UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR
MAHAMADOU FOFANA

DETECTION DES INTRUSIONS SUR LES RESEAUX VEHICULAIRES V2G EN UTILISANT LES RESEAUX DE NEURONES

AOUT 2025

REMERCIEMENTS

Je remercie Dieu le tout puissant de m'avoir donné la santé d'entamer et de terminer mon projet.

Je rends hommage à mon père qui n'est plus parmi nous aujourd'hui bien que mon cœur soit lourd de tristesse où que tu sois que ton âme repose en paix PAPA.

Je remercie aussi ma mère et mes frères pour leurs soutiens émotionnels avec leurs messages d'encouragement pour ne pas abandonner.

Les travaux de ce mémoire et les recherches ont pu se faire grâce à l'écoute et à la patience de l'encadrement de mon directeur de recherche le professeur Boucif Amar Bensaber. Je le remercie pour sa disponibilité.

Je suis également très reconnaissant envers mes collègues du laboratoire « LAMIA » de l'Université du Québec à Trois-Rivières qui ont collaboré à la réalisation de ce travail à travers nos différents échanges.

Enfin, je remercie tous mes frères, sœurs, proches amis, professeurs, camarades de classe et autres pour le soutien et le support moral durant ce mémoire.

ABREVIATIONS

V2G	Vehicle-to-Grid
V2H	Véhicule vers domicile
V2V	Vehicle-To-Vehicle
CICIDS2017	Canadian Institute for Cybersecurity Intrusion Detection Set 2017
DDoS	Attaque de déni de service distribué
VE	Véhicules électriques
NN :	Réseaux de neurones
CNN	Réseaux neuronaux Convolutifs
MLP	Perceptrons
NLP	Traitement du langage naturel
ML	Machine Learning
ATT	Variable Attaque
ANN	Réseaux de neurones artificiels
DL	Deep learning
SCMS	Smart Charging Management Systems
EVCC	Équipement d'alimentation des véhicules électrique
SECC	Contrôleur de communication d'équipement d'approvisionnement
SEV	Système de véhicules électrique
OCPP	Open Charge Point Protocol

LISTE DES TABLEAUX

LISTE DES FIGURES

Table des matières

RÉSUMÉ

Cette étude de recherche explore la technologie Vehicle-to-Grid (V2G) pour les véhicules électriques du point de vue de la cybersécurité. Étant donné l'importance cruciale de la cybersécurité dans les domaines des technologies de l'information (IT) et de l'informatique, cette étude examine en profondeur les risques, les vulnérabilités et les stratégies de protection au sein de l'écosystème V2G, où les véhicules électriques (VE) interagissent avec les réseaux électriques. En permettant des flux d'énergie bidirectionnels, la technologie V2G introduit de nouvelles possibilités pour la gestion de l'énergie et la stabilité du réseau ; cependant, elle expose également le système à des menaces de cybersécurité spécifiques qui peuvent compromettre à la fois les opérations du réseau et le fonctionnement des véhicules.

L'étude prend en compte divers secteurs de cyberattaques susceptibles de cibler les réseaux V2G, allant des attaques par déni de service distribué (DDoS) et des violations de données aux infiltrations de logiciels malveillants avancés. La méthodologie est élaborée avec un accent particulier sur les modèles d'apprentissage automatique capables d'identifier et de prédire ces menaces potentielles, renforçant ainsi le cadre de sécurité pour les véhicules électriques compatibles V2G. La méthodologie de recherche intègre plusieurs algorithmes d'apprentissage automatique, y compris la régression logistique, les arbres de décision, les forêts aléatoires et le gradient boosting, afin d'évaluer leur efficacité dans la classification et la prédiction des menaces de cybersécurité.

Parmi ceux-ci, l'algorithme du gradient boosting a montré un impact significatif en ce qui concerne la prédiction précise des attaques potentielles au sein des réseaux V2G, offrant des indications sur la probabilité et le type de menaces. La haute performance de cet algorithme suggère sa grande capacité prédictive, permettant ainsi des actions préventives pour atténuer les risques avant qu'ils n'affectent le système. Ces résultats soulignent l'importance des protocoles de cybersécurité robustes et du rôle de l'apprentissage automatique dans l'identification des vulnérabilités de l'infrastructure V2G. L'étude met en évidence l'importance de la surveillance continue et des améliorations algorithmiques à mesure que l'écosystème V2G évolue et se développe.

