



N° d'ordre : /2024

**Projet de Fin d'Etudes
Présenté en Vue de l'Obtention
du Diplôme de Master Sciences et Techniques
en Informatique Décisionnelle**

**Vers une gestion sécurisée de
l'énergie au sein des Smart Grids**

Réalisé par
Sana KHAYOU

Sous l'encadrement de
Abdellatif HAIR

Soutenue le **27 Juin 2024** devant le jury composé de :

- | | | |
|-----------------------|---|--|
| Mme. Najlae EL IDRISI | : | Professeur Chercheur, FST de Béni Mellal |
| Mr. Youssef SAADI | : | Professeur Chercheur, FST de Béni Mellal |
| Mr. Mohammed ERRITALI | : | Professeur Chercheur, FST de Béni Mellal |

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Avant-propos

Ce mémoire s'inscrit dans le cadre de l'obtention du diplôme de Master des sciences et techniques en Informatique Décisionnelle de la Faculté des Sciences et Techniques de Béni Mellal, relevant de l'Université Sultan Moulay Slimane. Il se concentre sur l'intégration de l'intelligence artificielle et de la Blockchain dans les réseaux électriques intelligents, également connus sous le nom de Smart Grids. L'idée de cette recherche est née du constat que les avancées en intelligence artificielle et en Blockchain peuvent significativement optimiser la gestion de l'énergie.

Ce mémoire représente l'aboutissement de plusieurs mois de recherches intensives et de travaux pratiques. Il reflète non seulement les connaissances acquises au cours de mon parcours universitaire, mais également les compétences techniques et analytiques développées à travers la résolution de problématiques complexes. Les Smart Grids, au cœur de ce projet, illustrent la convergence des avancées en technologies de l'information et de la communication avec les besoins énergétiques contemporains.

Je tiens à exprimer ma profonde gratitude envers l'ensemble de la communauté universitaire, en particulier aux professeurs et encadrants de la Faculté des Sciences et Techniques de Béni Mellal, pour leur soutien indéfectible et leurs conseils avisés tout au long de cette aventure académique. Leur expertise et leur disponibilité ont été essentielles à la réussite de ce projet.

Je remercie également mes collègues étudiants et tous ceux qui m'ont apporté leur aide précieuse et leurs encouragements.

Ce travail se veut une modeste contribution à la recherche dans le domaine des Smart Grids et aspire à inspirer d'autres étudiants et chercheurs à poursuivre des innovations dans ce domaine prometteur.

Dédicace

À mes chers parents, Mustapha KHAYOU et Hafida EL HANSALI,

Ce modeste travail est le fruit de votre amour, de votre tendresse et de vos sacrifices. Votre soutien indéfectible et vos conseils avisés ont été le phare qui a guidé mes pas. Aucune dédicace ne pourrait pleinement exprimer ma reconnaissance pour tout ce que vous avez fait pour moi depuis mon enfance. Puissiez-vous voir dans ces lignes le reflet de mon amour profond et de ma gratitude éternelle.

À ma sœur bien-aimée, Nada KHAYOU,

Pour ta complicité, ta tendresse et ton soutien, pour être ma confidente et mon amie. Ta présence dans ma vie est une source constante de joie et d'inspiration.

À toute ma famille,

Pour votre affection et votre encouragement, qui ont été mes piliers dans toutes les étapes de ma vie.

À toutes mes amies et à ceux que j'aime,

Pour votre amitié sincère, votre soutien et votre encouragement pendant ces deux années intenses de formation.

À mes Professeurs du Master Informatique Décisionnelle,

Pour leur enseignement précieux, leur patience et leur engagement envers notre réussite.

À toute la promotion Master ID 2024,

Pour les moments d'entraide, d'apprentissage partagé et de camaraderie qui ont enrichi cette expérience.

À l'administration de la Faculté des Sciences et Techniques de Béni Mellal,

Pour leur travail et leur dévouement qui ont facilité notre parcours éducatif.

Et à vous, chers lecteurs,

Pour l'intérêt que vous portez à ce travail et pour l'opportunité de partager mes recherches et mes réflexions.

Sana KHAYOU

Remerciements

Avec un profond respect et une gratitude infinie, j'adresse mes plus sincères remerciements à toutes les personnes bienveillantes ayant contribué à l'aboutissement de mon projet de fin d'études.

*En premier lieu, je remercie **Allah** pour Sa guidance et Son soutien essentiels tout au long de ce chemin académique. **Mes parents**, piliers de ma vie, reçoivent mon éternelle reconnaissance pour leur appui sans faille et leurs encouragements incessants.*

*Je tiens à exprimer mon respectueux hommage à mon encadrant, **Mr Abdellatif HAIR**, pour son accompagnement inestimable, sa disponibilité, ses conseils avisés et son apport intellectuel déterminant pour la réussite de ce projet.*

Je manifeste également ma considération envers les membres du jury pour leur attention et leurs observations pertinentes, ainsi que mes professeurs pour leur dévouement et leur accompagnement constant.

Enfin, j'adresse ma gratitude à tous ceux qui m'ont soutenu, de près ou de loin, pendant la réalisation de ce projet de fin d'études.

Résumé

Ce mémoire se consacre à l'étude, la conception et la mise en œuvre de techniques d'analyse de séries temporelles, de Deep Learning et de la Blockchain pour la détection de fraude dans les réseaux électriques intelligents. En utilisant les données de consommation électrique fournies par la State Grid Corporation of China, nous avons appliqué des modèles ARIMA et des réseaux de neurones convolutifs pour identifier des anomalies potentielles indiquant un vol d'énergie. Nous avons également exploré l'intégration théorique de la Blockchain et Deep Learning pour assurer la sécurité et l'intégrité des données. Nous avons procédé à une série d'analyses de données, incluant le nettoyage des données, les tests de stationnarité et l'optimisation des paramètres des modèles.

Les résultats obtenus ont été évalués à l'aide de diverses métriques de performance, démontrant l'efficacité relative des méthodes employées pour détecter les comportements frauduleux. Ce travail vise à contribuer à l'amélioration des systèmes de surveillance des réseaux intelligents en proposant des solutions robustes pour la détection précoce des fraudes.

Mots clés : Réseau intelligent, Vol d'énergie, Intelligence artificielle, Deep Learning, ARIMA, Blockchain, Détection de fraude.

Abstract

This thesis is dedicated to the study, design, and implementation of time series analysis, Deep Learning, and Blockchain techniques for fraud detection in smart electrical grids. Using electricity consumption data provided by the State Grid Corporation of China, we applied ARIMA models and convolutional neural networks to identify potential anomalies indicating energy theft. We also explored the theoretical integration of Blockchain to ensure data security and integrity. We conducted a series of data analyses, including data cleaning, stationarity tests, and model parameter optimization.

The results were evaluated using various performance metrics, demonstrating the relative effectiveness of the methods employed to detect fraudulent behaviors. This work aims to contribute to the improvement of smart grid monitoring systems by proposing robust solutions for early fraud detection.

Keywords: Smart grid, Energy theft, Artificial intelligence, Deep-learning, ARIMA, Blockchain, Fraud detection.

ملخص

ُخصصت هذه الأطروحة لدراسة وتصميم وتنفيذ تحليل السلسل الزمنية والتعلم العميق وتقنيات Blockchain للكشف عن الاحتيال في الشبكات الكهربائية الذكية. باستخدام بيانات استهلاك الكهرباء المقدمة من شركة State Grid Corporation الصينية، قمنا بتطبيق نماذج ARIMA و التعلم العميق لتحديد الحالات الشاذة المحتملة التي تشير إلى سرقة الطاقة. اقترحنا أيضًا نموذجًا يدمج بين التعلم العميق Blockchain لضمان أمن البيانات وسلامتها. أجرينا سلسلة من تحليل البيانات، بما في ذلك تنظيف البيانات واختبار الثبات وتحسين بaramترات النموذج.

تم تقييم النتائج باستخدام مقاييس أداء مختلفة، مما يدل على الفعالية النسبية للطرق المستخدمة للكشف عن السلوكيات الاحتيالية. يهدف هذا العمل إلى المساهمة في تحسين أنظمة مراقبة الشبكة الذكية من خلال اقتراح حلول قوية للكشف المبكر عن الاحتيال.

الكلمات المفتاحية: الشبكة الذكية، سرقة الطاقة، الذكاء الاصطناعي، التعلم العميق، ARIMA، Blockchain، كشف الاحتيال.

Table des matières

Liste des figures	13
Liste des tableaux	15
Glossaire	16
Liste des abréviations.....	18
Introduction générale	20

Chapitre I Fondements et Enjeux des Smart Grids

I.1 Introduction	23
I.2 Réseaux électriques intelligents (Smart Grid)	24
I.2.1 Définition des Smart Grids	24
I.2.2 Réseau Électrique Traditionnel vs Smart Grids.....	25
I.2.3 Architecture des Smart Grids.....	26
I.2.4 Architecture selon le Modèle SGAM	30
I.2.5 Les principales composantes d'un Smart Grid.....	32
I.2.5.1 Communication.....	33
I.2.5.2 Mesurage : Metering	34
I.3 La sécurité dans les Smart Grids	36
I.3.1 Menaces et Vulnérabilités dans les Smart Grids	36
I.3.1.1 Attaques par Injection de Fausses Données (FDIA).....	37
I.3.1.2 Attaque par Déni de Service (DoS).....	37
I.3.1.3 Attaque par Écoute (EVD)	38
I.3.1.4 Attaque par Usurpation d'Identité (IMP).....	38
I.3.1.5 Attaque par Rejeu (REP).....	39
I.3.1.6 Attaque par Répudiation (RPD)	39
I.3.1.7 Attaque par Compromission de Nœud (CMP).....	40
I.3.2 Vol d'Énergie	40
I.3.2.1 Types de vol d'énergie	41
I.3.2.2 Conséquences du Vol d'Énergie	42
I.4 Exigences de Sécurité pour les Réseaux de Comptage Intelligent	42

I.5 Conclusion.....	46
---------------------	----

Chapitre II Blockchain et Intelligence Artificielle au Service des Smart Grids

II.1 Introduction	47
II.2 Analyse des Séries Temporelles pour la Sécurité des Smart Grids	48
II.2.1 Introduction à l'analyse des séries temporelles	48
II.2.1.1 Définition d'une série temporelle.....	48
II.2.1.2 Composantes des Séries Temporelles	50
II.2.1.3 Types de données	51
II.2.1.4 Modélisations de Base pour les Séries Temporelles	51
II.2.2 Stationnarité	53
II.2.2.1 Tests de Stationnarité	54
II.2.3 Les modèles de séries temporelles	55
II.2.3.1 Modèle AutoRegressif (AR).....	55
II.2.3.2 Modèle de Moyenne Mobile (MA)	55
II.2.3.3 Modèle Autorégressif à Moyenne Mobile (ARMA)	55
II.2.3.4 Modèle Autorégressif Intégré à Moyenne Mobile (ARIMA)	56
II.2.3.5 Modèle Saisonnier ARIMA (SARIMA) :	56
II.2.4 Approche de Box et Jenkins	56
II.2.4.1 AutoCorrelation Function : ACF	57
II.2.4.2 Partial Autocorrelation Function : PACF	58
II.2.4.3 Théorie Générale de l'ACF et du PACF des Modèles ARIMA	58
II.2.5 Modélisation ARIMA et Détection des Attaques dans les Smart Grids....	59
II.3 Intelligence artificielle pour la Sécurité des Smart Grids	60
II.3.1 Définition.....	61
II.3.2 Les Modèles Deep Learning	62
II.3.2.1 Les réseaux neuronaux convolutionnels	63
II.3.2.2 Les Réseaux Neuronaux Récurrents (RNN).....	64
II.3.2.3 LSTM	65
II.3.3 Protection contre le vol d'énergie basé sur l'IA	66

II.3.3.1 Wide & Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids[15]	66
II.3.3.2 Electricity Theft Detection in Power Grids with Deep Learning and Random Forests [16]	68
II.4 Blockchain pour la Sécurité des Smart Grids	69
II.4.1 Technologie Blockchain	70
II.4.1.1 L'arbre de Merkle.....	71
II.4.1.2 Principe de fonctionnement de la Blockchain	72
II.4.1.3 Les protocoles de consensus	73
II.4.1.4 Les catégories de la Blockchain.....	75
II.4.2 Renforcement de la Sécurité des SG grâce à la Blockchain	76
II.4.2.1 Évolution vers un Réseau Intelligent Décentralisé	76
II.4.2.2 Motivations de l'Intégration de la Blockchain dans les SG	78
II.4.3 Sécurisation des Systèmes de Comptage Intelligent par une Architecture Blockchain Multitier [21]	78
II.4.3.1 Architecture proposée	79
II.4.3.2 Mécanisme de Fonctionnement	80
II.4.3.3 Mécanisme de Consensus.....	80
II.4.4 Cadre de protection des données basé sur la Blockchain distribuée pour les systèmes électriques modernes contre les cyberattaques [22]	80
II.4.4.1 Architecture proposée	81
II.4.4.2 Mécanisme de Fonctionnement :	81
II.4.4.3 Mécanisme de Consensus :.....	82
II.4.5 Analyse Comparative des Solutions proposées	83
II.5 Exploration des Approches Combinant Blockchain et Deep Learning pour la Sécurité des Smart Grids	84
II.5.1 Cadre de Protection de la Confidentialité Basé sur la Blockchain et le Deep Learning pour les Réseaux Électriques Intelligents [23].....	84
II.5.2 DeepCoin : Un Nouveau Cadre d'Échange d'Énergie Basé sur le Deep Learning et la Blockchain pour les Réseaux Intelligents [1]	85
II.5.3 Comparaison entre les solutions proposées.....	86
II.6 Conclusion.....	87

Chapitre III Élaboration des Approches pour Détecter les Fraudes dans les Smart Grids

III.1 Introduction	88
III.2 Architecture générale AMI	89
III.2.1 Smart Meter Layer	90
III.2.2 Concentrator Layer	90
III.2.3 Meter Data Management Layer	90
III.2.4 Enterprise Services.....	90
III.2.5 Relever le Défi du Vol d'Énergie dans l'AMI	91
III.3 Détection de la fraude dans les réseaux intelligents utilisant ARIMA	92
III.3.1 Paramétrage et Identification des Pics de Consommation	92
III.3.2 Étiquetage des Données	94
III.4 Détection de la fraude dans les réseaux intelligents utilisant Deep Learning... 	95
III.4.1 Choix du modèle DL utilisé.....	95
III.4.2 Collecte et Prétraitement des Données.....	96
III.4.3 Détection des Fraudes à l'aide du modèle CNN-1D	98
III.4.4 Détection des Fraudes à l'aide du modèle CNN-2D	99
III.4.5 Détection des Fraudes à l'aide du modèle CNN-LSTM	100
III.5 Intégration de la Blockchain et du Deep Learning pour la Détection des Fraudes	101
III.5.1 Vue d'ensemble de l'Architecture	102
III.5.2 Fonctionnalités et Infrastructure des Compteurs de Réseau NAN	103
III.5.2.1 Caractéristiques Fonctionnelles.....	103
III.5.2.2 Communication et Agrégation des Données	103
III.5.3 Blockchain niveau NAN	103
III.5.4 Blockchain niveau WAN/FAN.....	109
III.5.5 Réactions Automatisées via les Smart Contracts	111
III.6 Conclusion.....	112

Chapitre IV Résultats de la Détection des Fraudes dans les Smart Grids

IV.1 Introduction	114
IV.2 Environnement de travail.....	114
IV.2.1 Outils d'implémentations	114
IV.2.2 Configuration Matérielle	116

IV.3 Résultats de la détection des fraudes via le Deep Learning	116
IV.3.1 Analyse des Données	116
IV.3.2 Résultats CNN-1D.....	120
IV.3.2.1 Choix des Hyper paramètres du Modèle.....	120
IV.3.2.2 Résultats et métriques d'évaluation.....	121
IV.3.3 Résultats CNN-2D.....	122
IV.3.3.1 Choix des Hyper paramètres du Modèle.....	122
IV.3.3.2 Résultats et métriques d'évaluation.....	122
IV.3.4 Résultats CNN-LSTM	123
IV.3.4.1 Choix des Hyper paramètres du Modèle.....	123
IV.3.4.2 Résultats et métriques d'évaluation.....	124
IV.4 Time Series Analysis ARIMA	125
IV.4.1 Résultats et discussions	127
IV.5 Conclusion	128
Conclusion générale et perspectives	129
Bibliographie	132

Liste des figures

Figure I.1 : Architecture des Smart Grids.....	27
Figure I.2 : Home Area Network	28
Figure I.3 : Building Area Network	28
Figure I.4 Neighborhood Area Network	29
Figure I.5 : Wide Area Network	29
Figure I.6 : Architecture selon le Modèle SGAM	31
Figure I.7 : Vue d'ensemble du Smart Grid	33
Figure I.8 : Exigences de sécurité pour les SMNs	43
Figure II.1 : Consommation d'électricité sur toute la France	49
Figure II.2 : La Décomposition Additive	52
Figure II.3 : La Décomposition Multiplicative	53
Figure II.4 : Série temporelle stationnaire et non stationnaire	54
Figure II.5 : Organigramme de la méthode Box et Jenkins.....	57
Figure II.6 : ACF & PACF	58
Figure II.7 : Relations entre l'AI, ML et DL	62
Figure II.8 : Structure de RNN et LSTM	66
Figure II.9 : Wide & Deep Convolutional Neural Networks (CNN) framework	67
Figure II.10 : Flux de détection de vol d'énergie avec CNN-RF.....	69
Figure II.11 Architecture centralisée & Architecture décentralisée.....	70
Figure II.12 : Structure générale de la Blockchain	71
Figure II.13 Arbre de Merkle	72
Figure II.14 : Le principe de fonctionnement de la Blockchain	72
Figure II.15 : Architecture proposée basée sur quatre zones élémentaires AMI.	79
Figure II.16 : Réseau compteur-nœud	81
Figure II.17 : Structure des blocs	82
Figure III.1 : Architecture générale AMI.....	89

Figure III.2 : Représentation graphique des modèles de DL appliqués à la SG	96
Figure III.3 : Architecture CNN 1D.....	98
Figure III.4 : Architecture CNN -2D	100
Figure III.5 : Architecture CNN-LSTM	101
Figure III.6 : Architecture proposée pour la sécurisée de l'énergie dans les SG en utilisant la Blockchain et le DL	102
Figure III.7 : Structure de Bloc niveau NAN	106
Figure IV.1 : Consommation de deux Clients frauduleux	118
Figure IV.2 : Consommation de deux Clients non frauduleux	119
Figure IV.3 : Consommation sur Quatre Semaines.....	120
Figure IV.4 : Les courbes de perte et de précisionCNN-1D	122
Figure IV.5 : Les courbes de perte et de précisionCNN-2D	123
Figure IV.6 : Les courbes de perte et de précision CNN-LSTM	124
Figure IV.7 : Comparaison des Valeurs Réelles et Prédites avec ARIMA.....	126

Liste des tableaux

Table I.1 : Réseau électrique traditionnel VS Smart Grids.....	26
Table II.1 Théori générale de l'ACF et du PACF.....	58
Table II.2 : Résultats Wide & Deep CNN	68
Table II.3 : Résumé des scores de classification de CNN-RF.....	69
Table II.4 : Comparaison entre le Smart Grid et le Smart Grid décentralisé	77
Table II.5 : Objectifs communs de sécurité et solutions offertes par la Blockchain	78
Table II.6 : Comparaison des solutions Blockchain trouvées.	84
Table II.7 : Comparaison entre les solutions proposées DL & Blockchain.....	87
Table III.1 : Structure du bloc niveau FAN/WAN	111
Table IV.1 : Résultats CNN-1D.....	121
Table IV.2 : Résultats CNN-2D.....	122
Table IV.3 : Résultats CNN-LSTM	124
Table IV.4 : Résultats de l'approche ARIMA	127

Glossaire

Interopérabilité : L'interopérabilité est le catalyseur clé du réseau intelligent, permettant une harmonisation et une normalisation des interactions entre les différentes couches et acteurs du système électrique. Cette approche facilite les interactions et simplifie les cas d'usage, permettant ainsi de construire un réseau électrique intelligent, performant et résilient.

WiMax : désigne une norme de communication sans fil. Actuellement, il est principalement utilisé comme système de transmission et d'accès à Internet à haut débit, couvrant une large zone géographique. Ce terme est aussi utilisé comme une marque commerciale, similaire au Wi-Fi.

FAN : sont des infrastructures de communication conçues pour les réseaux extérieurs à grande échelle. Comme Internet sur un smartphone ou un ordinateur, un Field Area Network permet à des équipements industriels tels que les compteurs intelligents et les lampadaires de se connecter ensemble sur un réseau commun. Le programme de certification FAN, développé par l'Alliance Wi-SUN, garantit la compatibilité des appareils destinés aux services publics, aux promoteurs urbains et à d'autres fournisseurs de services. Cela facilite la mise en œuvre de villes intelligentes à grande échelle, de services publics intelligents et d'autres déploiements d'IoT.

Exactitude (Accuracy) : L'exactitude mesure la proportion d'observations correctement prédites par le modèle par rapport à l'ensemble des observations.

Erreur Quadratique Moyenne (RMSE) : Le RMSE évalue la différence entre les valeurs prédites et les valeurs réelles, et est couramment utilisé pour évaluer les performances des modèles de régression.

Erreur Absolue Moyenne (MAE) : Le MAE calcule l'erreur moyenne absolue entre les valeurs prédictes et les valeurs réelles.

F1-Score : Le F1-Score mesure la précision d'un test en calculant la moyenne harmonique de la précision (precision) et du rappel (recall), utile en particulier lorsque les classes sont déséquilibrées.

Aire Sous la Courbe ROC (AUC) : L'AUC évalue la capacité du modèle à distinguer entre les classes positives et négatives. La courbe ROC est un graphique de la sensibilité (recall) en fonction de la spécificité pour différents seuils de classification. L'AUC fournit une mesure agrégée de la performance sur tous les seuils de classification possibles.

Matrice de Confusion : Une matrice de confusion est un tableau qui montre les performances d'un modèle de classification, comparant les valeurs réelles avec les valeurs prédictes et affichant le nombre de prédictions correctes et incorrectes.

Liste des abréviations

SG : Smart Grids

HAN : Home Area Network

BAN : Building Area Network

NAN : Neighborhood Area Network

WAN : Wide Area Network

SGAM : Smart Grid Architecture Model

PLC : Power-Line Communication

DSL : Digital Subscriber Line

AMI : Advanced Metering Infrastructure

AMR : (Automatic Meter Reading

V2G : Vehicle to Grid

PMU : Phasor Measurement Units

PHEV : Plug-in Hybrid Electric Vehicles

SMN : Smart Metering Networks

WiMAX : Worldwide Interoperability for Microwave Access

SMN : Smart Metering Networks

FDIA : False Data Injection Attack

ADF : Augmented Dickey-Fuller

KPSS : Kwiatkowski Phillips Schmidt Shin

AR : AutoRegressive

MA : Moving Average

ARMA : AutoRegressive Moving Average

ARIMA : AutoRegressive Integrated Moving Average

SARIMA : Seasonal AutoRegressive Integrated Moving Average

ACF : AutoCorrelation Function

PACF : Partial AutoCorrelation Function

IA : Artificial Intelligence

ML : Machine Learning

DL : Deep Learning

CNN : Convolutional Neural Network

RNN : Recurrent Neural Network

LSTM : Long Short-Term Memory

ReLU : Rectified Linear Unit

RF : Random Forest

PoW: Proof of Work

PoS : Proof of Stake

PoA : Proof of Authority

PBFT : Practical Byzantine Fault Tolerance

PoEF : Proof-of-Efficiency

IOT : Internet of Things

VAE : Variational Autoencoder

MDMS : Meter Data Management System

ICS : Industrial Control System

SCADA : Supervisory Control and Data Acquisition

PLC : Programmable Logic Controller

AIC : Akaike Information Criterion

FAN : Field Area Networks

ROC : Receiver Operating Characteristic

AUC : Area Under the ROC Curve

MAE : Mean Absolute Error

RMSE : Root Mean Squared Error

Introduction générale

L'intégration de l'intelligence artificielle et de la technologie Blockchain dans les réseaux électriques intelligents représente une avancée majeure dans la gestion moderne de l'énergie. Ces innovations permettent d'optimiser la production, la distribution et la consommation d'électricité, tout en renforçant la sécurité et la résilience des infrastructures énergétiques. Ce mémoire se concentre sur l'application de ces technologies pour améliorer la détection des fraudes et la protection des réseaux contre diverses menaces, notamment les attaques par injection de fausses données. En particulier, il met en lumière l'analyse des séries temporelles avec les modèles ARIMA pour identifier les anomalies dans les données de consommation énergétique.

Au cours des dernières décennies, les systèmes énergétiques centralisés traditionnels, basés sur les combustibles fossiles, ont été confrontés à des défis majeurs tels que la transmission à longue distance, les émissions de carbone, la pollution environnementale et la crise énergétique. Pour bâtir une société durable face à ces défis, l'utilisation d'énergies renouvelables provenant de diverses sources, ainsi que l'amélioration de l'efficacité énergétique, sont deux solutions potentielles. Ces dernières années, le concept de réseau intelligent, intégrant des technologies de communication, un système d'alimentation interconnecté, des technologies de contrôle avancées et des compteurs intelligents, a été appliqué pour améliorer l'utilisation des sources d'énergie renouvelables et atténuer la crise énergétique.

Le concept de réseau intelligent a été introduit comme une nouvelle vision du réseau électrique conventionnel, visant à intégrer efficacement les technologies d'énergie verte et renouvelable. Le réseau intelligent connecté à Internet, également appelé Internet de l'énergie, se présente comme une approche innovante et bidirectionnelle de gestion de l'énergie et de l'information, permettant une gestion efficace des technologies énergétiques tout en assurant une sécurité optimale.

Cependant, l'intégration et la coordination d'un grand nombre de connexions croissantes posent un défi pour le système de réseau centralisé traditionnel. De plus, l'installation d'équipements autonomes, tels que les compteurs intelligents chez les clients, peut s'avérer critique en raison du risque de manipulation de données ou de dysfonctionnements prémedités.

En conséquence, le réseau intelligent évolue vers une topologie décentralisée, intégrant la technologie Blockchain pour pallier le problème de l'injection de fausses données. Par ailleurs, l'intelligence artificielle, associée à l'analyse des séries temporelles avec les modèles ARIMA, joue un rôle crucial dans la détection des vols ou manipulations d'énergie à partir des mesures transmises par les compteurs intelligents.

En conséquence, afin d'atteindre notre objectif de combiner l'IA et la Blockchain pour sécuriser les réseaux intelligents, notre travail est structuré en quatre chapitres :

- Le premier chapitre, « Fondements et Enjeux des Smart Grids », introduit les concepts fondamentaux des réseaux électriques intelligents, ou Smart Grids, et compare leur fonctionnement avec les réseaux traditionnels. Il examine également les principaux composants des Smart Grids, les défis de sécurité auxquels ils sont confrontés et les exigences spécifiques pour garantir la sécurité des réseaux de comptage intelligent.
- Le deuxième chapitre, « Blockchain et Intelligence Artificielle au Service des Smart Grids », explore l'application de la Blockchain et de l'intelligence artificielle, notamment le Deep Learning, dans les Smart Grids. Il présente une analyse des séries temporelles pour la sécurité des réseaux et décrit comment la Blockchain peut renforcer la fiabilité et la sécurité des transactions et des données dans les réseaux intelligents.

- Le troisième chapitre, intitulé « Élaboration des Approches pour Déetecter les Fraudes dans les Smart Grids », développe diverses méthodes de détection de fraude dans les réseaux intelligents. Il explore l'utilisation des modèles ARIMA pour l'analyse des séries temporelles et diverses architectures de Deep Learning. Chaque méthode est présentée en détail avec ses avantages et ses limitations. De plus, une approche intégrant la Blockchain et le Deep Learning est proposée pour analyser et renforcer la sécurité des réseaux intelligents contre les fraudes.
- Enfin, le quatrième chapitre, « Résultats de la Détection des Fraudes dans les Smart Grids », présente les résultats obtenus des différentes méthodes de détection de la fraude appliquées aux données des Smart Grids. Nous comparons les performances des modèles en termes de précision, rappel et autres métriques de performance, et discutons de l'efficacité de chaque méthode dans la détection des comportements frauduleux.
- La conclusion résume les principaux enseignements tirés des chapitres précédents et met en lumière les contributions majeures de cette étude. Elle aborde également les perspectives futures, suggérant des pistes de recherche et d'amélioration pour le développement des Smart Grids, en particulier en ce qui concerne l'intégration continue de la Blockchain et de l'intelligence artificielle pour renforcer la sécurité et l'efficacité des réseaux intelligents.

Chapitre I

Fondements et Enjeux des Smart Grids

I.1 Introduction

Les Smart Grids, ou réseaux électriques intelligents, transforment la gestion de l'énergie grâce à l'intégration des technologies de l'information et de la communication. Contrairement aux réseaux traditionnels, ils permettent une interaction bidirectionnelle entre producteurs et consommateurs, optimisant ainsi la production, la distribution et la consommation d'énergie. Cette interactivité améliore l'efficacité énergétique, la fiabilité et la résilience du réseau, tout en facilitant l'intégration des énergies renouvelables.

L'architecture des Smart Grids repose sur des composantes clés comme les systèmes de communication avancés, les compteurs intelligents et les systèmes de stockage d'énergie. Le modèle SGAM (Smart Grid Architecture Model) offre un cadre structuré pour développer ces systèmes, en insistant sur l'interopérabilité et la sécurité. Les systèmes de communication assurent une transmission rapide et fiable des données, tandis que les compteurs intelligents fournissent des informations en temps réel pour une gestion précise de la demande et une meilleure intégration des sources intermittentes.

Cependant, l'adoption des Smart Grids pose des défis de sécurité, notamment contre les cyberattaques et les fraudes. La protection des réseaux intelligents,

vulnérables à des menaces comme le vol d'électricité. Ce chapitre examine ces défis et propose des solutions pour renforcer la sécurité des Smart Grids, en mettant l'accent sur les exigences spécifiques des réseaux de comptage intelligent, telles que la confidentialité, l'intégrité, la fraîcheur, la disponibilité des données, la non-répudiation et l'authentification.

I.2 Réseaux électriques intelligents (Smart Grid)

I.2.1 Définition des Smart Grids

Les réseaux électriques intelligents, ou Smart Grids, sont des infrastructures énergétiques avancées qui intègrent les technologies de l'information et de la communication (TIC) pour révolutionner la gestion et la distribution de l'énergie électrique. Ces réseaux représentent une évolution critique par rapport aux systèmes traditionnels, en ce sens qu'ils facilitent une gestion dynamique et efficace de la production, de la transmission, de la distribution et de la consommation d'énergie. Les Smart Grids sont caractérisés par leur capacité à automatiser et à optimiser les flux d'énergie en temps réel grâce à une intégration poussée d'éléments programmables et interconnectés, tels que les compteurs intelligents et les dispositifs de réponse à la demande [1].

