètres actuels du réseau (poids et biais).

2. **Calcul de la fonction de perte**

   Une fois la sortie générée, l'étape suivante consiste à évaluer les performances du modèle en comparant ses prédictions aux valeurs cibles réelles issues des données d'apprentissage.

   Cela se fait à l'aide d'une fonction de perte (également appelée fonction de coût ou fonction objectif), qui quantifie l'écart entre les prédictions du réseau et la vérité terrain (ground truth).

   Parmi les fonctions de perte les plus courantes, on trouve :

   - **Erreur quadratique moyenne (MSE)**.
   - **Perte par entropie croisée (cross-entropy)**.

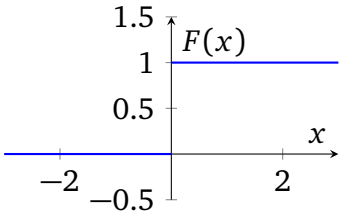
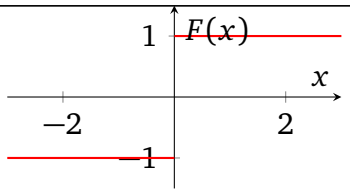
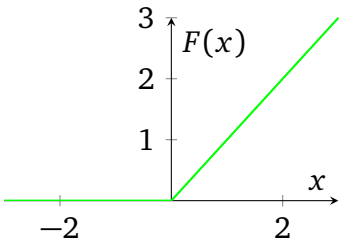
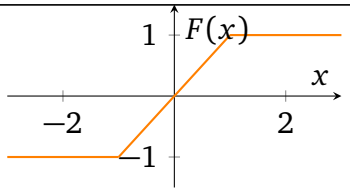
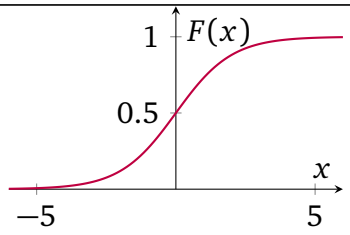
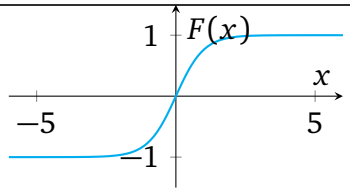
- **Perte de type charnière (hinge loss)**.
  L'objectif de l'entraînement est de minimiser cette fonction de perte, afin d'améliorer la précision du modèle.

3. **Rétropropagation (Backward Pass ou Backpropagation)**
   Après avoir calculé la fonction de perte, le réseau ajuste ses paramètres internes (les poids et les biais) pour réduire la perte et améliorer les performances.
   La rétropropagation consiste à calculer le gradient de la fonction de perte par rapport à chaque paramètre du réseau. Cela se fait en propagant l'erreur en sens inverse à travers les couches du réseau, en utilisant notamment la règle de la chaîne du calcul différentiel (chain rule).
   Les gradients obtenus sont ensuite utilisés pour mettre à jour les paramètres grâce à des algorithmes d'optimisation comme la descente de gradient ou ses variantes.
   En ajustant les paramètres de manière itérative sur la base des gradients calculés, le réseau apprend progressivement à effectuer des prédictions plus précises et à minimiser les erreurs [49].

II.4.3 Fonctions D'activation

Une fonction d'activation, aussi appelée fonction de transfert, permet de décider si un neurone doit être activé ou non en fonction des données qu'il reçoit. Elle prend une valeur réelle en entrée et retourne généralement un résultat compris entre 0 et 1. Plus la sortie est proche de 1, plus le neurone est considéré comme actif ; à l'inverse, une valeur proche de 0 indique que le neurone est inactif. Ces fonctions sont conçues pour réagir fortement à certaines entrées jugées « pertinentes » ou « utiles » pour la tâche à accomplir, et faiblement à celles qui le sont moins. Il existe plusieurs types de fonctions d'activation, chacune adaptée à un type de problème particulier : on retrouve notamment les fonctions binaires, linéaires, sigmoïdes, entre autres. Le tableau suivant présente les fonctions d'activation les plus utilisées dans les réseaux de neurones artificiels.

TABLE II.1 – Les différentes fonctions d'activation

Catégorie	**Type**	**Équation**	**Allure**
Seuil	Heaviside	$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$	
	Signe	$F(x) = \begin{cases} -1 & \text{si } x < 0 \\ 1 & \text{si } x > 0 \end{cases}$	
Linéaire	ReLU	$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	
	Saturé symétrique	$F(x) = \begin{cases} -1 & \text{si } x \leq -1 \\ x & \text{si } -1 < x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$	
Non linéaire	Sigmoïde	$F(x) = \frac{1}{1+e^{-x}}$	
	Tangente hyperbolique	$F(x) = \frac{2}{1+e^{-2x}} - 1$	

II.4.4 Types d'architectures de réseau de neurones

La manière dont un réseau est structuré dépend de l'algorithme d'apprentissage que vous envisagez d'utiliser. En général, on peut identifier trois classes de réseaux. Réseaux à propagation avant à une seule couche.

1. **Réseaux à une couche (feed-forward) :** Dans cette forme simple de réseau en couches, nous avons des nœuds d'entrée et une couche de sortie. Le signal se propage dans le réseau de manière linéaire, en partant de la couche d'entrée pour arriver à la couche de sortie. Il n'y a aucune connexion de retour ni de connexions transversales dans la couche de sortie.

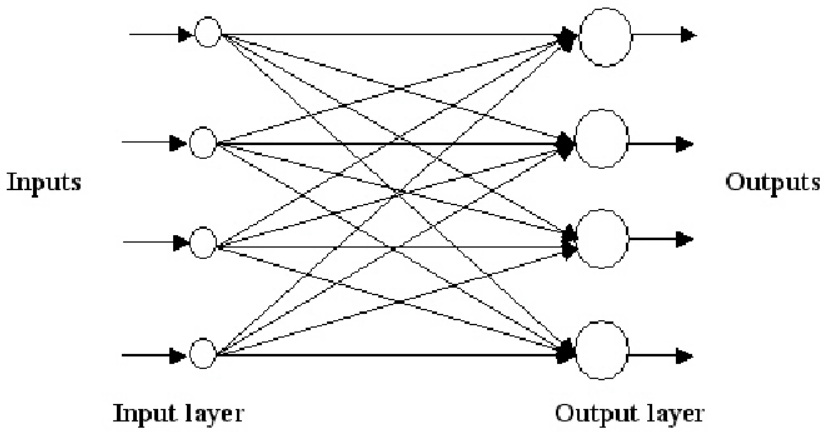


FIGURE II.4 – Réseaux à une couche (feed-forward).

2. **Réseaux à propagation avant multicouches :** Cette classe de réseaux à propagation avant diffère de la précédente par la présence d'une ou plusieurs couches de neurones cachés (couches cachées) entre les couches d'entrée et de sortie. Chaque couche reçoit des connexions de la couche précédente et envoie des connexions vers la suivante. Ainsi, la propagation du signal se fait de manière linéaire, sans cycles ni connexions transversales. Ce type d'architecture offre au réseau une perspective globale en augmentant les interactions entre les neurones.
3. **Réseaux récurrents (feed-back) :** La présence de cycles a un impact profond sur les capacités d'apprentissage du réseau ainsi que sur ses performances, en particulier en rendant le système dynamique.

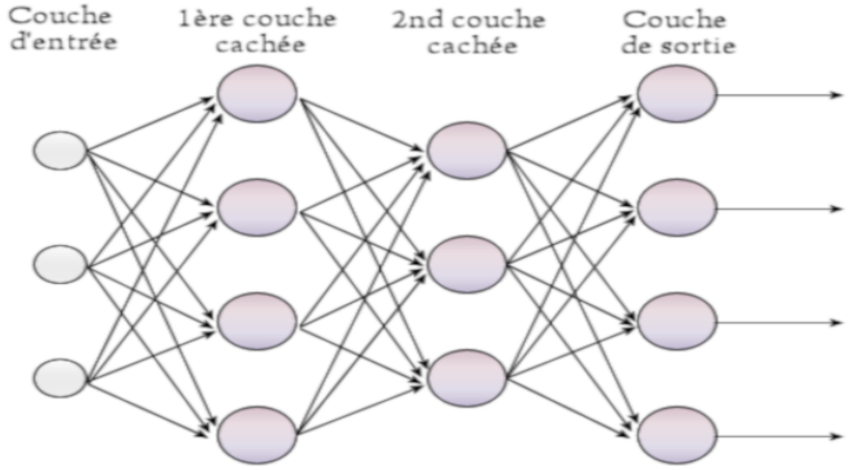


FIGURE II.5 – Réseaux récurrents (feed-back) [50].

II.4.5 Réseau de neurones à convolution

Convolutional Neural Network (CNN) (réseaux de neurones convolutifs)indique que le réseau emploi une opération mathématique appelée la convolution.Les réseaux de convolution sont un type spécialisé de réseaux neuronaux qui utilisent la convolution à la place de la multiplication matricielle générale dans au moins une de leurs couche,Qui se sont avérés très efficaces dans des domaines tels que la reconnaissance et la classification d'images et vidéos. Les CNN sont l'un des meilleurs algorithmes d'apprentissage pour faire l'opération de convolution qui aide à l'extraction de fonctionnalités utiles à partir de points de données corrélés localement. La sortie des noyaux convolutifs est ensuite affectée à l'unité de traitement non linéaire (fonction d'activation), qui non seulement aide à apprendre les abstractions, mais intègre également la non-linéarité dans l'espace des fonctionnalités. Cette non-linéarité génère différents modèles d'activations pour différentes réponses et facilite ainsi l'apprentissage des différences sémantiques dans les images.

La topologie de CNN est divisée en plusieurs étapes d'apprentissage composées d'une combinaison des couches convolutives, des unités de traitement non linéaires et des couches de sous-échantillonnage [51].

II.4.5.1 Fonctionnement des réseaux de neurones convolutifs :

CNN se distinguent des autres réseaux neuronaux par leurs performances supérieures avec des entrées de signaux d'image, de parole ou audio. Ils présentent les principaux types de couches :

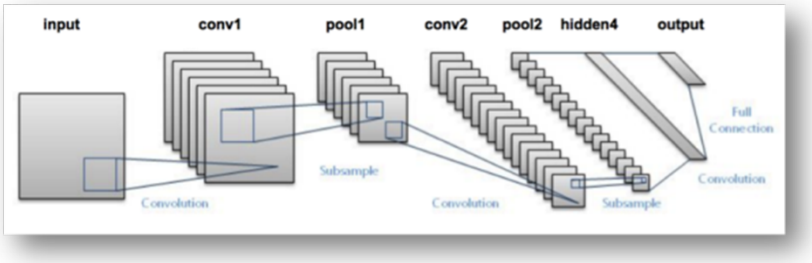


FIGURE II.6 – Les réseaux de neurones convolutifs.

- La couche de convolution (CONV) : qui traite les données d'un champ récepteur.
- La couche de pooling (POOL) : qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).Il existe deux principaux types de pooling :

  - Max pooling : à mesure que le filtre se déplace sur l'entrée, il sélectionne le pixel ayant la valeur maximale pour l'envoyer à la matrice de sortie. Soit dit en passant, cette approche a tendance à être plus souvent utilisée que l'average pooling [52].

  - Average pooling : à mesure que le filtre se déplace sur l'entrée, il calcule la valeur moyenne dans le champ réceptif pour l'envoyer à la matrice de sortie.
- La couche de correction (Relu) : souvent appelée par abus 'Relu' en référence à la fonction d'activation (Unité de rectification linéaire).
- La couche "entièrement connectée" (FC : Fully connected) : qui est une couche de type perceptron.

II.4.6 Réseaux de Neurones Récurrents :

Réseau de neurones récurrents (RNN) se distinguent des réseaux de neurones à propagation avant (feedforward) par la présence de connexions de rétroaction. Ces connexions permettent un flux d'information entre les neurones d'une même couche ou des neurones d'une couche supérieure vers une couche inférieure.

La présence de ces connexions de rétroaction permet aux RNN d'exécuter des tâches nécessitant une mémoire, car le réseau conserve une trace de son état précédent. Plus précisément, le réseau à l'instant t transmet à lui-même des informations qui seront utilisées à l'instant t+1 (en combinaison avec les entrées externes reçues à (t+1). Ainsi, le comportement du réseau dépend à la fois de l'entrée actuelle et de l'état passé du réseau, qui lui-même dépend des instants précédents [53].

II.4.7 Mémoire à long terme à court terme

Les réseaux de longue mémoire à court terme LSTM sont une extension pour les réseaux neuronaux récurrents, qui étend leur mémoire. Par conséquent, il est bien adapté pour ap-

prendre des expériences importantes qui ont des retards très longs entre les deux.

Les unités d'un LSTM sont utilisées comme unités de construction pour les couches d'un RNN, qui est alors souvent appelé un réseau LSTM.

Les LSTM permettent aux RNN de se souvenir de leurs intrants sur une longue période de temps. C'est parce que les LSTM contiennent leurs informations dans une mémoire, ce qui ressemble beaucoup à la mémoire d'un ordinateur parce que le LSTM peut lire, écrire et supprimer des informations de sa mémoire.

Cette mémoire peut être vue comme une cellule gated, où gated signifie que la cellule décide de stocker ou de supprimer des informations (par exemple si elle ouvre les portes ou non), en fonction de l'importance qu'elle attribue à l'information. L'attribution de l'importance se fait à travers des poids, qui sont également appris par l'algorithme. Cela signifie simplement qu'il apprend avec le temps quelle information est importante et laquelle ne l'est pas.