Mots clés : V2G, voiture électrique, machine learning, cybersécurité, l'intelligence artificielle

CHAPITRE 1 : INTRODUCTION

1.0 Introduction

Les voitures électriques représentent actuellement l'avenir de notre monde. La demande en électricité augmente à l'échelle mondiale dans le but de lutter contre le réchauffement climatique [1]. Dans les années à venir, plusieurs pays se tourneront vers les voitures électriques pour promouvoir le développement durable et réduire les émissions. De nombreux pays et parties prenantes estiment que les voitures électriques sont la solution pour encourager un développement durable. Cependant, le contexte de cette étude de recherche se base sur les voitures électriques V2G, ce qui signifie "Vehicle-to-Grid" en anglais. Plusieurs chercheurs dans ce domaine ont concentré leurs études sur les voitures électriques V2G afin de comprendre comment elles peuvent être vulnérables aux cyberattaques [1].

Tout d'abord, le concept du V2G est basé sur la compréhension de la charge intelligente, dont le but est de contrôler un programme de charge intelligent. Le taux de charge des batteries des véhicules électriques est ajusté en fonction du programme de charge intelligent, en augmentant ou en diminuant lorsque cela est nécessaire [2]. En résumé, il s'agit d'un flux d'énergie bidirectionnel entre le réseau et les véhicules électriques. La mise en œuvre du V2G est relativement peu coûteuse pour les fabricants de véhicules, car elle inclut un simple contrôleur pour gérer le taux de charge [3]. Le V2G peut fournir des services auxiliaires au réseau, tels que la réserve tournante et le contrôle du réseau électrique. Cela améliore la flexibilité des opérations du réseau électrique, qui est significativement renforcée grâce à ces services auxiliaires [2]. La conception du V2G pour les voitures électriques est représentée ci-dessous par la figure 1.

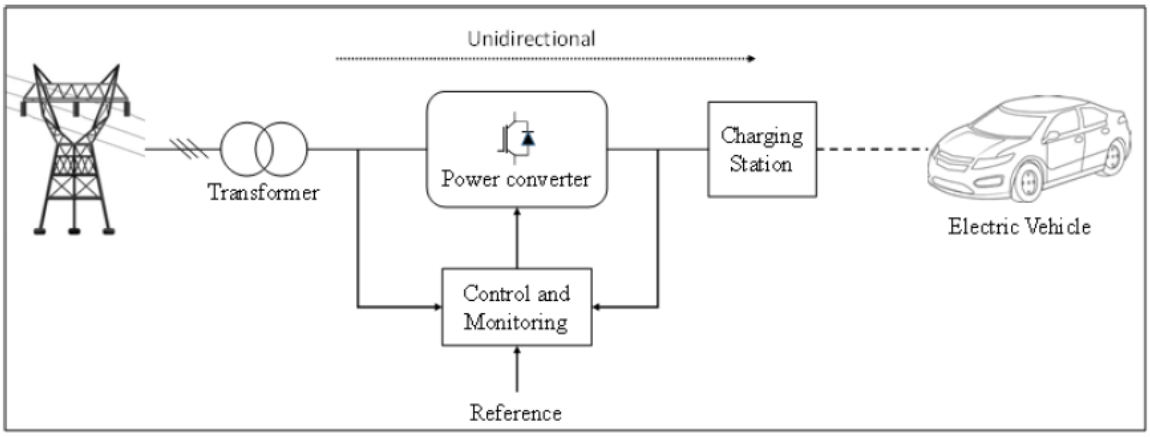


Figure 1 : Conception du V2G [2]