Les Smart Grids combinent les réseaux électriques de distribution conventionnels, tels que ceux gérés par l'ONER au Maroc, englobant l'ensemble de la chaîne d'approvisionnement – production, transport et distribution – avec des technologies avancées de l'information et de la communication. De plus, ils permettent un traitement de l'information de manière plus intelligente et optimisée.

Les réseaux intelligents peuvent être définir selon quatre caractéristiques :

- **Flexibilité** : Ils permettent de gérer plus finement l'équilibre entre production et consommation.
- **Fiabilité** : Ils améliorent l'efficacité et la sécurité des réseaux.
- **Accessibilité** : Ils favorisent l'intégration des sources d'énergies renouvelables sur l'ensemble du réseau.
- **Economie** : Ils apportent, grâce à une meilleure gestion du système, des économies d'énergie et une diminution des couts.

I.2.2 Réseau Électrique Traditionnel vs Smart Grids

Les réseaux électriques traditionnels et les Smart Grids diffèrent considérablement sur plusieurs aspects, comme détaillé ci-dessous [2] :

	Smart Grids	Réseaux Traditionnels
Fonctionnement	Reposent sur la digitalisation, offrant une gestion plus flexible et dynamique.	Basés sur la mécanisation, avec une structure rigide et centralisée.
Communication	Communication bidirectionnelle, permettant une interaction continue entre les fournisseurs et les consommateurs.	Communication unidirectionnelle, où l'information circule uniquement des fournisseurs vers les consommateurs.
Détection et Surveillance	Intègrent des capteurs le long du réseau, avec des systèmes avancés de surveillance tels que les compteurs intelligents (metering), les PMU (Phasor Measurement Units), et divers capteurs.	Utilisent un nombre limité de capteurs et des systèmes SCADA pour la surveillance.

Contrôle de la Consommation	Les consommateurs peuvent surveiller et contrôler leur consommation en temps réel, favorisant une utilisation plus efficiente de l'énergie	La consommation n'est pas contrôlée activement par les consommateurs.
Sécurité	Nécessitent des niveaux élevés de sécurité pour protéger contre les cyberattaques et les catastrophes naturelles.	Nécessitent peu de mesures de sécurité.
Stockage de l'Énergie	Permettent le stockage d'énergie, notamment grâce aux véhicules électriques (V2G, PHEV), améliorant la gestion des pics de demande	La génération, la transmission et la distribution d'énergie se font simultanément sans stockage significatif.
Mesure de la Consommation et Auto-guérison	La mesure de la consommation et l'auto-guérison se font automatiquement en temps réel, grâce à des systèmes intégrés de surveillance et de contrôle.	Les agents notent la consommation avec une intervention humaine.
Ressources Énergétiques	Utilisent une gamme diversifiée de sources d'énergie, incluant les combustibles fossiles, les énergies renouvelables, et l'énergie générée par les utilisateurs finaux.	Principalement basés sur des combustibles fossiles et des énergies renouvelables de manière limitée.

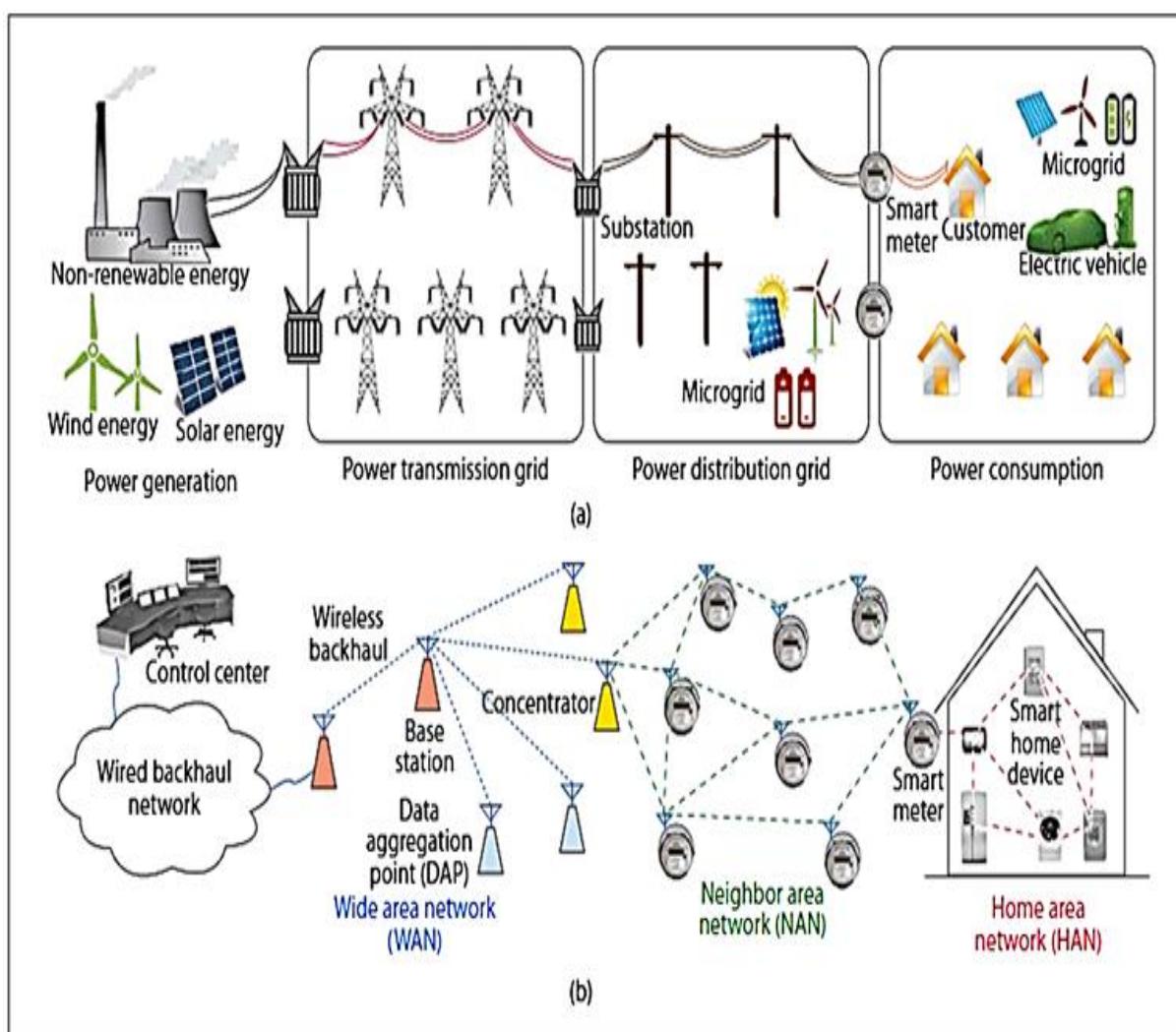
Table I.1 : Réseau électrique traditionnel VS Smart Grids

I.2.3 Architecture des Smart Grids

L'intégration des systèmes de réseaux électriques avec les systèmes de communication a considérablement amélioré la gestion et le contrôle des ressources énergétiques.

Le réseau intelligent repose sur deux interfaces informatiques : l'une placée à la source d'approvisionnement (comme une centrale électrique) et l'autre chez le client. Cette configuration permet une communication bidirectionnelle entre le fournisseur et le client, facilitant l'échange de données et l'exécution des commandes de contrôle de manière synchronisée. Les données recueillies par l'interface client sont envoyées au gestionnaire de réseau de distribution (GRD), qui ajuste l'approvisionnement en électricité en temps réel pour répondre aux besoins du client.

Comme illustré dans la Figure I.1 [3], les systèmes de communication des Smart Grids sont structurés de manière hiérarchique et comprennent plusieurs typologies de réseau



Home Area Network (HAN) : Le Home Area Network (HAN) est un réseau dédié reliant tous les appareils intelligents qui fonctionnent dans un réseau domestique. Un compteur intelligent sans fil est placé dans la maison du consommateur pour permettre la communication dans ce réseau. Ce compteur agit comme une passerelle HAN, gérant toutes les communications de données entre les appareils. Grâce à ce réseau, les consommateurs peuvent contrôler et surveiller l'énergie consommée et le flux de données entre les appareils électroménagers tels que les thermostats, les climatiseurs, les réfrigérateurs, les laveuses, les sécheuses et les cuisinières [4].

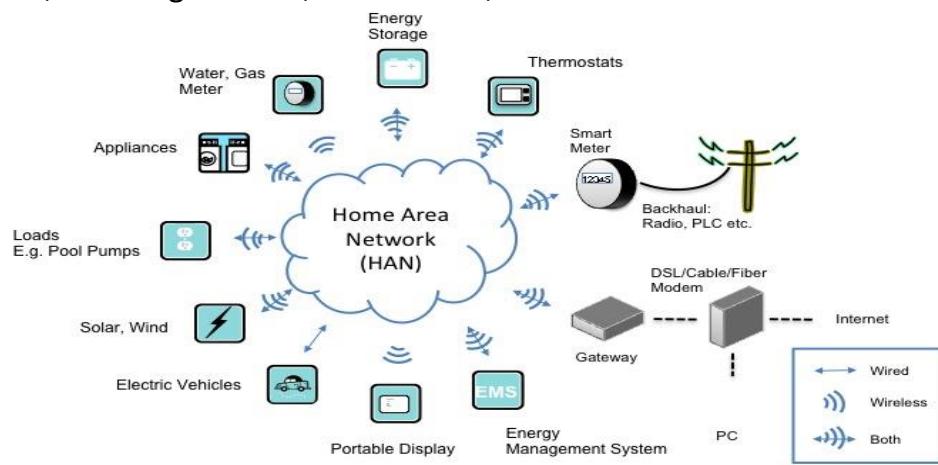


Figure I.2 : Home Area Network

Building Area Network (BAN) : Le Building Area Network (BAN) est un réseau installé dans un bâtiment pour agrandir, surveiller et fournir des informations sur la consommation d'énergie. La passerelle BAN collecte les données des HANs et les envoie au centre de collecte de données ou à la passerelle NAN. Le BAN joue un rôle crucial dans la gestion énergétique locale, permettant une supervision précise de la consommation d'énergie au niveau du bâtiment [5].

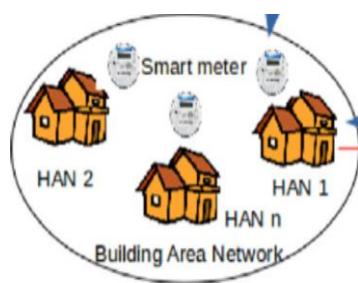


Figure I.3 : Building Area Network

Neighborhood Area Network (NAN) : Le Neighborhood Area Network (NAN) est constitué de points d'accès Wi-Fi et de réseaux locaux sans fil (WLAN). Ces réseaux permettent aux utilisateurs de se connecter à Internet et de transmettre les données énergétiques sur de petites zones géographiques. Le NAN relie plusieurs BANs à un collecteur de données central, facilitant ainsi la gestion énergétique de plusieurs bâtiments voisins. Les données collectées par les NANs sont ensuite transmises aux centres de contrôle via les réseaux WAN pour une analyse plus approfondie.

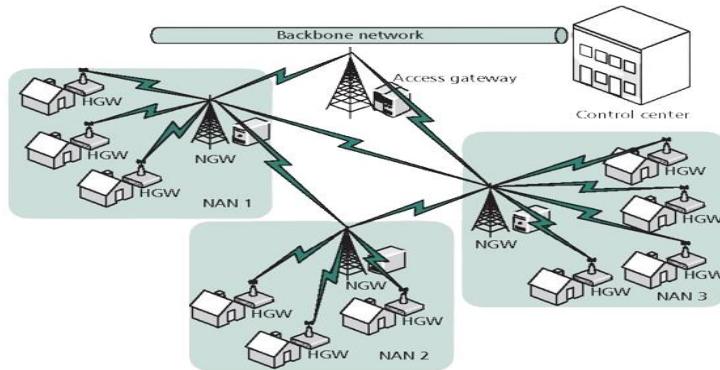


Figure I.4 Neighborhood Area Network

Wide Area Network (WAN) : Le Wide Area Network (WAN) couvre une vaste zone géographique et constitue la couche supérieure des réseaux de communication dans les Smart Grids. Il assure la transmission des données des passerelles NAN vers les centres de contrôle des services publics. Le WAN utilise diverses technologies de communication telles que WiMAX, le GSM 3G, la LTE et la fibre optique pour assurer une communication fiable et rapide sur de longues distances. Le WAN permet une intégration complète des différentes couches de réseau (HAN, BAN, NAN) en une infrastructure de communication unifiée et robuste.

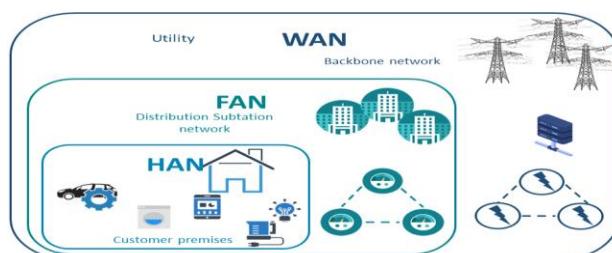


Figure I.5 : Wide Area Network

Fonctionnement Intégré des Réseaux

Les typologies de réseau HAN, BAN, NAN et WAN sont interconnectées pour assurer une communication fluide et efficace des données énergétiques. Les HANs collectent les données de consommation des appareils domestiques et les transmettent aux BANs. Les BANs agrègent ces données et les envoient aux NANs, qui les transmettent ensuite aux WANs pour une gestion centralisée. Cette structure en couches permet une surveillance continue et en temps réel de la consommation d'énergie, facilitant ainsi la détection rapide des anomalies. Ces réseaux de communication sont essentiels pour la collecte, l'analyse et la transmission des données, permettant une gestion optimisée de l'énergie.

Du point de vue des réseaux électriques, l'énergie peut être générée par diverses sources telles que le nucléaire, l'hydroélectricité, l'éolien et le solaire. L'énergie produite est distribuée via deux types de sous-stations :

- **Sous-station de Transmission** : Située près de la centrale de production, elle transmet de grandes quantités de tension à la sous-station de distribution.
- **Sous-station de Distribution** : Située près des zones industrielles ou résidentielles, elle convertit l'alimentation haute tension en moyenne tension avant de la distribuer aux consommateurs finaux.

I.2.4 Architecture selon le Modèle SGAM

L'architecture des Smart Grids selon le modèle de référence SGAM (Smart Grid Architecture Model) repose sur une approche d'interopérabilité entre différentes couches, chacune jouant un rôle distinct mais interconnecté. Le principe clé du SGAM est de permettre aux matériels, logiciels et protocoles différents de fonctionner ensemble et de partager des informations de manière harmonieuse.

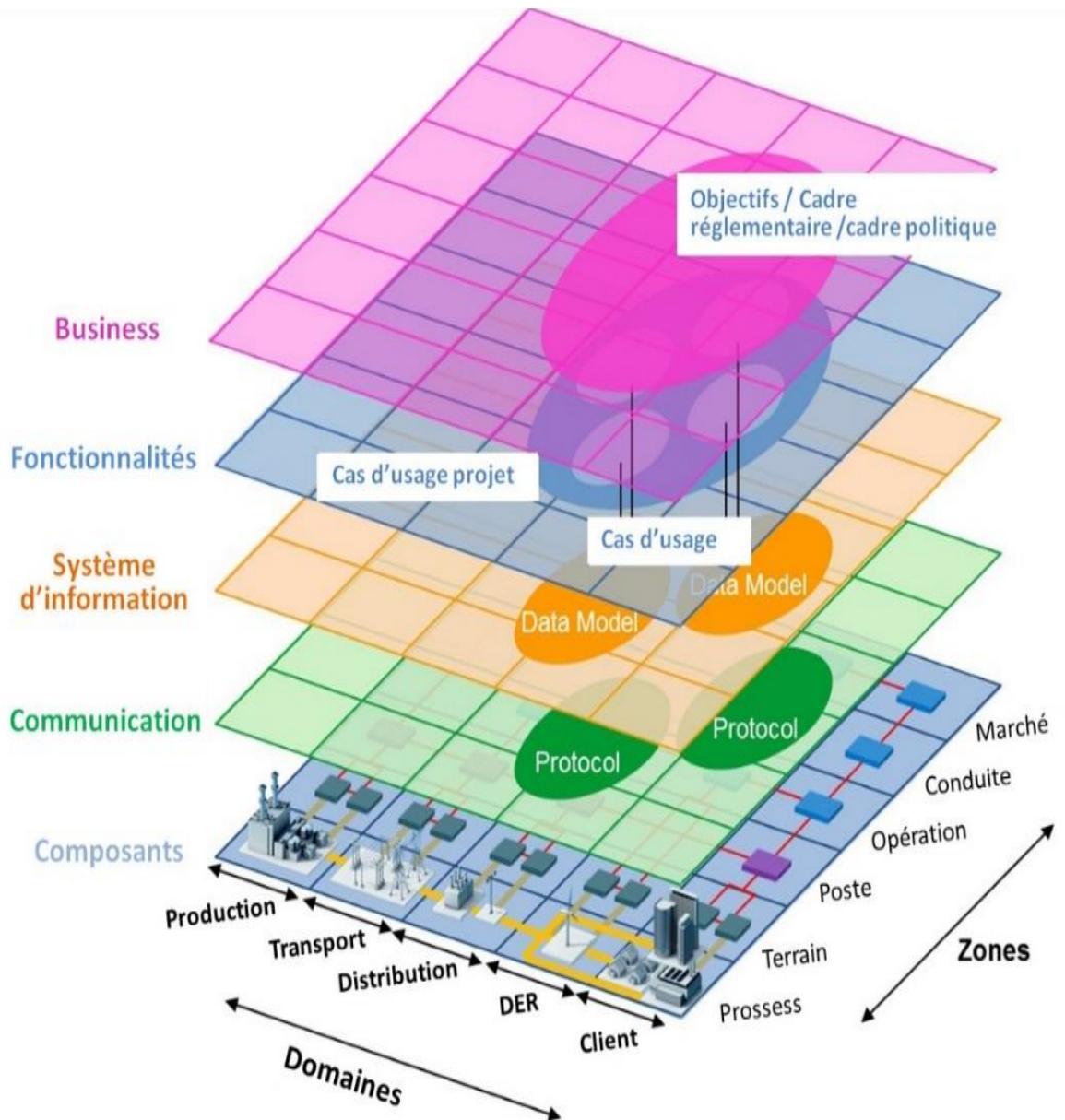


Figure I.6 : Architecture selon le Modèle SGAM

Comme illustré Figure I.6 dans la, les couches du Modèle SGAM

- **Couche Business**

Décrit les modèles d'affaires et les contraintes externes telles que la législation et les conditions de marché. Elle englobe les objectifs commerciaux, les politiques et le cadre réglementaire qui influencent les cas d'usage.

- **Couche Fonction**

Définit les fonctions et services associés aux cas d'usage, ainsi que leurs interactions d'un point de vue architectural. Cela inclut les processus nécessaires pour accomplir les objectifs commerciaux définis.

- **Couche Système d'Information**

Décrit les informations échangées entre les fonctions, services et composants. Cela comprend le type d'informations, les modèles de données utilisés, leur traitement et leur valorisation.

- **Couche Communication**

Décrit les protocoles de communication et les mécanismes d'échange entre les différents composants pour répondre aux besoins des couches supérieures. Cela assure la transmission sécurisée et fiable des données nécessaires à la gestion du réseau.

- **Couche Composants**

Représente la localisation physique des différents composants participant aux cas d'usage. Cela inclut les acteurs du système, les applications, et les équipements électriques et télécom du réseau.

I.2.5 Les principales composantes d'un Smart Grid

Figure I.7 illustre l'architecture d'un réseau électrique intelligent (Smart Grid), mettant en évidence les flux d'énergie et les communications sécurisées entre les différents composants du système. Elle montre comment les différents segments du réseau, incluant la génération, la transmission, la distribution et la consommation, sont intégrés via des infrastructures avancées de mesure (AMI) et des réseaux de communication (WAN, NAN, HAN).

Les données recueillies par les compteurs intelligents sont agrégées et analysées pour optimiser la gestion de l'énergie, assurant une interaction fluide entre les opérateurs, les fournisseurs de services, et les consommateurs [7].

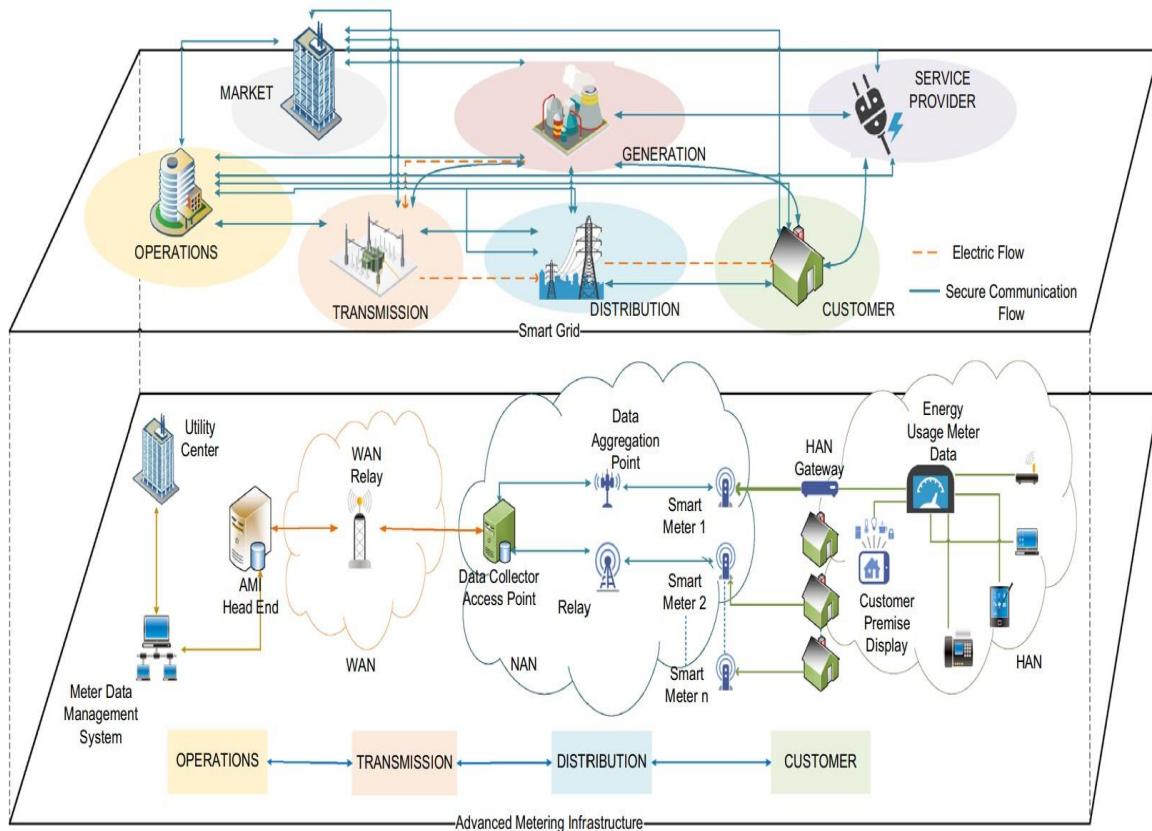


Figure I.7 : Vue d'ensemble du Smart Grid

I.2.5.1 Communication

- **Communication filaire**

La communication filaire est déjà largement déployée dans les réseaux conventionnels pour transmettre les informations de surveillance. Elle est préférée par les fournisseurs en raison de sa fiabilité et de son insensibilité aux interférences.

Différentes technologies de communication filaire incluent :

- *PLC (Power-Line Communication)* : Utilise des fréquences élevées pour transmettre les données sur les lignes électriques.

- *Communication optique* : Offre un taux élevé de transmission de données et est insensible aux interférences électromagnétiques.
- *DSL (Digital Subscriber Line)* : Utilise les lignes téléphoniques pour la transmission, évitant ainsi des coûts supplémentaires pour les supports de transmission.

- **Communication sans fil**

La communication sans fil est souvent la solution la plus appropriée dans les réseaux intelligents en raison de ses coûts économiques et de sa faisabilité technologique. Elle couvre de vastes zones, même celles non desservies par des lignes électriques. Les technologies sans fil, bien que sensibles à la bande passante limitée et aux interférences, offrent une flexibilité précieuse. Les principales technologies sans fil incluent :

- *Technologies NAN* : WiMAX, UMTS, LTE.
- *Technologies HAN* : WPAN, Satellite, Radio Cognitive.

I.2.5.2 Mesusage : Metering

- **AMI (Advanced Metering Infrastructure)**

L'AMI combine plusieurs technologies pour garantir la communication entre l'utilisateur et l'opérateur. Elle inclut le réseau électrique public, le réseau de communication et l'infrastructure de traitement des données. Les compteurs intelligents dans un système AMI communiquent des données de consommation énergétique aux services publics et aux consommateurs, permettant une sensibilisation accrue à la consommation d'énergie. Ces compteurs permettent également des fonctionnalités avancées comme la tarification en temps réel, la détection des pertes de puissance, la surveillance de la qualité de l'alimentation, et la gestion à distance [2].

- **AMR (Automatic Meter Reading) :**

L'AMR fonctionne en mode simplex, permettant la lecture des états et des alarmes de consommation des utilisateurs, puis transmet ces informations aux opérateurs via des supports de communication. Cela permet aux opérateurs de générer des factures basées sur des données analogiques transformées en numériques [2].

- **Compteurs intelligents (Smart meters)**

Les compteurs intelligents permettent aux opérateurs de collecter automatiquement les informations de consommation pour détecter les pointes, gérer la consommation et mettre en place des stratégies de facturation. Pour les consommateurs, ils offrent la possibilité de mesurer et de consulter leur consommation énergétique, contrôlant ainsi leur utilisation et les coûts associés. Ils facilitent la communication bidirectionnelle, l'équilibrage de la charge, la stabilisation de la tension, et bien plus encore [2].

- **PMU: Phasor Measurement Units**

Les PMU mesurent les phasors ou ondes électriques sur un réseau synchronisé à partir de la même source temporelle, garantissant des valeurs en temps réel des tensions et des courants. Ils sont utilisés pour des applications avancées de surveillance, de contrôle, et de protection du réseau [2].

- **Capteurs**

Les capteurs, utilisés dans les stations de détection, convertissent les signaux analogiques en valeurs numériques lisibles. Ces données sont ensuite traitées par des unités centralisées pour déterminer des valeurs significatives, permettant d'évaluer l'état du système.

- **Stockage**

Les technologies de stockage d'énergie comprennent les batteries à flux, les ultra-condensateurs, les roues volantes, l'hydroélectricité pompée, le stockage d'énergie magnétique supraconducteur et l'air comprimé.

- **V2G : Vehicle to Grid**

Les véhicules équipés de batteries rechargeables peuvent fournir de l'énergie aux réseaux pendant les heures de pointe. Les opérateurs peuvent se connecter à ces véhicules directement ou via des parkings qui transmettent les besoins aux véhicules garés.

- **PHEV: Plug-in Hybrid Electric Vehicles**

Les PHEV sont des véhicules hybrides utilisant à la fois du pétrole et de l'électricité. Ils passent à l'électricité lorsque les prix du carburant sont élevés ou lorsque le niveau de la batterie est suffisant [2].

I.3 La sécurité dans les Smart Grids

I.3.1 Menaces et Vulnérabilités dans les Smart Grids

Les Smart Grids, en intégrant des technologies avancées de communication et de gestion des données, sont exposés à une multitude de menaces de sécurité. Ces menaces peuvent provenir de diverses sources et cibler différents aspects du réseau, allant des attaques physiques aux cyberattaques sophistiquées. Il est crucial de comprendre les différents types d'attaques pour mettre en place des mesures de sécurité efficaces. Cette section explore en détail les principales menaces et vulnérabilités auxquelles les Smart Grids sont confrontés, en mettant un accent particulier sur les attaques par injection de fausses données, les attaques par déni de service, les attaques par écoute, les attaques par usurpation d'identité, les attaques par répudiation et les attaques par compromission de nœud.

I.3.1.1 Attaques par Injection de Fausses Données (FDIA)

Les attaques par injection de fausses données (False Data Injection Attack, FDIA) représentent un type d'attaque cybernétique particulièrement dangereuse qui vise à manipuler l'intégrité des données en introduisant de fausses informations dans le réseau. L'objectif est de tromper le centre de contrôle pour qu'il prenne des décisions erronées en matière d'analyse de contingence, de répartition de puissance et de processus de facturation. Dans les réseaux de comptage intelligent (SMNs), la FDIA se produit lorsque l'attaquant parvient à intercepter et à altérer le contenu des données avant qu'elles ne soient envoyées au réseau. Étant donné que les SMNs utilisent des médiums de communication sans fil et à diffusion, les données transmises par les compteurs intelligents via ces médiums peuvent facilement être capturées ou interceptées par l'attaquant. Ce type d'attaque peut être extrêmement nuisible, en particulier lorsque plusieurs attaquants collusent pour injecter de fausses données dans le réseau. À notre connaissance, la détection des FDI collusaires est l'une des attaques les plus difficiles à identifier dans les SMNs authentifiés en raison de l'absence ou de l'insuffisance de mécanismes de détection de mauvais comportements dans les schémas de sécurité actuels. Par conséquent, lorsque ces nœuds sont compromis ou détournés avec succès, l'injection de fausses données peut être effectuée facilement par l'attaquant[8].

L'attaque FDIA ne se contente pas d'endommager les données d'origine, elle peut également entraîner d'autres dommages tels que la surcharge de la demande de puissance, l'instabilité du réseau, des pannes de courant, voire l'autodestruction d'un générateur de puissance.

I.3.1.2 Attaque par Déni de Service (DoS)

L'attaque par déni de service (DoS) est courante dans les communications filaires ou sans fil. L'attaque DoS la plus connue dans les SMNs est l'attaque par brouillage.

Cette attaque survient dans les réseaux sans fil lorsqu'un brouilleur tente de perturber les fréquences radio utilisées par les compteurs intelligents en émettant d'autres signaux radio pour interrompre le processus de transmission des données. Pour atténuer cette attaque, certains auteurs ont utilisé une approche théorique des jeux appelée jeu stochastique à somme nulle. Dans ce jeu, le capteur (compteur intelligent) et le brouilleur rivalisent pour transmettre les données au centre de contrôle, tandis que le brouilleur essaie de bloquer les canaux de communication disponibles. En plus de l'attaque par brouillage, les attaques par inondation de paquets et par perte de paquets sont d'autres exemples d'attaques qui entrent dans la catégorie DoS [8].