Dans un LSTM vous avez trois portes : entrée, oublier et sortie porte. Ces portes déterminent s'il faut ou non laisser entrer une nouvelle entrée (porte d'entrée), supprimer l'information car elle n'est pas importante (oublier la porte) ou la laisser influencer la sortie au pas de temps courant (porte de sortie). Les portes d'un LSTM sont analogiques, sous la forme de sigmoïdes, ce qui signifie qu'elles vont de 0 à 1. Le fait qu'elles soient analogiques leur permet de faire une rétropropagation avec elles.

Les problèmes posés par la disparition des gradients sont résolus grâce à LSTM, car les gradients sont assez raides et l'entraînement est relativement court et la précision élevée [54].

II.4.7.1 Architecture du LSTM

Les architectures LSTM sont un type de réseau de neurones récurrents conçus pour résoudre le problème de l'apprentissage des dépendances à long terme dans les séquences de données. Elles reposent sur une cellule mémoire qui conserve les informations pertinentes au fil du temps, et dont le comportement est régulé par trois portes principales : la porte d'entrée, la porte d'oubli et la porte de sortie. Ces portes jouent un rôle crucial dans la gestion du flux d'informations.

- Porte d'entrée : elle contrôle quelles nouvelles informations doivent être ajoutées à la cellule mémoire. Elle prend en compte l'entrée actuelle et l'état caché précédent pour décider ce qui est pertinent à retenir.

- Porte d'oubli : elle détermine quelles informations de la cellule mémoire doivent être effacées. Cela permet au réseau de se débarrasser des informations devenues inutiles ou obsolètes.

- Porte de sortie : elle contrôle quelles informations de la cellule mémoire doivent être transmises à l'état caché, c'est-à-dire à la sortie immédiate du LSTM.

Ce mécanisme de portes permet aux réseaux LSTM de filtrer, conserver ou oublier les informations de manière sélective, ce qui leur donne la capacité de modéliser des dépen-

dances temporelles longues dans des séquences, comme celles présentes dans le langage naturel ou les séries temporelles.

Le réseau dispose également d'un état caché, qui agit comme une mémoire à court terme. Cet état est mis à jour à chaque étape à l'aide de l'entrée courante, de l'état caché précédent et de l'état actuel de la cellule mémoire, assurant ainsi une continuité et une adaptation dynamique de la mémoire au fil du temps.

II.4.7.2 Fonctionnement du LSTM

L'architecture LSTM possède une structure en chaîne qui contient quatre réseaux neuronaux et différents blocs de mémoire appelés **cellules** .

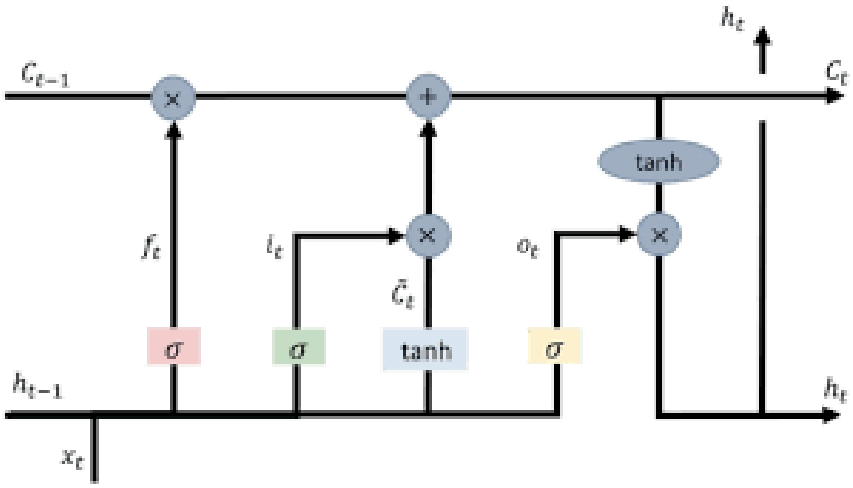


FIGURE II.7 – Schéma d'une cellule mémoire d'un LSTM.

L'information est conservée par les cellules et les manipulations de la mémoire sont effetuées par les portes,Dans une cellule LSTM, ce sont les portes qui régulent de manière dynamique le flux d'information : elles permettent de décider quelles informations doivent être mémorisées, oubliées ou sorties. Parmi ces portes, on retrouve :

1)Porte de l'oubli

L'information devenue inutile dans l'état de la cellule est supprimée grâce à la porte d'oubli. Deux entrées, *x t* (entrée à l'instant donné) et *h t-1* (sortie de la cellule précédente), sont transmises à la porte et multipliées par des matrices de pondération, puis un biais est ajouté. Le résultat est transmis à une fonction d'activation qui produit une sortie binaire. Si, pour un état de cellule donné, la sortie est égale à 0, l'information est oubliée ; pour une sortie égale à 1, l'information est conservée pour une utilisation ultérieure.

- Une valeur proche de 0 indique que l'information correspondante doit être oubliée.
- Une valeur proche de 1 indique qu'elle doit être conservée pour les prochaines étapes.

L'équation de la Porte de l'oubli :

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \tag{II.1}$$

- où :
  - W_i représente la matrice de pondération associée à la porte d'entrée ;
  - $[h_{t-1}, x_t]$ désigne la concaténation de l'état caché précédent h_{t-1} et de l'entrée actuelle x_t .
  - b_i est le biais associé à la porte d'entrée .
  - σ est la fonction d'activation sigmoïde.

Cette porte permet au LSTM de gérer efficacement sa mémoire en éliminant les informations obsolètes, ce qui est essentiel pour l'apprentissage des dépendances à long terme dans les séquences.

2)Porte d'entrée

L'ajout d'informations utiles à l'état de la cellule est assuré par la porte d'entrée. Tout d'abord, l'information est régulée à l'aide de la fonction sigmoïde et les valeurs à mémoriser sont filtrées, comme pour la porte d'oubli, à l'aide des entrées *h t-1* et *x t* . Ensuite, un vecteur est créé à l'aide de la fonction *tanh* , qui produit une sortie comprise entre -1 et +1, contenant toutes les valeurs possibles de h t-1 et *x t* . Enfin, les valeurs du vecteur et les valeurs régulées sont multipliées pour obtenir l'information utile. L'équation de la porte d'entrée est :

L'équation de la porte d'entrée est :

$$je_t = \sigma\big(W_{je} \cdot [h_{t-1}, x_t] + b_{je}\big) \tag{II.2}$$

L'équation du vecteur candidat d'état cellulaire est :

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{II.3}$$

Nous multiplions l'état précédent par f_t, ce qui permet d'oublier certaines informations de l'état antérieur. Ensuite, nous ajoutons $i_t \cdot \tilde{C}_t$, ce qui représente les nouvelles valeurs candidates pondérées par le degré auquel nous décidons de les retenir. **L'état de la cellule à l'instant t est donné par :**

$$C_t = f_t \odot C_{t-1} + i_t \odot \hat{C}_t \tag{II.4}$$

où :

- $\odot$ désigne la multiplication élément par élément (produit de Hadamard) ;
- σ est la fonction d'activation sigmoïde ;
- tanh est la fonction d'activation tangente hyperbolique ;

- W_i, W_c sont les matrices de poids associées ;
- b_i, b_c sont les biais correspondants.

3)Porte de sortie

La porte de sortie permet d'extraire les informations utiles de l'état actuel de la cellule pour les présenter en sortie. Un vecteur est d'abord généré par l'application de la fonction tanh à l'état de la cellule. Ensuite, l'information est régulée par la fonction sigmoïde et filtrée selon les valeurs à mémoriser à l'aide des entrées h_{t-1} et x_t. Enfin, les valeurs du vecteur et les valeurs régulées sont multipliées pour être envoyées en sortie et en entrée à la cellule suivante [55].
L'équation de la porte de sortie est :

$$ou_t = \sigma\big(W_{ou} \cdot [h_{t-1}, x_t] + b_{ou}\big) \tag{II.5}$$

II.4.8 Évaluation des performances :

Pour évaluer la qualité d'un modèle,on utilise généralement plusieurs mesures de performances.parmi lesquelles figurent les plus courantes :

II.4.8.1 Matrice de confusion :

La matrice de confusion,aussi connue sous le nom de tableau de contingence,constitue un outil fondamental en apprentissage automatique pour l'évaluation des performances des modèles de classification. Elle permet de mesurer la fréquence à laquelle les prédictions d'un modèle correspondent ou ne correspondent pas à la réalité en comparant les prédictions avec les étiquettes réelles [56].

Pour calculer une matrice de confusion, on a besoin d'un ensemble de données de test et un autre de validation qui contient les valeurs des résultats obtenus.
la matrice de confusion contient quatre quantités :

- TP (true positive) : le modèle détecte correctement une anomalie réelle.
- TN (true negative) : le modèle confirme correctement qu'il n'y a pas d'anomalie.
- FP (false positive) : le modèle signale une anomalie alors que tout est normal (fausse alerte).
- FN (false negative) : le modèle ne détecte pas une anomalie réelle (anomalie ignorée).

TABLE II.2 – Matrice de confusion

Predicted	**Real**	
	Positive	**Negative**
Positive	TP	FP
Negative	FN	TN

À partir de la matrice de confusion, on peut calculer plusieurs indicateurs qu'on appelle métriques d'évaluation. Ces métriques permettent de mieux comprendre comment le modèle se comporte, en analysant ses performances sous différents aspects, notamment en tenant compte des erreurs qu'il peut commettre. Parmi les plus utilisées, on retrouve : l'exactitude (accuracy), la précision (precision), le rappel (recall) et la mesure F1 (F1-score).

- **L'exactitude (Accuracy) :**
  C'est la mesure la plus simple. Elle représente le pourcentage de prédictions correctes faites par le modèle sur l'ensemble des prédictions, calculée comme suit :

$$\text{Accuracy} = \frac{TP + TN}{N}$$

- **La précision (Precision) :**
  Mesure la proportion de vrais positifs parmi les éléments prédits comme positifs par le modèle.
  Elle est particulièrement utile dans les cas où les faux positifs sont coûteux. Elle est calculée comme suit :

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

- **Le rappel (Recall) :**
  Le rappel mesure la proportion de vrais positifs parmi les éléments qui étaient effectivement positifs.
  Cette mesure est particulièrement utile dans les cas où les faux négatifs sont coûteux. Elle est calculée comme suit :

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

- **F1-score :**
  C'est la moyenne harmonique de la précision et du rappel. Par conséquent, ce score prend en compte à la fois les faux positifs et les faux négatifs. Intuitivement, ce n'est pas aussi facile à comprendre que la précision, mais F1 est généralement plus utile que la précision, surtout si vous avez une distribution de classe inégale. La précision fonctionne mieux si les faux positifs et les faux négatifs ont un coût similaire. Si le

coût des faux positifs et des faux négatifs est très différent, il vaut mieux regarder à la fois Précision et Rappel.

$$\text{F1-score} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

II.4.8.2 Métriques d'évaluation :

Voici quelques mesures utilisées pour évaluer le niveau d'exactitude :

- **Spécificité** : également appelée TNR (True Negative Rate), cette mesure indique la proportion de données normales correctement identifiées comme telles. Elle est donnée par la formule :

$$\text{TNR} = \frac{TN}{TN + FP}$$

- **Précision** : également appelée valeur prédictive positive (PPV), elle mesure la proportion d'anomalies correctement identifiées parmi tous les points prédits comme étant des anomalies. Elle est définie par la formule :

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

  Cette mesure quantifie la pertinence des éléments détectés comme anomalies.

- **Rappel** : aussi appelé sensibilité ou TPR (True Positive Rate), il indique la proportion d'anomalies qui ont été correctement détectées. Il est défini par la formule :

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

  Cette mesure évalue la capacité du modèle à détecter les anomalies existantes.

- **Fallout** : également appelé FPR (False Positive Rate), il mesure la proportion de faux positifs parmi l'ensemble des points normaux. Il est calculé selon la formule :

$$\text{FPR} = \frac{FP}{FP + TN}$$

- **Taux de faux négatifs (FNR)** : il représente la proportion d'anomalies qui n'ont pas été détectées. Il est donné par :

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

II.5 La Déffirence entre Deep Learning et Machine Learning

Plusieurs différences majeures séparent ces deux concepts. Le Deep Learning requiert de plus larges volumes de données d'entraînement, mais apprend de son propre environnement et de ses erreurs.Au contraire, le Machine Learning permet l'entraînement sur des jeux de données moins vastes, mais requiert davantage d'intervention humaine pour apprendre et corriger ses erreurs.

Dans le cas du Machine Learning, un humain doit intervenir pour labelliser les données et indiquer leurs caractéristiques. Un système Deep Learning tente au contraire d'apprendre ces caractéristiques sans intervention humaine.Par exemple, pour la reconnaissance faciale, le programme de Deep Learning apprend d'abord à détecter et reconnaître les bordures et les lignes du visage. Il apprend ensuite les parties les plus importantes des visages, et finalement la représentation générale des visages. Ceci requiert d'immenses volumes de données, mais la probabilité de réussite augmente au fil de l'entraînement.

L'approche est radicalement différente. Les algorithmes de Machine Learning tendent à séparer les données en plusieurs parties, qui sont ensuite combinées pour proposer un résultat ou une solution. De leur côté, les systèmes Deep Learning considèrent un problème dans son entièreté.