Cependant, pour comprendre la conception du V2G des voitures électriques, les chercheurs dans ce domaine ont conclu que le V2G facilite le renvoi d'énergie vers le réseau électrique à partir des batteries des véhicules électriques [4]. Cette technologie offre un avantage considérable pour les voitures électriques, car elle permet une intégration accrue des énergies renouvelables dans le réseau. Le V2G représente une avancée majeure, non seulement pour les véhicules électriques, mais aussi pour la gestion de l'énergie à l'échelle globale [5]. En effet, en utilisant les voitures électriques comme unités de stockage d'énergie, le V2G permet de lisser les fluctuations de production des énergies renouvelables, telles que l'énergie solaire et éolienne, qui sont par nature intermittente. Cela aide à stabiliser le réseau, à éviter les coupures de courant et à optimiser l'utilisation de l'énergie. De plus, les propriétaires de véhicules peuvent bénéficier financièrement en vendant l'énergie stockée dans leurs batteries aux heures de pointe, lorsque la demande en électricité est élevée [5].

Il est indiscutable que les voitures électriques représentent l'avenir du transport moderne, grâce aux nombreux bénéfices qu'elles apportent à la société [6]. Cependant, il est essentiel de comprendre que pour assurer le bon fonctionnement des véhicules électriques, des problèmes de sécurité peuvent surgir en lien avec l'utilisation de leurs systèmes électriques. Une violation de sécurité sur les dispositifs de communication dans l'écosystème des véhicules électriques peut avoir un impact significatif sur les secteurs de l'Énergie et des Transports. La sécurisation de cet écosystème est donc d'une importance capitale, car ces deux secteurs sont généralement considérés comme des infrastructures nationales critiques [6].

Le problème de sécurité est particulièrement crucial, car il dépend largement de la robustesse des infrastructures de sécurité des véhicules électriques. La vulnérabilité des systèmes électriques des voitures peut être exploitée par des pirates informatiques, mettant en

danger non seulement le bon fonctionnement des véhicules, mais également celui de tout le système de transport [7]. En effet, des cyberattaques visant les voitures électriques pourraient entraîner des conséquences graves telles que la prise de contrôle à distance des véhicules, des perturbations dans l'approvisionnement en électricité, ou encore la compromission des données sensibles des utilisateurs. Une attaque coordonnée pourrait même perturber l'ensemble du réseau électrique, affectant non seulement les voitures électriques, mais aussi d'autres secteurs dépendants de l'énergie [8].

Ainsi, il est impératif de développer des solutions de cybersécurité solides pour protéger les véhicules électriques et leur écosystème contre ces menaces. Cela inclut la mise en place de protocoles de communication sécurisés, l'authentification renforcée des dispositifs connectés, et la surveillance en temps réel des anomalies [9]. Ces mesures doivent être accompagnées d'une collaboration étroite entre les gouvernements, les fabricants de véhicules et les experts en sécurité pour garantir que les infrastructures critiques restent protégées face à l'évolution des menaces. Finalement, assurer la sécurité des véhicules électriques n'est pas seulement une question technique, mais aussi une question stratégique pour garantir la confiance des utilisateurs et préserver la continuité des services de transport et d'énergie, tout en soutenant la transition vers un avenir plus durable et connecté [9].

Cependant, les attaques sur les voitures électriques sont réellement liées au fonctionnement du V2G qui e un système vulnérable. Dans le cas d'une infrastructure de recharge, il est essentiel de prendre en compte les vulnérabilités associées aux systèmes et réseaux numériques. Les véhicules électriques (VE) sont davantage exposés à une cyberattaque en raison des réseaux connectés, comme les applications mobiles et les plateformes basées sur le cloud [8]. Les attaques peuvent prendre différentes formes, allant de l'accès non autorisé et des violations de données à l'insertion de logiciels malveillants ou de ransomwares. Un acteur malveillant peut avoir différentes intentions, allant de la quête de profits financiers à la prise de contrôle de fonctions essentielles, en mettant en péril les mécanismes de sécurité du véhicule ou en manipulant des informations confidentielles [8].