I.3.1.3 Attaque par Écoute (EVD)

L'attaque par écoute (EVD) est une autre attaque courante dans les SMNs. Elle consiste à écouter et enregistrer secrètement les communications de données des compteurs intelligents voisins. L'EVD peut ou non être dangereuse ; cela dépend des motivations de l'écouteur. Cependant, l'EVD peut violer la vie privée en réalisant des actions illégitimes telles que le vol d'informations, la fraude d'identité et le profilage des comportements humains. Actuellement, plusieurs schémas se concentrent sur la détection et la lutte contre l'EVD. Cependant, presque tous les schémas mettent davantage l'accent sur l'amélioration des techniques de cryptage des données, nécessitant une mémoire supplémentaire et une puissance de calcul élevée pour leur exécution [8].

I.3.1.4 Attaque par Usurpation d'Identité (IMP)

L'attaque par usurpation d'identité (IMP) ou attaque de l'homme du milieu se produit lorsqu'un compteur intelligent malveillant s'empare de l'identité d'autres compteurs intelligents légitimes. En se faisant passer pour un compteur intelligent légitime, un attaquant peut recevoir et altérer le contenu des messages reçus.

Pire encore, l'IMP est utilisée par les usurpateurs se faisant passer pour plusieurs identités de compteurs intelligents fictifs réalisant diverses actions illégales, introduisant une menace plus sévère connue sous le nom d'attaque de Sybil. L'attaque de Sybil a la capacité de paralyser ou de dégrader le fonctionnement du réseau de compteurs intelligents en colludant pour lancer des attaques DoS telles que la perte et l'inondation de paquets. L'IMP peut se produire non seulement dans les communications entre compteurs, mais aussi dans les communications entre compteurs et appareils intelligents, ce qui peut entraîner des scénarios comme une demande excédant l'offre, une surfacturation ou même une coupure d'électricité [8].

I.3.1.5 Attaque par Rejeu (REP)

L'attaque par rejeu (REP) tente de perturber la sécurité en stockant ou enregistrant des données non autorisées et en retransmettant les données plus tard après avoir effectué des modifications sur les données d'origine. Par exemple, la transmission de données entre compteurs intelligents peut être capturée ou interceptée par un attaquant, puis rejouée après avoir effectué des modifications. L'attaque REP peut être prévenue en utilisant des techniques de nonce telles que le timestamp ou le numéro de séquence des messages. Cependant, l'utilisation de numéros de séquence nécessite des canaux de communication fiables, tandis que l'utilisation de timestamp nécessiterait l'échange de messages supplémentaires, ce qui peut entraîner une surcharge de communication [8].

I.3.1.6 Attaque par Répudiation (RPD)

L'attaque par répudiation (RPD) peut être définie comme le déni de participation à la communication. La fonction de non-répudiation dans les réseaux de comptage intelligent vise principalement à garantir que les consommateurs ou les services publics ne pourront pas nier avoir envoyé et/ou reçu leurs données de comptage authentifiées pour éviter toute responsabilité.

Le schéma proposé utilise une méthode de génération de signature unique pour protéger les données des compteurs intelligents contre les attaques par répudiation. Ce schéma a la capacité de contrer les attaques par répudiation tout en aidant à réduire la consommation d'énergie et le coût de calcul. Cependant, l'utilisation d'une génération de signature unique pose un autre problème appelé faille de signature. Cette faille est liée à l'utilisation d'une seule clé pour tous les processus d'authentification. Si cette clé est compromise, la sécurité de tout le réseau sera en danger puisque seule une signature est utilisée pour l'ensemble du processus de communication [8].

I.3.1.7 Attaque par Compromission de Nœud (CMP)

L'attaque par compromission de nœud (CMP) est l'attaque la plus intéressante et la plus difficile à gérer dans les SMNs. Elle implique l'accès physique au nœud du compteur intelligent pour prendre le contrôle de la communication et obtenir un accès non autorisé aux données sensibles du nœud, telles que les informations cryptographiques. Cette attaque est particulièrement dangereuse car, même si un seul nœud est compromis, les clés partagées pourraient être révélées, permettant ainsi à l'attaquant de participer aux processus de chiffrement et de déchiffrement. Dans le pire des cas, l'attaquant pourrait également injecter de fausses données dans le réseau après avoir modifié les données chiffrées d'origine. Pour atténuer cette attaque, un schéma basé sur la gestion de la confiance par réputation est proposé. Ce schéma utilise un algorithme de confiance basé sur la règle de la majorité pour détecter les nœuds malveillants qui donnent de fausses lectures sur les données de consommation [8].

I.3.2 Vol d'Énergie

Le vol d'énergie, considéré comme une pratique criminelle, consiste à utiliser de l'énergie sans en payer le prix ou en payant moins que ce qui est dû, en manipulant ou en contournant les compteurs.

Ce phénomène peut se manifester par la manipulation des compteurs, le contournement des lignes d'alimentation, les irrégularités de facturation, etc. Les entreprises de services publics sont confrontées à cette menace majeure, exacerbée par l'émergence des réseaux intelligents (SG) et les nombreuses applications associées, telles que le commerce d'énergie, qui reposent sur des environnements en réseau vulnérables. La convergence de divers systèmes de puissance hérités avec les technologies Internet (comme les déploiements ICS et les compteurs intelligents) a non seulement amélioré les infrastructures de comptage, mais a également ouvert la porte à des cyberattaques visant à voler de l'énergie. Ces vecteurs d'attaque exploitent de nombreuses vulnérabilités inhérentes à ces systèmes, créant ainsi un défi complexe pour les entreprises de services publics.

I.3.2.1 Types de vol d'énergie

En général, le vol d'énergie peut être mis en œuvre par une variété de techniques exploitant à la fois des propriétés physiques et des données ou de la communication du réseau actuel. Par conséquent, la catégorisation adéquate des types de vol d'énergie est une tâche très complexe. Pour relever ce défi et structurer de manière appropriée ce travail, nous identifions deux classes distinctes de vol d'énergie [9] :

- **Vol d'énergie sans données** : Il s'agit de la manipulation physique des composants de puissance par des techniques telles que l'obstruction et le contournement des compteurs électromécaniques, le hameçonnage de câbles, ainsi que la modification des circuits des compteurs.
- **Vol d'énergie basé sur les données** : Il s'agit de manipuler et d'altérer les données de communication et/ou de consommation générées et/ou enregistrées par tout dispositif de comptage en réseau (par exemple, les compteurs intelligents), de gestion (par exemple, les systèmes SCADA) et de contrôle (par exemple, les PLC), ainsi que les logiciels de facturation (par

exemple, les applications mobiles des services publics), dans le but de rapporter de fausses informations de consommation à l'autorité de distribution d'énergie.

Les deux classes ciblent le flux bidirectionnel d'énergie ou de données entre différents points d'agrégation du réseau et ont attiré une attention considérable de la communauté de recherche ainsi que de la société en général. De plus, ces types se sont révélés applicables à tous les niveaux d'agrégation au sein d'un réseau intelligent. Ainsi, le vol d'énergie peut être mis en œuvre dans l'infrastructure de production d'énergie, le réseau de transmission et de distribution (T&D), ainsi qu'au niveau des consommateurs finaux.

I.3.2.2 Conséquences du Vol d'Énergie

- **Compagnies de Services Publics** : Perte de revenus et déficits financiers [10].
- **Qualité de l'Alimentation Électrique** : Surcharge du système, détérioration de la qualité de l'approvisionnement, et risque de coupures [10].
- **Consommateurs Honnêtes** : Augmentation des factures pour compenser les pertes, impactant financièrement les consommateurs honnêtes [10].

I.4 Exigences de Sécurité pour les Réseaux de Comptage Intelligent

En raison des propriétés uniques des réseaux de comptage intelligent (SMNs), telles que la communication bidirectionnelle, il est difficile et complexe de sécuriser et protéger les données des compteurs intelligents contre les attaques. Bien qu'il existe de nombreuses solutions de sécurité pour les réseaux filaires et sans fil, elles ne peuvent pas simplement être directement implémentées dans les SMNs. Par conséquent, il est crucial d'explorer les exigences de sécurité pour ce réseau afin d'assurer la sécurité des communications de données dans les SMNs.

Dans cette section, nous mettons en lumière et discutons plusieurs exigences de sécurité pour les SMNs. Comme illustré dans la figure ci-dessous, il y a six exigences de sécurité essentielles : la confidentialité des données, l'intégrité des données, la fraîcheur des données, la disponibilité des données, la non-réputation et l'authentification, qui sont des aspects importants à considérer pour garantir la sécurité des communications de données dans les SMNs [8].

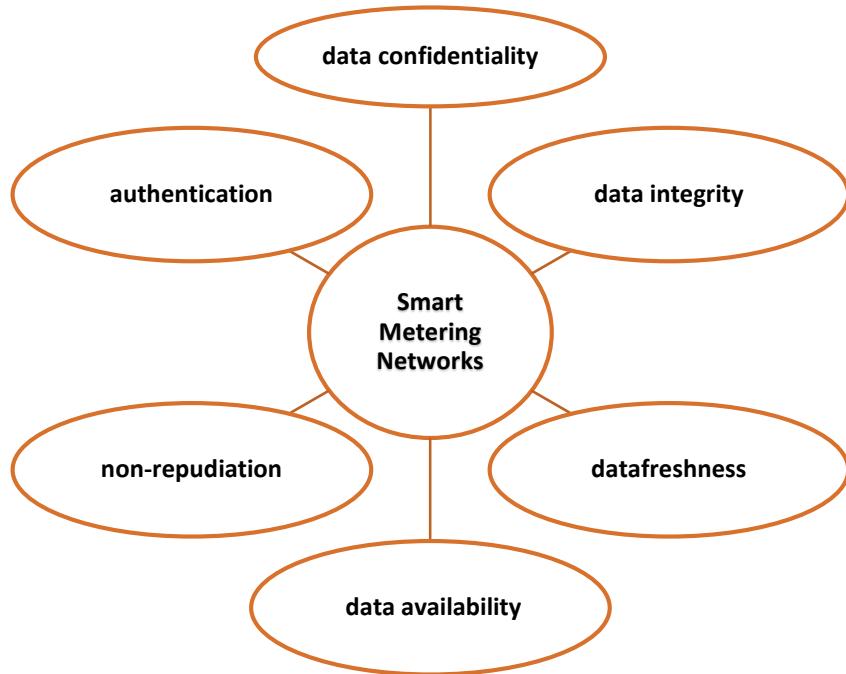


Figure I.8 : Exigences de sécurité pour les SMNs

- **Confidentialité des Données**

La confidentialité des données doit être préservée pour garantir que le contenu des données transmises ne sera jamais exposé à des parties non autorisées. Dans les SMNs, la confidentialité des données est une exigence essentielle pour protéger les données contre des attaques telles que l'écoute clandestine (EVD) et les attaques de l'homme du milieu (MIM), et pour préserver les informations privées des utilisateurs. La confidentialité des données peut être assurée en utilisant des techniques de cryptage des données, telles que le cryptage symétrique ou asymétrique.

- **Intégrité des Données**

Lors du transfert de données sur le réseau, l'expéditeur doit s'assurer que le destinataire reçoit des données authentiques, similaires à celles envoyées initialement. L'objectif de l'intégrité des données dans la communication de comptage intelligent est de garantir que le contenu des données d'origine n'a pas été modifié ou altéré accidentellement ou malicieusement pendant le processus de transmission. Pour garantir l'intégrité des données dans les SMNs, des fonctions de hachage ou des codes d'authentification de message (MAC) peuvent être ajoutés aux données cryptées afin que toute modification non autorisée des données d'origine puisse être détectée.

- **Fraîcheur des Données**

La fraîcheur des données est une autre exigence de sécurité importante pour garantir que les données transmises sont récentes et actuelles. Dans les SMNs, la fraîcheur des données peut être obtenue en utilisant des techniques de nonce qui peuvent être représentées sous forme de compteur, de timestamp ou de numéro de séquence de message généré aléatoirement à l'aide d'un générateur de nombres aléatoires. Pour éviter toute altération par un attaquant, le nonce sera crypté avec le message avant d'être envoyé à sa destination. Cette exigence vise à protéger les données contre les attaques de rejet.

- **Disponibilité des Données**

Si la confidentialité est associée à la vie privée, la disponibilité des données est associée à la survivabilité. Dans les SMNs, la disponibilité des données assure que le réseau est opérationnel et que les données sont accessibles même en présence d'une attaque par déni de service (DoS). Une attaque DoS pourrait être lancée à n'importe quelle couche des SMNs, telles que les attaques par brouillage qui peuvent perturber les fonctions des couches physiques et de contrôle d'accès au support.

À la couche réseau, un attaquant pourrait interrompre ou détruire le protocole de routage, tandis qu'à la couche application, un attaquant pourrait désactiver des services importants tels que la diffusion de réseau et les services de gestion des clés.

- **Non-Répudiation**

La non-répudiation est une exigence de sécurité visant à garantir qu'un expéditeur ou un destinataire ne peut pas nier avoir envoyé ou reçu un message. Cette exigence est essentielle pour détecter l'existence d'un attaquant tentant de lancer des attaques par injection de fausses données, des inondations ou des attaques par rejet. Il existe de nombreuses façons de satisfaire à cette exigence, l'une d'elles étant l'utilisation de la cryptographie à clé publique.

- **Authentification**

L'authentification est le processus permettant de déterminer ou de vérifier si une personne ou un objet est bien celui qu'il prétend être. L'authentification peut être divisée en authentification d'entité et authentification de données. L'authentification d'entité permet à un récepteur de vérifier si les données reçues proviennent bien de la source correcte. Dans ce cas, un attaquant ne peut pas participer ou se joindre à des activités dans le réseau ciblé car il n'a pas le privilège d'accéder au réseau. Sans authentification d'entité, un attaquant pourrait se faire passer pour un compteur intelligent légitime, obtenant ainsi un accès non autorisé à des données sensibles. D'autre part, l'authentification des données permet à un récepteur de vérifier que les données reçues sont similaires à celles qui ont été transmises.

Ces exigences de sécurité sont cruciales pour assurer la protection des données et la fiabilité des communications dans les réseaux de comptage intelligent, garantissant ainsi leur fonctionnement sécurisé et efficace.

I.5 Conclusion

En conclusion, les Smart Grids représentent une avancée technologique majeure dans le secteur de l'énergie, offrant des bénéfices substantiels en termes d'efficacité, de fiabilité et de durabilité. En intégrant des technologies avancées comme les systèmes de communication et les compteurs intelligents, ils permettent une gestion plus dynamique et réactive de l'énergie. Cependant, cette sophistication technologique entraîne également des défis complexes, notamment en matière de sécurité. La protection contre les cyberattaques et la fraude est essentielle pour maintenir la confiance des utilisateurs et la stabilité du réseau.

L'étude des menaces et des vulnérabilités spécifiques aux Smart Grids a mis en lumière la nécessité de solutions robustes pour détecter et prévenir les fraudes. Les méthodes traditionnelles de sécurité peuvent ne pas suffire face à la complexité croissante des attaques, ce qui nécessite l'adoption de technologies innovantes. Dans ce contexte, le recours au Deep Learning, Time Series Analysis et à la Blockchain émerge comme une approche prometteuse. La Blockchain, avec sa capacité à fournir une traçabilité transparente et sécurisée des transactions, et le Deep Learning, avec ses capacités avancées d'analyse et de détection de patterns, peuvent jouer un rôle crucial dans la sécurisation des Smart Grids.

Le chapitre suivant sera donc dédié à la technologie de la Blockchain, aux modèles de prédiction et de détection basés sur le Deep Learning, ainsi qu'à l'analyse des séries temporelles.

Chapitre II

Blockchain et Intelligence Artificielle au Service des Smart Grids

II.1 Introduction

Le développement des Smart Grids représente une avancée considérable dans la gestion et la distribution de l'énergie. La mise en place de ces réseaux intelligents exige une cohésion et une collaboration étroite entre tous les acteurs du réseau.

Cependant, plus le nombre d'acteurs impliqués augmente, plus la complexité du projet s'accroît, tant en matière d'organisation que d'exploitation des données. Cette complexité technologique nécessite la mise en œuvre d'une stratégie de défense exhaustive, capable de couvrir l'ensemble des menaces et vulnérabilités, qu'il s'agisse de détection ou de mesures proactives. La pérennité d'un réseau intelligent repose donc fondamentalement sur la synergie entre ses divers acteurs.

C'est dans ce cadre que des solutions technologiques telles que la Blockchain et le Deep Learning s'avèrent indispensables. Ce chapitre est consacré à l'étude de ces technologies, en mettant en lumière leur application pour renforcer la sécurité et l'efficacité des Smart Grids.

Nous nous attarderons sur trois domaines clés : l'analyse des séries temporelles, l'utilisation de la Blockchain, et l'application du Deep Learning. Chacune de ces technologies offre des solutions uniques pour améliorer la détection des fraudes et optimiser la gestion des réseaux électriques intelligents. Nous présenterons également les différentes solutions proposées dans la littérature pour leur intégration et leur mise en œuvre dans les Smart Grids.

II.2 Analyse des Séries Temporelles pour la Sécurité des Smart Grids

II.2.1 Introduction à l'analyse des séries temporelles

II.2.1.1 Définition d'une série temporelle

Une série temporelle, également appelée série chronologique ou chronique, est une suite finie (x_1, \dots, x_n) de données indexées par le temps. L'indice temporel peut varier selon les cas, allant de la minute à l'année. Le nombre n désigne la longueur de la série. Il est souvent très utile de représenter cette série temporelle sur un graphique construit de la manière suivante : en abscisse, le temps ; en ordonnée, la valeur de l'observation à chaque instant. Pour des raisons de lisibilité, les points obtenus sont reliés par des segments de droite, formant ainsi une ligne brisée.

Une série temporelle se définit donc comme une séquence de points de données ou d'observations enregistrées à différents intervalles de temps réguliers. En général, une série temporelle est constituée de points de données pris à des intervalles de temps équidistants. La fréquence des données enregistrées peut être horaire, quotidienne, hebdomadaire, mensuelle, trimestrielle ou annuelle. La prévision des séries temporelles consiste à utiliser un modèle statistique pour prédire les valeurs futures d'une série temporelle à partir des résultats passés.

Exemple : Figure II.1 présente la consommation d'électricité sur toute la France. Une donnée toutes les demi-heures, pendant 35 jours en juin-juillet 1991. Soit 1680 données. On y voit bien deux composantes périodiques correspondant à la journée (48 unités de temps) et à la semaine, ainsi qu'un effet « week-end » massif.[11]

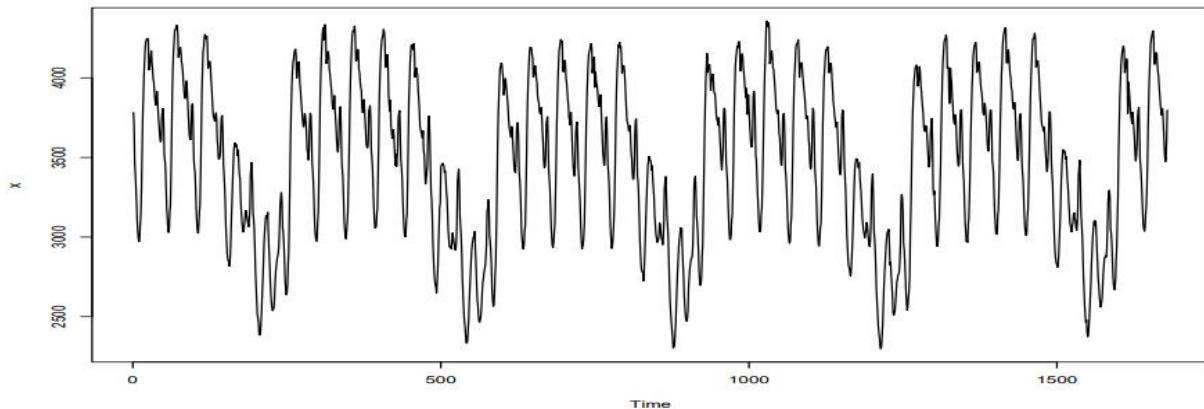


Figure II.1 : Consommation d'électricité sur toute la France

L'analyse des séries temporelles englobe des méthodes statistiques permettant d'examiner les données de séries temporelles. Ces méthodes nous permettent d'extraire des statistiques significatives, des motifs et d'autres caractéristiques des données. Les séries temporelles sont souvent visualisées à l'aide de graphiques en ligne. Ainsi, l'analyse des séries temporelles implique la compréhension des aspects inhérents des données temporelles, afin de créer des prévisions significatives et précises.

Les applications des séries temporelles se retrouvent dans les domaines des statistiques, de la finance ou des affaires. Un exemple courant de données de séries temporelles est la valeur de clôture quotidienne d'un indice boursier comme le NASDAQ ou le Dow Jones. D'autres applications courantes incluent les prévisions de ventes et de demande, la prévision météorologique, l'économétrie, le traitement du signal, la reconnaissance de formes, la prédiction des tremblements de terre et la prédiction de consommation d'électricité.

II.2.1.2 Composantes des Séries Temporelles

Les séries temporelles se composent de plusieurs éléments clés qui aident à analyser les tendances et à faire des prévisions précises. Ces composantes incluent la tendance, la saisonnalité, le cycle et les variations irrégulières.

- **Tendance (Tend)**

La tendance (T_t) indique la direction générale des données de la série temporelle sur une longue période. Elle peut être croissante (à la hausse), décroissante (à la baisse) ou horizontale (stationnaire). La tendance montre comment les données évoluent globalement au fil du temps, indépendamment des fluctuations saisonnières ou des cycles économiques.

- **Saison (Seasonality)**

La composante saisonnière (S_t) représente une tendance qui se répète régulièrement en fonction du moment, de la direction et de l'ampleur. Par exemple, une augmentation de la consommation d'eau en été en raison des conditions météorologiques chaudes est un phénomène saisonnier. Cette composante est cyclique et se reproduit à des intervalles de temps réguliers, tels que les jours, les mois ou les années.

- **Composante Cyclique**

Les composantes cycliques sont des tendances qui ne se répètent pas selon une période fixe. Un cycle fait référence aux périodes de hausses et de baisses, souvent observées dans les cycles économiques. Contrairement à la saisonnalité, ces cycles ne montrent pas de variation saisonnière spécifique, mais se produisent généralement sur une période de 3 à 12 ans, en fonction de la nature de la série temporelle.

- **Variation Irrégulière**

Les variations irrégulières, ou bruits (e_t), sont les fluctuations dans les données de la série temporelle qui deviennent évidentes lorsque les tendances et les variations cycliques sont supprimées. Ces variations sont imprévisibles, erratiques et peuvent ou non être aléatoires. Elles représentent les anomalies ou les événements inattendus qui ne suivent pas les tendances ou les cycles observés.

II.2.1.3 Types de données

Comme mentionné précédemment, l'analyse des séries temporelles est l'analyse statistique des données de séries temporelles. Les données de séries temporelles se caractérisent par des enregistrements effectués à différents moments ou intervalles de temps. Il existe trois types principaux de données de séries temporelles :

- Données de séries temporelles : Les observations des valeurs d'une variable enregistrées à différents points dans le temps sont appelées données de séries temporelles.
- Données transversales : Il s'agit des données d'une ou plusieurs variables enregistrées à un même instant.
- Données combinées : Elles résultent de la combinaison de données de séries temporelles et de données transversales.

II.2.1.4 Modélisations de Base pour les Séries Temporelles

Les séries temporelles peuvent être décomposées en différentes composantes pour en faciliter l'analyse et la prévision. Deux types principaux de décompositions sont la décomposition additive et la décomposition multiplicative.

- La Décomposition Additive

La décomposition additive est l'une des méthodes de base pour analyser les séries temporelles. Elle se définit comme suit :

$$X_t = m_t + S_t + U_t, \quad 1 \leq t \leq T$$

Où :

- m_t est la composante tendancielle déterministe, qui donne le comportement de la variable observée sur le long terme (croissance ou décroissance linéaire, quadratique, etc.).
- S_t est une suite périodique correspondant à une composante saisonnière (quotidienne, Trimestrielle et annuelle).
- U_t représente une composante irrégulière et aléatoire. Cette composante est importante en pratique car elle représente les erreurs de mesure ou les anomalies. U_t est souvent auto corrélée, ce qui signifie que la covariance entre U_t et U_{t+h} est non nulle.

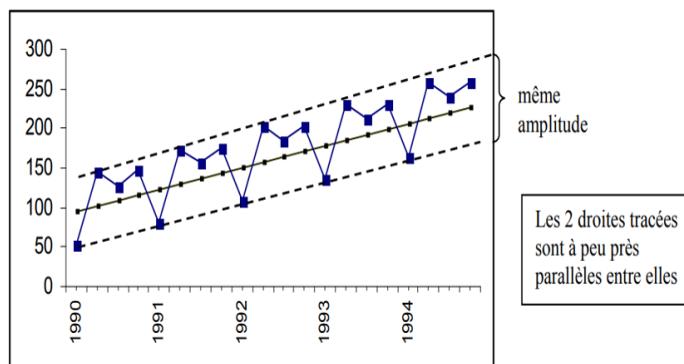


Figure II.2 : La Décomposition Additive

- La Décomposition Multiplicative

La décomposition multiplicative est une autre méthode pour analyser les séries temporelles, définie comme suit :

$$X_t = m_t S_t U_t, 1 \leq t \leq T$$

Ici, les composantes m_t et S_t sont de la même forme que pour le modèle additif, et la composante irrégulière U_t a pour moyenne 1.

Par une transformation logarithmique, cette décomposition peut être ramenée à une décomposition additive. La décomposition multiplicative est particulièrement utile lorsqu'on observe une variation linéaire des effets saisonniers.

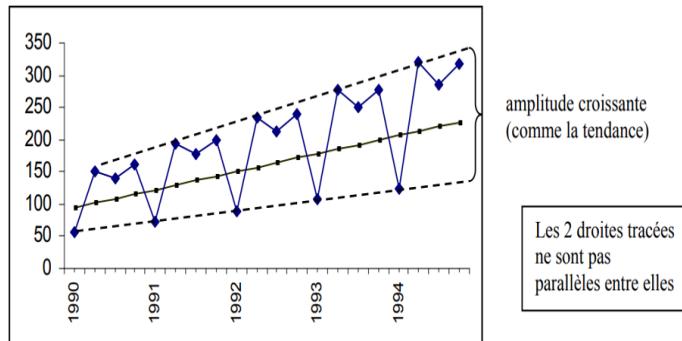


Figure II.3 : La Décomposition Multiplicative

II.2.2 Stationnarité

Une série temporelle est dite stationnaire si ses propriétés statistiques, comme la moyenne, la variance et la covariance, ne varient pas dans le temps. La stationnarité est une condition essentielle pour appliquer de nombreux modèles de prévision, car elle garantit que les caractéristiques de la série restent constantes au fil du temps, facilitant ainsi les prédictions.

- **Moyenne** : La moyenne d'une série stationnaire est constante au cours du temps.
- **Variance** : La variance, qui mesure la dispersion des valeurs autour de la moyenne, est également constante.
- **Covariance** : La covariance entre les valeurs à différents temps dépend uniquement de la différence de temps et non des temps eux-mêmes.

La stationnarité permet de faire des prédictions plus précises, car les propriétés futures de la série peuvent être estimées à partir des propriétés passées. On dit qu'une suite de variables aléatoires ($\dots, X_{-1}, X_0, X_1, \dots$) est stationnaire si les espérances sont constantes

$$E(X_n) = \mu \quad \forall n$$

Et si les covariances sont stables par translation du temps c'est-à-dire, pour tout h ,

$$\text{Cov}(X_n, X_{n+h}) = : \sigma(h) \forall n,$$

Où la covariance est définie par :

$$\begin{aligned} \text{Cov}(X_n, X_{n+h}) &= E(X_n X_{n+h}) - E(X_n)E(X_{n+h}) \\ &= E((X_n - E X_n)(X_{n+h} - E X_{n+h})) \end{aligned}$$

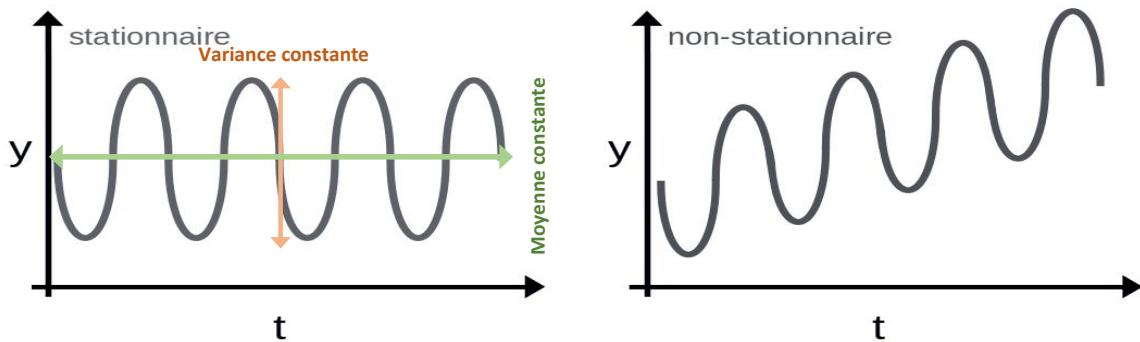


Figure II.4 : Série temporelle stationnaire et non stationnaire

II.2.2.1 Tests de Stationnarité

- Test Racine unitaire (Non-Stationnarité)

- *Test de Dickey Fuller augmenté(ADF)*
- *Test Phillips Perron*

L'hypothèse nulle est que la série a été générée par un processus présentant une racine unitaire, et donc, qu'elle n'est pas stationnaire.

$\text{Test d'hypothèse} \begin{cases} H_0 & \text{n'est pas stationnaire} \\ H_1 & \text{stationnaire} \end{cases}$
--

- Test stationnarité

- *Test KPSS*

Le test KPSS part de l'hypothèse nulle que la série est stationnaire. Si la p-value est inférieure au niveau de signification, on rejette l'hypothèse nulle et conclut que la série est non stationnaire.