Le Machine Learning nécessite un temps d'entraînement plus court, mais son niveau de précision est plus faible. Le Deep Learning permet à la machine de réaliser des corrélations complexes et non linéaires entre les données.L'entraînement Deep Learning est beaucoup plus long à cause de l'importante quantité de données à traiter, et des nombreux paramètres et formules mathématiques impliqués. Un système Machine Learning peut être entraîné en quelques secondes ou quelques heures, tandis que le Deep Learning peut nécessiter des semaines.

Enfin, le Machine Learning permet l'entraînement sur un CPU (unité de traitement centrale) tandis que le Deep Learning requiert un GPU (unité de traitement graphique). Ce puissant hardware est indispensable pour traiter les larges volumes de données et effectuer les calculs complexes des algorithmes.Compte tenu de leurs différences, le Machine Learning et le Deep Learning sont utilisés pour différentes applications. Le Machine Learning est exploité par les programmes prédictifs de la finance ou de la météo, les identificateurs de spam dans les emails, ou encore les programmes visant à concevoir des traitements personnalisés pour les malades.

Le Deep Learning est utilisé pour les recommandations des services de streaming, la reconnaissance faciale, mais également pour les véhicules autonomes. Grâce aux réseaux de neurones, les voitures sont capables de déterminer les objets à éviter, de reconnaître les feux tricolores et les panneaux, et de savoir quand accélérer ou ralentir [57].

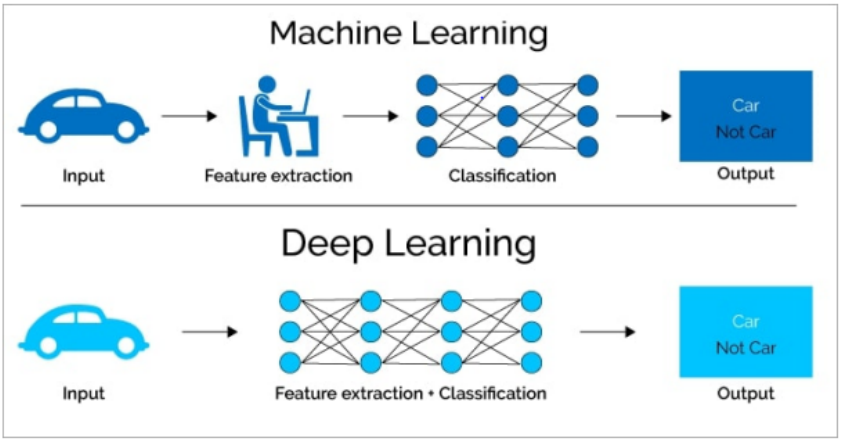


FIGURE II.8 – déffirence entre Deep et Machine Learning.

II.6 Conclusion

Dans ce chapitre, nous avons exploré les fondements théoriques et les principales méthodes de l'intelligence artificielle appliquées aux réseaux de télécommunication, en particulier le *Machine Learning* et le *Deep Learning*.

Nous avons commencé par expliquer comment fonctionne le *Machine Learning*, en présentant ses différents types d'apprentissage (supervisé, non supervisé, par renforcement et semi-supervisé), ainsi que ses principaux domaines d'application, notamment pour la classification et la prédiction dans les systèmes de communication.

Ensuite, nous avons approfondi le *Deep Learning*, en montrant pourquoi cette approche est devenue incontournable pour traiter de grandes quantités de données et gérer la complexité des réseaux mobiles. Nous avons exploré les réseaux de neurones, en faisant le lien avec le fonctionnement du neurone biologique, avant de détailler plusieurs architectures importantes, comme les réseaux convolutifs (CNN) et les réseaux de neurones artificiels (ANN), avec une attention particulière portée à la mémoire à long terme (LSTM), très adaptée pour modéliser des données séquentielles dans le temps.

Enfin, nous avons abordé les méthodes permettant d'évaluer les performances des modèles, en présentant des outils clés tels que la matrice de confusion et des métriques comme la précision, le rappel et le F1-score. Ces éléments seront essentiels pour analyser les résultats des modèles développés dans la suite de ce travail.

Dans l'ensemble, ce chapitre fournit une base théorique solide qui permettra de construire des modèles intelligents pour classifier la qualité de service (QoS), détecter les anomalies et optimiser les performances des réseaux 5G, comme nous le verrons dans le chapitre suivant.

CHAPITRE III

DÉTECTION D'ANOMALIES DANS LES RÉSEAUX 5G

III.1 Introduction

Dans ce chapitre nous appliquons et comparons les performances de trois architectures d'apprentissage profond : ANN,CNN et les réseaux de neurones récurrents de type LSTM. Ces modèles sont évalués selon leur capacité à prédire ou classifier les niveaux de QoS à partir des données d'entrée, permettant ainsi une évaluation intelligente de l'état du réseau.

L'objectif de cette étude est double : d'une part, mesurer l'efficacité de chaque modèle en termes de précision, de temps d'entraînement et de capacité de généralisation ; d'autre part, identifier l'architecture la plus adaptée à la modélisation des performances 5G dans une optique de maintenance proactive ou d'optimisation du service.

Les résultats expérimentaux présentés en fin de chapitre permettront non seulement de comparer les algorithmes entre eux, mais aussi d'éclairer les choix méthodologiques pour des travaux futurs autour de la qualité de service dans les réseaux de nouvelle génération.

III.2 Motivation

Avec l'essor des réseaux 5G, les opérateurs mobiles font face à de nouveaux défis en matière de gestion et de maintien de la qualité de service (QoS). La 5G introduit des cas d'usage exigeants, comme la réalité augmentée, les véhicules autonomes,IA ou encore l'Internet des objets (IoT), qui nécessitent une latence ultra-faible, une bande passante élevée et une allocation dynamique des ressources. Dans ce contexte, une surveillance fine et intelligente de la QoS devient essentielle.

Les méthodes traditionnelles de supervision, souvent basées sur des règles fixes ou une analyse manuelle, montrent rapidement leurs limites face à la complexité croissante du réseau et à la diversité des applications utilisées. Une dégradation non détectée de la QoS peut impacter directement la qualité d'expérience (QoE) des utilisateurs finaux, entraînant

des pertes économiques et une insatisfaction client.

III.3 Objectives

Notre objectif est d'exploiter des approches basées sur l'apprentissage automatique, en particulier les algorithmes profonds pour analyser les données réseau et identifier automatiquement les variations de performance. Bien que ces modèles soient généralement appliqués à des bases de données représentatives des conditions réseau réelles, dans cette étude, nous avons utilisé des données simulées afin de reproduire différents scénarios de qualité de service ; Ce choix est motivé par l'absence d'une base de données gratuite et accessible répondant aux besoins spécifiques de notre étude. Dans ce travail nous visons à anticiper les dégradations de service et à recommander des ajustements adaptés. Cela permettra d'assister les administrateurs réseau dans la prise de décision, d'améliorer la réactivité du système, et de garantir un niveau de QoS conforme aux exigences des applications 5G critiques.donc le but de ce memoire est de développer un système intelligent permettant de tester et d'évaluer automatiquement la QoS offerte par le réseau 5G, en se basant sur l'analyse des données collectées durant l'utilisation du réseau. Pour cela, plusieurs algorithmes d'apprentissage profond ANN, CNN et LSTM sont appliqués à une base de données composée de mesures réseau représentatives.

Chaque échantillon de données est classé selon deux niveaux de qualité :

low : indique une qualité de service faible ou insuffisante. Cela peut être dû à une latence élevée, une bande passante insuffisante, un signal faible ou un mauvais équilibre dans l'allocation des ressources. Cette classification permet d'identifier les zones ou périodes où la QoS est dégradée et nécessite une attention particulière.

high : reflète une bonne qualité de service, répondant aux exigences des applications en cours. Aucune action corrective n'est nécessaire dans ce cas.

III.4 Description de la base de données

Dans le cadre de ce projet, nous avons utilisé une base de données publique intitulée «5G Quality of Service», disponible sur la plateforme Kaggle au lien suivant :https ://www.kaggle.com/ Cette base a été créée par l'utilisateur omarsobhy14 et a pour objectif principal de fournir un environnement simule permettant d'étudier et de prédire la qualité de service (QoS) dans les réseaux mobiles de cinquième génération (5G), en particulier dans le contexte de l'allocation des ressources réseaux [58].

III.4.1 Types de données utilisée

Dans le cadre de cette étude, nous avons exploité un ensemble de données simulées, structuré de manière à reproduire fidèlement divers scénarios de qualité de service (QoS)

dans un environnement réseau. Ce jeu de données comprend plusieurs attributs clés, chacun jouant un rôle spécifique dans l'évaluation et l'analyse des performances du réseau,Les attributs principaux de cette base de données sont les suivants :

- **Timestamp** : représente l'horodatage de la mesure, permettant de suivre l'évolution des performances du réseau dans le temps. Ce champ est essentiel pour l'analyse temporelle et le suivi de la QoS sur des périodes spécifiques.
- **User_ID** : identifiant unique de l'utilisateur ou de l'équipement connecté au réseau. Ce champ permet de différencier les sessions utilisateurs et de réaliser des analyses individuelles ou groupées.
- **Application_Type** : type d'application utilisée (par exemple : vidéo en streaming, appel vocal, jeu en ligne, navigation web). Cette information est cruciale pour adapter les ressources réseau, chaque application ayant des exigences QoS différentes.
- **Signal_Strength (puissance du signal)** : indique la qualité du signal reçu par l'appareil utilisateur. Une valeur faible peut indiquer une dégradation du service ou un éloignement de l'antenne.
- **Latency (latence)** : temps de réponse du réseau, mesuré en millisecondes. Ce paramètre est essentiel pour les applications temps réel comme les jeux en ligne ou les visioconférences.
- **Required_Bandwidth (bande passante requise)** : volume de données que l'application ou l'utilisateur demande pour fonctionner de manière optimale, exprimé en Mbps (mégabits par seconde).
- **Allocated_Bandwidth (bande passante allouée)** : quantité de bande passante réellement attribuée par le réseau à l'utilisateur, pouvant être inférieure à la demande en cas de congestion.
- **Resource_Allocation** : représente le niveau ou le type de ressources réseau (fréquences, canaux, blocs de temps) allouées à l'utilisateur. Cet indicateur reflète les décisions prises par le planificateur du réseau en fonction des priorités et des disponibilités.

III.4.2 Prétraitement des données

Avant l'entraînement des modèles de *Deep Learning*, un certain nombre d'étapes de prétraitement ont été appliquées :

- **Nettoyage** : Suppression des données manquantes ou incohérentes.
- **Normalisation** : Les variables ont été normalisées à l'aide de la méthode *Z-score* afin d'assurer une convergence plus rapide et plus stable lors de l'entraînement.
- **Encodage** : L'étiquette `Signal_Level` a été convertie en format catégoriel pour la classification.

III.5 Expérimentations et discutions

Dans cette section, nous présentons les différentes expérimentations menées dans le cadre de la mise en œuvre et de l'évaluation de notre système de détection d'anomalies pour les réseaux 5G. L'environnement de développement utilisé est MATLAB, qui offre des outils performants pour le traitement des données, l'entraînement de modèles d'apprentissage automatique et l'analyse des performances.

Nous détaillerons successivement les processus d'entraînement pour trois types d'architectures de réseaux de neurones : ANN, CNN, et le réseau de neurones récurrent à LSTM . Chacune de ces approches a été testée afin de comparer leur efficacité dans le cadre de la détection d'anomalies, en mettant l'accent sur les performances obtenues et les observations relevées au cours des simulations.

III.5.1 Matlab

MATLAB, acronyme de "MATrix LABoratory", est un langage de programmation propriétaire et un environnement de calcul numérique développé par MathWorks. Il a été créé par Cleve Moler, informaticien et mathématicien, et présenté au public pour la première fois en 1984, lors de la conférence Automatic Control à Las Vegas.

Conçu principalement pour les applications scientifiques, techniques et d'ingnierie, MATLAB prend en charge plusieurs paradigmes de programmation. Il permet notamment :

- les opérations matricielles avancées,
- la visualisation de données et la création de graphiques de fonctions,
- l'implémentation d'algorithmes,
- la conception d'interfaces utilisateur et l'interfaçage avec d'autres langages.

Bien que MATLAB soit principalement orienté vers le calcul numérique, il offre également des fonctionnalités de deep learning via des programmes spécialisés. En complément, Simulink, un module séparé, propose un environnement graphique pour la modélisation, la simulation et l'analyse de systèmes dynamiques et embarqués, très prisé dans les domaines de l'automobile, de l'aéronautique ou de l'électronique.

D'ici 2020, on estimait que plus de 4 millions de personnes utilisaient MATLAB à travers le monde. Toutefois, . Enfin, le nom MATLAB fait aussi référence à la fenêtre de commande bien connue (»), utilisée comme shell mathématique interactif, ainsi qu'à l'exécution de scripts écrits dans des fichiers texte [59].
En 2021, MATLAB a renforcé l'importation de modèles TensorFlow/PyTorch (R2021a) et introduit une planification parallèle et des outils NLP dans R2021b. L'année 2022 a marqué une avancée vers l'IA embarquée avec R2022a, et l'optimisation des modèles explicables (XAI) via R2022b.
En 2023, MATLAB a accentué son orientation vers les réseaux 5G/6G (R2023a) et les technologies de vision avancées (R2023b).