La structure du mémoire est organisée en 6 chapitres distincts que nous allons aborder dans cette étude de recherche. Le chapitre 1 introduit le concept des voitures électriques, du système V2G et des questions de cybersécurité. Le chapitre 2 présente le contexte, la problématique générale ainsi que les objectifs de cette étude. Le chapitre 3 présente l’état de l’art. Le chapitre 4 est consacré à la méthodologie utilisant l’apprentissage automatique

(Machine Learning) avec des analyses algorithmiques. Le chapitre 5, présente l'évaluation des modèles et la comparaison des résultats, les expériences menées pour évaluer les modèles décrits dans le chapitre 4 et compare les résultats finaux obtenus à partir des différentes approches. Enfin, le chapitre 6, présente la conclusion et synthétise les travaux réalisés avec les résultats obtenus, en identifiant les perspectives potentielles pour des développements futurs.

CHAPITRE 2 : LES VEHICULES ELECTRIQUES ET LE RESEAU V2G

2. Introduction

Cette section de l'étude vise à appréhender les véhicules électriques et les réseaux de neurones. L'apparition et la propagation des véhicules électriques (VE) ont marqué le tournant technologique du XXIe siècle, se présentant comme une solution écologique aux véhicules traditionnels à combustion interne.

Les véhicules électriques ont pour objectif de diminuer les émissions de carbone et de réduire la dépendance aux combustibles fossiles, tout en proposant des bénéfices économiques à long terme aux consommateurs. Dans le même temps, les avancées dans le domaine de l'intelligence artificielle (IA) ont incorporé les réseaux de neurones artificiels au centre des technologies de pointe. Inspirés du fonctionnement du cerveau humain, ces réseaux de neurones ont la capacité d'acquérir et de s'adapter à des ensembles de données complexes, ce qui les rend indispensables pour différentes applications, notamment dans le domaine des véhicules électriques.

2.1 Historique des véhicules électriques.

Le développement des véhicules électriques n'est pas récent. En réalité, le véhicule électrique est présent depuis plus de 100 ans et a une histoire de développement fascinante qui perdure jusqu'à nos jours (Fig.2). À la fin des années 1800, la France et l'Angleterre ont été les pionnières dans le développement du véhicule électrique. C'est seulement en 1895 que les Américains ont pris conscience des véhicules électriques. À la fin des années 1890 et au début des années 1900, de nombreuses innovations ont suivi et l'intérêt pour les véhicules à moteur a connu une augmentation significative [10].

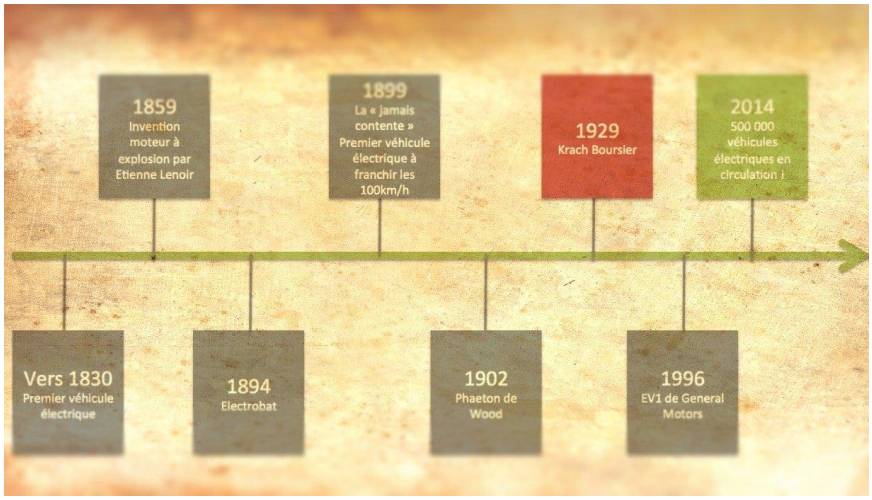


Figure 2 : Historique des voitures électriques. [10]