<i>Test d'hypothèse</i>	$\begin{cases} H_0 & \text{stationnaire} \\ H_1 & \text{n'est pas stationnaire} \end{cases}$
-------------------------	--

Remarque : P – Value < α → Rejeter H_0 ou α : Niveau de signification (0,05)

II.2.3 Les modèles de séries temporelles

Différents modèles statistiques permettent de capturer les caractéristiques et les tendances des séries temporelles, chacun offrant des perspectives uniques pour analyser et prédire les valeurs futures. Voici les principaux types de modèles utilisés dans l'analyse des séries temporelles :

II.2.3.1 Modèle AutoRégressif (AR)

Le modèle autorégressif, noté AR(p), modélise une série temporelle en fonction de ses valeurs passées. La relation peut être exprimée comme suit :

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t$$

Où $\phi_0, \phi_1, \dots, \phi_p$ sont les paramètres du modèle et ϵ_t est une erreur aléatoire.

II.2.3.2 Modèle de Moyenne Mobile (MA)

Le modèle MA(q) modélise une série temporelle en fonction des erreurs passées. La relation peut être formulée comme suit :

$$X_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \epsilon_t$$

Où μ est la moyenne de la série, $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients et ϵ_t est une erreur aléatoire.

II.2.3.3 Modèle Autorégressif à Moyenne Mobile (ARMA)

Le modèle ARMA (p, q) combine les modèles AR et MA pour modéliser des séries temporelles stationnaires en utilisant à la fois les valeurs passées et les erreurs passées

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

Pour mettre en place ce modèle, il faut s'assurer de la stationnarité de la série temporelle sinon un passage au log peut suffire.

II.2.3.4 Modèle Autorégressif Intégré à Moyenne Mobile (ARIMA)

Lorsque la série n'est pas stationnaire, le modèle ARIMA (p, d, q) est utilisé pour la différencier et la rendre stationnaire. La composante "intégrée" (I) du modèle ARIMA indique l'ordre de différenciation nécessaire pour obtenir la stationnarité.

En considérant : ARIMA (1, 1, 1) alors transformation $Z_t = X_t - X_{t-1}$

$$Z_t = \alpha_0 + \alpha_1 Z_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$$

Où Z_t représente la série différenciée.

II.2.3.5 Modèle Saisonnier ARIMA (SARIMA) :

Le modèle SARIMA est une extension du modèle ARIMA pour traiter les composantes saisonnières d'une série temporelle. Il est représenté par $SARIMA (p, d, q)(P, D, Q, s)$ où les termes (P, D, Q) représentent les paramètres de la composante saisonnière et " s " indique la période de saisonnalité.

II.2.4 Approche de Box et Jenkins

La méthodologie de Box et Jenkins est une approche systématique pour identifier, estimer et diagnostiquer des modèles ARIMA pour les séries temporelles. Elle se compose de quatre étapes principales :

- **Identification Provisoire** : Utilisation des données historiques pour identifier un modèle ARIMA approprié en examinant les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF).

- **Estimation** : Estimation des paramètres du modèle provisoirement identifié à partir des données historiques.
- **Contrôle Diagnostique** : Vérification de l'adéquation du modèle en utilisant des tests statistiques et des diagnostics sur les résidus.
- **Prévisions** : Utilisation du modèle final pour prévoir les valeurs futures de la série temporelle.

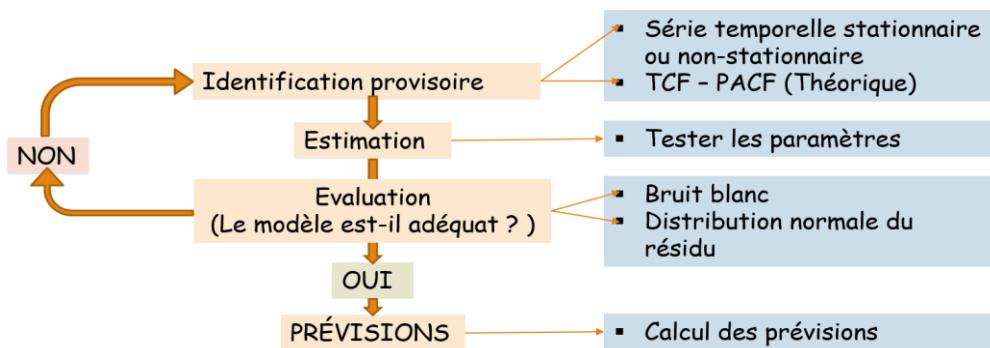


Figure II.5 : Organigramme de la méthode Box et Jenkins

Les graphiques ACF et PACF sont essentiels pour déterminer l'ordre des modèles AR, MA et ARIMA.

II.2.4.1 AutoCorrelation Function : ACF

La fonction d'autocorrélation (ACF) mesure la corrélation entre les valeurs d'une série temporelle à différents décalages.

Elle commence à un décalage de 0, correspondant à la corrélation de la série temporelle avec elle-même, ce qui donne une corrélation de 1. Le graphique ACF permet de déterminer si la série temporelle observée est un bruit blanc, si une observation est liée à une observation adjacente ou à une observation éloignée, et si la série peut être modélisée par un modèle de moyenne mobile (MA) et son ordre.

II.2.4.2 Partial Autocorrelation Function : PACF

La fonction d'autocorrélation partielle (PACF) mesure la corrélation entre les valeurs d'une série temporelle en éliminant l'impact des autres valeurs décalées. Elle répond à des questions sur la modélisation de la série temporelle par un modèle autorégressif (AR) et l'ordre du modèle. La PACF montre la corrélation supplémentaire expliquée par chaque terme décalé successif.

Les deux fonctions ACF et PACF commencent à un décalage de 0, avec une corrélation de 1. La différence entre l'ACF et la PACF réside dans l'inclusion ou l'exclusion des corrélations indirectes dans le calcul. Les graphiques montrent une zone bleue représentant l'intervalle de confiance à 95 %, indiquant le seuil de signification. Tout ce qui se trouve dans la zone bleue est statistiquement proche de zéro, tandis que ce qui est en dehors est statistiquement non nul.

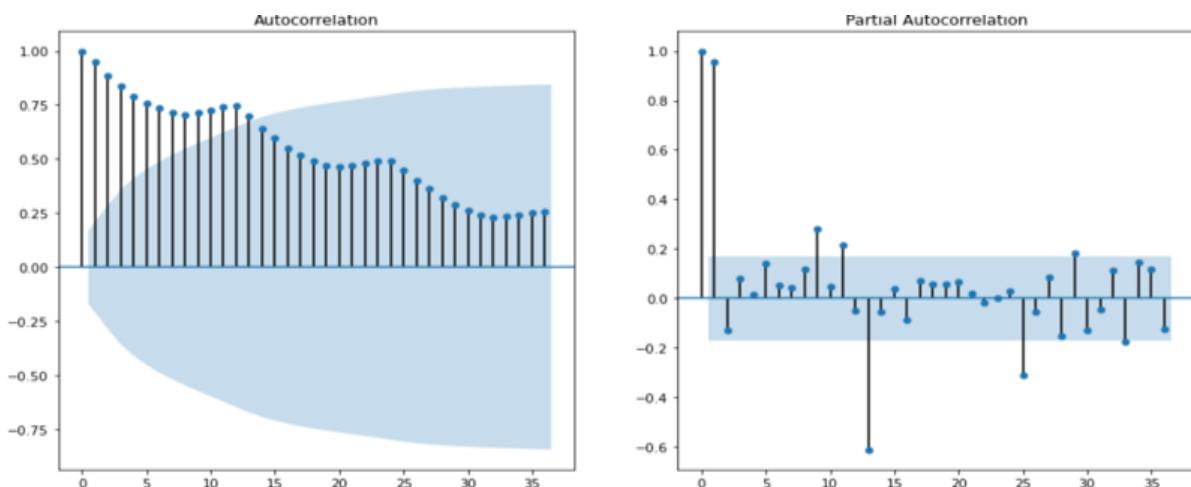


Figure II.6 : ACF & PACF

II.2.4.3 Théorie Générale de l'ACF et du PACF des Modèles ARIMA

	AR(p)	MA(q)	ARMA (p, q)
ACF	Tends vers 0	S'annule après l'ordre q	Tends vers 0
PACF	S'annule après l'ordre p	Tends vers 0	Tends vers 0

Table II.1 Théorie générale de l'ACF et du PACF

II.2.5 Modélisation ARIMA et Détection des Attaques dans les Smart Grids

Dans l'article [12] les auteurs soulignent que les modèles ARMA, couramment utilisés dans la littérature sur la détection d'anomalies, ne conviennent pas aux comportements de consommation non stationnaires observés chez la plupart des consommateurs. En utilisant des méthodes automatisées d'ajustement de modèles, ils démontrent que la différenciation de premier ordre permet de rendre ces relevés faiblement stationnaires, justifiant ainsi l'emploi des modèles ARIMA pour la validation des consommations. Leur étude évalue également l'efficacité de la prévision ARIMA dans le contexte d'un modèle d'attaque spécifique, où les relevés de compteurs intelligents sont modifiés pour voler de l'électricité. Ils proposent des contrôles supplémentaires sur la moyenne et la variance des relevés pour réduire significativement la quantité d'électricité volée. Leur évaluation, basée sur un jeu de données réel de 450 compteurs, montre que ces contrôles peuvent réduire le vol d'électricité jusqu'à 77,46 %. Leurs travaux fournissent des algorithmes spécifiques que les compagnies d'électricité peuvent utiliser pour vérifier l'intégrité des relevés rapportés par les compteurs intelligents, contribuant ainsi à renforcer la sécurité des infrastructures critiques.

L'intervalle de confiance ARIMA fournit une limite aux mesures et sert de bon détecteur de mesures invalides pour les compteurs défectueux. Cependant, ces limites ne sont pas suffisantes pour détecter les attaques où l'attaquant a une connaissance complète du système.

Les auteurs considèrent un modèle d'attaque spécifique dans lequel l'attaquant vole de l'électricité à un voisin pour un gain monétaire. Dans ce modèle, la consommation de l'attaquant à l'instant t est A_t et celle de son voisin est X_t .

L'attaquant compromet le compteur intelligent de son voisin pour signaler une consommation $X'_t > X_t$ tout en déclarant sa propre sous-consommation en ajustant sa lecture à $A'_t = A_t - (X'_t - X_t)$. Ainsi, il vole une quantité $(X'_t - X_t)$ à son voisin, tout en évitant la vérification de l'équilibre effectuée par le réseau de distribution électrique, puisque la somme des relevés rapportés $(X'_t + A'_t)$ correspond à la consommation mesurée $(X_t + A_t)$.

Sans le mécanisme de détection ARIMA en place, l'attaquant pourrait voler une quantité arbitraire d'électricité, limitée uniquement par les contraintes physiques du système de distribution électrique. Plus précisément, les lignes de distribution sont classées en fonction du courant maximal qu'elles peuvent transporter. Si la demande de l'attaquant augmente, le courant dans les lignes de distribution augmente également, générant de la chaleur sous forme de pertes I^2R , où I est le courant et R est la résistance. Si ce courant dépasse les limites thermiques normales, les lignes peuvent subir des dommages entraînant des coupures de courant ou d'autres pannes d'équipement, ce qui constitue une indication évidente de consommation anormale. Par conséquent, l'attaquant essaierait d'éviter la détection en maintenant sa consommation dans ces limites physiques.

II.3 Intelligence artificielle pour la Sécurité des Smart Grids

Afin de garantir un service fiable et prévisible dans le réseau électrique, il est crucial d'évaluer la fiabilité des composants essentiels et des sous-stations. Un des principaux défis auxquels les réseaux intelligents sont confrontés est le vol d'énergie, qui occasionne des pertes financières considérables pour les entreprises de services publics. Vu l'immensité des données à analyser, une inspection manuelle des vols d'énergie est impraticable.

C'est pourquoi les techniques d'apprentissage automatique, les méthodes statistiques et l'analyse comportementale sont indispensables pour détecter efficacement toute tentative de détournement d'énergie dans les réseaux intelligents.

II.3.1 Définition

L'ambition de l'intelligence artificielle (IA) réside dans la reproduction de l'intelligence humaine au sein des systèmes informatiques. En informatique, l'IA englobe l'étude des "agents intelligents", des entités capables de percevoir leur environnement et d'initier des actions pour optimiser leurs chances de réaliser des objectifs précis. Le domaine de l'apprentissage automatique (machine learning - ML) se focalise quant à lui sur le développement et l'application de méthodes permettant l'acquisition de connaissances à partir de jeux de données. Le ML est largement employé dans une diversité de secteurs tels que la reconnaissance vocale, la vision par ordinateur, l'analyse de texte, les jeux vidéo, les sciences médicales et le cyber sécurité [13].

L'apprentissage profond (Deep Learning - DL) représente une sous-discipline de l'apprentissage automatique qui se distingue par son aptitude à traiter efficacement des données non structurées. Actuellement, les techniques de Deep Learning surpassent les approches conventionnelles en matière d'apprentissage automatique. Les modèles de Deep Learning s'inspirent de l'architecture et du fonctionnement du système nerveux et du cerveau humains. Ces modèles organisent les unités de traitement en couches d'entrée, de cachées et de sortie. Au sein de chaque couche, les nœuds ou unités sont interconnectés avec ceux de la couche inférieure, et chaque connexion se voit attribuer un poids. Les unités agrègent les entrées en les multipliant par leurs poids respectifs.

La Figure II.7 illustre la corrélation entre l'IA, le ML et le DL, soulignant que le machine Learning et le Deep Learning constituent des sous-domaines de l'intelligence artificielle [13].

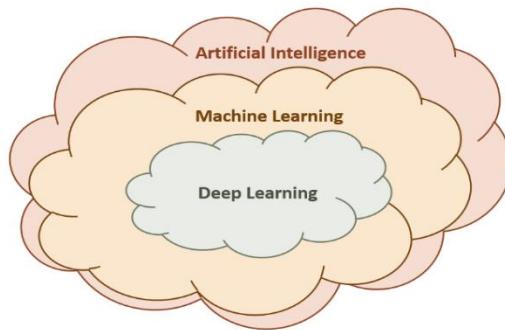


Figure II.7 : Relations entre l'AI, ML et DL

II.3.2 Les Modèles Deep Learning

L'apprentissage profond : consiste à apprendre des représentations hiérarchiques de données en utilisant des architectures avec de multiples couches cachées. Grâce aux avancées des infrastructures de calcul haut performance, les techniques d'apprentissage profond utilisant des réseaux neuronaux profonds ont gagné en popularité. Dans un algorithme de Deep Learning, les données traversent plusieurs couches, chaque couche extrayant progressivement des caractéristiques et transmettant l'information à la couche suivante. Les premières couches extraient des caractéristiques de bas niveau, lesquelles sont ensuite combinées par les couches ultérieures pour former une représentation globale.

À l'ère de l'apprentissage profond, une vaste gamme de méthodes et d'architectures a été développée. Ces modèles peuvent être classés en deux grandes catégories : les approches discriminatives (supervisées) et génératives (non supervisées). Parmi les modèles discriminatifs, on trouve deux groupes principaux : les réseaux neuronaux convolutionnels (CNN) et les réseaux neuronaux récurrents (RNN). Les approches génératives comprennent divers modèles tels que les réseaux antagonistes génératifs (GAN) et les autoEncodeurs (AE). Dans les sections suivantes, nous proposons une étude complète des différents types de modèles d'apprentissage profond.

II.3.2.1 Les réseaux neuronaux convolutionnels

Les réseaux neuronaux convolutionnels (CNN) constituent une classe avancée de modèles d'apprentissage profond, largement appliqués à diverses tâches telles que la détection d'objets, la reconnaissance vocale, la vision par ordinateur, la classification d'images et la Bio-informatique. Ils ont également prouvé leur efficacité dans la prédition de séries temporelles. Contrairement aux méthodes traditionnelles, les CNN apprennent et reconnaissent automatiquement les caractéristiques des données sans nécessiter d'extraction manuelle. Inspirés par la perception visuelle, leurs composants principaux incluent les couches de convolution, de Pooling et entièrement connectées [13].

Les couches de convolution des CNN capturent les corrélations sémantiques des caractéristiques spatiales grâce à des opérations de convolution sur des données multidimensionnelles. La carte des caractéristiques d'un CNN comprend plusieurs filtres répartis spatialement, tandis que les opérations de Pooling réduisent la taille des cartes de caractéristiques. Les couches entièrement connectées combinent ces caractéristiques extraites avec une matrice de poids et un biais, suivies d'une fonction d'activation [13].

Cependant, les CNN présentent certaines limitations, comme leur incapacité à capturer des caractéristiques temporelles et des temps d'entraînement prolongés. Des variantes améliorées, telles que les réseaux de convolution temporelle (TCN) et les modules de compression et excitation (CNN-SE), ont été développées pour surmonter ces limitations. Malgré les défis liés à la profondeur et au nombre de couches, les CNN restent des outils puissants pour l'extraction de caractéristiques spatiales et l'implémentation distribuée [14].

Les CNN se classent en deux catégories principales en fonction de la nature des données d'entrée : les CNN 1D et les CNN 2D.

- **CNN 1D (Convolutional Neural Networks 1D)**

Les CNN 1D sont principalement utilisés pour traiter des données séquentielles ou des séries temporelles, telles que des signaux ou des données textuelles. Ils appliquent des filtres de convolution le long d'une dimension pour extraire des motifs locaux dans les séquences d'entrée. Par exemple, pour la détection de fraudes électriques, un CNN 1D peut analyser les relevés de consommation d'électricité pour identifier des anomalies. Les caractéristiques extraites sont ensuite aplatis et passées à des couches entièrement connectées pour la classification finale.

- **CNN 2D (Convolutional Neural Networks 2D)**

Les CNN 2D traitent des données structurées en deux dimensions, comme les images. Les filtres de convolution dans les CNN 2D glissent le long des deux dimensions pour extraire des motifs locaux. Les caractéristiques extraites sont également aplatis et passées à des couches entièrement connectées pour la classification finale.

II.3.2.2 Les Réseaux Neuronaux Récursifs (RNN)

Les réseaux neuronaux récurrents (RNN) représentent une catégorie sophistiquée de modèles d'apprentissage profond, caractérisés par leur mémoire interne, qui leur permet de capturer les dépendances séquentielles des données. Contrairement aux réseaux neuronaux classiques, qui traitent les entrées comme des entités indépendantes, les RNN considèrent l'ordre temporel des entrées, les rendant ainsi adaptés aux tâches impliquant des informations séquentielles. Grâce à une boucle récurrente, les RNN appliquent la même opération à chaque élément d'une série, où chaque calcul actuel dépend de l'entrée courante ainsi que des calculs précédents. Cette capacité à exploiter des informations contextuelles est particulièrement précieuse dans des domaines tels que le traitement du langage naturel, la classification vidéo et la reconnaissance vocale.

Par exemple, dans la modélisation linguistique, la compréhension des mots précédents d'une phrase est cruciale pour prédire le mot suivant, une tâche dans laquelle les RNN excellent en raison de leur nature récurrente.

Les RNN sont spécifiquement conçus pour traiter des données séquentielles, où la sortie du réseau est rétro alimentée à l'entrée. Le traitement récursif des RNN comporte des couches cachées avec une boucle de rétroaction, fournissant des informations pertinentes sur les états antérieurs. Pour une séquence d'entrées et de sorties, les états cachés sont calculés à l'aide d'une fonction non linéaire, telle que la fonction sigmoïde. La Figure II.8 montre l'architecture du modèle proposé.

Cependant, une des principales limitations des RNN simples est leur mémoire à court terme, ce qui restreint leur capacité à retenir des informations sur de longues séquences. Pour pallier cette limitation, des variantes plus avancées des RNN ont été développées, telles que le Long Short-Term Memory (LSTM).

II.3.2.3 LSTM

Le Long Short-Term Memory (LSTM) est une variante avancée des Réseaux Neuronaux Récurrents (RNN) qui résout le problème de la capture des dépendances à long terme. Introduit initialement par [32] en 1997 et amélioré par [38] en 2013, le LSTM a gagné une popularité significative dans la communauté de l'apprentissage profond. Comparé aux RNN standards, les modèles LSTM se sont avérés plus efficaces pour conserver et utiliser les informations sur de longues séquences.

Dans un réseau LSTM, l'entrée actuelle à un instant spécifique et la sortie de l'instant précédent sont intégrées dans l'unité LSTM, qui génère ensuite une sortie transmise à l'instant suivant. La couche cachée finale du dernier instant, parfois en conjonction avec toutes les couches cachées, est généralement utilisée à des fins de classification. L'architecture globale d'un réseau LSTM est illustrée dans la Figure II.8 [14].

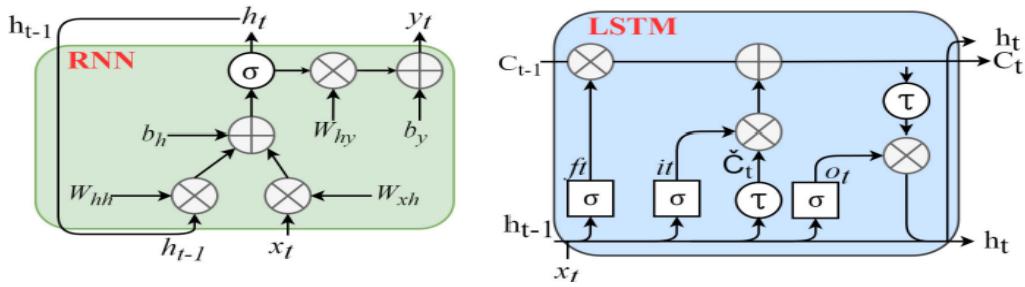


Figure II.8 : Structure de RNN et LSTM

II.3.3 Protection contre le vol d'énergie basé sur l'IA

De nombreuses études récentes ont mis en lumière les vulnérabilités des Smart Grids et proposé des solutions de sécurité avancées. L'apprentissage en profondeur s'est révélé particulièrement prometteur pour détecter les intrusions en identifiant des modèles complexes et des anomalies dans les données. Ces avancées sophistiquées améliorent la résilience et la sécurité des réseaux intelligents. Nous citons quelques exemples de ces travaux dans la section suivante.

II.3.3.1 Wide & Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids[15]

Zheng et al. présentent une méthode innovante pour la détection du vol d'électricité dans les réseaux intelligents en utilisant un modèle de réseaux neuronaux convolutifs profonds et larges (Wide & Deep Convolutional Neural Networks, CNN). Cette approche vise à surmonter les limitations des méthodes existantes en capturant à la fois les caractéristiques globales des données de consommation électrique sur une dimension (1-D) et les caractéristiques périodiques et non périodiques sur deux dimensions (2-D) [15].

- Architecture Proposée

L'architecture du modèle Wide & Deep CNN [15] se compose de deux composants principaux :

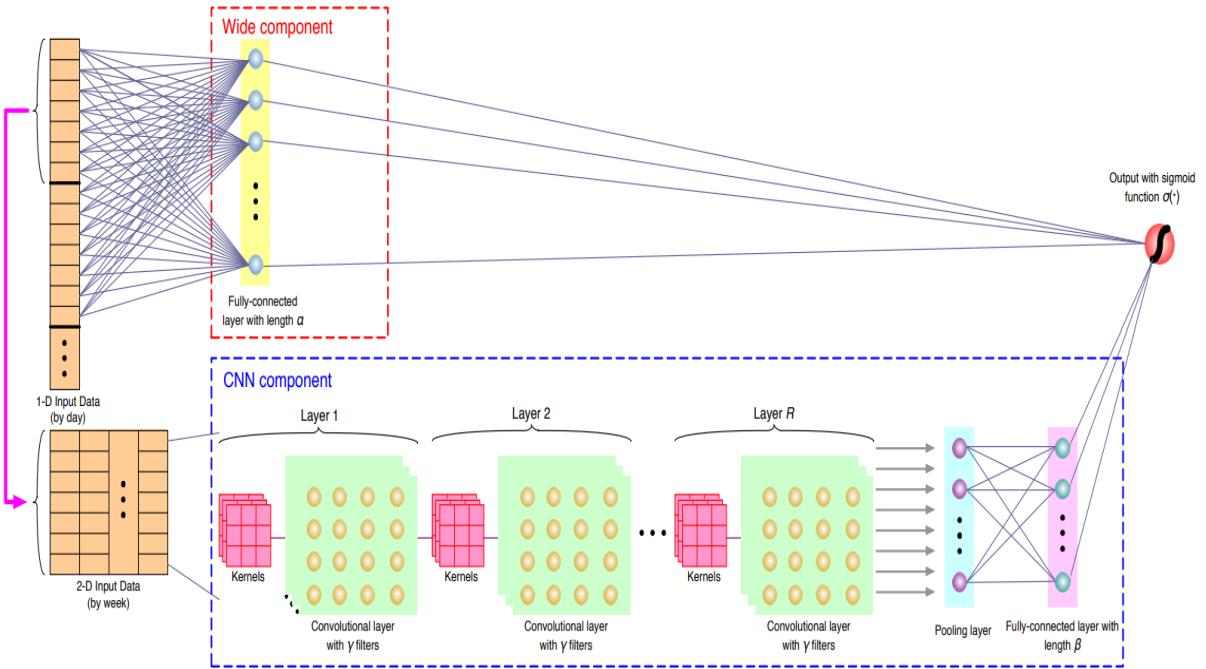


Figure II.9 : Wide & Deep Convolutional Neural Networks (CNN) framework

- Composant Large (Wide Component) :

Ce composant utilise une couche entièrement connectée de réseaux neuronaux pour extraire les caractéristiques globales des données de consommation électrique en une dimension (1-D). Chaque neurone de cette couche calcule son propre score en appliquant une équation de pondération et de biais aux données d'entrée 1-D. La fonction d'activation utilisée est ReLU (Rectified Linear Unit), qui active uniquement les valeurs positives afin de prévenir le surapprentissage.

- Composant Profond CNN (Deep CNN Component) :

Ce processus transforme les données de consommation électrique 1-D en données 2-D hebdomadaires pour capturer les caractéristiques périodiques. Il comprend plusieurs couches convolutionnelles, une couche de pooling et une couche entièrement connectée. Les couches convolutionnelles extraient les caractéristiques périodiques à l'aide de filtres 3x3, tandis que la couche de pooling réduit le nombre de paramètres en sélectionnant les valeurs maximales.

Enfin, la couche entièrement connectée fusionne ces caractéristiques pour obtenir une représentation finale des données.

Les deux composants, Large et Profond, sont combinés en utilisant une somme pondérée de leurs sorties, puis alimentés dans une fonction de perte logistique pour la formation et la prédiction conjointes. Le modèle effectue une propagation avant et arrière pour optimiser les paramètres générés par les deux composants[15].

- **Résultats**

Les expériences, utilisant diverses proportions de données d'entraînement (50%, 60%, 70%, 80%), ont évalué les performances en termes de courbe ROC (AUC). Les résultats pour le modèle Wide & Deep CNN sont présentés dans le tableau ci-dessous :

Training ratio	50%	60%	70%	80%
AUC	0.7760	0.7922	0.7860	0.7815

Table II.2 : Résultats Wide & Deep CNN

II.3.3.2 Electricity Theft Detection in Power Grids with Deep Learning and Random Forests [16]

L'article propose un modèle hybride novateur, le CNN-RF, qui associe un réseau neuronal à convolution (CNN) et une forêt aléatoire (RF) pour détecter automatiquement le vol d'électricité. Le processus de détection, illustré dans Figure II.10, se divise en trois étapes principales : l'analyse et le prétraitement des données, la génération des ensembles d'entraînement et de test, et la classification à l'aide du modèle CNN-RF. L'analyse des données inclut la séparation des utilisateurs résidentiels et non résidentiels, ainsi que la prise en compte des variations saisonnières.

Les données sont ensuite prétraitées pour nettoyer les valeurs manquantes et les biais. Les ensembles d'entraînement et de test sont équilibrés à l'aide de l'algorithme SMOT. La classification se fait par l'extraction de caractéristiques complexes avec le CNN, suivie de la classification avec la forêt aléatoire [16].

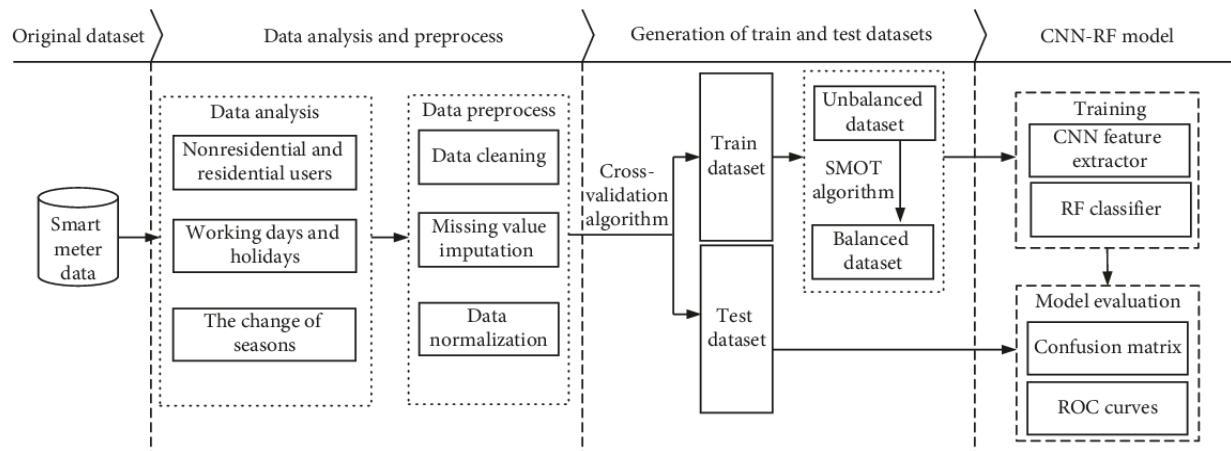


Figure II.10 : Flux de détection de vol d'énergie avec CNN-RF

Le tableau suivant montre les résultats de classification du modèle montrent une haute précision, un rappel élevé et un score F1 significatif pour les deux classes où la classe 0 est un modèle d'anomalie et la classe 1 est un modèle normal.

	Classe 0	Classe 1	Moyenne/Total
Précision	0.97	0.98	0.97
Rappel	0.96	0.98	0.97
Score F1	0.96	0.98	0.97

Table II.3 : Résumé des scores de classification de CNN-RF

II.4 Blockchain pour la Sécurité des Smart Grids

La transition vers les réseaux intelligents pose un défi aux autorités, passant d'une structure centralisée à une plus décentralisée. La technologie des réseaux intelligents permet d'orchestrer de multiples connexions, nécessitant une adaptation des méthodes de gestion.

La Blockchain émerge comme une solution avec ses caractéristiques uniques de gestion des données. Contrairement à une architecture centralisée où une tierce partie contrôle les interactions, la Blockchain offre une architecture décentralisée, permettant des transactions directes entre les participants. La figure illustre cette transition et explore les implications de la Blockchain dans les réseaux intelligents en pleine évolution.