Enfin, les versions R2024a et R2024b ont consolidé ces progrès avec le Satellite Communications Toolbox, le support des formes d'onde 6G, des approches de contrôle non linéaires, et une meilleure intégration pour la modélisation système avec System Composer (MathWorks, 2020–2024) [60]. Dans ce projet, nous avons utilisé MATLAB accompagné de la Deep Learning Toolbox pour développer, entraîner et évaluer des modèles de classification

III.5.2 Résultats de l'apprentissage et du processus d'entraînement du réseau de neurones Artificiel

Afin d'évaluer les performances du réseau de neurones artificiel développé, il est essentiel d'analyser en premier lieu le déroulement de son entraînement. Cette phase constitue une étape cruciale du processus, car elle permet d'optimiser les poids internes du modèle à partir des données d'apprentissage. Dans ce qui suit, nous détaillons les différentes étapes du processus d'entraînement, les choix architecturaux retenus ainsi que les paramètres d'optimisation utilisés.

III.5.2.1 Processus d'entraînement du réseau de neurones Artificiel

Le processus comprend le chargement des données, la normalisation, l'équilibrage des classes, la division en sous-ensembles, la définition de l'architecture du réseau de neurones, la configuration des options d'entraînement et l'entraînement du modèle, comme suit :
a)Chargement des données : Les données sont chargées à partir d'un fichier CSV contenant des paramètres de qualité de service (QoS) du réseau 5G.
b) Prétraitement des données : Les caractéristiques (Signal Strength, Latency, Required Bandwidth, Allocated Bandwidth, Resource Allocation) sont normalisées à l'aide de la fonction zscore, afin de centrer et réduire les valeurs. Cela permet d'améliorer la convergence de l'apprentissage, d'assurer une contribution équitable de chaque caractéristique et de stabiliser les gradients lors de l'entraînement.
c) Équilibrage des classes : Un équilibrage est effectué en sous-échantillonnant les classes majoritaires pour obtenir un nombre équivalent d'exemples par classe, réduisant ainsi les biais du modèle envers une classe spécifique.
Division des données : Les données équilibrées sont divisées en trois ensembles distincts :
70 % pour l'entraînement (XTrain, YTrain).
15 % pour la validation (XVal, YVal).
15 %pour le test (XTest, YTest) Cette répartition permet d'évaluer la performance du modèle de manière fiable.
e) Définition de l'architecture du réseau de neurones : Le modèle définie par une séquence de couches spécifiques, chacune ayant un rôle distinct dans le traitement des données et la classification finale.

1.Couche d'entrée : 5 caractéristiques.

2.Couche entièrement connectée : 32 neurones, fonction d'activation ReLU.

3.Couche de Dropout : 30% .
4.Couche entièrement connectée : 16 neurones, fonction d'activation ReLU.
5.Couche de Dropout : 20 % .
6.Couche de sortie : Nombre de neurones correspondant aux classes cibles, fonction d'activation Softmax.
7.Couche de classification finale.

f) Définition des options d'entraînement : Le tableau suivant présente les différents paramètres utilisés pour configurer l'entraînement du réseau de neurones :

TABLE III.1 – Paramètres d'entraînement du modèle

Paramètre	**Valeur**	**Description**
Algorithme d'optimisation	Adam	Utilise l'algorithme Adam pour l'optimisation des poids.
MaxEpochs	100	Nombre maximal d'époques (passages complets sur les données d'entraînement).
MiniBatchSize	150	Taille des mini-lots utilisés à chaque itération d'entraînement.
InitialLearnRate	0.001	Taux d'apprentissage initial (taille des ajustements des poids).
Shuffle	every-epoch	Mélange les données d'entraînement à chaque époque pour améliorer la généralisation.

Ces paramètres définissent les principales conditions de l'apprentissage du modèle. L'algorithme Adam est utilisé pour optimiser les poids du réseau. L'entraînement s'effectue sur un maximum de 100 époques, avec des mini-lots de taille 150 à chaque itération. Le taux d'apprentissage initial est fixé à 0,001, et les données sont mélangées à chaque époque pour favoriser la généralisation du modèle.

G) Entraînement du réseau de neurones : Le réseau de neurones est entraîné sur les données d'apprentissage (XTrain et YTrain) à l'aide de la fonction trainNetwork.

III.5.2.2 Analyse de la progression de l'entraînement

Dans la figure suivante Évolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau ANN :

La courbe (a) : représentant l'évolution de la précision, met en évidence une amélioration continue des performances du modèle au fil des itérations. La précision sur les données de validation augmente rapidement et atteint un niveau proche de 98.18 % dès l'itération 332, ce qui reflète une excellente capacité de généralisation. La précision sur l'entraînement suit une tendance similaire, traduisant une progression régulière de l'apprentissage.

la courbe (b) : illustrant la fonction de perte, elle montre une diminution constante aussi bien pour les données d'entraînement que de validation. Cette évolution indique que

le modèle apprend efficacement en réduisant progressivement l'erreur. L'absence d'écart significatif entre les deux courbes suggère que le modèle ne présente pas de surapprentissage et que le processus d'optimisation est bien maîtrisé.

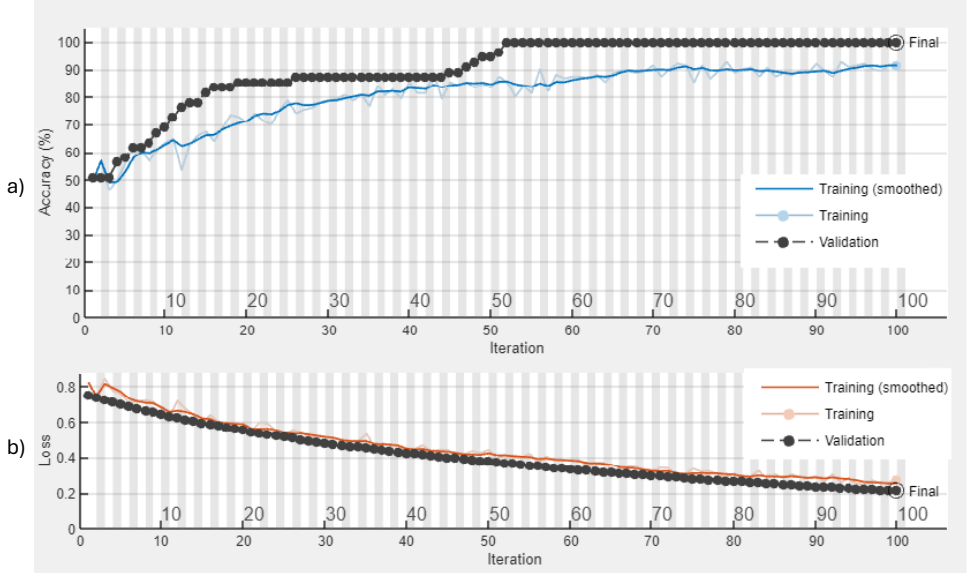


FIGURE III.1 – Évolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau ANN : (a) Accuracy, (b) Loss.

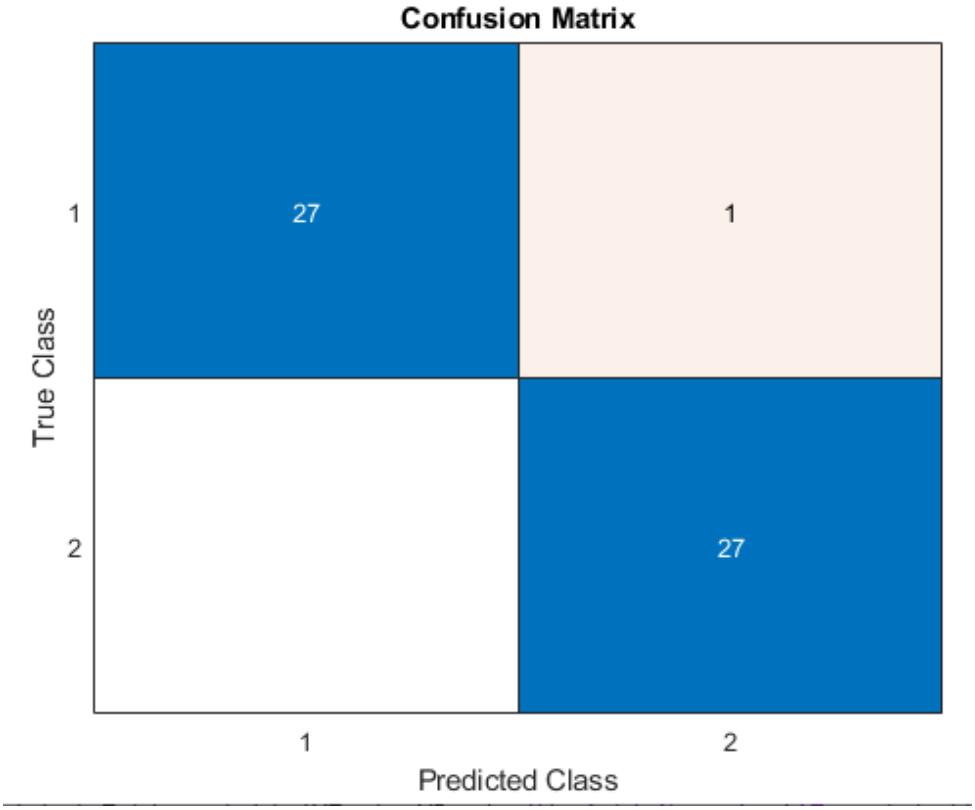


FIGURE III.2 – matrice de confusion(ANN)

III.5.2.3 Analyse de la matrice de confusion

En analysant les résultats de la matrice de confusion obtenue, on constate que le modèle présente une excellente capacité de classification. Il a correctement identifié 27 cas de la classe 2 comme appartenant à cette même classe, ce qui correspond aux vrais positifs (TP). Ce résultat indique une forte aptitude à reconnaître les cas positifs. De plus, le modèle a correctement classé 27 cas de la classe 1 comme étant bien de la classe 1, représentant ainsi les vrais négatifs (TN), c'est-à-dire les prédictions négatives exactes dans un contexte binaire où la classe 1 est considérée comme négative. En ce qui concerne les erreurs, un seul cas de la classe 1 a été incorrectement prédit comme appartenant à la classe 2, soit un faux positif (FP), ce qui reste négligeable et témoigne d'une bonne précision. Fait remarquable, aucun cas de la classe 2 n'a été mal classé comme appartenant à la classe 1, soit aucun faux négatif (FN). Cela démontre une sensibilité parfaite (recall de 100 %, indiquant que le modèle n'a omis aucun cas pertinent. En somme, ces résultats montrent que le modèle est particulièrement performant dans la reconnaissance des cas positifs tout en limitant les erreurs de classification. Cette combinaison d'une forte sensibilité et d'un faible taux d'erreurs en fait un outil fiable, notamment dans des contextes où chaque détection est essentielle.

III.5.2.4 Métriques d'évaluation du ANN

Le modèle évalué dans ce cadre a démontré des performances remarquables en matière de classification. Il affiche tout d'abord un taux de vrais positifs (TVP) de **100 %**, ce qui signifie qu'il a correctement identifié l'intégralité des cas appartenant à la classe positive. Une telle sensibilité parfaite est particulièrement précieuse dans les domaines critiques, notamment pour la détection d'anomalies, où aucun cas positif ne doit être négligé.

Le taux de faux positifs (TFP) s'élève à seulement **3,57 %**, indiquant que très peu de cas négatifs ont été incorrectement classés comme positifs. Ce faible pourcentage limite efficacement les fausses alertes et contribue à la fiabilité globale du système.

Par ailleurs, le modèle présente un taux de vrais négatifs (TVN) de **96,43 %**, ce qui témoigne d'une excellente capacité à reconnaître et rejeter correctement les cas négatifs. Cette spécificité élevée renforce la pertinence du modèle dans les environnements où il est essentiel de ne pas surévaluer la présence d'anomalies.

Le taux de faux négatifs (TFN) est de **0 %**, ce qui signifie qu'aucun cas positif n'a été omis par le modèle. Cette capacité à détecter tous les cas réellement positifs constitue un atout majeur dans les contextes sensibles où la non-détection pourrait entraîner des conséquences importantes.

Enfin, le score F1 atteint **98,18 %**, reflétant un excellent équilibre entre la précision (la proportion de prédictions positives correctes) et le rappel (la capacité à détecter tous les cas positifs). Ce score élevé atteste de la robustesse et de l'efficacité du modèle dans la classification des cas d'intérêt.

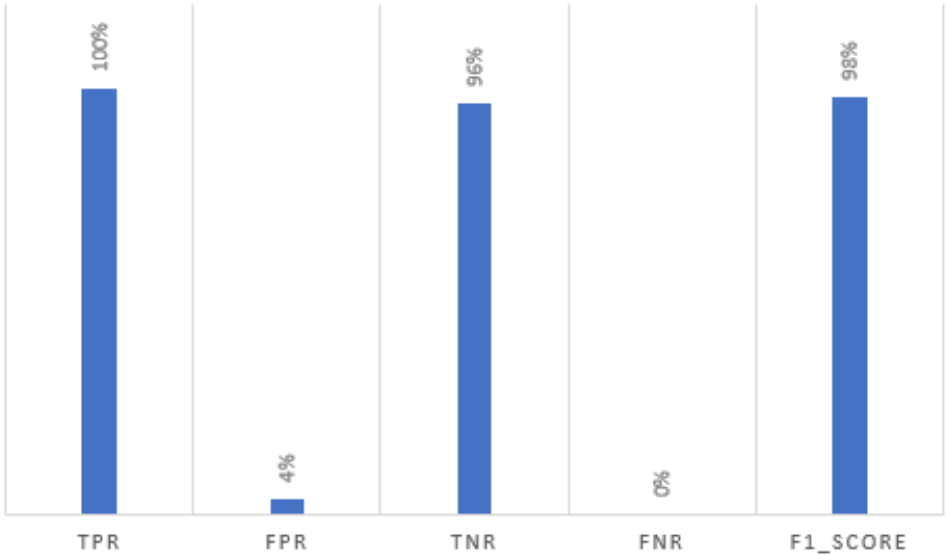


FIGURE III.3 – Mesures de Performance d'un Modèle de Classification ANN.