En 1935, les véhicules électriques avaient presque disparu. Au cours des années qui ont suivi, de 1935 à 1960, le développement, des véhicules électriques et leur utilisation dans le domaine du transport personnel ont connu une période difficile. Les années 1960 et 1970 ont été marquées par l'importance de l'utilisation de véhicules à carburant alternatif afin de diminuer les problèmes liés aux émissions d'échappement des moteurs à combustion interne et de diminuer la dépendance au pétrole brut étranger importé. Plusieurs efforts ont été déployés pour concevoir des véhicules électriques fonctionnels depuis les années 1960 jusqu'à nos jours. Les efforts de développement des véhicules électriques ont été renforcés par plusieurs mesures législatives et réglementaires [10]. L'une des principales mesures prises est la loi sur la qualité de l'air de 1990, l'amendement de 1992 à la politique énergétique et les réglementations émises par le California Air Resources Board. Plusieurs États ont émis des exigences plus rigoureuses concernant les émissions atmosphériques et les réglementations qui exigent des réductions de l'utilisation de l'essence, ainsi que des exigences concernant les véhicules à émission zéro. Par le biais du Partenariat pour une nouvelle génération de véhicules, les trois grands constructeurs automobiles (Chrysler, Ford et General Motors) ainsi que le Département de l'Énergie des États-Unis et plusieurs sociétés de conversion de véhicules, ont joué un rôle actif dans le développement des véhicules électriques [11].

2.2 Le réseau V2G

La technologie Vehicle-to-Grid, également connue sous le nom de V2G, est une technologie qui permet d'extraire de l'énergie du réseau électrique depuis la batterie d'un véhicule électrique (VE). La technologie V2G permet de décharger la batterie d'un véhicule électrique en se basant sur divers signaux, comme la production ou la consommation d'énergie

dans les environs. La technologie V2G offre une recharge bidirectionnelle, permettant non seulement de recharger la batterie du véhicule électrique, mais également d'exploiter l'énergie conservée dans le véhicule et de la retourner au réseau électrique. Même si les mots « recharge bidirectionnelle » et « V2G » sont fréquemment confondus, il y a une légère distinction entre ces derniers [12].

Les batteries pour véhicules électriques (VE) constituent sans doute le mode de stockage d'énergie le plus lucratif, car elles ne requièrent pas d'investissement additionnel en matériel. Grâce au V2G, l'utilisation de la capacité des batteries est jusqu'à 10 fois plus performante qu'avec le type traditionnel de recharge intelligente [13]. La technologie Vehicle-to-Grid nous offre la possibilité d'exploiter au mieux la flotte de véhicules déjà en place. Dans le monde, jusqu'à 250 millions de véhicules électriques pourraient arriver d'ici 2030. Cela indique que nous aurons approximativement 250 millions d'unités de stockage d'énergie sur roues. Effectivement, des études indiquent qu'à la fin de cette décennie, les batteries pour véhicules électriques devraient pouvoir satisfaire aux exigences de stockage d'énergie sur le court terme [14].

En pratique, la technologie V2G nécessite des infrastructures spéciales et des protocoles de communication pour coordonner l'échange d'énergie entre les véhicules électriques et le réseau électrique. Bien que la technologie V2G présente un grand potentiel pour améliorer l'efficacité énergétique et la durabilité du réseau électrique, elle est encore en phase de développement et fait l'objet de recherches continues pour surmonter les défis techniques, réglementaires et commerciaux [15].

2.3 Généralités sur les réseaux de neurones

Les réseaux de neurones sont également connus sous le nom de réseaux de neurones artificiels. Ils s'inspirent du fonctionnement du système nerveux humain, qui est constitué de milliards de cellules nerveuses interconnectées appelées neurones. Ils sont conçus pour imiter ces processus biologiques et sont devenus un pilier central de l’intelligence artificielle moderne. Ces réseaux informatiques complexes ont la capacité d’apprendre à partir de données, de reconnaître des modèles, de prendre des décisions et même de résoudre des problèmes que l’on pensait autrefois insolubles avec les méthodes de programmation traditionnelles [16]**.**

a. Principes des réseaux de neurones artificiels

Les réseaux de neurones artificiels ~~qui~~ sont inspirés du fonctionnement du cerveau et du système nerveux humains. Le perceptron multicouche est le type de réseau de neurones le plus populaire [17].

b. Architecture et fonctionnement des réseaux à couches

Les réseaux de neurones artificiels sont constitués de couches de neurones reliées entre elles, chacune de ces couches comportant plusieurs neurones. Chaque neurone étant une unité de calcul autonome.