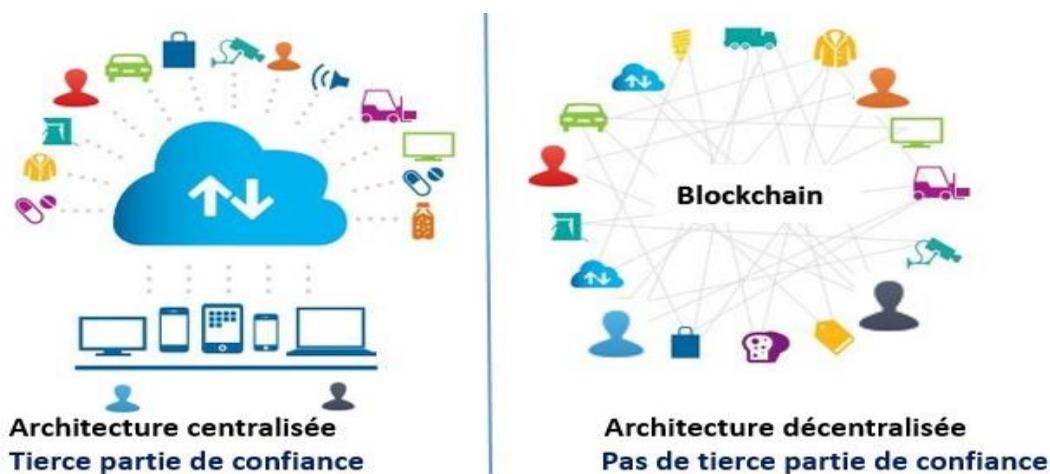


Figure II.11 Architecture centralisée & Architecture décentralisée

Dans cette section, nous cherchons à définir cette technologie et à étudier son application dans les réseaux intelligents.

II.4.1 Technologie Blockchain

La Blockchain se définit comme une chaîne composée de nombreux blocs contenant des informations, ou comme un registre distribué constitué d'une collection de blocs qui enregistrent divers types de données ou des informations sur les transactions. Chaque bloc est relié aux autres par une chaîne, chaque bloc se référant au hachage cryptographique des données des blocs précédents, comme illustré par la Figure II.12. Dans le réseau Blockchain, les blocs nouvellement générés sont continuellement ajoutés à la chaîne à intervalles réguliers, et cette chaîne est répliquée parmi les membres du réseau [17].

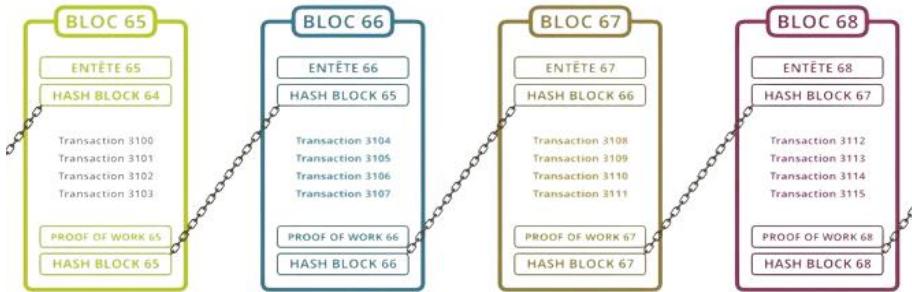


Figure II.12 : Structure générale de la Blockchain

Chaque bloc peut également inclure un horodatage, un nonce, un arbre de hachage appelé arbre de Merkle, ainsi que des scripts de contrats intelligents. Le hachage et l'arbre de Merkle permettent de vérifier que le contenu du bloc n'a pas été modifié, garantissant ainsi l'intégrité des données. Cette technologie assure un transfert sécurisé de propriétés, d'argent et de données sans nécessiter d'intermédiaires tels que les gouvernements ou les banques.

II.4.1.1 L'arbre de Merkle

L'arbre de Merkle, proposé par Ralph Merkle en 1979, est une méthode permettant de vérifier l'intégrité des données en les divisant en blocs, en appliquant des hachages à chaque bloc et en organisant ces hachages sous forme d'un arbre binaire jusqu'à obtenir une racine unique. Utilisé couramment dans la technologie Blockchain, l'arbre de Merkle garantit que toute altération des données peut être immédiatement détectée. Cependant, bien qu'il vérifie l'intégrité, il ne garantit pas la sécurité intrinsèque des données, notamment contre les falsifications avec des hachages reconstruits. Pour pallier ces limites, une nouvelle méthode, le SM-Tree, est proposée. Adaptée aux environnements de surveillance vidéo basés sur la Blockchain, cette méthode assure une vérification efficace des données, empêche leur manipulation et permet des mises à jour delta pour une synchronisation rapide et économique en bande passante [18].

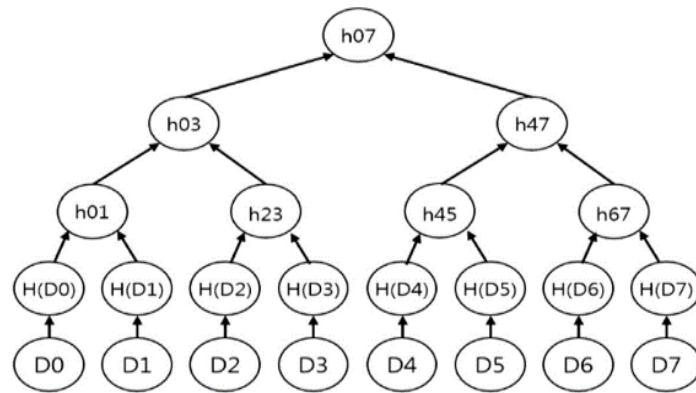


Figure II.13 Arbre de Merkle

II.4.1.2 Principe de fonctionnement de la Blockchain

La Blockchain fonctionne en suivant un processus standardisé. Une transaction initiée par un utilisateur est diffusée à tous les participants du réseau. Les nœuds vérifient la transaction via des mécanismes de hachage. Après vérification, la transaction est ajoutée à un nouveau bloc, qui est ensuite intégré à la chaîne existante, garantissant transparence et immutabilité. Les hachages sécurisent la Blockchain, mais des ordinateurs puissants pourraient permettre des manipulations. Divers algorithmes de consensus sont donc utilisés pour prévenir ces menaces. Lorsqu'une transaction est formulée, elle est enregistrée dans le bloc en cours. Les nœuds de validation vérifient chaque transaction avant d'ajouter le bloc à la chaîne. Une fois validée, le bloc est horodaté et ajouté à la chaîne, finalisant ainsi la transaction.

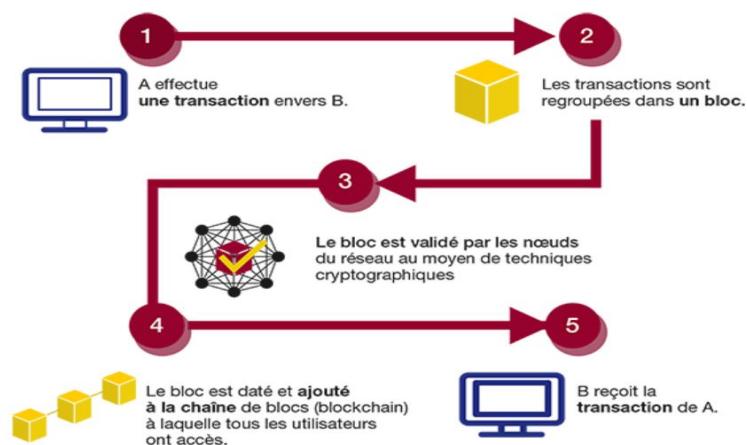


Figure II.14 : Le principe de fonctionnement de la Blockchain

II.4.1.3 Les protocoles de consensus

Pour assurer le bon fonctionnement d'un registre distribué, il est impératif de disposer d'un processus de validation des transactions, connu sous le nom de consensus, qui repose sur des mécanismes ou algorithmes spécifiques.

Les mécanismes de consensus, également appelés protocoles ou algorithmes déterminent qui peut ajouter de nouveaux blocs à la Blockchain et comment les nœuds parviennent à un accord sur le prochain bloc à ajouter. Ces mécanismes sont essentiels pour garantir le bon fonctionnement d'une Blockchain, notamment dans le domaine des crypto-monnaies, car un choix inapproprié pourrait exposer la Blockchain à diverses attaques.

- **Proof-of-work (PoW)**

PoW : Preuve de Travail est adopté par Bitcoin, Ethereum,... . PoW sélectionne un nœud pour créer un nouveau bloc à chaque tour de consensus par compétition de puissance de calcul. Dans la compétition, les nœuds participants doivent résoudre un casse-tête cryptographique. Le nœud qui résout d'abord le casse-tête a le droit de créer un nouveau bloc. Il est très difficile de résoudre un casse-tête PoW. Les nœuds doivent ajuster la valeur du nonce pour obtenir la réponse correcte, ce qui nécessite beaucoup de puissance de calcul. Un attaquant malveillant peut renverser un bloc dans une chaîne, mais à mesure que les blocs valides dans la chaîne augmentent, la charge de travail s'accumule également, donc renverser une longue chaîne nécessite une énorme quantité de puissance de calcul. PoW appartient aux protocoles de consensus à finalité probabiliste car il garantit une cohérence éventuelle [19].

- **Proof-of-stake (PoS)**

PoS : Preuve d'Enjeu, la sélection à chaque tour du nœud qui crée un nouveau bloc dépend de l'enjeu détenu plutôt que de la puissance de calcul. Bien que les nœuds doivent encore résoudre un casse-tête SHA256.

La différence avec PoW est que les nœuds n'ont pas besoin d'ajuster le nonce plusieurs fois, mais la clé pour résoudre ce casse-tête est le montant de l'enjeu (pièces). Ainsi, PoS est un protocole de consensus économe en énergie, qui utilise un système d'incitation à la monnaie interne au lieu de consommer beaucoup de puissance de calcul pour parvenir à un consensus. PoS est également un protocole de consensus à finalité probabiliste [19].

- **Proof-of-authority (PoA)**

Preuve d'Autorité (PoA) : La PoA est spécifiquement conçue pour les Blockchains autorisées. Dans ce protocole, les participants doivent vérifier leur identité au sein du réseau avant d'obtenir l'autorisation de publier un bloc. Contrairement à la PoS, où la possession de pièces ou d'actifs est nécessaire, la PoA considère l'identité d'un participant comme sa mise en jeu. De plus, elle suppose que les autorités sont présélectionnées et dignes de confiance pour la publication de blocs. Cette approche facilite également la détection d'autorités malveillantes, permettant une notification rapide de toute activité malveillante aux autres nœuds. Parité Ethereum est un exemple de plateforme développée sur la base de la PoA [20].

- **Practical Byzantine Fault Tolerance (PBFT)**

Tolérance aux Fautes Byzantines Pratique (PBFT) : Le PBFT aborde le Problème des Généraux Byzantins dans des environnements asynchrones. Il fonctionne sous l'hypothèse qu'au moins les deux tiers de tous les nœuds sont honnêtes. Le protocole implique plusieurs phases :

- Sélection d'un nœud principal pour servir de leader pour la création et la validation de blocs.
- Génération d'un nouveau bloc par le leader lorsqu'il reçoit une demande de l'utilisateur.
- Diffusion du bloc à d'autres nœuds participants pour vérification et audit.
- Audit des données du bloc par chaque nœud et diffusion des résultats d'audit avec un hash aux autres nœuds pour comparaison.

- Consensus entre les nœuds sur le bloc candidat, suivi de l'envoi de commentaires au leader contenant les résultats d'audit et de comparaison.
- Finalisation du bloc par le leader, en l'incorporant dans la chaîne, lorsqu'il reçoit l'accord d'au moins les deux tiers des nœuds.

II.4.1.4 Les catégories de la Blockchain

- **Permissioned vs Permissionless**

Selon la manière dont la Blockchain est restreinte pour participer à la création de nouveaux blocs et accéder au contenu des blocs, elle peut être avec ou sans permission. Dans une chaîne sans permission (permissionless), n'importe qui peut rejoindre le réseau Blockchain et participer à la création d'un nouveau bloc. En revanche, dans une chaîne avec permission (permissioned), seuls des nœuds prédéfinis et autorisés peuvent effectuer cette tâche [20].

- **Privée vs Publique**

La Blockchain peut également être catégorisée comme publique ou privée. Les Blockchains publiques sont véritablement décentralisées et sans permission. Elles permettent une participation ouverte et la possibilité pour quiconque de maintenir une copie de la chaîne. Ce type de Blockchain comprend généralement un grand nombre d'utilisateurs anonymes. À l'opposé, dans les Blockchains privées, certains utilisateurs sélectionnés, prédéfinis et de confiance sont autorisés à valider et participer à la publication de nouveaux blocs. Les autres utilisateurs publics ou autorisés du réseau sont restreints à la lecture des données des blocs. Contrairement aux chaînes publiques, les chaînes privées peuvent être partiellement décentralisées. De plus, un autre type de chaîne privée est appelé Blockchain consortium ou fédérée, qui est également une chaîne avec permission. Dans ce type de Blockchain, plusieurs organisations forment un consortium pour maintenir la Blockchain et assurer la transparence entre les participants [20].

- **On-chain vs Off-chain**

Les transactions Blockchain peuvent être on-chain ou off-chain, et elles diffèrent de plusieurs manières. Les transactions disponibles sur la Blockchain, visibles par tous les utilisateurs du réseau, sont des transactions on-chain. Ces transactions sont confirmées par un nombre suffisant de participants et impliquent l'enregistrement des détails de la transaction sur un bloc approprié, ainsi que la transmission des informations de base et essentielles à l'ensemble du réseau Blockchain [20].

II.4.2 Renforcement de la Sécurité des SG grâce à la Blockchain

Les systèmes de sécurité actuels des réseaux intelligents reposent largement sur des modèles centralisés où une autorité unique gère divers services tels que la facturation, la surveillance et les échanges d'énergie. Cependant, cette approche présente des défis dans un contexte où les réseaux évoluent vers des structures décentralisées. La transition vers des réseaux intelligents décentralisés et automatisés vise à améliorer les interactions entre les composants du réseau. Dans ce contexte, Blockchain offre une opportunité pour faciliter cette transition.

II.4.2.1 Évolution vers un Réseau Intelligent Décentralisé

Dans le modèle traditionnel centralisé, les différentes composantes des réseaux intelligents sont connectées à des entités centralisées qui surveillent, traitent les données et fournissent des signaux de contrôle. Cependant, cette centralisation présente des limites en termes de flexibilité, de capacité de traitement et de résilience face aux attaques. Ainsi, la transition vers un modèle décentralisé devient essentielle pour rendre les réseaux plus dynamiques, intelligents et résilients. Cette évolution implique également une transformation de l'infrastructure vers des topologies décentralisées et entièrement automatisées.

Le tableau ci-dessus résume les différences entre le modèle traditionnel centralisé et le futur modèle décentralisé des réseaux intelligents. Cette comparaison met en lumière les avantages d'un système décentralisé, notamment en termes de résilience, d'efficacité énergétique et de flexibilité dans la gestion des ressources distribuées.

Caractéristiques	Smart Grid	Smart Grid Décentralisé
Source d'énergie	Plus d'énergies renouvelables et intégration aux réseaux centralisés	Intégration de diverses ressources énergétiques distribuées
Technologies	Intégration de technologies avancées de détection et de contrôle	Surveillance en temps réel, ajustement automatique du contrôle et optimisation
Marché de l'énergie	Interdépendance des intermédiaires et marchés centralisés	Possibilité pour les utilisateurs de générer et partager leur propre énergie via le Peer-to-Peer (Voir Annexe 1)
Communication	Utilisation de technologies de communication avancées	Dominé par l'internet de l'énergie pour le partage de l'énergie et de l'information en continu
Contrôle	Dépendance du système régional de contrôle du système	Accès en douceur à des ressources énergétiques distribuées massives
Résilience	Vulnérabilité à un point unique de défaillance	Résilience contre un point de défaillance unique

Table II.4 : Comparaison entre le Smart Grid et le Smart Grid décentralisé

II.4.2.2 Motivations de l'Intégration de la Blockchain dans les SG

Le tableau ci-dessous résume ces objectifs communs de sécurité et explique comment la Blockchain peut les réaliser :

Objectif	Comment la Blockchain peut le réaliser
Confidentialité	Utilisation de techniques cryptographiques
Intégrité	Structure de données protégée cryptographiquement, enregistrements signés
Authentification	Enregistrements signés par les clés privées des utilisateurs
Auditabilité	Enregistrements/transactions publiques dans la Blockchain
Autorisation et contrôle d'accès	Définition par l'utilisateur via des contrats intelligents et des certificats d'attributs
Confiance	Algorithmes de consensus distribués
Disponibilité	Architecture distribuée permettant la réPLICATION de la Blockchain
Automaticité	Communication et échange de valeurs de pair à pair via des contrats intelligents
Transparence	Maintien d'un registre distribué immuable

Table II.5 : Objectifs communs de sécurité et solutions offertes par la Blockchain

II.4.3 Sécurisation des Systèmes de Comptage Intelligent par une Architecture Blockchain Multitier [21]

La solution [21] propose une architecture de sécurité pour les systèmes de comptage intelligent basée sur une architecture Blockchain multi-niveaux. Cette solution vise à garantir l'intégrité et la sécurité des données de consommation énergétique collectées par les compteurs intelligents (SM).

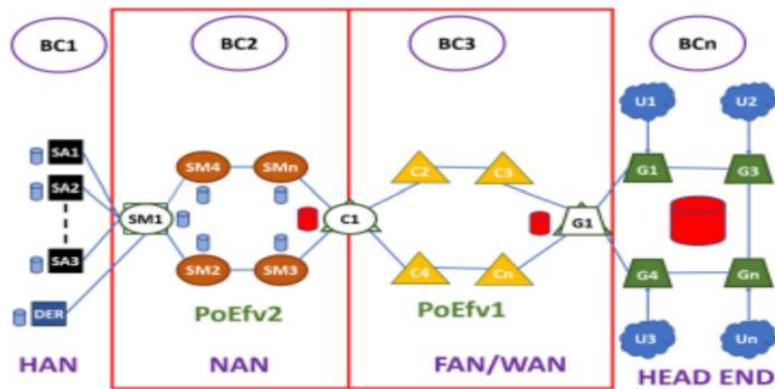


Figure II.15 : Architecture proposée basée sur quatre zones élémentaires AMI.

II.4.3.1 Architecture proposée

L'architecture illustrée sur Figure II.15 est divisée en quatre niveaux élémentaires :

- **HAN (Home Area Network)** : Les appareils électriques dans un réseau domestique mesurent la consommation d'énergie, qui est ensuite stockée dans une Blockchain locale au niveau des SM. Cette Blockchain collecte les données en temps réel et les intègre dans une base de données embarquée.
- **NAN (Neighborhood Area Network)** : Les SM rapportent les informations agrégées de consommation et de production au niveau des dispositifs du HAN. Ces informations sont résumées sous forme de transactions énergétiques, qui sont ensuite validées et stockées dans la Blockchain NAN.
- **FAN/WAN (Field Area Network/Wide Area Network)** : À ce niveau, la Blockchain est constituée de tous les centres de données (DCs) d'une région. La densité des nœuds et la distance entre eux déterminent combien de niveaux de Blockchain supplémentaires doivent être implémentés.
- **Data Center (Niveau N)** : Les serveurs de tête de l'AMI concentrent toutes les transactions des différents niveaux de Blockchain. Cette Blockchain pourrait être énorme et les données ne sont jamais effacées.

II.4.3.2 Mécanisme de Fonctionnement

- *Blockchain dans le HAN (Niveau 1)* : Tous les appareils mesurent la consommation d'énergie, et les données sont stockées dans la Blockchain du HAN.
- *Blockchain dans le NAN (Niveau 2)* : Les SM agrègent les données du HAN et les enregistrent dans la Blockchain du NAN.
- *Blockchain dans le FAN/WAN (Niveau 3)* : Les DCs regroupent les données des niveaux précédents.
- *Blockchain dans le Data Center (Niveau N)* : Les données sont centralisées et conservées dans une Blockchain traditionnelle.

II.4.3.3 Mécanisme de Consensus

L'algorithme de consensus, appelé Proof-of-Efficiency (PoEf), récompense les nœuds pour une consommation d'énergie efficace. Les transactions de production d'énergie qui ne respectent pas les critères de qualité sont rejetés.

II.4.4 Cadre de protection des données basé sur la Blockchain distribuée pour les systèmes électriques modernes contre les cyberattaques [22]

La solution [22] propose une architecture innovante utilisant la Blockchain pour protéger les données dans les systèmes de distribution d'énergie modernes, visant spécifiquement à contrer les cyberattaques. Cette architecture se distingue par une approche décentralisée qui renforce la sécurité et améliore la résilience face aux attaques.

II.4.4.1 Architecture proposée

- **Reconfiguration du Réseau**

- Les données sont collectées en temps réel et la couche de communication est isolée de l'Internet pour minimiser les risques de cyberintrusion.
- Les compteurs intelligents sont géographiquement dispersés et équipés pour collecter, émettre et traiter les données.
- Des chemins de communication spécifiques relient chaque paire de nœuds pour garantir une transmission sécurisée des données.
- Seuls les compteurs autorisés peuvent acquérir des données, transformant le réseau en une Blockchain privée.
- Les interactions entre nœuds reposent sur des mécanismes de consensus sans intervention humaine [22].

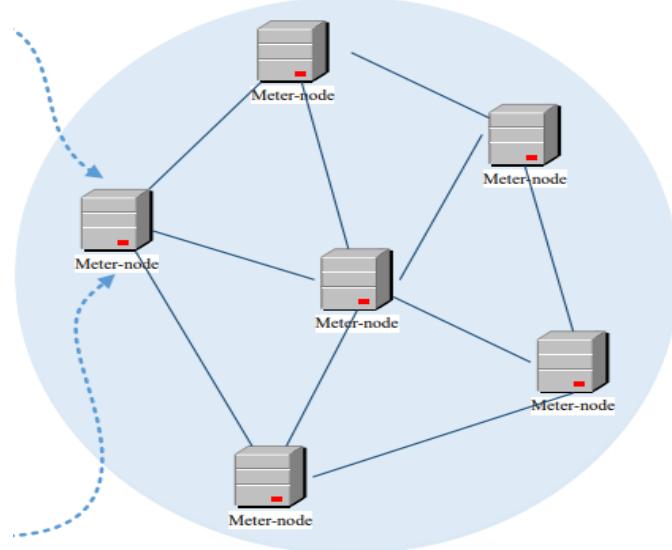


Figure II.16 : Réseau compteur-nœud

II.4.4.2 Mécanisme de Fonctionnement :

- Stockage Distribué des Données : Les données sont stockées dans un grand livre distribué sous forme de blocs connectés.

- Vérification et Validation : Les données sont vérifiées par un mécanisme de vote avant d'être accumulées et extraites.
- Intégrité des Données : Les blocs contiennent des éléments tels que le numéro de bloc, le contenu des données, l'horodatage, les résultats de hachage précédent et actuel, et une solution de nonce pour résoudre les puzzles cryptographiques.

II.4.4.3 Mécanisme de Consensus :

- Structure des Blocs : Chaque bloc contient le numéro de bloc, les données, l'horodatage, les résultats de hachage précédents et actuels, et une solution de nonce [22].

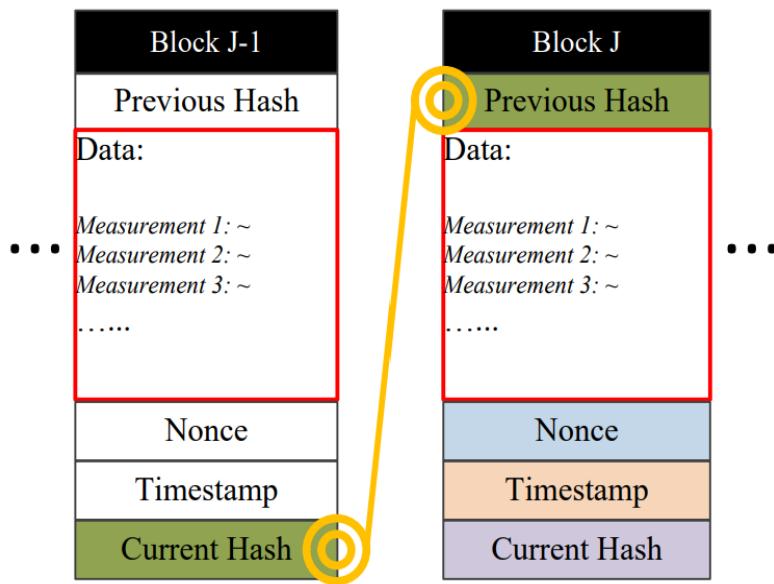


Figure II.17 : Structure des blocs

- Prétraitement et Hachage : Le processus de prétraitement combine plusieurs éléments pour générer le FinalHash via l'algorithme SHA256.
- Résolution des Puzzles : Les nœuds fonctionnent comme des mineurs pour résoudre les puzzles et valider les blocs.

- Diffusion et Validation : Une fois le nonce trouvé, il est diffusé aux autres nœuds pour validation, suivi d'un vote distribué basé sur l'adresse pour confirmer le résultat.

II.4.5 Analyse Comparative des Solutions proposées

Le tableau suivant présente une comparaison détaillée des deux solutions proposées

Critère	[21]	[22]
Type de Blockchain	Privée	Privée
Niveaux d'utilisation de la Blockchain	Quatre niveaux : HAN, NAN, FAN/WAN, Data Center	Trois niveaux principaux : collecte de données, validation et stockage
Algorithme de consensus	Proof-of-Efficiency (PoEf)	Consensus distribué adapté à chaque niveau
Nœuds de la Blockchain	HAN : appareils électriques, NAN : Smart meters, FAN/WAN : DC, Data Center: serveurs	Collecte des données : smart meters, Validation : nœuds spécialisés, Stockage : serveurs de Blockchain
Nœuds validateurs	Choisis en fonction de leurs capacités matérielles lors de la première mise en fonction	Nœuds pré-autorisés, validation distribuée
Sécurité des données	Haut niveau de sécurité et d'intégrité des données à chaque niveau	Protection robuste contre les cyberattaques et amélioration de la résilience du système
Stockage des données	Données stockées dans des Blockchains adaptées à chaque niveau	Données validées stockées dans une Blockchain distribuée

Interopérabilité	Haute interopérabilité entre différents niveaux et autres services publics	Interopérabilité avec divers composants de réseaux intelligents
Efficacité énergétique	Récompense les nœuds pour une consommation d'énergie efficace	Optimise la consommation d'énergie en fonction des performances antérieures

Table II.6 : Comparaison des solutions Blockchain trouvées.

II.5 Exploration des Approches Combinant Blockchain et Deep Learning pour la Sécurité des Smart Grids

L'association de l'intelligence artificielle et de la Blockchain a le potentiel de transformer en profondeur la technologie des Smart Grids, en particulier dans le domaine de la sécurité. La détection des attaques de vol d'énergie dans un Smart Grid représente un défi considérable. Toutefois, l'implémentation de l'IA et de la Blockchain permet le développement de nouveaux mécanismes de détection, assurant une sécurité optimale pour les utilisateurs des réseaux intelligents. Cette section se penche sur différentes solutions possibles pour concevoir des contre-mesures de détection des attaques dans les Smart Grids, en utilisant le machine learning et la technologie de la Blockchain.

II.5.1 Cadre de Protection de la Confidentialité Basé sur la Blockchain et le Deep Learning pour les Réseaux Électriques Intelligents [23]

Une solution novatrice pour améliorer la détection des attaques dans les réseaux électriques intelligents. Cette approche repose sur deux niveaux de techniques de protection de la confidentialité.

- Le premier niveau développe une technique améliorée de proof-of-work (PoW) pour authentifier les enregistrements de données et prévenir les attaques par empoisonnement. Cette méthode assure l'intégrité des données, en garantissant qu'elles ne sont ni altérées ni compromises durant leur transmission ou stockage.
- Le deuxième niveau utilise un autoencodeur variationnel (VAE) pour convertir les données originales en un format codé, réduisant ainsi les risques d'attaques par inférence. Le VAE masque les caractéristiques sensibles des données, rendant plus difficile pour les attaquants de tirer des informations exploitables.

Pour la détection d'anomalies, une technique d'apprentissage profond basée sur les réseaux de neurones LSTM (Long Short-Term Memory) est appliquée. Les LSTM sont particulièrement efficaces pour analyser les séries temporelles, telles que les données de consommation énergétique collectées au fil du temps. L'évaluation des données se fait avant et après l'application des techniques de protection de la confidentialité, assurant une détection robuste des anomalies même après la transformation des données.

II.5.2 DeepCoin : Un Nouveau Cadre d'Échange d'Énergie Basé sur le Deep Learning et la Blockchain pour les Réseaux Intelligents [1]

La solution présente un cadre innovant pour les réseaux intelligents, combinant le Deep Learning et la Blockchain. Baptisé DeepCoin, ce Framework intègre deux schémas distincts.

- Le schéma basé sur la Blockchain comprend cinq phases essentielles :
 - Phase d'installation,
 - Phase d'accord,
 - Phase de création d'un bloc,
 - Phase de consensus,
 - Phase de changement de vue.

Ces phases sont conçues pour assurer un échange d'énergie pair-à-pair fiable, reposant sur un algorithme pratique de tolérance aux pannes byzantines (PBFT), permettant d'atteindre un débit élevé et de garantir la sécurité et la fiabilité des transactions énergétiques. Pour prévenir les attaques, la solution propose la génération de blocs en utilisant des signatures courtes et des fonctions de hachage, renforçant ainsi la sécurité contre les tentatives de fraude et les intrusions.

- L'autre schéma, basé sur le Deep Learning, propose un système de détection d'intrusion (IDS) utilisant des réseaux neuronaux récurrents (RNN).

Les RNN sont appliqués pour détecter les attaques réseau et les fraudes dans le réseau énergétique basé sur la Blockchain. Cette approche permet une surveillance continue et proactive du réseau, assurant ainsi une protection renforcée contre les cybers menaces.