III.5.3 Processus d'entraînement du réseau de neurones convolutif

Le processus comprend le chargement des données, la normalisation, l'équilibrage des classes, la division en sous-ensembles, la définition de l'architecture du CNN, la configuration des options d'entraînement et l'entraînement du modèle, comme suit :

a) Chargement des données : Les données sont importées à partir d'un fichier CSV contenant des paramètres de QoS du réseau 5G.

b) Prétraitement des données : Les caractéristiques (Signal Strength, Latency, Required Bandwidth, Allocated Bandwidth, Resource Allocation) sont normalisées à l'aide de la fonction `zscore` pour standardiser les valeurs. Cela facilite la convergence de l'entraînement, améliore la stabilité numérique et assure une contribution équilibrée de chaque paramètre.

c) Équilibrage des classes : Les classes majoritaires sont sous-échantillonnées afin de garantir une distribution équilibrée entre les différentes catégories, ce qui réduit le biais du modèle envers les classes les plus fréquentes.

d) Division des données : Les données équilibrées sont divisées en trois ensembles distincts :

70 % pour l'entraînement (`XTrain, YTrain`)

15 % pour la validation (`XVal, YVal`).

15 % pour le test (`XTest, YTest`). Cette répartition permet d'évaluer la performance du modèle de manière fiable.

e) Définition de l'architecture du CNN : Le modèle CNN est conçu pour traiter les données tabulaires transformées en un format compatible avec les couches convolutives. L'architecture typique comprend :

1. **Couche d'entrée**

- Taille d'entrée : `1 × 5 × 1` (adaptée aux 5 caractéristiques du signal)

2. **Première couche convolutive**
   - Nombre de filtres : `16`
   - Taille des filtres : `1 × 2`
   - Fonction d'activation : `ReLU`
3. **Première couche de Max Pooling**
   - Fonction : réduction de la dimension spatiale
4. **Deuxième couche convolutive**
   - Nombre de filtres : `32`
   - Fonction d'activation : `ReLU`
5. **Deuxième couche de Max Pooling**
   - Fonction : réduction supplémentaire de la dimension spatiale
6. **Couche de flatten (aplatissement)**
   - Fonction : conversion des cartes de caractéristiques en vecteur linéaire
7. **Couches entièrement connectées (denses)**
   - Nombre : une ou deux couches
   - Dropout intégré pour éviter le surapprentissage
8. **Couche de sortie**
   - Nombre de neurones : égal au nombre de classes
   - Fonction d'activation : `Softmax` (pour la classification multi-classes)

f) Définition des options d'entraînement : Les principales options d'entraînement sont :

TABLE III.2 – Paramètres d'entraînement du modèle CNN

Paramètre	**Valeur**
Algorithme d'optimisation	Adam
Taux d'apprentissage initial	0.001
Nombre d'époques	100
Taille du mini-batch	150
Fréquence de validation	Toutes les quelques itérations
Critère d'arrêt anticipé	Basé sur la perte de validation
Affichage	Progression de l'entraînement et précision/perte

g) Entraînement du réseau de neurones : Le réseau de neurones est entraîné sur les données d'apprentissage (`XTrain` et `YTrain`) à l'aide de la fonction `trainNetwork` de MATLAB, en utilisant l'architecture définie et les options d'entraînement précédemment spécifiées.

III.5.3.1 Analyse de la progression de l'entraînement

Dans la figure suivante Évolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau CNN :

La courbe (a) : montre clairement que la précision augmente petit à petit avec les époques. La précision sur les données de validation atteint environ 91% vers l'époque 90, ce qui montre que le modèle généralise plutôt bien, même sur des données qu'il n'a jamais vues. La précision sur l'ensemble d'apprentissage suit la même tendance, ce qui prouve que l'apprentissage est stable et cohérent.

La courbe (b) : qui montre la fonction de perte, indique une baisse continue, aussi bien pour les données d'entraînement que de validation. Cela veut dire que le modèle apprend efficacement et réduit l'erreur au fil des itérations. Et comme il n'y a pas d'écart important entre les deux courbes, on peut dire que le modèle n'est pas en overfitting et que l'optimisation est bien maîtrisé.

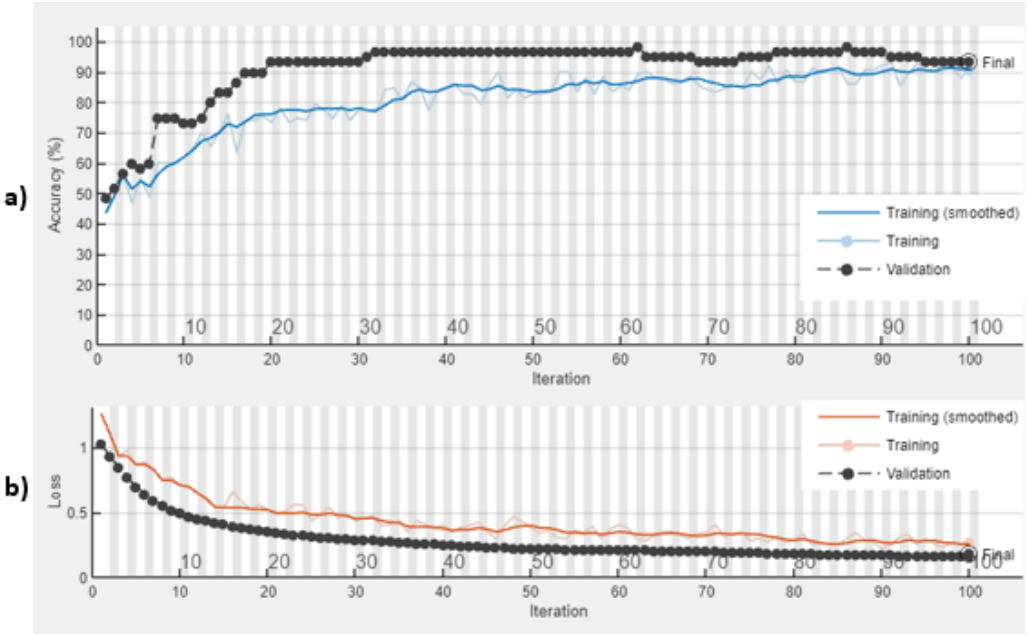


FIGURE III.4 – Evolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau CNN : (a) Accuracy, (b) Loss..

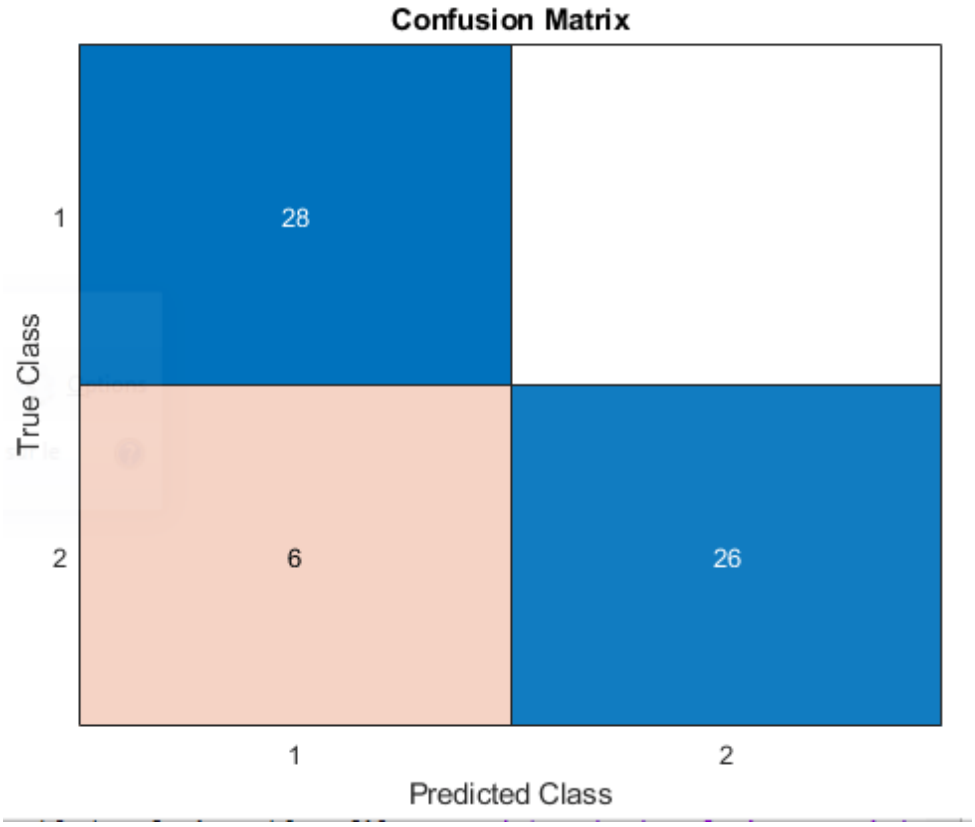


FIGURE III.5 – matrice de confusion(CNN)

III.5.3.2 La matrice de confusion

L'analyse de la matrice de confusion permet de mieux comprendre le comportement du modèle face aux différentes classes. Sur l'ensemble des prédictions, le modèle a correctement identifié **26 instances** appartenant à la classe 1, ce qui correspond aux **vrais positifs (TP)**. Ces résultats indiquent que le modèle reconnaît avec précision une grande partie des cas positifs.

De plus, **28 instances** de la classe 0 ont été correctement classées comme telles, représentant les **vrais négatifs (TN)**. Ce bon taux de reconnaissance des cas négatifs montre une excellente capacité du modèle à éviter les erreurs de classification dans cette classe.

Aucun **faux positif (FP)** n'a été observé : le modèle n'a classé aucun cas négatif (classe 0) à tort comme positif (classe 1). Cette absence totale de fausses alertes est un indicateur particulièrement rassurant, notamment dans des contextes où les surdétections peuvent entraîner des conséquences indésirables.

En revanche, 6 cas appartenant réellement à la classe 1 ont été prédits comme appartenant à la classe 0, ce qui constitue les faux négatifs (FN). Cela montre que le modèle a manqué certains cas positifs, révélant une légère faiblesse dans la détection complète de la classe d'intérêt. Cette lacune pourrait affecter la sensibilité globale, en particulier dans des situations où la détection exhaustive des cas positifs est critique.

En résumé, le modèle présente une excellente capacité à éviter les fausses alertes et à reconnaître les classes de manière précise, en particulier les négatives. Bien que quelques erreurs subsistent dans la détection des cas positifs, ses performances restent **globalement fiables** et **particulièrement adaptées aux contextes où la priorité est donnée à la réduction des faux positifs**.

III.5.3.3 Métriques d'évaluation du CNN

L'analyse des performances du modèle **CNN**, appliqué à la classification des données de qualité de service (QoS) dans un réseau 5G, repose sur plusieurs métriques issues de la matrice de confusion. Les résultats obtenus montrent que le modèle présente des performances globalement solides et équilibrées.

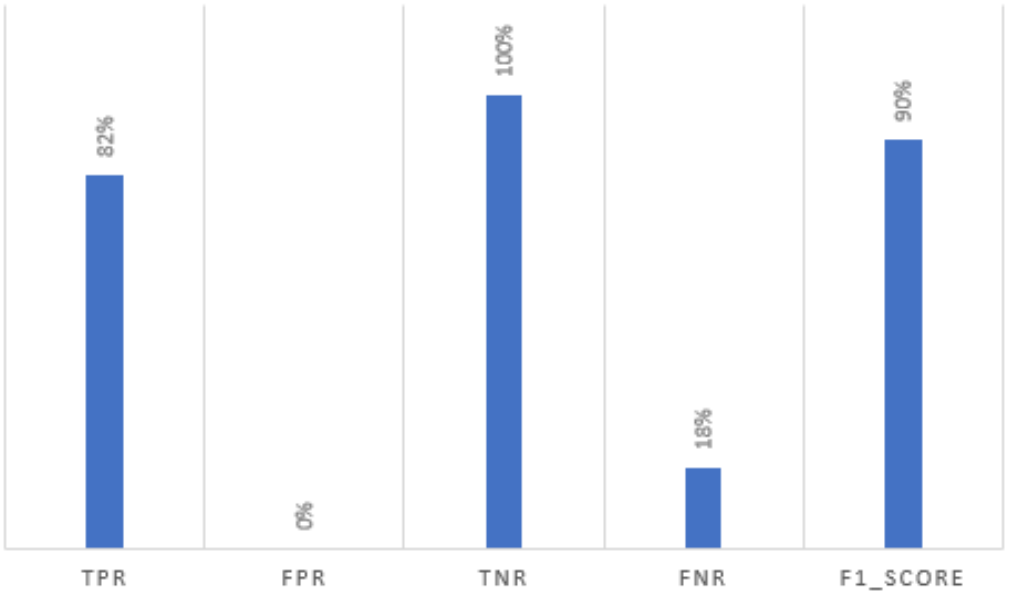
Tout d'abord, le taux de vrais positifs (TVP) atteint **82,35 %**, ce qui témoigne d'une bonne sensibilité. Cela signifie que le modèle parvient à identifier correctement une grande majorité des instances appartenant à la classe d'intérêt, démontrant ainsi son efficacité en détection.

En ce qui concerne le taux de faux positifs (TFP), celui-ci est remarquable puisqu'il est de **0,00 %**. Aucune instance négative n'a été classée à tort comme positive, ce qui souligne la capacité du modèle à éviter les fausses alertes, un critère essentiel dans des systèmes sensibles.

Le taux de vrais négatifs (TVN) est également excellent, atteignant **100 %**, ce qui traduit une spécificité parfaite. Le modèle reconnaît donc avec exactitude tous les cas négatifs, renforçant la fiabilité globale de la classification.

Cependant, un point d'amélioration peut être noté au niveau du **taux de faux négatifs (TFN)**, qui s'élève à **17,65 %**. Cela signifie qu'une partie des cas positifs n'a pas été détectée, ce qui peut être problématique si ces cas représentent des situations critiques.

Enfin, le **score F1** du modèle est de **90,32 %**, ce qui reflète un bon compromis entre la **précision** (la fiabilité des prédictions positives) et le **rappel** (la capacité à détecter l'ensemble des cas positifs). Un tel score est considéré comme excellent dans la plupart des applications réelles, confirmant ainsi la robustesse du modèle CNN dans ce contexte de classification QoS en environnement 5G.



H

FIGURE III.6 – Mesures de Performance d'un Modèle de Classification cnn

III.5.4 Processus d'entraînement du réseau de neurones reseau neurons Long Short Term Memory

Le processus comprend le chargement des données, la normalisation, la division des ensembles, la préparation des données pour LSTM, la définition de l'architecture du LSTM, la configuration des options d'entraînement, et l'évaluation du modèle :

a) Chargement des données : Les données de qualité de service QoS sont importées depuis un fichier CSV (5G QoS.csv). La variable cible Signal Level est convertie en variable catégorielle pour une tâche de classification. Une nouvelle variable Utilization est calculée à partir du rapport entre bande passante requise et bande passante allouée.

b) Prétraitement des données : Les caractéristiques utilisées pour entraîner le modèle sont (Signal Strength, Latency, Required Bandwidth, Allocated Bandwidth, Resource Allocation et Utilization.) Ces paramètres sont normalisés à l'aide de la fonction normalize, équivalente à une standardisation de type z-score, afin d'assurer une contribution équilibrée de chaque caractéristique et d'améliorer la stabilité numérique durant l'entraînement du modèle.

c) Équilibrage des classes : Contrairement au modèle CNN décrit, ce code LSTM ne comporte pas de mécanisme explicite d'équilibrage des classes (ni sur échantillonnage, ni sous échantillonnage). Si l'équilibrage est nécessaire, il pourrait être ajouté à l'étape de prétraitement.

d) Division des données Les données sont divisées en trois ensembles :

70 % pour l'entraînement.

15 % pour la validation.

15 % pour le test.

Deux appels à cvpartition permettent de séparer les ensembles tout en respectant les proportions cibles.

e)Définition de l'architecture du modèle LSTM : Les données d'entrée sont d'abord transformées en séquences à l'aide de la fonction num2cell, ce qui permet de formater chaque observation sous forme de cellule contenant une séquence. Cette étape est essentielle, car les réseaux de neurones LSTM sont conçus pour traiter des données séquentielles, telles que des séries temporelles ou des flux ordonnés de caractéristiques. dans ce programme est structurée comme suit :

1. **sequenceInputLayer**
    - Accepte des séquences avec **6 caractéristiques** par pas de temps : *Signal Strength, Latency, Required Bandwidth, Allocated Bandwidth, Resource Allocation, Utilization*.
2. **lstmLayer(64, 'OutputMode', 'sequence')**
    - Première couche LSTM avec **64 unités de mémoire**.
    - Produit une sortie **à chaque pas de temps** pour capturer l'information temporelle complète.
3. **dropoutLayer(0.3)**
    - Désactive aléatoirement **30 %** des connexions pour réduire le surapprentissage (régularisation).
4. **lstmLayer(32, 'OutputMode', 'last')**
    - Deuxième couche LSTM avec **32 unités**.
    - Ne conserve que la **dernière sortie** de la séquence : résumé global pour la classification.
5. **dropoutLayer(0.3)**
    - Seconde couche de régularisation pour améliorer la généralisation du modèle.
6. **fullyConnectedLayer(numClasses)**
    - Couche dense avec **un neurone par classe**.
    - Convertit la représentation en sortie des LSTM en scores de classe.
7. **softmaxLayer**
    - Transforme les scores en **probabilités** pour chaque classe (classification multiclasse).
8. **classificationLayer**
    - Calcule la **fonction de perte catégorielle** (cross-entropy) utilisée pour l'apprentissage.

f) Définition des options d'entraînement : Le modèle est entraîné avec trainNetwork selon les paramètres suivants :

TABLE III.3 – Paramètres d'entraînement du modèle LSTM avec descriptions

Paramètre	**Valeur**	**Description**
Optimiseur	Adam	Algorithme d'optimisation adaptatif combinant les avantages de RMSProp et de Momentum.
Taux d'apprentissage	0.0005	Contrôle la vitesse de mise à jour des poids du réseau. Une valeur plus petite permet une convergence plus stable.
Nombre d'époques	100	Nombre de fois que l'ensemble complet d'entraînement est parcouru pendant l'apprentissage.
Époques maximales	100	Limite supérieure du nombre d'époques, identique ici au nombre défini.
MaxMiniBatchSize	32	Nombre maximum d'échantillons traités simultanément lors d'une itération. Impacte la stabilité et la vitesse d'entraînement.
Fréquence de validation	1 (chaque époque)	Indique la fréquence à laquelle la performance du modèle est évaluée sur les données de validation.

g) Entraînement et évaluation du modèle : Le réseau est entraîné sur les données d'entraînement (XTrainLSTM, YTrainVal) et évalué sur l'ensemble de test (XTestLSTM, YTest).

III.5.4.1 Analyse de la progression de l'entraînement

Dans la figure suivante, on observe l'évolution de la précision d'entraînement, de validation, et de la perte au cours des itérations du réseau LSTM :

La courbe (a) : montre clairement que la précision augmente progressivement avec les époques. La précision sur les données de validation atteint environ 96,67 % vers l'époque 112, ce qui montre que le modèle généralise plutôt bien, même sur des données qu'il n'a jamais vues. La précision sur l'ensemble d'apprentissage suit la même tendance, ce qui prouve que l'apprentissage est stable et cohérent.

La courbe (b) : qui montre la fonction de perte, indique une baisse continue, aussi bien pour les données d'entraînement que de validation. Cela signifie que le modèle apprend efficacement et réduit l'erreur au fil des itérations. Et comme il n'y a pas d'écart important entre les deux courbes, on peut dire que le modèle n'est pas en overfitting et que l'optimisation est bien maîtrisée.

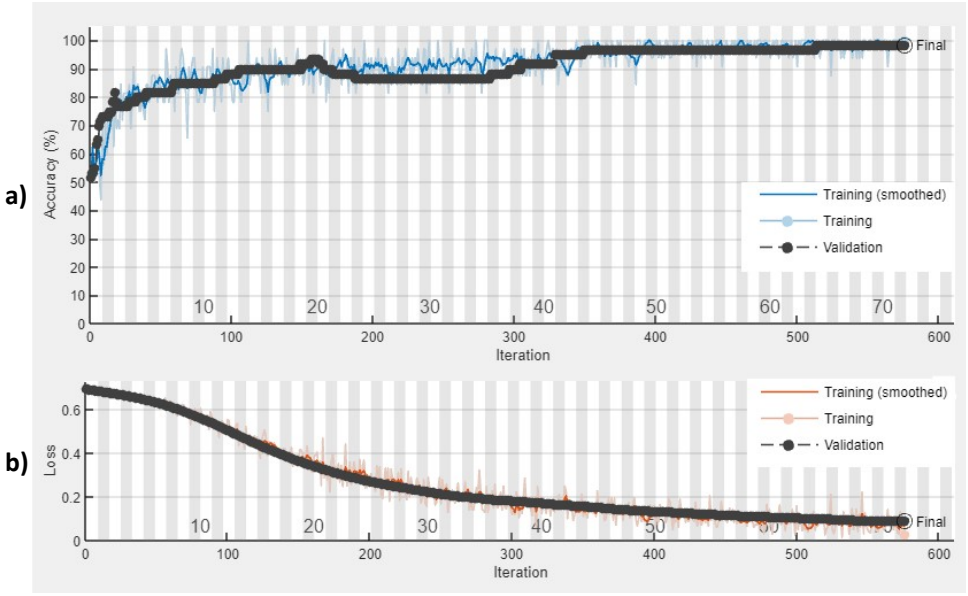


FIGURE III.7 – Evolution de la précision d'entraînement, de validation, et de perte au cours des itérations du réseau LSTM : (a) Accuracy, (b) Loss.

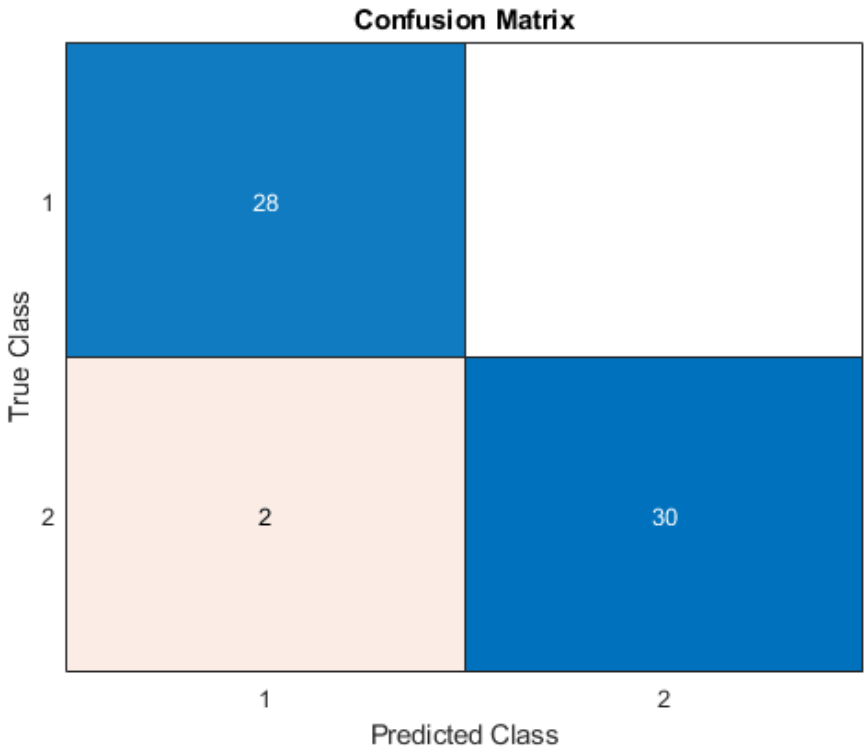


FIGURE III.8 – matrice de confusion (LSTM).

III.5.4.2 La matrice de confusion

En étudiant les résultats de la matrice de confusion obtenue, on observe que le modèle présente une excellente capacité de classification, en particulier pour la classe cible. Il a correctement identifié 30 cas de la classe 2 comme appartenant à cette même classe, ce qui correspond aux vrais positifs (TP) et témoigne d'une forte aptitude à reconnaître les instances positives. Par ailleurs, 28 cas de la classe 1 ont également été correctement classés, représentant les vrais négatifs (TN). Cela signifie que le modèle distingue efficacement les exemples négatifs, considérés ici comme la classe 1 dans un contexte binaire.

Fait notable, aucun cas de la classe 1 n'a été incorrectement attribué à la classe 2, ce qui indique zéro faux positif (FP). Cette absence totale d'erreur dans l'identification des cas négatifs reflète une précision maximale dans ce domaine. Toutefois, le modèle a commis 2 erreurs de classification en identifiant des cas de la classe 2 comme étant de la classe 1 ; ces faux négatifs (FN) affectent légèrement le rappel (ou sensibilité), bien que ce dernier demeure élevé.

Globalement, le modèle démontre une robustesse remarquable, capable de détecter avec grande exactitude les cas positifs tout en minimisant les fausses alertes. Cette performance équilibrée entre précision et sensibilité le rend particulièrement adapté aux applications critiques, où une détection fiable et une faible tolérance aux erreurs sont essentielles

III.5.4.3 Métriques d'évaluation du LSTM

L'évaluation du modèle LSTM, appliqué à la classification des données de QoS dans un réseau 5G, met en évidence des performances particulièrement élevées, fondées sur plusieurs métriques extraites de la matrice de confusion. Le taux de vrais positifs (TVP) atteint 93,33 %, ce qui reflète une excellente sensibilité du modèle, capable d'identifier avec grande précision les instances appartenant à la classe d'intérêt. En parallèle, le taux de faux positifs (TFP) est de 0,00 %, indiquant qu'aucune instance négative n'a été classée à tort comme positive ; cette performance démontre une capacité remarquable à *éviter les fausses alertes*, gage d'une grande fiabilité.

Le taux de vrais négatifs (TVN) s'élève quant à lui à 100 %, soulignant une *spécificité parfaite* : toutes les instances négatives ont été correctement identifiées. Cette propriété est cruciale dans les systèmes où une mauvaise classification des cas négatifs pourrait compromettre l'efficacité globale. Toutefois, le modèle présente un taux de faux négatifs (TFN) de 6,67 %, ce qui signifie qu'un petit nombre de cas positifs n'ont pas été détectés et ont été classés à tort comme négatifs. Bien que cette proportion soit faible, elle mérite une attention particulière dans les situations où les cas positifs ont une importance critique.