II.5.3 Comparaison entre les solutions proposées

Le tableau suivant met en lumière les différences clés et les points forts de chaque approche, facilitant ainsi une compréhension comparative des mécanismes

La solution	[23]	[1]
Type de Blockchain	Privé	Privé
Algorithme de consensus	Un algorithme de minage de PoW amélioré (PoW) qui n'exige pas une grande puissance de calcul	L'algorithme pratique de Byzantine tolérance aux pannes (PBFT)
Méthode de Deep Learning	Apprentissage profond LSTM (Long Short-Term Memory) appliqué sur les données codées avec VAE	Réseaux neuronaux récurrents RNN (Recurrent Neural Networks)
Datasets	Power System dataset et UNSW-NB15 qui comprend une combinaison d'enregistrements	CICID 2017 Power System dataset et Bot-IoT

	normaux et d'attaques actuels	
Méthode de collaboration entre DL et la Blockchain	Pendant l'exécution de l'algorithme de consensus, utilisation d'un auto-encodeur variationnel (VAE) sur les données provenant du power système ensuite classer ces données par un module de détection d'anomalie basé sur LSTM	Un IDS utilise des réseaux neuronaux récurrents (RNN) pour vérifier que les trames s'exécutant sur le réseau de transaction énergétique sont conformes à un ensemble de règles

Table II.7 : Comparaison entre les solutions proposées DL & Blockchain

II.6 Conclusion

Dans ce chapitre, nous avons exploré en détail les fondements cruciaux de la sécurité des Smart Grids, en mettant en avant l'analyse des séries temporelles, le Deep Learning et la Blockchain. Chacune de ces technologies offre des solutions novatrices pour renforcer la résilience et la robustesse des réseaux électriques intelligents face aux cyberattaques. L'analyse des séries temporelles, à travers des techniques comme ARIMA, se révèle essentielle pour prévoir les tendances de consommation électrique et détecter les anomalies, assurant ainsi l'intégrité des données et une gestion efficace des réseaux. De même, le Deep Learning, en particulier avec des réseaux neuronaux récurrents tels que LSTM, améliore la précision et la rapidité des systèmes de sécurité en identifiant les intrusions. La technologie Blockchain, offrant une infrastructure décentralisée et sécurisée, réduit les risques de manipulations malveillantes et renforce la confiance dans les systèmes de distribution d'énergie.

Dans le prochain chapitre, nous concevrons notre propre solution de détection des vols d'énergie, en mettant en œuvre les concepts étudiés jusqu'à présent de manière concise et efficace.

Chapitre III

Élaboration des Approches pour Déetecter les Fraudes dans les Smart Grids

III.1 Introduction

Le développement rapide des technologies de réseaux intelligents a entraîné une intégration croissante de dispositifs connectés, augmentant ainsi la vulnérabilité des systèmes énergétiques aux cyberattaques.

L'objectif principal de ce chapitre est de présenter et d'analyser les méthodes avancées pour la détection de fraude et la protection des Smart Grids, en mettant un accent particulier sur l'utilisation des technologies de la Blockchain, du Deep Learning et de l'analyse de séries temporelles (ARIMA). Nous explorerons comment ces technologies peuvent être combinées pour renforcer la sécurité et la fiabilité des réseaux intelligents, tout en examinant les travaux connexes pertinents dans ce domaine.

En particulier, nous nous concentrerons sur une méthode innovante intégrant des modèles ARIMA pour la détection de fraude, assurant l'intégrité et la traçabilité des données via la Blockchain, et optimisant la détection des comportements frauduleux grâce au Deep Learning.

Nous détaillerons les étapes d'analyse, les techniques de détection des pics de consommation, et la méthodologie utilisée pour identifier la fraude, offrant ainsi une protection complète et efficace contre les menaces modernes.

III.2 Architecture générale AMI

L'infrastructure de mesure avancée (AMI) et le système de gestion des données de compteur (MDMS) sont essentiels au fonctionnement des réseaux intelligents. L'AMI collecte et transmet les données des compteurs intelligents entre les appareils et le MDMS, facilitant ainsi la collecte, le stockage et la gestion des données. Ces systèmes permettent une tarification, une surveillance et une conservation en temps réel, et réagissent dynamiquement aux changements de conditions du réseau. La figure ci-dessous illustre l'architecture de l'AMI [24], composée de plusieurs couches interconnectées :

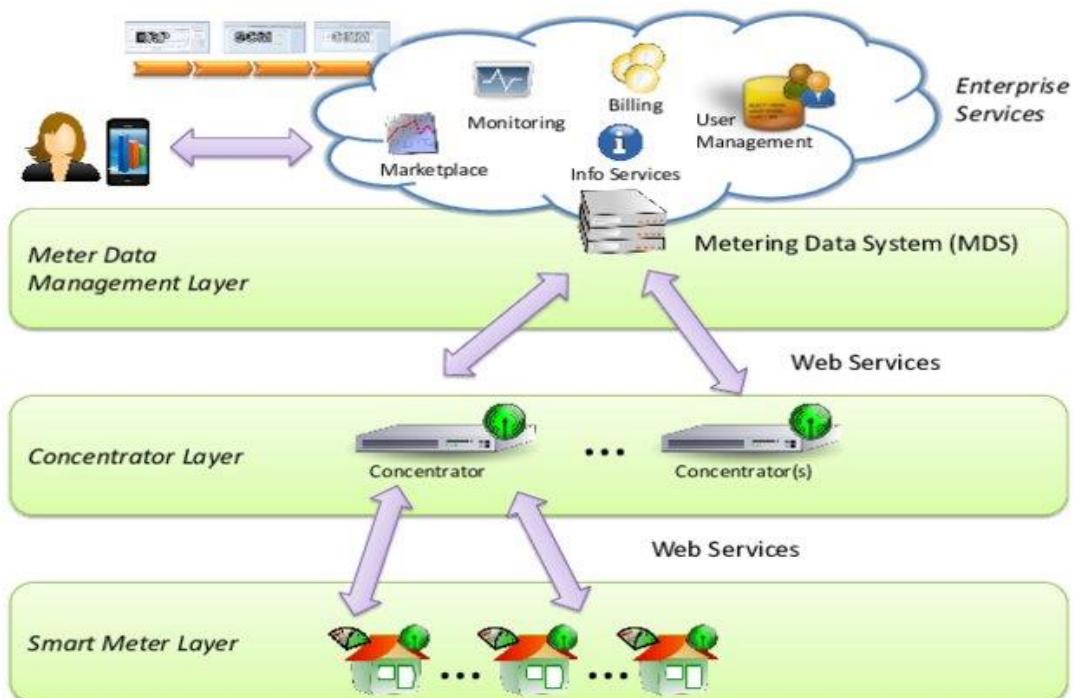


Figure III.1 : Architecture générale AMI

III.2.1 Smart Meter Layer

Située au niveau des utilisateurs finaux, cette couche inclut les compteurs intelligents installés dans les habitations. Ces compteurs enregistrent la consommation d'énergie et peuvent détecter des événements comme des dysfonctionnements ou des tentatives de manipulation physique [24].

III.2.2 Concentrator Layer

Cette couche comprend les concentrateurs qui agrègent les données des compteurs intelligents et les transmettent au système de gestion des données. Les concentrateurs servent de passerelles, garantissant une communication efficace et sécurisée entre les compteurs et les systèmes centraux [24].

III.2.3 Meter Data Management Layer

Située au sommet de l'architecture, cette couche englobe le système de gestion des données de compteur (MDMS). Le MDMS collecte, stocke et analyse les données reçues des concentrateurs. Il joue un rôle crucial dans la gestion des informations de comptage, facilitant la surveillance en temps réel, la tarification dynamique et la détection des anomalies [24].

III.2.4 Enterprise Services

Cette couche comprend divers services de l'entreprise tels que la surveillance, la facturation, la gestion des utilisateurs et les services d'information. Elle permet une interaction directe avec les utilisateurs via des portails web ou des applications mobiles, offrant des informations détaillées sur la consommation d'énergie et d'autres services.

L'Infrastructure de Mesure Avancée (AMI) révolutionne la manière dont l'électricité est mesurée, consommée et même distribuée. Les compteurs intelligents numériques rapportent à distance non seulement des données détaillées sur la consommation d'énergie, mais aussi des journaux d'événements indiquant des dysfonctionnements, des mauvaises configurations et des tentatives de manipulation physique. Ces capacités de surveillance, associées à l'agrégation de données à grande échelle d'AMI, promettent de réduire de manière significative le problème du vol d'énergie, particulièrement répandu dans les pays en développement.

Cependant, l'AMI accroît également la surface d'attaque que les entreprises de services publics doivent protéger en introduisant de nouvelles menaces cybernétiques sur des dispositifs physiquement accessibles. Les tests de pénétration ont révélé des vulnérabilités dans les compteurs intelligents pouvant mener à des fraudes énergétiques discrètes. De plus, la lecture à distance des compteurs élimine la visite mensuelle des techniciens pour enregistrer les consommations et inspecter visuellement les compteurs.

III.2.5 Relever le Défi du Vol d'Énergie dans l'AMI

L'infrastructure de mesure avancée (AMI) est particulièrement vulnérable aux actes de vol d'énergie, étant donné qu'elle repose sur des systèmes de comptage intelligents et des réseaux de communication avancés pour surveiller et gérer la consommation d'énergie. Le vol d'énergie dans l'AMI peut être abordé de plusieurs manières, notamment [9] :

- **Détection et Prévention Physiques** : Renforcer la sécurité physique des compteurs et des câbles, et utiliser des dispositifs anti-fraude pour prévenir les manipulations.
- **Surveillance des Données** : Utiliser des techniques avancées de détection d'anomalies pour identifier les schémas de consommation suspects.

- **Sécurisation de la Communication** : Mettre en place des protocoles de communication sécurisés pour protéger les données de consommation et de facturation contre les altérations.

Pour limiter la portée de notre projet et nous concentrer uniquement sur le vol d'énergie, nous allons nous focaliser sur :

- La détection de vol d'énergie résultant de la manipulation des compteurs intelligents sur le terrain (attaques internes).

Nous posons les hypothèses suivantes :

- Les données traitées et envoyées par un centre de données (DC) ne sont pas falsifiées, car les DCs dans le réseau WAN sont résistants aux manipulations et donc fiables.
- Les données sont stockées dans le centre de contrôle de manière sécurisée.

III.3 Détection de la fraude dans les réseaux intelligents utilisant ARIMA

L'approche que nous avons adoptée utilise le modèle ARIMA (AutoRegressive Integrated Moving Average) pour analyser les données de consommation électrique et identifier les pics de consommation anormaux qui pourraient indiquer une fraude.

III.3.1 Paramétrage et Identification des Pics de Consommation

Etape 1 : Collecte et Prétraitement des données

Les données relatives à la consommation électrique sont recueillies à partir de compteurs intelligents déployés chez les clients, incluant les relevés horaires. Une fois collectées, ces données sont soumises à un processus de nettoyage. Par la suite, elles sont transformées afin d'être adéquatement préparées pour une analyse selon le modèle ARIMA.

Etape 2 : Paramétrage du modèle ARIMA

Le modèle ARIMA repose sur trois paramètres essentiels, à savoir :

- p : représentant le nombre de termes autorégressifs (AR),
- d : indiquant le degré de différenciation pour rendre la série temporelle stationnaire,
- q : désignant le nombre de termes de moyenne mobile (MA).

La détermination de ces paramètres se fait via une technique d'optimisation appelée recherche par grille. Cette méthode consiste à explorer systématiquement une plage de valeurs pour p , d , et q fin de sélectionner la combinaison qui minimise un critère de performance, tel que le critère d'information d'Akaike (AIC).

Etape 3 : Détection des Pics de Consommation

Une fois le modèle ARIMA correctement ajusté, nous procédons à l'identification des pics de consommation en utilisant plusieurs méthodes de détermination de seuil :

- **Méthode 90e Percentile**

Le 90e percentile, P_{90} , est une valeur telle que 90% des données de consommation sont inférieures ou égales à cette valeur. Pour un ensemble de données triées $\{x_1, x_2, \dots, x_n\}$ le 90e percentile est situé à la position $p = 0.90 \times (n + 1)$. Si p n'est pas un entier, une interpolation linéaire entre les valeurs adjacentes est appliquée.

- **Méthode Moyenne + Écart Type**

Le seuil est fixé à la moyenne de la consommation plus deux fois l'écart type. $Seuil = \mu + 2\sigma$. Où La moyenne (μ) et l'écart type (σ) de la consommation. Les pics sont identifiés comme les valeurs excédant ce seuil.

- **Modèle de Mélange Gaussien (GMM)**

Un modèle de mélange gaussien est ajusté aux données pour capturer les différentes composantes de la distribution de consommation. Le seuil est basé sur les moyennes des composantes du GMM. En général, nous prenons la moyenne (μ) la plus élevée des composantes pour définir le seuil des pics.

- **Validation Croisée**

Les données de consommation sont binarisées (transformées en 0 et 1) en fonction de différents seuils. Un classificateur (un arbre de décision) est utilisé pour évaluer les performances de ces seuils en utilisant la validation croisée à 5 plis (cv=5).

Etape 4 : Analyse comparative des prédictions ARIMA et des données effectives de consommation.

Pour chaque client, nous procédons à une comparaison entre les valeurs réelles de consommation obtenues des compteurs intelligents (Smart Meters) et les valeurs prédites par le modèle ARIMA pour chaque pic de consommation identifié. Les mesures prises en compte dans cette analyse comparative incluent :

III.3.2 Étiquetage des Données

Les clients sont étiquetés comme frauduleux ou non-frauduleux en fonction de critères spécifiques, basés sur les analyses des pics de consommation et des différences de consommation moyenne :

- **Nombre de pics :**

Un client est considéré comme potentiellement frauduleux si le nombre de pics de consommation observés dépasse la moyenne du nombre de pics calculés dans l'ensemble des données. Mathématiquement, si N_{pics} pics représente le nombre de pics pour un client donné et \bar{N}_{pics} est la moyenne des nombres de pics pour tous les clients, alors :

$$Etiquette = \begin{cases} Frausuleux & si N_{pics} > \bar{N}_{pics} \\ Non-frauduleux & Sinon \end{cases}$$

Cette méthodologie permet d'analyser les données de consommation électrique en utilisant le modèle ARIMA et diverses techniques de seuil pour identifier les pics de consommation anormaux. En combinant l'analyse des différences entre les valeurs réelles et prédictives, ainsi que les critères spécifiques d'étiquetage, les clients présentant des comportements de consommation suspects sont étiquetés comme potentiellement frauduleux.

Cette approche systématique de classification des clients en fonction de leur comportement de consommation facilite la détection des anomalies et des fraudes potentielles dans les réseaux intelligents, permettant aux fournisseurs d'électricité de cibler plus efficacement leurs investigations et leurs actions de prévention de la fraude.

III.4 Détection de la fraude dans les réseaux intelligents utilisant Deep Learning

III.4.1 Choix du modèle DL utilisé

Lors de l'examen des performances des différentes approches pour la détection des anomalies dans les réseaux intelligents, les modèles de Deep Learning, en particulier les réseaux de neurones convolutionnels (CNN), se distinguent nettement par rapport aux autres techniques de Deep Learning. Cette supériorité des CNN est principalement due à leur capacité à extraire automatiquement des caractéristiques pertinentes à partir de données complexes, comme les séries temporelles de consommation d'énergie, et à identifier des schémas anormaux ou frauduleux. Les CNN sont en mesure de détecter à la fois des motifs locaux et globaux dans les données, ce qui est crucial pour identifier des comportements inhabituels dans les habitudes de consommation d'énergie.

En outre, l'utilisation des réseaux de neurones récurrents de type LSTM (Long Short-Term Memory), spécialisés dans le traitement des séquences de données, permet de capturer les dépendances temporelles à long terme présentes dans les séries temporelles de consommation d'énergie. Les LSTM sont particulièrement efficaces pour traiter des données séquentielles et sont capables de conserver des informations sur de longues périodes, ce qui est essentiel pour détecter des fraudes pouvant s'étendre sur une longue durée.

La figure ci-dessous illustre la répartition des différentes approches de Deep Learning appliquées aux réseaux intelligents entre 2015 et 2021. Les CNN (37.8%) et les LSTM (15.1%) y figurent parmi les modèles les plus utilisés, ce qui confirme la pertinence de notre choix pour cette étude.

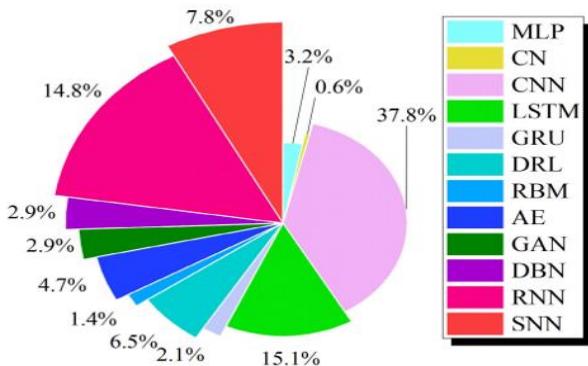


Figure III.2 : Représentation graphique des modèles de DL appliqués à la SG

III.4.2 Collecte et Prétraitement des Données

Les données de consommation énergétique sont collectées via des Smart Meters (SM) installés chez les consommateurs, puis centralisées dans des Data Concentrator (DC). Pour le prétraitement des données, plusieurs tâches sont considérées telles que le nettoyage des données (résolution des valeurs aberrantes), l'imputation des valeurs manquantes et la transformation des données.

- **Nettoyage des données**

Pour traiter les anomalies de consommation d'énergie causées par des événements spéciaux tels que les vacances, les anniversaires ou les célébrations, nous avons utilisé une méthode basée sur la règle empirique des trois écart-types. Cette méthode permet de détecter et de remplacer les valeurs aberrantes comme décrit par la formule suivante :

$$\hat{X}_{i,t} = \text{avg}(X_{i,t}) + 3\sigma(X_{i,t}), \quad F(X_{i,t}) = \begin{cases} \hat{X}_{i,t} & \text{Si } X_{i,t} > \hat{X}_{i,t} \\ X_{i,t} & \text{Sinon} \end{cases}$$

$X_{i,t}$ Représente la consommation d'électricité du consommateur i sur une période donnée.

- **Imputation des valeurs manquantes**

Les données de consommation d'électricité peuvent contenir des valeurs manquantes dues à des problèmes de stockage ou des défaillances des compteurs intelligents. Pour résoudre ce problème, nous avons utilisé Interpolation linéaire avec une limite de 2 par consécutifs. La formule suivante est utilisée pour imputer les valeurs manquantes :

$$F(X_{i,t}) = \begin{cases} \frac{X_{i,t-1} + X_{i,t+1}}{2} & \text{Si } X_{i,t} \in NaN \\ X_{i,t} & \text{Sinon} \end{cases}$$

$X_{i,t}$ Represente la consommation d'électricité du consommateur i sur une période donnée

- **Normalisation des données**

La normalisation des données est essentielle car les réseaux de neurones sont sensibles à la diversité des données. Nous avons appliqué la méthode de normalisation calculée selon la formule suivante :

$$F(X_{i,t}) = \frac{X_{i,t} - \min(X_{i,t})}{\max(X_{i,t}) - \min(X_{i,t})}$$

Où $\min(\cdot)$ et $\max(\cdot)$ représentent respectivement les valeurs minimale et maximale sur une journée.

III.4.3 Détection des Fraudes à l'aide du modèle CNN-1D

Le modèle CNN 1D commence par une couche d'entrée qui accepte les données sous forme de séries temporelles quotidiennes. Les données sont ensuite traitées par une couche de convolution 1D, permettant l'extraction des caractéristiques locales à travers les séries temporelles.

Cette couche de convolution est suivie par une couche de Flattening qui aplatis les données pour créer un vecteur de caractéristiques unidimensionnel. Ensuite, le modèle comporte plusieurs couches entièrement connectées successives, chacune activée par une fonction ReLU. Enfin, la couche de sortie, avec une fonction d'activation sigmoïde, permet de classifier les données d'entrée en prédictant la probabilité de fraude. Cette architecture est particulièrement adaptée pour capturer les motifs temporels locaux dans les données de consommation d'énergie, facilitant ainsi la détection des comportements anormaux.

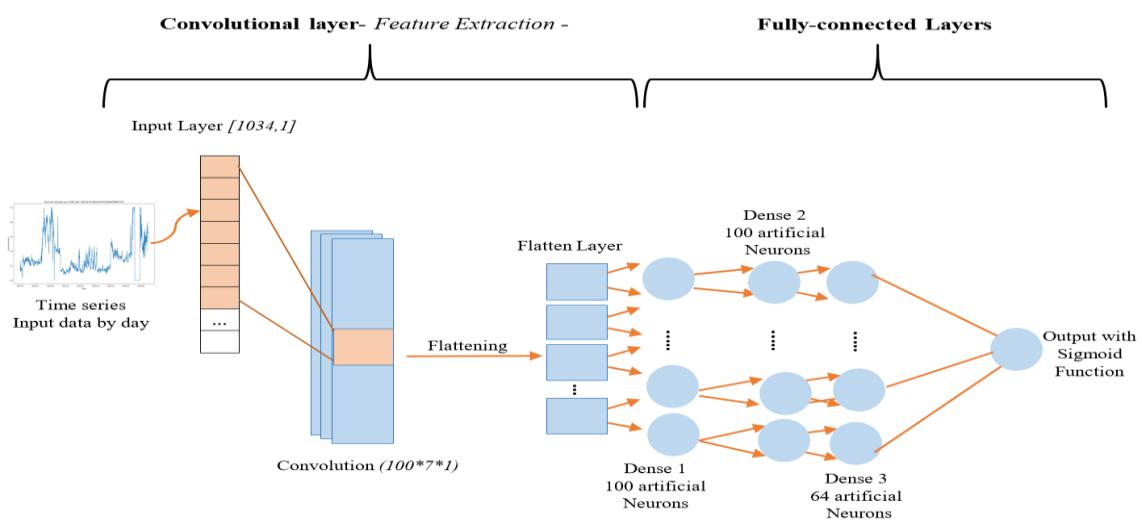


Figure III.3 : Architecture CNN 1D

III.4.4 Détection des Fraudes à l'aide du modèle CNN-2D

Le modèle CNN 2D traite les données de consommation d'énergie sous forme de séries temporelles hebdomadaires, permettant une analyse plus globale des comportements de consommation. La couche d'entrée accepte les données sous forme de matrices 2D. Les données sont ensuite traitées par une couche de convolution 2D, permettant l'extraction des caractéristiques spatiales sur les séries temporelles. Une fois les caractéristiques extraites, les données passent par une couche de flattening pour transformer la matrice en un vecteur unidimensionnel. Ce vecteur est ensuite traité par plusieurs couches entièrement connectées, chaque couche étant activée par une fonction ReLU.

La couche de sortie, avec une fonction d'activation sigmoïde, permet de classifier les données d'entrée en prédictant la probabilité de fraude. Cette architecture permet de capturer les motifs spatiaux et temporels dans les données de consommation d'énergie, offrant une vue d'ensemble des comportements de consommation et facilitant ainsi la détection des fraudes.

Prétraitement des données en séquences de sept jours

Pour adapter les séries temporelles au modèle CNN2D, nous devons ajouter une dimension supplémentaire. Voici comment cela se fait :

- Les données de série temporelle sont segmentées en séquences de 7 jours, représentées comme des images unidimensionnelles où chaque point de données correspond à une valeur sur une période de temps donnée.
- Ces images unidimensionnelles sont transformées en tenseurs pour être utilisées dans le modèle CNN2D.
 - La première dimension du tenseur : Représente le nombre de canaux. Dans ce cas, chaque canal correspond à une séquence de 7 jours.

- La deuxième dimension du tenseur : Indique la taille de chaque séquence, c'est-à-dire le nombre d'intervalles de temps dans chaque séquence.
- La troisième dimension du tenseur : Ajoutée pour assurer la compatibilité avec le modèle CNN2D.

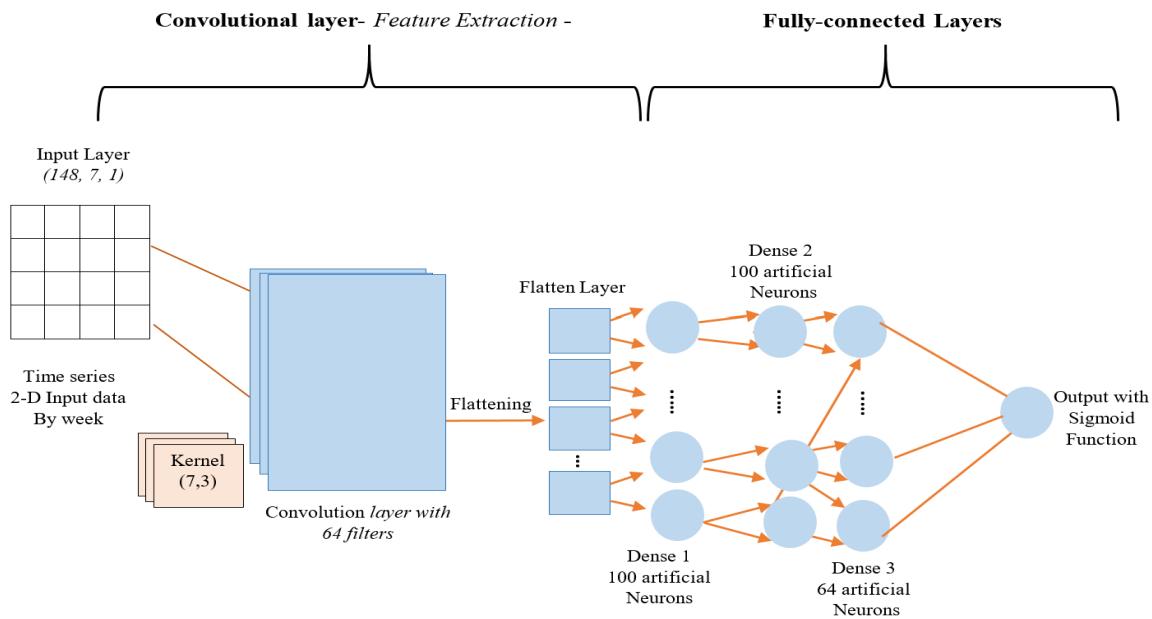


Figure III.4 : Architecture CNN -2D

III.4.5 Détection des Fraudes à l'aide du modèle CNN-LSTM

Le modèle CNN-LSTM combine des couches convolutionnelles et des couches récurrentes pour exploiter à la fois les caractéristiques locales et les dépendances temporelles des données de consommation d'énergie. La couche d'entrée accepte les séries temporelles quotidiennes, qui sont ensuite traitées par une couche LSTM, capturant les dépendances temporelles à long terme. Après la couche LSTM, une couche Dropout est appliquée pour prévenir le surapprentissage. En parallèle, les données sont également passées par une série de couches de convolution 1D, chacune suivie par une couche de batch Normalization et une activation ReLU.

Une couche de global Average pooling est ensuite appliquée pour réduire la dimensionnalité tout en conservant les caractéristiques importantes. Enfin, les sorties des couches LSTM et convolutionnelles sont concaténées et passées par une couche dense avec une activation sigmoïde pour la classification finale. Cette architecture hybride permet de capturer efficacement les motifs locaux et les dépendances temporelles, améliorant ainsi la précision de la détection des fraudes.

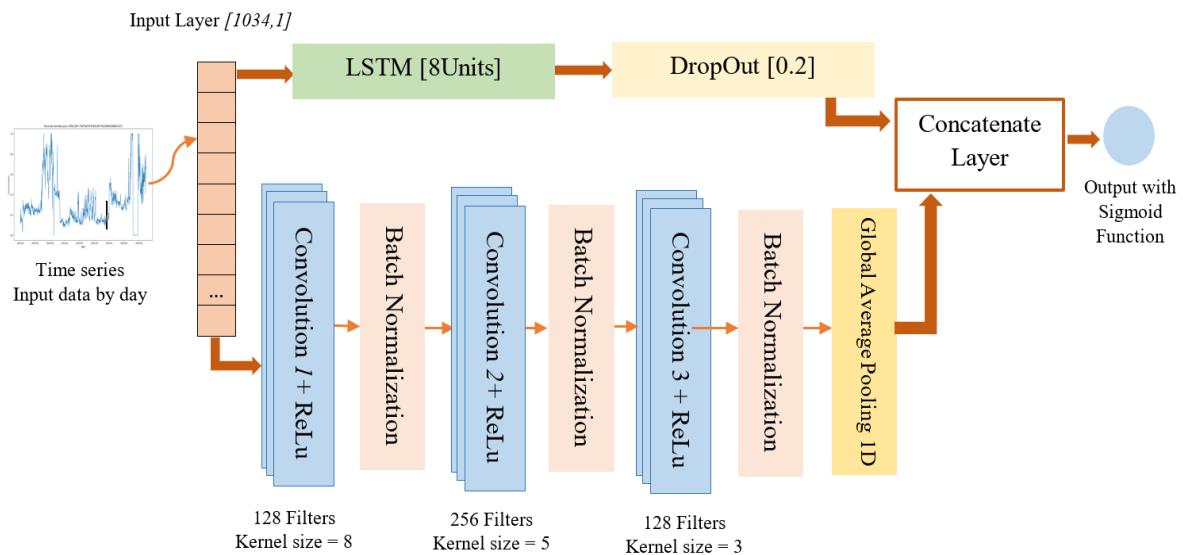


Figure III.5 : Architecture CNN-LSTM

III.5 Intégration de la Blockchain et du Deep Learning pour la Détection des Fraudes

Nous proposons une solution intégrée combinant la Blockchain et le Deep Learning pour détecter et prévenir les fraudes de manière sécurisée et efficace. Cette approche vise à renforcer la sécurité, améliorer l'efficacité opérationnelle, et garantir la transparence des transactions énergétiques, tout en permettant une détection précise et en temps réel des anomalies.

III.5.1 Vue d'ensemble de l'Architecture

L'architecture globale comprend les éléments suivants :

- Smart Meters (SM) : Dispositifs installés chez les consommateurs pour mesurer la consommation d'énergie en temps réel.
- Réseau NAN (Neighborhood Area Network) : Réseau local pour l'agrégation des données des smart meters.
- Concentrateurs de Données (DC) : Dispositifs qui agrègent, traitent et analysent les données des SM.
- Réseau WAN/FAN (Wide Area Network/Field Area Network) : Réseau qui connecte les DC au centre de contrôle.
- Centre de Contrôle : Infrastructure centrale qui supervise et gère l'ensemble du réseau.
- Blockchain : Technologie utilisée pour enregistrer les transactions de manière transparente et sécurisée.
- Smart Contracts : Automatise les transactions et vérifie les conditions.