Enfin, le score F1 atteint 96,66 %, traduisant un *équilibre optimal entre la précision et le rappel*. Ce score élevé confirme la robustesse du modèle LSTM, capable de maintenir une performance stable et fiable, tant dans l'identification des cas positifs que dans la minimisation des erreurs de classification.

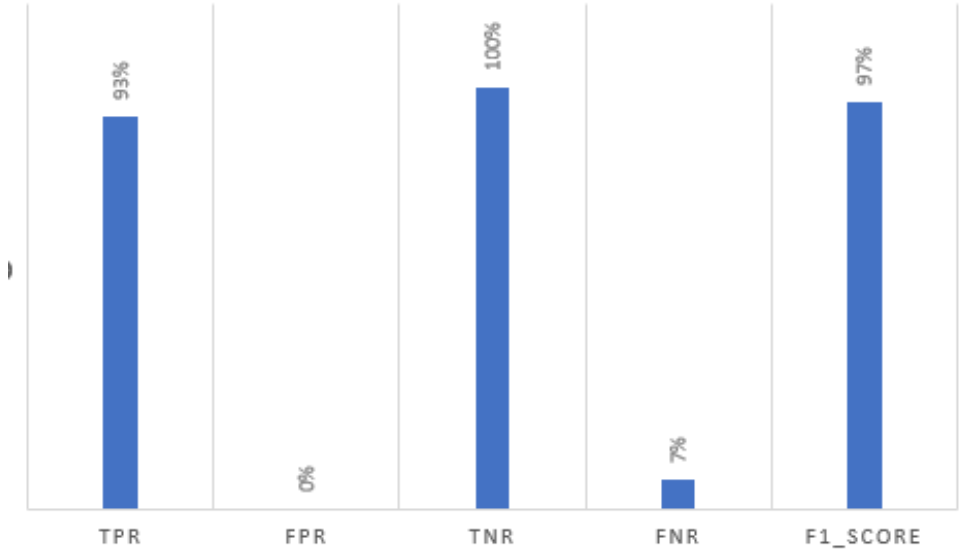


FIGURE III.9 – Mesures de Performance d'un Modèle de Classification LSTM.

III.6 Comparaison détaillée des performances des modèles CNN, ANN et LSTM

La figure ci-dessus présente une comparaison des performances de trois modèles d'apprentissage profond à travers plusieurs métriques d'évaluation. Les indicateurs mesurés incluent la précision (**Accuracy**), le taux de vrais positifs (**TPR**), le taux de faux positifs (**FPR**), le taux de vrais négatifs (**TNR**), le taux de faux négatifs (**FNR**) et le **score F1**. Ces métriques permettent d'évaluer la capacité de chaque modèle à classer correctement les données, à détecter les anomalies ou événements d'intérêt, et à limiter les erreurs. Cette visualisation met en évidence les différences de performances entre les trois architectures, soulignant les points forts et faibles de chacune selon les critères choisis.

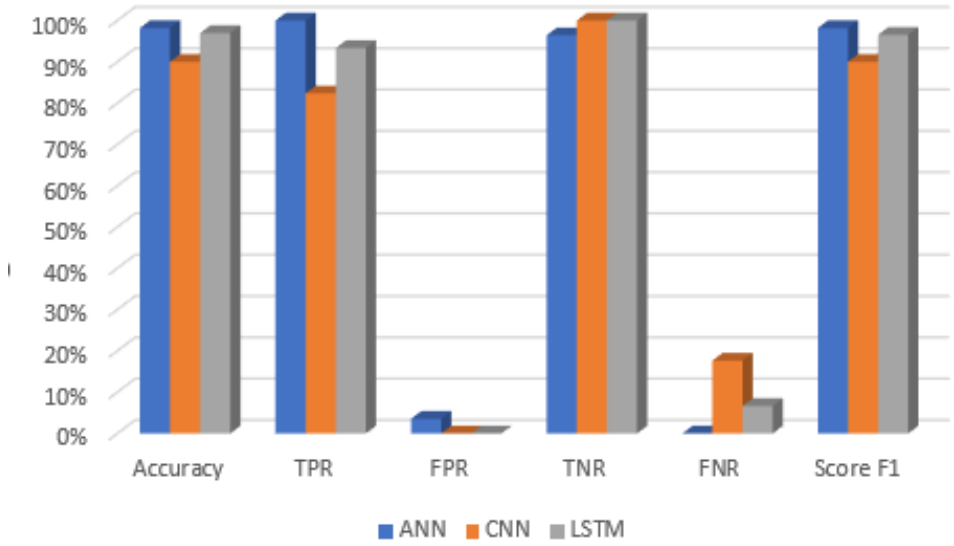


FIGURE III.10 – comparaison des Performances .

1. Exactitude (Accuracy)

Le CNN affiche la plus faible exactitude parmi les trois modèles, avec seulement 90 % de bonnes prédictions globales. L'ANN et le LSTM montrent des performances nettement supérieures, très proches l'une de l'autre, avec environ 96,67 % de précision globale. Le ANN est légèrement meilleur que LSTM avec pourcentage de 98,18 % sur ce critère.

3. Taux de vrais positifs (TPR)

L'ANN atteint la sensibilité maximale (100 %), détectant tous les cas positifs. Le LSTM suit avec un très bon taux de 93,33 %. Le CNN est loin derrière, avec seulement 82,35 %, ce qui indique qu'il rate une proportion non négligeable de cas positifs, ce qui peut être problématique selon l'application.

4. Taux de faux positifs (FPR)

Les modèles CNN et LSTM ne génèrent aucun faux positif, garantissant qu'aucun cas négatif n'est mal classé comme positif. En revanche, l'ANN présente un taux de faux positifs légèrement supérieur (3,57 %), ce qui peut engendrer quelques fausses alertes, un compromis à prendre en compte selon les exigences du système.

5. Taux de vrais négatifs (TNR)

CNN et LSTM obtiennent une spécificité parfaite, ils classent tous les cas négatifs correctement. L'ANN, avec 96,43 %, reste très performant mais présente une petite marge d'erreur, cohérente avec son taux de faux positifs plus élevé.

6. Taux de faux négatifs (FNR)

Le CNN est pénalisé par un taux de faux négatifs élevé (17,65 %), ce qui signifie qu'il manque près d'un cinquième des cas positifs. L'ANN est parfait ici, aucun cas positif manqué. Le LSTM fait aussi très bien avec seulement 6,67 % de faux négatifs, mais reste moins bon que l'ANN sur ce point.

7. Score F1

Le score F1 synthétise la précision et le rappel. L'ANN obtient le meilleur score (98,18 %), suivi du LSTM (96,66 %) qui reste très performant. Le CNN arrive nettement derrière (90,32 %), confirmant ses performances plus faibles, surtout sur la détection des positifs

III.6.1 Discussion des resultats

l'ANN offre une performance globale très robuste, tandis que le LSTM apporte un avantage notable dans la réduction des erreurs de classification négative, conférant aux deux modèles des avantages complémentaires importants. Le CNN reste le moins performant, notamment à cause de ses lacunes en sensibilité :

1)Réduction des faux positifs : Le LSTM et le CNN affichent un taux de faux positifs nul, ce qui signifie qu'aucun cas négatif n'a été incorrectement classé comme positif. En comparaison, l'ANN présente un taux légèrement plus élevé de faux positifs, bien que faible, indiquant un risque minimal mais existant de fausses alertes.

2)Performance équilibrée : L'ANN atteint une sensibilité parfaite, détectant tous les cas positifs sans faux négatifs, ce qui lui permet d'avoir le meilleur équilibre global entre précision et rappel, illustré par son score F1 le plus élevé. Le LSTM suit de près, avec un excellent score F1 et une très bonne sensibilité, tandis que le CNN montre un déséquilibre notable, manquant une part importante des cas positifs, ce qui impacte négativement son score F1.

On conclut que l'ANN et le LSTM offrent les meilleures performances globales, chacun avec ses points forts : l'ANN se distingue par un équilibre parfait entre précision et rappel, tandis que le LSTM assure une excellente détection avec très peu d'erreurs.

III.6.2 Comparaison par nombre époques

Le tableau III.4 présente une comparaison des performances des trois architectures de réseaux de neurones suivantes : ANN, CNN et LSTM, appliquées à une tâche de classification binaire. Pour chaque modèle, nous reportons la précision obtenue lorsque la perte de validation est minimale, ainsi que le nombre d'époques nécessaires pour atteindre cette performance.

Les résultats indiquent que le modèle ANN présente la meilleure efficacité, atteignant une précision de 98,18 % en seulement 332 époques. En comparaison, le CNN atteint également 90 % de précision mais en 91 époques, tandis que le LSTM atteint 96,67 % en 112 époques. Cela montre que le ANN est non seulement plus précis, mais aussi plus rapide à converger dans ce contexte binaire.

TABLE III.4 – Précision de validation et nombre d'époques de convergence pour chaque modèle dans une tâche de classification binaire.

Modèle	Précision (%)	Époques
ANN	98,18	332
CNN	90	91
LSTM	96,67	112

Abréviation : Époques = nombre d'époques où la perte de validation est minimale.

III.6.3 Comparaison avec des travaux antérieurs

D'autres travaux ont mis en œuvre des algorithmes d'apprentissage automatique classiques pour l'analyse de la qualité de service dans les réseaux 5G. En se basant sur des indicateurs tels que la latence, la bande passante et l'allocation des ressources, ces études ont exploité des modèles comme Random Forest ou SVM pour classifier les niveaux de service ou détecter des anomalies. L'une des expériences notables a permis d'atteindre une précision de 96,5 % en utilisant la validation croisée à cinq plis (5-fold cross-validation), ce qui démontre l'efficacité des méthodes supervisées dans la détection des dégradations du service [61]. Parmi les études comparatives accessibles en ligne, deux travaux disponibles sur la plateforme Kaggle se distinguent par leur pertinence. Le premier est celui de Ioannis Panagiotopoulos [62], où plusieurs modèles d'apprentissage automatique (Random Forest, SVM, KNN, XGBoost) ont été évalués sur des données QoS simulées, avec une précision moyenne maximale avoisinant les 96,7 %. Le second travail est celui de Ioannis Bartsiokas [63], qui a comparé plusieurs techniques supervisées et obtenu également des performances proches de 96,5 % pour les meilleurs modèles classiques, bien que sans recourir à des architectures profondes comme LSTM ou CNN. Ces expériences confirment l'efficacité des modèles classiques dans des contextes simples ou bien maîtrisés, mais montrent aussi leurs limites face à des données plus dynamiques ou bruitées. La figure suivante synthétise les résultats moyens issus de ces deux travaux ainsi que ceux obtenus dans notre étude. En comparaison, notre modèle de réseau de neurones artificiels (ANN) a enregistré la meilleure performance avec une précision de 98,18 %, suivi par le modèle LSTM avec 96,67 %, tandis que le CNN a atteint 90 %. Ces résultats soulignent la supériorité des approches basées sur le deep learning, notamment l'ANN, dans la capture des relations complexes et non linéaires présentes dans les données QoS des réseaux 5G.

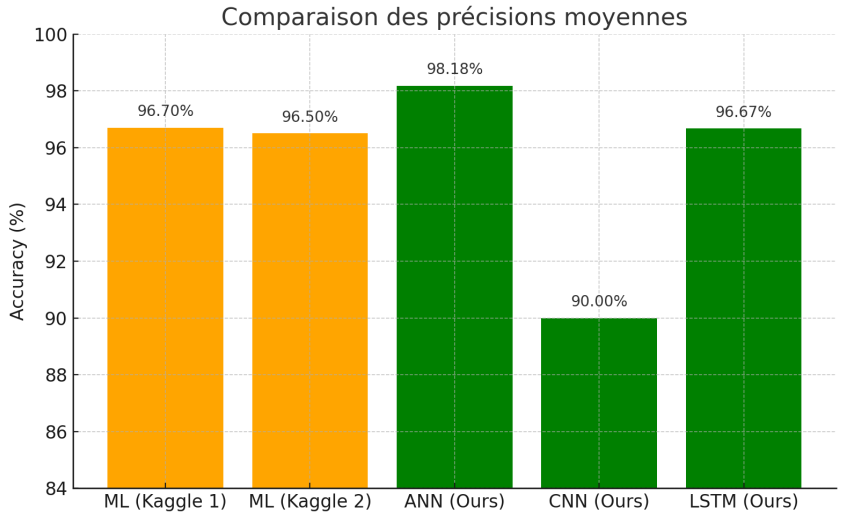


FIGURE III.11 – Comparaison des précisions moyennes entre notre travail et des travaux antérieurs.

III.7 Conclusion

Ce chapitre a exploré la détection d'anomalies et la classification de la qualité de service (QoS) dans les réseaux 5G à l'aide de modèles d'apprentissage profond, notamment ANN, CNN et LSTM. Parmi ces modèles, l'ANN a obtenu les meilleurs résultats, avec une sensibilité de 100 %et un F1-score de 98,18 %, démontrant un excellent équilibre entre précision et rappel.

Le LSTM a également montré de bonnes performances, notamment aucun faux positif et une bonne capacité de généralisation, grâce à sa structure adaptée aux données temporelles. Le CNN, bien qu'efficace sur certains aspects, reste en retrait à cause d'une sensibilité plus faible et d'un taux plus élevé de faux négatifs.

En conclusion, l'ANN est le modèle le plus performant pour la détection d'anomalies dans les réseaux 5G, selon les résultats obtenus.

CONCLUSION GÉNÉRALE

Ce mémoire a mis en évidence l'impact majeur de la cinquième génération de réseaux mobiles (5G) sur l'écosystème des communications modernes, en soulignant ses avancées technologiques ainsi que les défis complexes qu'elle engendre en matière de gestion et de qualité de service (QoS). Face à ces enjeux, l'intégration de l'intelligence artificielle, et plus particulièrement du deep learning, s'est révélée une approche pertinente pour automatiser la surveillance et la détection des anomalies dans les réseaux 5G.

Dans ce contexte, nous avons proposé un cadre basé sur l'utilisation de réseaux de neurones profonds pour la classification de la QoS. Parmi les modèles testés ANN,LSTM,CNN . Le réseau de neurones artificiels (ANN) a démontré les meilleures performances globales, avec un score F1 remarquable de 98,18 %, attestant de sa capacité à équilibrer précision et rappel de manière optimale. Ce résultat met en évidence la pertinence de l'ANN comme outil fiable pour détecter les dégradations de service et anticiper les anomalies de manière proactive.

En conclusion, notre étude confirme l'efficacité des approches basées sur le deep learning dans la gestion intelligente des réseaux mobiles de nouvelle génération. L'implémentation de tels modèles ouvre la voie à des systèmes de surveillance plus autonomes, réactifs et adaptés aux exigences croissantes de la 5G. Les perspectives futures pourront s'appuyer sur ces fondations pour développer des solutions encore plus robustes et intégrées. Pour les travaux futurs, nous proposons de développer un système de détection d'anomalies capable d'identifier les problèmes survenant dans un système de supervision en temps réel et de manière dynamique dans un réseau mobile réel.

BIBLIOGRAPHIE

[1] Hamza Yahiaoui and Redouane Lounnas. Étude des performances des réseaux 4g lte. Mémoire de master académique, Université Mouloud Mammeri de Tizi-Ouzou, Tizi-Ouzou, Algérie, 2017.

[2] Timo Halonen, Javier Romero, and Juan Melero. *GSM, GPRS and EDGE performance : evolution towards 3G/UMTS*. John Wiley & Sons, 2004.

[3] Ismaël Houssein. Comprendre le réseau gsm (partie 2), novembre 2011. Consulté le 1 juin 2025.

[4] Auteur inconnu. Migration 2g vers la 4g. `https://fr.scribd.com/document/283731400`. Mémoire de fin d'études, consulté en avril 2025.

[5] Technologue Pro. Architecture d'un réseau radio mobile gsm, 2025. Consulté le : 15 février 2025.

[6] C. Demoulin and M. Van Droogenbroeck. Principes de base du fonctionnement du réseau gsm. *Revue de l'AIM*, 4, 2004.

[7] M. E. Ketfi. Infrastructure d'un réseau gsm. Mémoire de master 2 – itlc, Université de M'sila, non précisé. [Consulté le : 15 février 2025].

[8] M. Messaoui. Développement d'un outil d'aide à la gestion des capacités des équipements bss en mapbasic sous le sig mapinfo, 2011.

[9] A. Ali and C. Slimane. Configuration d'un centre de commutation mobile (msc) dans un réseau gsm. Mémoire de fin d'études, Université Mouloud Mammeri, 2008.

[10] P. Pattem. Research on detection and prevention of rogue base stations in the 5g network. Mémoire de fin d'études, Université de l'Alberta, 2022.

[11] Ajay R. Mishra. *Advanced Cellular Network Planning and Optimisation : 2G/2.5G/3G... Evolution to 4G*. John Wiley & Sons, Chichester, UK, 2006.

[12] H. Kheddar. From 2g to 4g mobile network : Architecture and key performance indicators. *arXiv preprint*, 2022. arXiv :2210.00642.

[13] O. Hasnaoui. Umts. Mémoire de fin d'études, Université Djillali Liabès, 2006.

[14] Hocine Benhalla. Chapitre 01 : Introduction aux systèmes d'exploitation, 2016. Université Abou Bekr Belkaid - Tlemcen, Cours sur les systèmes d'exploitation.

[15] R. Bonasse and H. L. Lissouck. Simulation de l'interopérabilité des nouvelles architectures d'émetteurs en système radio mobile. Mémoire de fin d'études, ENSET, 2014. Master I Professionnel en Réseaux Télécommunications.

[16] H. G. Alban. Simulation de l'interopérabilité des nouvelles architectures d'émetteurs en système radio mobile. Mémoire de fin d'études, ENSET, 2014. Master I Professionnel en Réseaux Télécommunications.

[17] M. Hasni. Rapport de projet de fin d'études optimisation globale de la capacité d'un réseau umts par la méthode d'antenna-ret, 2006.

[18] Alban Hounton. Études des performances des réseaux 4g, 2015. Mémoire de Licence Professionnelle, UATM GASA FORMATION.

[19] E. Jammazi. Optimisation d'un réseau pilote 4g pour tunisie télécom. Master's thesis, Projet de fin d'études en Réseaux et Communications, June 2013.

[20] H. Bouchentouf and R. B. Stambouli. Étude des performances des réseaux 4g (lte). Mémoire de master en télécommunication, option réseaux mobiles et services (rms), Université de Tlemcen, June 2013.

[21] Efort. LTE + SAE = EPS, Principes et Architecture, 2009. Consulté en 2009.

[22] E. M. Roche and W. Townes. La prochaine génération de téléphonie mobile (5g) et ses implications (infrastructure, réglementation). *Netcom. Réseaux, communication et territoires*, 32(1/2) :139–162, 2018.

[23] Autorité de Régulation des Communications Électroniques et des Postes (Arcep). Les enjeux de la 5g, March 2017. Consulté en mars 2017.

[24] D. Kaoutar and D. Warda. Étude de massive mimo pour 5g. Master's thesis, Université Hamma Lakhdar - Eloued, 2019. Mémoire de Master.

[25] Ricon Mobile Inc. Communication Technology of the Future 5G. `https://riconmobile.com/public/blog/communication-technology-of-the-future-5g`, 2025. [En ligne, consulté le 9 juin 2025].

[26] M. BellaHSENE and I. MEDJADI. Étude et ingénierie du déploiement du réseau 5g nsa dans la bande c. Spécialité : Systèmes de Télécommunications, Université SAAD DAHLAB BLIDA 1, Département d'Électronique, 2023. Mémoire de Master,.

[27] Andrea Detti, Stefano Salsano, and Nicola Blefari-Melazzi. Functional architecture. In *5G Italy White eBook : From Research to Market*. CNIT, 2018. [Online]. Available : `https://www.5gitaly.eu/2018/wp-content/uploads/2019/01/5G-Italy-White-eBook-Functional-architecture.pdf`.

[28] IPLOOK Technologies Co., Ltd. Fonction de référentiel réseau (NRF). https://fr.iplook.com/productos/5gc/nrf.html, 2025. [En ligne, consulté le 9 juin 2025].

[29] IoT Industriel. Réseau 5g : L'essentiel à savoir, 2025. [Online]. Available : https://iotindustriel.com/autres/technologies/4g-5g/reseau-5g-lessentiel-a-savoir/.

[30] R. Chataut and R. Akl. Massive mimo systems for 5g and beyond networks—overview, recent trends, challenges, and future research direction. *Sensors*, 20(10) :2753, 2020.

[31] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. Massive mimo for next generation wireless systems. *IEEE Communications Magazine*, 52(2) :186–195, 2014.

[32] Chao-Kai Wen, Wan-Ting Shih, and Shi Jin. Deep learning for massive mimo csi feedback. *IEEE Wireless Communications Letters*, 9(1) :60–64, 2020.

[33] Zafar Abbas, Nadeem Javaid, Zahoor Ali Khan, Umar Qasim, and Abdullah A Alrajeh. Channel estimation techniques for massive mimo systems : A survey. *IEEE Access*, 5 :19000–19020, 2017.

[34] Noha Hassan and Xavier Fernando. Massive mimo wireless networks : An overview. *Electronics*, 6(3) :63, 2017.

[35] Autorité de régulation des communications électroniques et des postes (ARCEP). Les enjeux de la 5g. Technical report, RÉPUBLIQUE FRANÇAISE, March 2017. Consulté en avril 2025.

[36] Assane Ngom. *Conception de petits réseaux d'antennes reconfigurables ou "Small Cells" pour le standard 5G*. PhD thesis, COMUE Université Côte d'Azur (2015-2019) ; Université Cheikh Anta Diop (Dakar), 2019. [En ligne, consulté le 9 juin 2025].

[37] Faker Skandrani. 5g sa vs 5g nsa : Quelles différences ?, April 2022. Consulté le 20 avril 2025.

[38] telecomHall Forum. 5G SA vs 5G NSA vs 5G DSS. https://www.telecomhall.net/t/5g-sa-vs-5g-nsa-vs-5g-dss/18225/3, 2022. [En ligne, consulté le 9 juin 2025].

[39] Haissam Haidar-Ahmad, Mohamed Alamin Hamid Salih, Alison Luangrath, Théo Sede, and Lucas Zugaj. Intelligence artificielle et imagerie médicale, 2022. Consulté le 9 juin 2025.

[40] Jérémy Robert. Machine learning : Définition, fonctionnement, utilisations, novembre 2020. Consulté le 12 mai 2025.

[41] Indranath Chatterjee. *Machine Learning and Its Application : A Quick Guide for Beginners*. Bentham Science Publishers, 2021.

[42] R. Chrystal. Five machine learning types to know. https://www.ibm.com/think/topics/machine-learning-types, 2025. Consulté le 12 mai 2025.

[43] J. Zhang, X. Liu, and H. Wang. Deep learning and machine learning in engineering. In *MATEC Web of Conferences*, volume 277, page 02035, 2019.

[44] Accelirate. Uipath's strategy for ai powered automation at work. https://www.accelirate.com/uipaths-strategy-for-ai-powered-automation-at-work/, November 2023.

[45] E-SPIN Group. Nom de l'article ici, 2025. Consulté le 21 mai 2025.

[46] Data Bird. Réseaux de neurones : Comprendre leur fonctionnement, 2022. [Consulté le : 17 mai 2025].

[47] D Cat laz. Neurone.svg - schéma d'un neurone typique. https://fr.wikipedia.org/wiki/Fichier:Neurone.svg, 2022. Image disponible sous licence Creative Commons Attribution – Partage dans les Mêmes Conditions 4.0 International (CC BY-SA 4.0).

[48] Pat Nakamoto. *Neural Networks & Deep Learning : Deep Learning explained to your granny – A visual introduction for beginners who want to make their own Deep Learning Neural Network*. Independently published, 2019.

[49] Jeslur Rahman. Simply understanding artificial neural networks (ann) : Deep learning. *Medium*, 2024. Published on February 10, 2024.

[50] Timothy Baldwin and Wei Xu. Deep learning and neural networks : The new ai, 2017. Consulté le 9 juin 2025.

[51] Mohamed Abd Elmoumen Djaballah. Système de prédiction de la consommation d'énergie basé deep learning. Mémoire de master, Université de 8 Mai 1945 – Guelma, Département d'Informatique, Algérie, 2021.

[52] IBM. Qu'est-ce qu'un réseau de neurones convolutifs ? https://www.ibm.com/fr-fr/think/topics/convolutional-neural-networks, 2025. [Consulté le : 17 mai 2025].

[53] Ottavio Calzone. An intuitive explanation of lstm, 2022. [Consulté le : 17 mai 2025].

[54] Oumaima Hourrane. Réseaux neuronaux récurrents et lstm. *arXiv preprint arXiv :2505.01234*, 2025. Faculty of Sciences Ben M'Sik, Hassan II University, Casablanca, Morocco.

[55] Sourabh Padhy. Lstm | derivation of back propagation through time. *GeeksforGeeks*, 2025. Published on April 5, 2025.

[56] Matrice de confusion : qu'est-ce que c'est et comment l'utiliser ? https://datascientest.com/matrice-de-confusion, n.d. [Consulté le : 16 mai 2023].

[57] Jérémy Robert. Mixture of experts : L'approche qui pourrait façonner l'avenir de l'ia. *DataScientest*, 2020. Publié le 13 mars 2025.

[58] Omar Sobhy. 5G Quality of Service. https://www.kaggle.com/datasets/omarsobhy14/5g-quality-of-service, 2021. Dataset disponible sur Kaggle, consulté le 3 juin 2025.

[59] Fouad Sabry. *Logiciel Aphélie : Unlocking Vision : Explorer les profondeurs du logiciel Aphelion*. Un Milliard De Personnes Informées, 2024. Disponible sur Booktopia : https://www.booktopia.com.au/aphelion-software-fouad-sabry/ebook/6610000569045.html.

[60] Inc. The MathWorks. Matlab and simulink release notes. MathWorks, Natick, MA, USA, 2025. Consulté le 22 mai 2025.

[61] Omar Sobhy. 5G Quality of Service. https://www.kaggle.com/datasets/omarsobhy14/5g-quality-of-service, 2023. Accessed : 2025-06-14.

[62] Ioannis Panagiotopoulos. Mcml 24/25 lab1 - introduction to ml with qos data. https://www.kaggle.com/code/ioannispan/mcml-24-25-lab1-intro, 2024. Consulté en juin 2025.

[63] Ioannis Bartsiokas. Amct - introductory lab 08.11.2023. https://www.kaggle.com/code/ioannisbartsiokas/amct-introductory-lab-8-11-2023, 2023. Consulté en juin 2025.