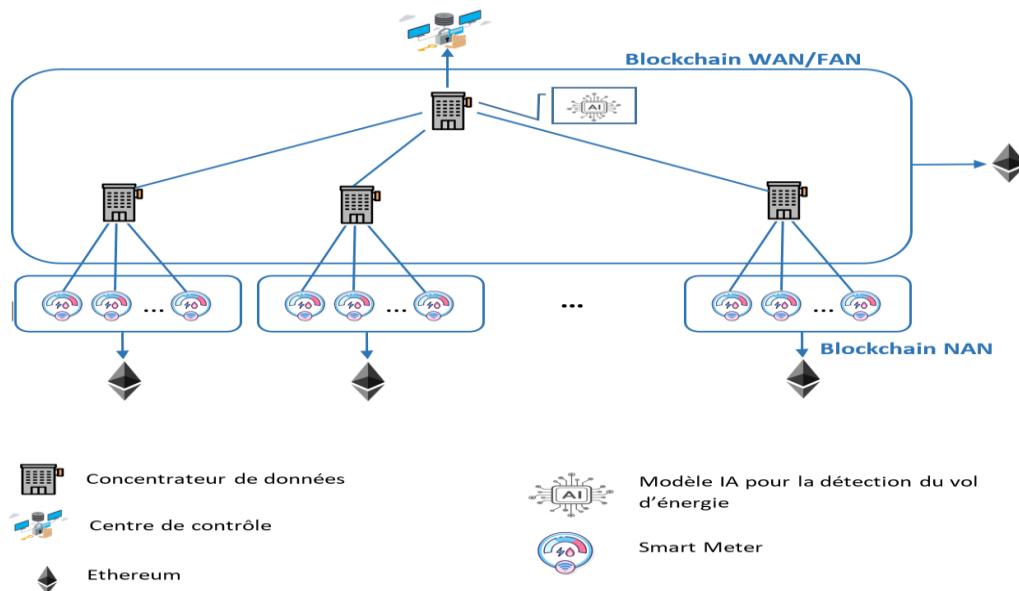


Figure III.6 : Architecture proposée pour la sécurisée de l'énergie dans les SG en utilisant la Blockchain et le DL

III.5.2 Fonctionnalités et Infrastructure des Compteurs de Réseau NAN

III.5.2.1 Caractéristiques Fonctionnelles

- **Adresse Unique** : Chaque compteur a une adresse unique pour l'identification.
- **Génération de Clés** : Logiciel spécifique pour générer une clé publique et une clé privée.
- **Mémoire Vive et Matériel de Calcul** : Pour le stockage temporaire et le traitement des données.
- **Dispositif de Collecte et Traitement des Données** : Pour enregistrer et analyser les données de consommation.
- **Émetteur et Récepteur de Signaux** : Pour la communication avec d'autres compteurs et le DC.

III.5.2.2 Communication et Agrégation des Données

- **Réseau NAN** : Contient un nombre N de smart meters distribués géographiquement. Chaque paire de compteurs peut communiquer par des canaux de communication (filaire ou sans fil).
- **Centre de Contrôle** : Contient les serveurs de l'entreprise pour stocker les données de tous les clients.
- **Chemin de Communication** : Relie chaque paire de nœuds dans chaque réseau (entre DCs dans le WAN et les serveurs dans le centre de contrôle).

III.5.3 Blockchain niveau NAN

Étape 1 : Collecte et Transmission des Données

Les smart meters (compteurs intelligents) mesurent la consommation d'énergie en temps réel et enregistrent ces données à des intervalles réguliers.

Pour garantir la confidentialité et la sécurité des données transmises, chaque smart meter chiffre les données collectées avant de les envoyer aux nœuds du réseau NAN.

- **Chiffrement des Données**

Les smart meters collectent les données de consommation énergétique :

$$M = \text{mesures_énergétiques}$$

Le smart meter chiffre les mesures collectées M en utilisant la clé publique du concentrateur de données (DC).

$$C = CH_{\text{CléPublique_DC}}(M)$$

- **Construction du Message**

Le smart meter combine le texte chiffré C avec sa clé publique pour former un message m

$$m = \text{CléPublique} \parallel C$$

- **Hachage du Message**

Le smart meter calcule le hash du message m en utilisant une fonction de hachage $H(m)$.

$$H(m) = \text{HashFunction}(m)$$

- **Signature du Message**

Le smart meter signe le hash $H(m)$ avec sa clé privée pour obtenir la signature S .

$$S = CH_{\text{CléPrivé_SM}}(H(m))$$

- **Construction du Message Final**

Le smart meter combine le message m avec la signature S pour former le message à envoyer D .

$$D = m \parallel S$$

- **Transmission des données chiffrées et de la Signature**

Une fois chiffrées et signées, les données D sont transmises aux nœuds du réseau NAN. Les données circulent sur le réseau, ce qui empêche tout accès non autorisé pendant le transfert.

Étape 2 : Validation et agrégation des données

- **Réception et Déchiffrement de la Signature**

À la réception, chaque nœud du réseau NAN déchiffre la signature S_{en} utilisant la clé publique du smart meter émetteur pour obtenir le hash H' .

$$H' = CH_{CléPublique_SM}(S)$$

- **Hachage du Message**

Le nœud hache le message m pour obtenir le hash $H(m)$.

$$H(m) = HashFunction(m)$$

- **Comparaison des Hashes**

Si $H(m) = H'$, cela signifie que la signature est vérifiée et que l'intégrité du message m est confirmée.

$$H(m) = H'$$

Étape 3 : Crédit et ajout du Bloc à la Blockchain

Les transactions validées sont agrégées en un nouveau bloc. Un Merkle Tree est utilisé pour garantir l'intégrité des données.

$$MR - D = Merkle_Root(données_validées)$$

- **Création du Bloc**

Chaque bloc contient les transactions validées, un identifiant unique, un timestamp, le hash du bloc précédent et le Merkle Root.

Champ du bloc	Signification
ID	Identifiant unique du bloc.
Timestamp	Date et heure de création du bloc.
Data	Transactions validées, incluant les mesures de consommation énergétique et les clés publiques des smart meters.
Merkle Root Data (MR-D)	Root du Merkle Tree permettant de vérifier l'intégrité des données.
Hash Précédent	Hash du bloc précédent dans la chaîne, assurant la continuité de la Blockchain.
Hash	Hash du bloc courant, calculé à partir de l'ensemble des données du bloc.

Figure III.7 : Structure de Bloc niveau NAN

Étape 4 : Ajout à la Blockchain après validation par consensus

Une fois le bloc créé, il doit être validé par un mécanisme de consensus avant d'être ajouté à la Blockchain. Les mécanismes de consensus les plus couramment utilisés sont le Proof of Stake (PoS) et le Proof of Authority (PoA).

- **Proof of Stake (PoS)**

Le Proof of Stake est un mécanisme de consensus où les validateurs sont choisis pour créer des blocs en fonction de la quantité de crypto-monnaie qu'ils possèdent et sont prêts à "staker" (mettre en jeu) comme garantie. Contrairement au Proof of Work (PoW), qui nécessite des ressources informatiques intensives pour résoudre des puzzles cryptographiques, le PoS est beaucoup plus économique en énergie.

⇒ Processus de Validation :

- Sélection des Validateurs : Les validateurs sont choisis en fonction de la quantité de crypto-monnaie qu'ils possèdent. Plus un validateur possède de crypto-monnaie, plus ses chances de valider un bloc, sont élevées

$$P(\vartheta i) = \frac{S_i}{\sum_j S_j}$$

Où :

- $P(\vartheta i)$: est la probabilité de sélection du validateur i
- S_i : Est la quantité stakée par i
- $\sum S_j$: est la somme des quantités stakées par tous les validateurs.
- Création et Validation du Bloc :
 - Le validateur sélectionné crée un nouveau bloc avec les transactions validées.
 - Le bloc est signé par le validateur pour garantir son authenticité.
 - Les autres validateurs du réseau vérifient le bloc créé et, s'il est validé par une majorité, il est ajouté à la Blockchain.
- Récompense et Punitio :
 - Le validateur qui crée le bloc reçoit une récompense sous forme de crypto-monnaie.
 - En cas de comportement malveillant (ex : double spending), le validateur peut perdre une partie ou la totalité de ses mises en jeu (slashing).

- **Proof of Stake (PoS)**

Le Proof of Authority est un mécanisme de consensus où les validateurs sont des entités pré-approuvées et réputées qui valident les blocs.

Contrairement au PoS et au PoW, le PoA repose sur l'identité et la réputation des validateurs, rendant ce mécanisme très efficace et rapide.

Processus de Validation :

- Sélection des Validateurs :
 - Les validateurs sont des entités connues et approuvées par l'ensemble du réseau. Ils sont généralement choisis en fonction de leur réputation et de leur identité vérifiable.
- Création et Validation du Bloc :
 - Un validateur pré-approuvé crée un nouveau bloc avec les transactions validées.
 - Le bloc est signé par le validateur pour garantir son authenticité.
 - Les autres validateurs du réseau vérifient le bloc créé. Une majorité simple suffit souvent pour valider le bloc.
- Recompense et PunitioN :
 - Les validateurs peuvent recevoir des frais de transaction comme récompense.
 - En cas de comportement malveillant, le validateur peut être retiré de la liste des validateurs approuvés.
- **Justification du Choix du Consensus**

Pourquoi Proof of Stake (PoS)?

Le PoS est choisi pour sa capacité à fournir un mécanisme de consensus sécurisé et énergétiquement efficace. Il évite les problèmes de consommation d'énergie excessive associés au PoW tout en offrant une incitation financière pour les validateurs à bien se comporter.

De plus, le PoS permet une participation plus démocratique au consensus, car tout utilisateur possédant une crypto-monnaie peut participer à la validation des blocs.

Pourquoi Proof of Authority (PoA)?

Le PoA est choisi pour sa rapidité et son efficacité. Dans un réseau où les validateurs sont des entités de confiance (par exemple, des entreprises d'énergie ou des institutions reconnues), le PoA offre une solution pratique pour valider rapidement les transactions sans nécessiter une consommation énergétique élevée. Il est particulièrement adapté aux environnements où la rapidité des transactions est cruciale.

III.5.4 Blockchain niveau WAN/FAN

Étape 5 : Agrégation et prétraitement des données par le DC

Un concentrateur de données fait partie de la Blockchain de son réseau FAN/WAN, ainsi que de celle de son réseau NAN. Dans le réseau NAN, il peut uniquement déchiffrer les données à l'aide de sa clé privée, ce qui lui permet ensuite d'utiliser un système de détection de vol d'énergie basé sur le Deep Learning, utilisant les mesures des différents compteurs intelligents (smart meters).

- Prétraitement et Transmission des Données**

Le prétraitement des données inclut leur nettoyage, normalisation, et la suppression des doublons et des lignes contenant des valeurs nulles. Les valeurs manquantes sont traitées par interpolation linéaire, et les valeurs aberrantes sont ajustées en fonction de la règle des trois écart-types. Après ce prétraitement, les données sont analysées par le modèle CNN2D (Figure III.4) pour détecter des anomalies et des fraudes. Les résultats doivent être transmis au centre de contrôle pour une validation et une action ultérieure. Cette transmission se fait via le réseau WAN/FAN.

Les résultats de l'analyse, incluant les anomalies détectées, sont encapsulés dans des messages structurés. Chaque message contient les informations suivantes :

- **Détails de l'Anomalie** : Description de l'anomalie (Vol d'énergie), y compris les données de consommation suspectes.
- **Horodatage** : Le moment où l'anomalie a été détectée.
- **Signature Numérique** : Une signature numérique pour garantir l'authenticité des résultats.

Les messages sont chiffrés pour assurer la confidentialité pendant la transmission. Les messages chiffrés sont envoyés via le réseau WAN/FAN au centre de contrôle.

$$M_{chiffré} = E_{clé_publique_centre_de_contrôle}(M)$$

Étape 6 : Validation et enregistrement sur la Blockchain FAN/WAN

Une fois que les résultats sont reçus par le centre de contrôle, ils doivent être validés et enregistrés sur la Blockchain pour garantir leur intégrité et assurer une traçabilité immuable

- **Déchiffrement des résultats**

Le centre de contrôle déchiffre les messages reçus en utilisant sa clé privée :

$$M = D_{clé_privée_centre_de_contrôle}(M_{chiffré})$$

- **Validation des résultats**

- Le centre de contrôle vérifie la signature numérique pour s'assurer que les résultats proviennent bien des DC et qu'ils n'ont pas été altérés.
- Comparaison des Hashes pour confirmer l'intégrité des données.

- **Enregistrement sur la Blockchain**

Les résultats validés sont encapsulés dans un nouveau bloc. Ce bloc contient les informations suivantes :

Champ du bloc	Signification
ID du bloc	Un identifiant unique pour le bloc.
Timestamp	Le moment où le bloc est créé.
Données	Les résultats validés de l'analyse
Hash du bloc précédent	Pour maintenir la continuité de la chaîne.
Merkle Root	Pour assurer l'intégrité des données au sein du bloc.

Table III.1 : Structure du bloc niveau FAN/WAN

Le bloc est ajouté à la Blockchain après validation par consensus (Proof of Stake ou Proof of Authority).

III.5.5 Réactions Automatisées via les Smart Contracts

Les smart contracts sont des programmes autonomes déployés sur la Blockchain qui exécutent automatiquement des actions basées sur les résultats des analyses. Ces actions peuvent inclure l'ajustement des factures, l'envoi d'alertes aux consommateurs et opérateurs, et l'application de pénalités pour comportements frauduleux.

- **Ajustement des Factures**

- Automatisation des Ajustements : Les smart contracts peuvent ajuster automatiquement les factures des consommateurs en fonction des anomalies détectées.
- Recalcule des Factures : En cas de détection de fraude, le smart contract peut recalculer la facture en tenant compte des pénalités et ajustements nécessaires.

- **Envoi d'Alertes**

- Détection d'Anomalies : Lorsqu'une fraude ou une anomalie est détectée, des alertes automatiques peuvent être envoyées aux consommateurs et aux opérateurs.

- Canaux de Communication : Les alertes peuvent être envoyées par email, SMS ou d'autres canaux de communication intégrés.
- **Application de Pénalités**
 - Imposition Automatique : Les smart contracts peuvent imposer automatiquement des pénalités aux consommateurs coupables de fraude.
 - Formes de Pénalités : Les pénalités peuvent inclure des frais supplémentaires, une suspension temporaire des services, ou d'autres mesures disciplinaires définies dans le contrat.

Pourquoi utiliser des Smart Contracts ?

- **Automatisation** : Les smart contracts permettent d'automatiser les transactions et les réponses aux anomalies, réduisant ainsi le besoin d'intervention humaine.
- **Transparence et Fiabilité** : Les actions exécutées par les smart contracts sont immuables et vérifiables, garantissant une transparence totale.
- **Efficacité** : Les smart contracts exécutent les actions rapidement et sans erreur, ce qui améliore l'efficacité opérationnelle du réseau.

III.6 Conclusion

Ce chapitre a exploré des méthodes avancées pour renforcer la sécurité des Smart Grids, notamment l'analyse de séries temporelles (ARIMA), le Deep Learning et la Blockchain. L'intégration de ces technologies offre une défense robuste contre le vol d'énergie et les cyberattaques. L'analyse ARIMA détecte les anomalies dans les données de consommation, facilitant une détection précoce des comportements suspects. Le Deep Learning améliore la précision de la détection des fraudes grâce à ses capacités d'apprentissage automatique. La Blockchain assure la traçabilité et l'intégrité des données avec son registre décentralisé et immuable.

Dans le chapitre suivant, nous présenterons les résultats obtenus grâce à la mise en œuvre des approches ARIMA et Deep Learning, en fournissant une analyse approfondie de leur efficacité, de leurs avantages et de leurs limitations

Chapitre IV

Résultats de la Détection des Fraudes dans les Smart Grids

IV.1 Introduction

Ce dernier chapitre se focalise sur la présentation, l'analyse et l'évaluation des résultats obtenus grâce à diverses approches déployées pour la détection des fraudes dans les Smart Grids, mettant l'accent particulièrement sur l'analyse des données temporelles avec ARIMA ainsi que sur les techniques avancées d'apprentissage profond.

IV.2 Environnement de travail

IV.2.1 Outils d'implémentations

- Scikit-learn (sklearn)



Scikit-learn est une bibliothèque de machine learning en Python qui fournit une multitude d'algorithmes pour la classification, la régression, le clustering et la réduction de dimensionnalité.

- **Python**



Python est un langage de programmation interprété, de haut niveau et polyvalent, reconnu pour sa simplicité, sa lisibilité et sa syntaxe claire. Il supporte divers paradigmes de programmation, tels que procédural, orienté objet et fonctionnel.

Python est particulièrement prisé dans de nombreux domaines, notamment le développement web, l'analyse de données, l'automatisation et, surtout, l'intelligence artificielle.

- **NumPy**



NumPy

NumPy, abréviation de Numerical Python, est une bibliothèque fondamentale pour le calcul scientifique en Python.

Elle propose des structures de données puissantes, telles que les tableaux multidimensionnels (ndarray), et des fonctions mathématiques pour effectuer des opérations rapides sur ces tableaux.

- **SciPy**



SciPy

Scientific Python, est une bibliothèque qui s'appuie sur NumPy pour offrir des outils supplémentaires dédiés aux mathématiques et à l'ingénierie. Elle comprend des modules pour l'optimisation, l'interpolation, l'algèbre linéaire, et bien plus encore.

- **Pandas**



Pandas est une bibliothèque destinée à la manipulation et à l'analyse des données.

Elle propose des structures de données comme les DataFrames pour travailler avec des données tabulaires, ainsi que des outils pour lire, écrire et transformer ces données.

- **Matplotlib**



Matplotlib est une bibliothèque de traçage qui permet de créer des visualisations en deux dimensions en Python.

- **TensorFlow**



TensorFlow est une bibliothèque de calcul numérique et de machine learning développée par Google. Elle est particulièrement utilisée pour créer et entraîner des modèles de Deep Learning.

TensorFlow permet de définir et d'exécuter des graphes de flux de données où les nœuds représentent des opérations mathématiques et les arêtes représentent des tableaux de données multidimensionnels (tensors).

IV.2.2 Configuration Matérielle

Les expériences sont implémentées en Python 3.8 sur un ordinateur portable Huawei MateBook 14s, doté d'un processeur Intel Core i7-11370H de 11ème génération fonctionnant à une fréquence de base de 3,30 GHz. Cet appareil est équipé de 16,0 Go de RAM installée. Toutefois, il convient de noter que cet ordinateur ne possède pas de carte graphique dédiée pour le traitement parallèle intensif, ce qui pourrait limiter ses performances pour certaines tâches d'apprentissage profond.

IV.3 Résultats de la détection des fraudes via le Deep Learning

IV.3.1 Analyse des Données

- **Choix du Dataset**

Pour cette étude, nous avons utilisé un jeu de données réaliste sur la consommation d'électricité publié par la State Grid Corporation of China.

Ce dataset couvre la consommation d'électricité de 42 372 clients sur une période de 1 035 jours, du 1er janvier 2014 au 31 octobre 2016. Chaque enregistrement dans le dataset représente la consommation quotidienne d'un client, identifiée par un numéro unique de consommateur (CONS_NO) [15].

La structure du dataset est organisée en série temporelle. Chaque ligne correspond à la consommation d'un client pour une journée donnée, et les colonnes représentent les différentes dates de consommation sur la période de 1 035 jours. En d'autres termes, chaque ligne contient la consommation d'électricité d'un client pour chaque jour de la période étudiée. Les principales colonnes du dataset comprennent :

- **FLAG** : Indicateur de l'état des données.
 - Un FLAG de 1 signifie que le client est frauduleux (vol d'électricité)
 - Tandis qu'un FLAG de 0 indique que le client n'est pas frauduleux.
 - **CONS_NO** : Numéro unique du consommateur.
 - **2014-01-01 00:00:00 à 2016-10-31 00:00:00** : Colonnes représentant la consommation quotidienne d'électricité pour chaque date sur la période étudiée.
- **Nettoyage des Données et Prétraitement**

Le nettoyage et le prétraitement des données sont des étapes cruciales pour garantir la qualité et l'intégrité des données avant l'analyse. Ces étapes incluent la gestion des valeurs manquantes, la normalisation des données, et la suppression des valeurs aberrantes afin d'obtenir des séries temporelles cohérentes et utilisables pour nos modèles d'analyse. (III.4.2, III.4.2)

• **Génération d'ensembles de données d'entraînement et de test**

Pour évaluer les performances de la méthodologie, l'ensemble de données prétraité est divisé en un ensemble de données d'entraînement et un ensemble de données de test à l'aide de l'algorithme de validation croisée.

- L'ensemble de données d'entraînement, composé de 31,681 consommateurs, dont 26,401 normaux et 5,280 frauduleux, est utilisé pour former les paramètres de notre modèle.
- L'ensemble de données de test est utilisé pour évaluer la capacité du modèle à se généraliser à de nouveaux échantillons de clients invisibles.

Étant donné que les consommateurs frauduleux sont nettement plus nombreux que les consommateurs non frauduleux, la nature déséquilibrée de l'ensemble de données peut avoir un impact négatif majeur sur les performances des méthodes d'apprentissage automatique supervisé. Pour réduire ce biais, l'algorithme de technique de sur-échantillonnage minoritaire synthétique (SMOTE) est utilisé pour équilibrer le nombre de consommateurs frauduleux et non frauduleux dans l'ensemble de données d'entraînement.

- **Consommation des Clients avec Fraude**

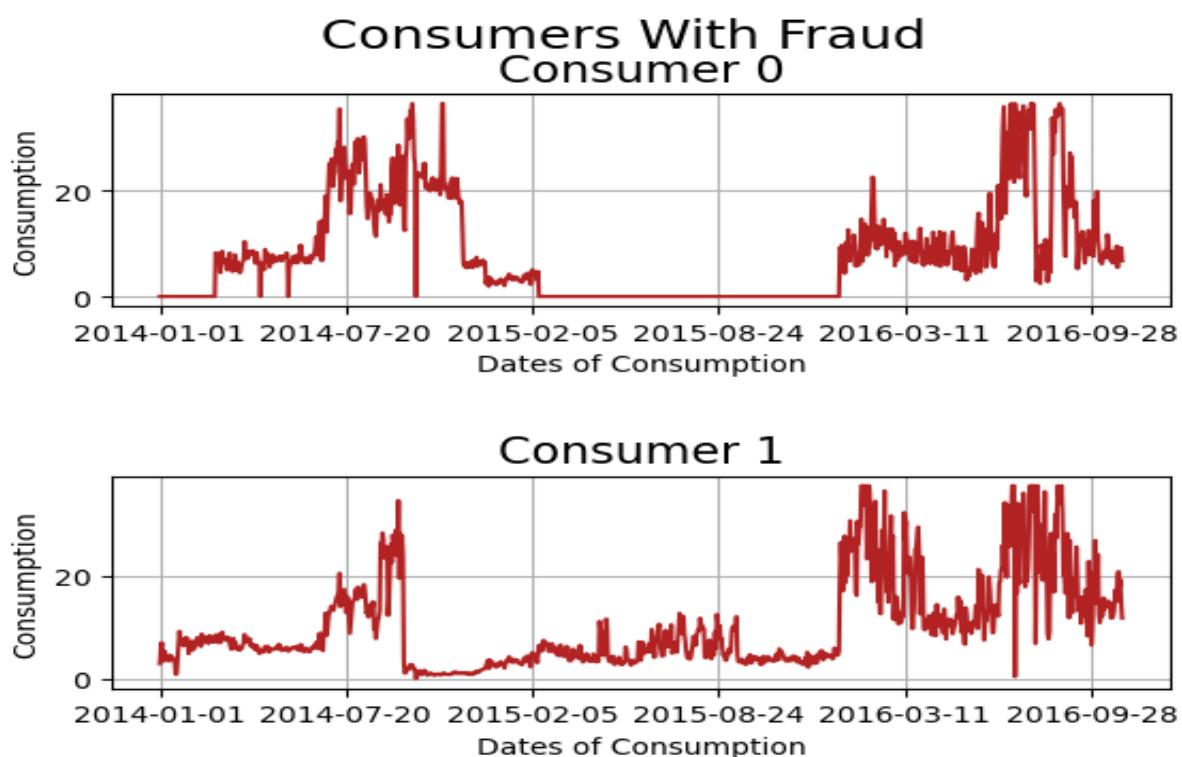


Figure IV.1 : Consommation de deux Clients frauduleux

Les figures montrent des schémas de consommation irréguliers avec des pics et des creux significatifs, ce qui indique des manipulations potentielles ou des comportements anormaux. Ces irrégularités peuvent être des indicateurs de fraude, nécessitant une analyse plus approfondie pour confirmer les soupçons.

- **Consommation des Clients sans Fraude**

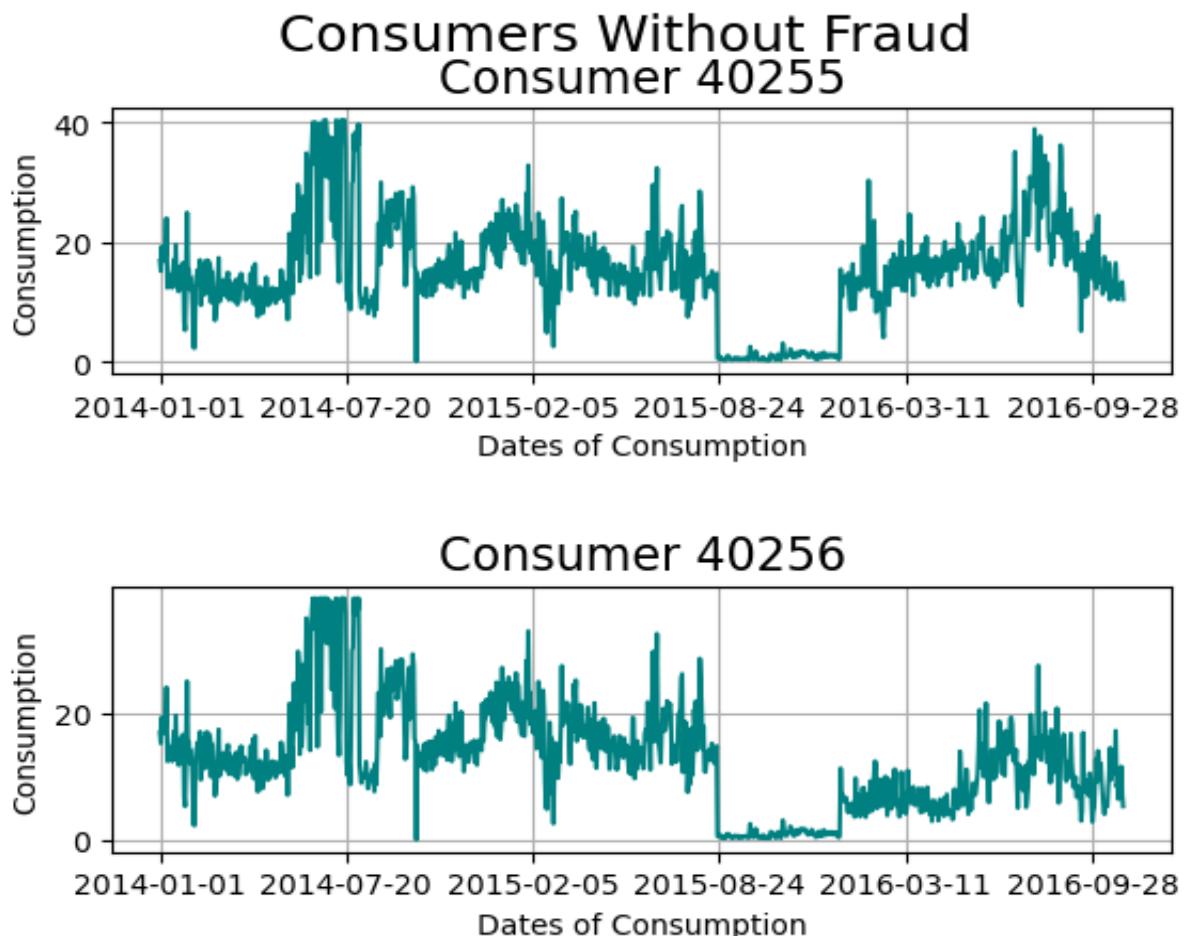


Figure IV.2 : Consommation de deux Clients non frauduleux

Les clients sans fraude montrent des schémas de consommation plus stables et réguliers, sans les fluctuations extrêmes observées chez les clients frauduleux. La stabilité des schémas de consommation est un indicateur de comportements normaux, en contraste avec les anomalies observées chez les clients suspects.

- Consommation sur Quatre Semaines

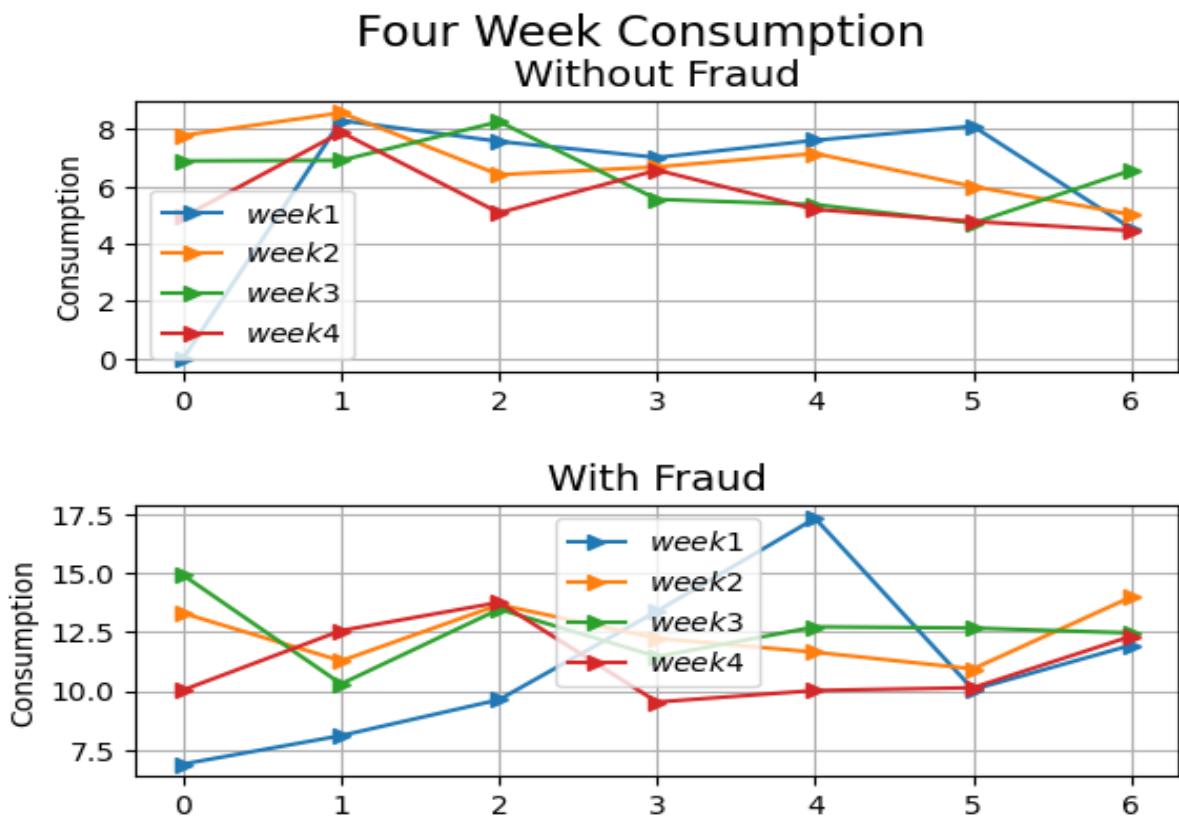


Figure IV.3 : Consommation sur Quatre Semaines

La comparaison des consommations hebdomadaires montre que les clients frauduleux ont des variations plus importantes d'une semaine à l'autre comparé aux clients non frauduleux. Les irrégularités hebdomadaires chez les clients frauduleux sont plus prononcées, suggérant des comportements de consommation anormaux.

IV.3.2 Résultats CNN-1D

IV.3.2.1 Choix des Hyper paramètres du Modèle

Optimisateur : Adam, avec un taux d'apprentissage de 0,01 L'optimiseur Adam a été choisi pour sa capacité à résoudre efficacement des problèmes d'apprentissage en profondeur en utilisant des grands modèles et ensembles de données

Fonction de Perte : Binary Cross-Entropy Utilisée pour la classification binaire, la fonction de perte binaire cross-entropy mesure la différence entre les étiquettes vraies et les prédictions du modèle.

Fonction d'Activation : ReLU et une fonction Sigmoid convertit le score en une probabilité entre 0 et 1, adaptée à la classification binaire.

Époques d'Entraînement : 150 Le modèle a été entraîné sur 150 époques, assurant une convergence stable et une performance optimale sans signes d'overfitting.

IV.3.2.2 Résultats et métriques d'évaluation

Les métriques	Accuracy	RMSE	MAE	F1	AUC	Confusion Matrix
Résultats	92.3958	0.2757	0.07604	[95.9398, 40.1737]	73.6665	[[6510, 92], [459, 185]]

Table IV.1 : Résultats CNN-1D

Le modèle atteint une précision de 92.3958% avec une RMSE de 0.2757, indiquant une faible erreur quadratique moyenne. Le score F1 montre une performance robuste dans la détection des fraudes. L'AUC de 63.6665% suggère que le modèle est raisonnablement bon pour distinguer entre les classes frauduleuses et non frauduleuses. La matrice de confusion indique une bonne performance générale avec peu de faux positifs et de faux négatifs.

Les courbes de perte et de précision montrent une convergence rapide avec une stabilisation de la perte et une précision qui reste élevée tout au long de l'entraînement. Cela indique que le modèle apprend efficacement à partir des données d'entraînement.

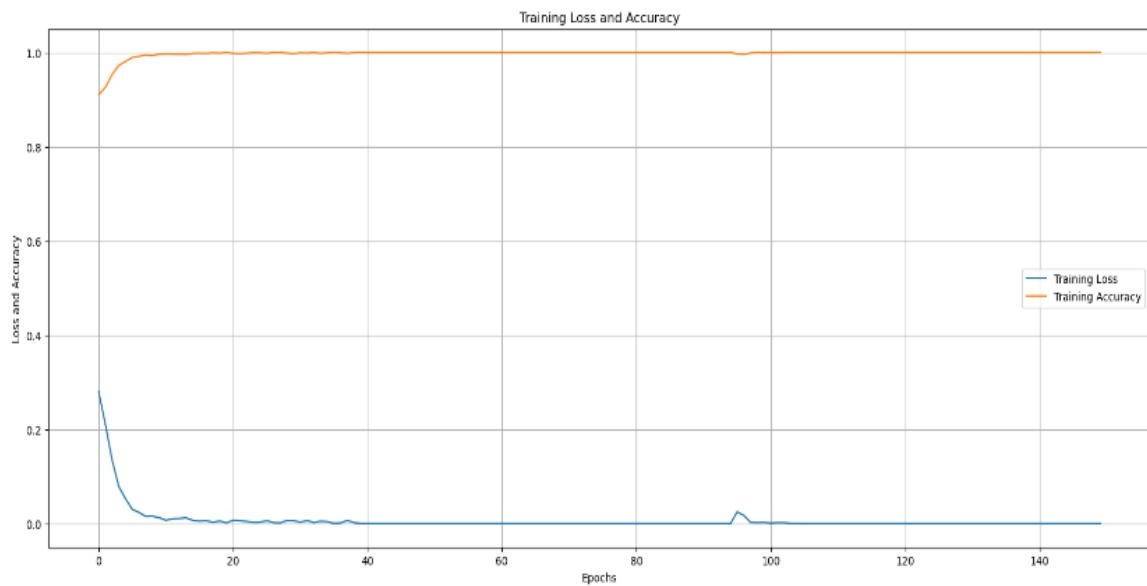


Figure IV.4 : Les courbes de perte et de précisionCNN-1D

IV.3.3 Résultats CNN-2D

IV.3.3.1 Choix des Hyper paramètres du Modèle

Optimisateur : Adam, avec un taux d'apprentissage de 0,001

Fonction de Perte : Binary Cross-Entropy

Fonction d'Activation : ReLU et La fonction Sigmoid est utilisée dans la couche de sortie.

Époques d'Entraînement : 150

IV.3.3.2 Résultats et métriques d'évaluation

Les métriques	Accuracy	RMSE	MAE	F1	AUC	Confusion Matrix
Résultats	93.1824	0.2611	0.0681	[96.3649, 45.2328]	65.4295	[[6548, 54], [440, 204]]

Table IV.2 : Résultats CNN-2D

Le modèle atteint une précision de 93.1824% avec une RMSE de 0.2611, indiquant une faible erreur quadratique moyenne.

Le score F1 montre une performance robuste dans la détection des fraudes. L'AUC de 65.4295% suggère que le modèle est raisonnablement bon pour distinguer entre les classes frauduleuses et non frauduleuses. La matrice de confusion indique une bonne performance générale avec peu de faux positifs et de faux négatifs.

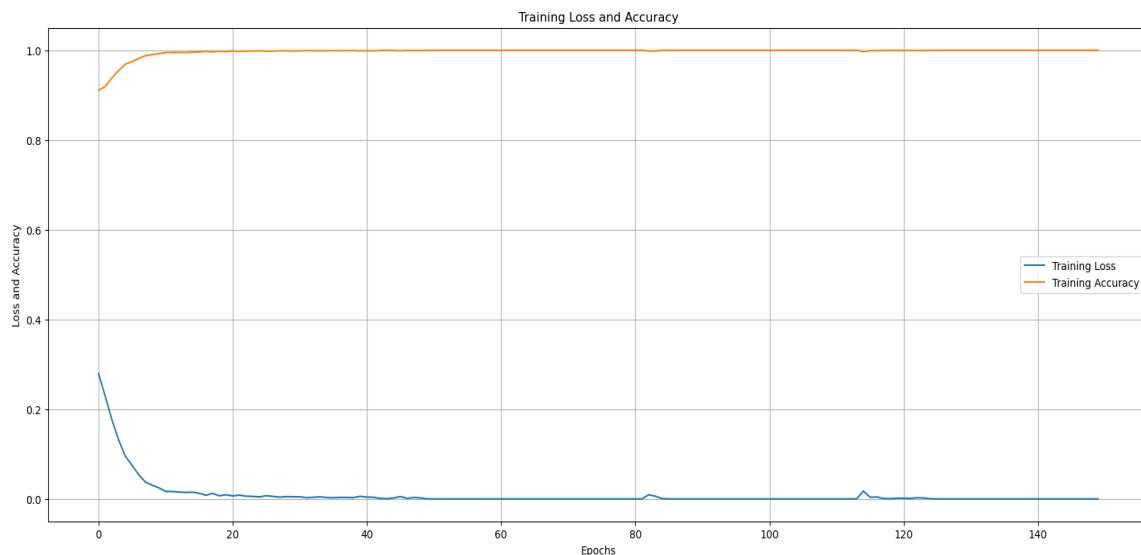


Figure IV.5 : Les courbes de perte et de précisionCNN-2D

Les courbes de perte et de précision montrent une convergence rapide avec une stabilisation de la perte et une précision qui reste élevée tout au long de l'entraînement. Cela indique que le modèle apprend efficacement à partir des données d'entraînement.

IV.3.4 Résultats CNN-LSTM

IV.3.4.1 Choix des Hyper paramètres du Modèle

Optimisateur : Adam, avec un taux d'apprentissage de 0,001

Fonction de Perte : Binary Cross-Entropy

Fonction d'Activation : ReLU et La fonction Sigmoid est utilisée dans la couche de sortie.

Époques d'Entraînement : 150

IV.3.4.2 Résultats et métriques d'évaluation

Les métriques	Accuracy	RMSE	MAE	F1	AUC	Confusion Matrix
Résultats	89.95	0.3169	0.1004	[94.4756, 44.5966]	69.8932	[[6225, 377], [351, 293]]

Table IV.3 : Résultats CNN-LSTM

Le modèle atteint une précision de 89.95% avec une RMSE de 0.3169, indiquant une faible erreur quadratique moyenne. Le score F1 montre une performance robuste dans la détection des fraudes. L'AUC de 69.8932% suggère que le modèle est raisonnablement bon pour distinguer entre les classes frauduleuses et non frauduleuses. La matrice de confusion indique une bonne performance générale avec peu de faux positifs et de faux négatifs.

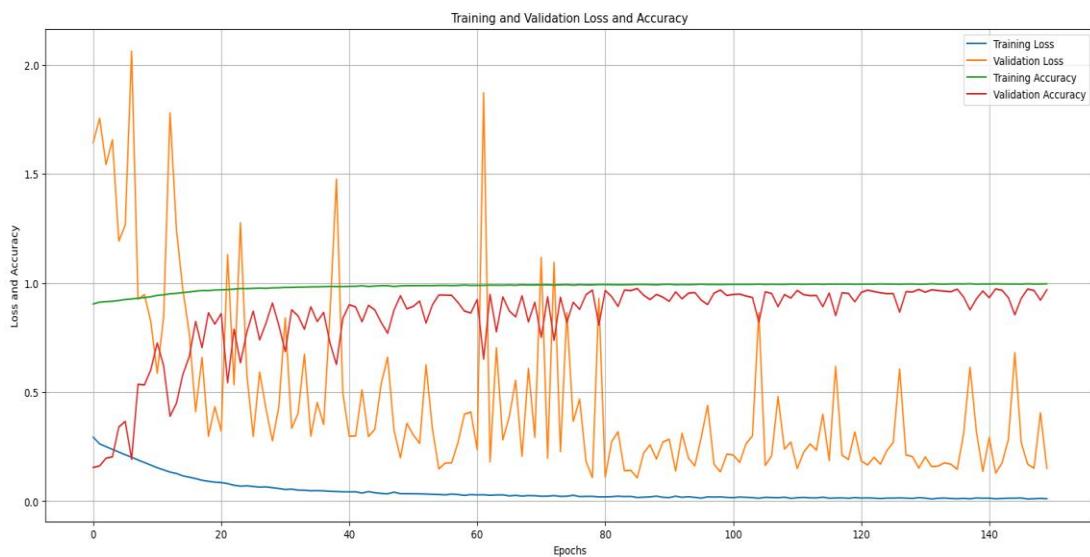


Figure IV.6 : Les courbes de perte et de précision CNN-LSTM

La perte d'entraînement et de validation démontre également une bonne convergence, mais avec des pics de volatilité. La précision de validation reste relativement stable après les premières époques, indiquant une bonne généralisation malgré la volatilité observée.

Les modèles CNN-1D, CNN-2D et CNN-LSTM ont été comparés en termes de précision, RMSE, MAE, F1-score et AUC. Le modèle CNN-2D a montré la meilleure performance globale, avec la plus haute précision, la plus faible erreur quadratique moyenne et le meilleur F1-score, indiquant une détection des fraudes plus efficace. De plus, le CNN-2D a obtenu une meilleure capacité discrétionnaire mesurée par l'AUC. En résumé, bien que les trois modèles soient performants, le CNN-2D se distingue comme le plus efficace pour la détection de fraudes dans les réseaux intelligents.

IV.4 Time Series Analysis ARIMA

Nous avons utilisé la même dataset fourni par la State Grid Corporation of China (<http://www.sgcc.com.cn/>). Nous avons sélectionné un échantillon de 100 clients identifiés comme frauduleux pour appliquer le modèle ARIMA et effectuer une étude comparative. Pour illustrer notre approche, nous avons sélectionné un client étiqueté comme frauduleux pour une analyse approfondie des séries temporelles en utilisant le modèle ARIMA. Voici les étapes détaillées de notre analyse :

- **Test de Stationnarité**

Pour confirmer l'aptitude des séries temporelles à l'application d'ARIMA, nous avons effectué un test de stationnarité. Les résultats sont les suivants :

```
ADF Statistic: -3.930470
P-value: 0.001822
Number of lags used: 14
Number of observations used: 1019
La série est probablement stationnaire
(P-value>0.05)
(Rejeter l'hypothèse nulle d'une unité racine).
```

- **Détermination des Paramètres Optimaux**

Nous avons utilisé des méthodes d'auto-corrélation et de corrélation partielle pour identifier les meilleurs paramètres pour le modèle ARIMA.

```
Best (p, d, q) parameters: (4, 0, 4)
Best AIC: -2132.683857223343
```

- **Application de l'ARIMA**

Nous avons appliqué le modèle ARIMA (4, 0, 4) aux données de consommation du client sélectionné. Les résultats obtenus sont les suivants :

```
Mean Absolute Error: 0.048948807075058956
Mean Squared Error: 0.007325312975369499
Root Mean Squared Error: 0.08558804224521963
R-squared: 0.8469356574862319
```

Les résultats du modèle ARIMA (4, 0, 4) montrent une bonne correspondance entre les valeurs réelles et prédictées, avec une forte capacité de prédiction illustrée par un R-squared de 0.8469.

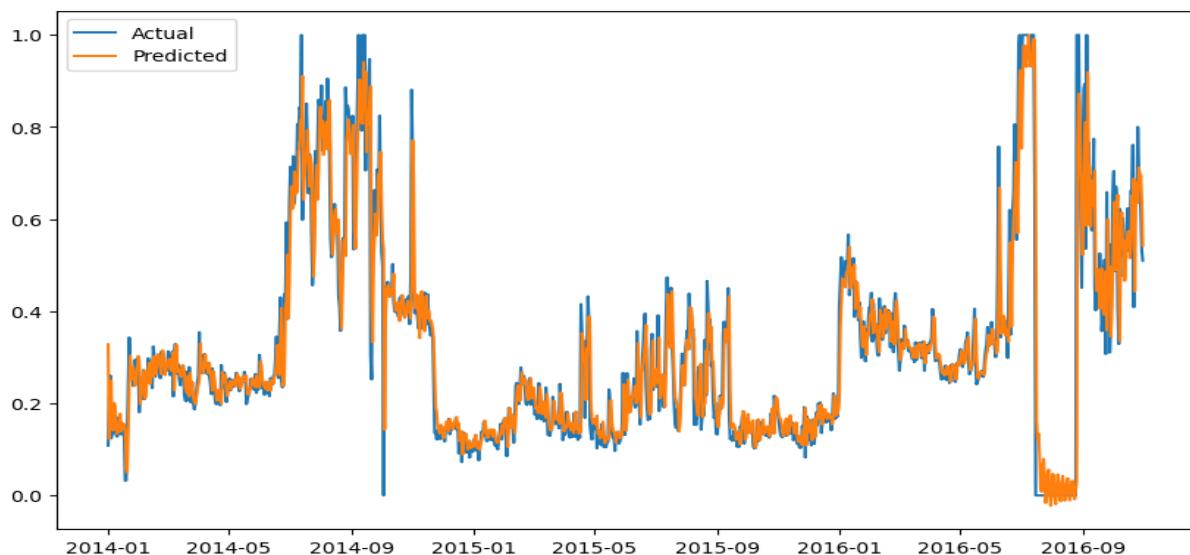


Figure IV.7 : Comparaison des Valeurs Réelles et Prédites avec ARIMA

Les étapes détaillées ci-dessus pour un client seront appliquées de manière systématique à l'ensemble des 100 clients frauduleux sélectionnés. Pour chaque client, nous procéderons comme suit :

- **Test de Stationnarité**
- **Détermination des Paramètres Optimaux**
- **Application du Modèle ARIMA**

Nous avons également effectué un test de validation sur un dataset séparé en retirant les étiquettes (FLAG) initiales. Les données de consommation ont ensuite été étiquetées en estimation, et une comparaison entre les étiquettes estimées et les données réelles a été réalisée pour évaluer la performance de notre modèle.

IV.4.1 Résultats et discussions

Dans cette section, nous présentons les résultats de notre analyse visant à détecter la fraude dans les réseaux intelligents à l'aide du modèle ARIMA. Nous avons appliqué diverses méthodes de détermination de seuil pour identifier les pics de consommation anormaux et comparer les performances de chaque méthode. Le tableau suivant présente un résumé détaillé des résultats obtenus pour chaque méthode de détection des pics de consommation, incluant les seuils calculés, les matrices de confusion et les principales métriques de performance.

Méthode	Seuil Calculé	Précision Accuracy	Rappel Recall	Vrais Négatifs (TN)	Faux Positifs (FP)	Faux Négatifs (FN)	Vrais Positifs (TP)
90e Percentile	0.622	0.49	0.42	1653	1700	187	138
Moyenne + Ecart-type	0.793	0.46	0.56	1360	1686	128	165
Mélange Gaussien	0.347	0.45	0.39	1604	1907	208	135
Validation Croisée	0.847	0.47	0.55	1414	1619	129	160

Table IV.4 : Résultats de l'approche ARIMA

L'application de différentes méthodes de seuil pour la détection de fraude énergétique a révélé des performances variées.

La méthode du 90e percentile a obtenu une précision de 49% et un rappel de 42%, mais a généré de nombreux faux positifs, indiquant une mauvaise différenciation entre clients frauduleux et non frauduleux. La méthode Moyenne + Écart-type a montré une précision légèrement inférieure de 46% mais un rappel plus élevé de 56%, indiquant une meilleure identification des cas de fraude avec toujours un nombre élevé de faux positifs. Le modèle de mélange gaussien a présenté une des performances les moins satisfaisantes avec une précision de 45% et un rappel de 39%, marquée par de nombreux faux positifs et faux négatifs. Enfin, la validation croisée a affiché une précision de 47% et un rappel de 55%, se montrant la plus équilibrée en termes de détection de fraude en réduisant le nombre de faux négatifs, bien que les faux positifs restent nombreux.

Les résultats obtenus ne sont pas très satisfaisants en raison des limitations matérielles rencontrées. Nous avons rencontré des problèmes avec le matériel utilisé, ce qui nous a empêchés de tester l'approche sur un échantillon plus large de clients. Initialement appliquée à 100 clients, nous avons envisagé d'augmenter ce nombre à 200, mais l'ordinateur a bloqué après trois jours d'exécution sans parvenir à obtenir des résultats. Ces limitations matérielles ont significativement entravé notre capacité à optimiser et valider les modèles de manière plus exhaustive.

IV.5 Conclusion

Ce chapitre a exploré l'application de diverses méthodes de Deep Learning pour détecter la fraude énergétique dans les réseaux intelligents. Nous avons évalué des modèles comme les CNN et les architectures hybrides CNN-LSTM sur des données de consommation électrique. Les résultats ont montré que chaque méthode présente des avantages et des défis uniques, notamment en termes de précision, de rappel et de capacité à réduire les faux positifs. Ces conclusions soulignent la nécessité de poursuivre la recherche pour développer des modèles plus robustes et efficaces pour la détection des anomalies dans les réseaux énergétiques.

Conclusion générale et perspectives

L'exploration approfondie des techniques avancées pour la détection de la fraude dans les réseaux intelligents a particulièrement mis en lumière l'analyse des séries temporelles et l'utilisation des modèles de Deep Learning.

L'analyse des séries temporelles, en particulier avec le modèle ARIMA, a révélé des résultats prometteurs. Il est toutefois nécessaire de continuer à optimiser les paramètres pour chaque client afin de maximiser la précision des prédictions. L'application de l'ARIMA à un échantillon de 100 clients a offert une base solide pour comprendre les schémas de consommation et identifier les anomalies potentielles. L'extension de cette analyse à l'ensemble des 40 256 clients du dataset permettra de couvrir une diversité plus grande de comportements de consommation et de renforcer la robustesse des résultats. Chaque client présente des schémas de consommation uniques, influencés par divers facteurs tels que les habitudes de vie, la taille du foyer et les équipements utilisés. En appliquant le modèle ARIMA à tous les clients, nous pourrons identifier des motifs de consommation spécifiques et des anomalies potentielles avec une précision accrue. Une optimisation itérative des paramètres du modèle ARIMA pour chaque client est cruciale pour maximiser la précision des prédictions, impliquant l'ajustement des paramètres (p, d, q) en fonction des données spécifiques de chaque client afin de capturer au mieux les tendances et les schémas de consommation individuels.

Parallèlement, les modèles de Deep Learning, incluant les CNN et LSTM, ont montré un potentiel significatif pour capturer des motifs complexes dans les données de consommation, améliorant ainsi la capacité de détection des fraudes.

Toutefois, ces modèles nécessitent des architectures plus complexes et des ensembles de données plus vastes pour réaliser pleinement leur potentiel. Entraîner des modèles de Deep Learning avec des architectures plus complexes et des ensembles de données plus vastes permettra d'améliorer la détection des fraudes.

Explorer des techniques d'apprentissage par transfert pour tirer parti des modèles pré-entraînés sur des ensembles de données similaires réduira le temps d'entraînement et améliorera la performance des modèles en utilisant des connaissances déjà acquises sur des tâches similaires. Intégrer des méthodes d'ensemble learning, telles que le boosting ou le bagging, pour combiner les prédictions de plusieurs modèles augmentera la robustesse et la précision des résultats.

La proposition théorique d'intégration de la Blockchain offre une voie prometteuse pour assurer la sécurité et l'intégrité des données dans les réseaux intelligents. Passer de la conception théorique à la mise en œuvre pratique de la Blockchain est une étape cruciale pour assurer la sécurité et l'intégrité des données de consommation dans les réseaux intelligents. Cela inclut la mise en place d'une infrastructure Blockchain capable de gérer les transactions énergétiques de manière sécurisée et transparente. Développer et tester des smart contracts permettront d'automatiser la détection et la gestion des anomalies en temps réel. Ces smart contracts pourront ajuster automatiquement les factures des consommateurs en fonction des anomalies détectées et imposer des pénalités en cas de fraude avérée. Explorer l'implémentation de mécanismes de consensus adaptés, tels que Proof of Stake (PoS) ou Proof of Authority (PoA), garantira la validation des transactions énergétiques avec un minimum de ressources, optimisant ainsi la consommation énergétique.

Pour évaluer la performance des modèles dans un environnement pratique, il est essentiel d'effectuer des tests en conditions réelles avec des données opérationnelles.

Cela permettra de vérifier la robustesse et la fiabilité des modèles en situation réelle. Collaborer avec des opérateurs de réseaux intelligents pour obtenir des retours d'expérience et affiner les modèles en fonction des retours terrain est crucial. Cette collaboration permettra d'adapter les modèles aux besoins spécifiques des opérateurs et d'améliorer leur efficacité. De plus, déployer des prototypes en milieu contrôlé pour observer le comportement des modèles et ajuster les paramètres en conséquence permettra de valider les modèles avant leur déploiement à grande échelle.

Enfin, affiner les méthodes d'étiquetage des données pour réduire les erreurs de classification est primordial. L'amélioration des techniques d'étiquetage permettra d'augmenter la précision des modèles et de réduire les faux positifs et faux négatifs. Développer des techniques de validation croisée plus sophistiquées améliorera la robustesse des modèles prédictifs. La validation croisée permettra de tester les modèles sur différentes partitions des données et d'évaluer leur performance de manière plus rigoureuse. Établir des benchmarks et des standards de comparaison pour évaluer les performances des différents modèles et méthodes de manière objective est également essentiel. La mise en place de benchmarks permettra de comparer les résultats obtenus avec différentes approches et de choisir les modèles les plus performants.

En somme, ce travail pose les bases d'une approche hybride prometteuse pour la détection de la fraude dans les réseaux intelligents. Les perspectives proposées visent à renforcer la précision et l'efficacité des méthodes développées, en tirant parti des avancées technologiques récentes et des opportunités offertes par les technologies de la Blockchain et du Deep Learning. En mettant en œuvre ces améliorations, nous pouvons espérer développer une solution plus robuste et efficace pour lutter contre la fraude énergétique, tout en assurant la sécurité et la transparence des données de consommation.

Bibliographie

- [1] M. A. Ferrag et L. Maglaras, « DeepCoin: A Novel Deep Learning and Blockchain-Based Energy Exchange Framework for Smart Grids », *IEEE Trans. Eng. Manag.*, vol. 67, n° 4, p. 1285-1297, nov. 2020, doi: 10.1109/TEM.2019.2922936.
- [2] H. El Makhtoum et Y. Bentaleb, « Réseaux Electriques Intelligents (Smart Grids) », in *Colloque sur les Objets et systèmes Connectés - COC'2021*, MARSEILLE, France: IUT d'Aix-Marseille, mars 2021. Consulté le: 21 mai 2024. [En ligne]. Disponible sur: <https://hal.science/hal-03593721>
- [3] N. Raza, M. Q. Akbar, A. A. Soofi, et S. Akbar, « Study of Smart Grid Communication Network Architectures and Technologies », *J. Comput. Commun.*, vol. 07, n° 03, p. 19-29, 2019, doi: 10.4236/jcc.2019.73003.
- [4] A. C. Santos et R. L. C. Canato, « SMART GRID: SECURITY CHALLENGES », vol. 2, n° 1, 2020.
- [5] L. Yan, Y. Chang, et S. Zhang, « A lightweight authentication and key agreement scheme for smart grid », *Int. J. Distrib. Sens. Netw.*, vol. 13, p. 155014771769417, févr. 2017, doi: 10.1177/1550147717694173.
- [6] M. Uslar, *The DISCERN tool support for knowledge sharing in large Smart Grid projects*. 2016.
- [7] S. Desai, R. Alhadad, N. Chilamkurti, et A. Mahmood, « A survey of privacy preserving schemes in IoE enabled Smart Grid Advanced Metering Infrastructure », *Clust. Comput.*, vol. 22, mars 2019, doi: 10.1007/s10586-018-2820-9.
- [8] « Attacks, Vulnerabilities and Security Requirements in Smart Metering Networks », *KSII Trans. Internet Inf. Syst.*, vol. 9, n° 4, avr. 2015, doi: 10.3837/tiis.2015.04.013.
- [9] A. Althobaiti, A. Jindal, A. K. Marnerides, et U. Roedig, « Energy Theft in Smart Grids: A Survey on Data-Driven Attack Strategies and Detection Methods », *IEEE Access*, vol. 9, p. 159291-159312, 2021, doi: 10.1109/ACCESS.2021.3131220.
- [10] M. Ahmed *et al.*, « Energy Theft Detection in Smart Grids: Taxonomy, Comparative Analysis, Challenges, and Future Research Directions », *IEEECAA J. Autom. Sin.*, vol. 9, n° 4, p. 578-600, 2022, doi: 10.1109/JAS.2022.105404.
- [11] « Maîtrise d'Économétrie Cours de Séries Temporelles Années 1999 à 2004, A. Philippe,M.-C. Viano. »
- [12] V. Badrinath Krishna, R. Iyer, et W. Sanders, *ARIMA-Based Modeling and Validation of Consumption Readings in Power Grids*. 2016, p. 210. doi: 10.1007/978-3-319-33331-1_16.
- [13] F. M. Shiri, T. Perumal, N. Mustapha, et R. Mohamed, « A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU ».

- [14] M. Massaoudi, H. Abu-Rub, S. S. Refaat, I. Chihi, et F. S. Oueslati, « Deep Learning in Smart Grid Technology: A Review of Recent Advancements and Future Prospects », *IEEE Access*, vol. 9, p. 54558-54578, 2021, doi: 10.1109/ACCESS.2021.3071269.
- [15] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, et Y. Zhou, « Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids », *IEEE Trans. Ind. Inform.*, vol. 14, n° 4, p. 1606-1615, avr. 2018, doi: 10.1109/TII.2017.2785963.
- [16] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, et Q. Zhao, « Electricity Theft Detection in Power Grids with Deep Learning and Random Forests », *J. Electr. Comput. Eng.*, vol. 2019, p. 1-12, oct. 2019, doi: 10.1155/2019/4136874.
- [17] « (2) (PDF) An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends ». Consulté le: 9 juin 2024. [En ligne]. Disponible sur: https://www.researchgate.net/publication/318131748_An_Overview_of_Blockchain_Technology_Architecture_Consensus_and_Future_Trends#fullTextFileContent
- [18] « Blockchain based privacy preserving multimedia intelligent video surveillance using secure Merkle tree | Multimedia Tools and Applications ». Consulté le: 9 juin 2024. [En ligne]. Disponible sur: <https://link.springer.com/article/10.1007/s11042-020-08776-y>
- [19] S. Zhang et J.-H. Lee, « Analysis of the main consensus protocols of blockchain », *ICT Express*, vol. 6, n° 2, p. 93-97, juin 2020, doi: 10.1016/j.icte.2019.08.001.
- [20] M. B. Mollah *et al.*, « Blockchain for Future Smart Grid: A Comprehensive Survey », *IEEE Internet Things J.*, vol. 8, n° 1, p. 18-43, janv. 2021, doi: 10.1109/JIOT.2020.2993601.
- [21] J. C. Olivares-Rojas, E. Reyes-Archundia, J. A. Gutierrez-Gneccchi, J. Cerda-Jacobo, et J. W. Gonzalez-Murueta, « A Novel Multitier Blockchain Architecture to Protect Data in Smart Metering Systems », *IEEE Trans. Eng. Manag.*, vol. 67, n° 4, p. 1271-1284, nov. 2020, doi: 10.1109/TEM.2019.2950410.
- [22] G. Liang, S. R. Weller, F. Luo, J. Zhao, et Z. Y. Dong, « Distributed Blockchain-Based Data Protection Framework for Modern Power Systems Against Cyber Attacks », *IEEE Trans. Smart Grid*, vol. 10, n° 3, p. 3162-3173, mai 2019, doi: 10.1109/TSG.2018.2819663.
- [23] M. Keshk, B. Turnbull, N. Moustafa, D. Vatsalan, et K.-K. R. Choo, « A Privacy-Preserving-Framework-Based Blockchain and Deep Learning for Protecting Smart Power Networks », *IEEE Trans. Ind. Inform.*, vol. 16, n° 8, p. 5110-5118, août 2020, doi: 10.1109/TII.2019.2957140.
- [24] S. Karnouskos, P. Silva, et D. Ilic, *Assessment of high-performance smart metering for the web service enabled smart grid era*. 2011, p. 144. doi: 10.1145/1958746.1958768.