Un perceptron multicouche comme un réseau de neurones artificiels est composé de trois couches ou plus. Les couches sont successives et ont des fonctions spécifiques dans le traitement de l'information.

- La couche d'entrée : C'est la première couche où les données brutes sont présentées sous forme de vecteurs pour expliquer le phénomène à analyser.
- La couche cachée : C'une couche intermédiaire entre la couche d'entrée et la couche de sortie, qui est le cœur du perceptron et elle est utilisée pour effectuer des calculs intermédiaires. Un réseau peut avoir une ou plusieurs couches cachées.
- La couche de sortie est la dernière couche d'un réseau de neurones qui donne le résultat final du calcul interne [18].

Chacune de ces couches contient des nœuds, appelés neurones qui se comportent comme des unités de calcul autonome. Les neurones d'une même couche n'ont aucune connexion entre eux [18].

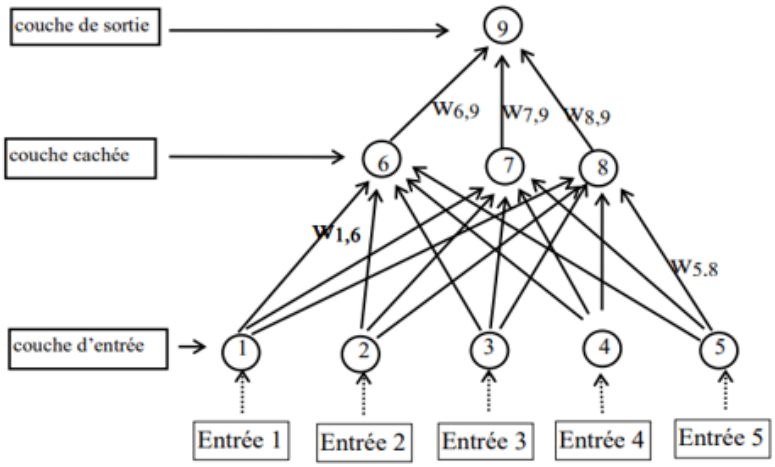


Figure 3 : Architecture d'un réseau à couches.

Les réseaux de neurones (fig.3) font partie de la grande famille des techniques d'apprentissage automatique et supportent les deux types d'apprentissages que sont l'apprentissage supervisé et l'apprentissage non supervisé :

- *L'apprentissage supervisé*

L'apprentissage supervisé utilise des données étiquetées pour entraîner l'algorithme à traiter les entrées et obtenir le résultat attendu. Des modifications peuvent être apportées pour optimiser le fonctionnement de l'algorithme. L'apprentissage supervisé peut être utilisé lorsque nous avons des données étiquetées. Lorsque les données de sortie à prédire sont des catégories ou des classes, il s'agit d'un problème de classification. Dans le cas où ces données sont des valeurs quantitatives, on parle d'un problème de régression [19].

- *L'apprentissage non supervisé :*

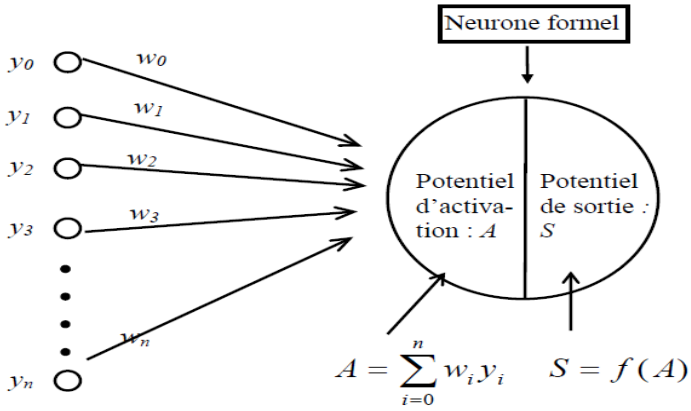
En apprentissage non supervisé, un réseau de neurones analyse des données non étiquetées en utilisant une fonction pour s'adapter et déterminer le résultat de la tâche. Cette méthode s'appuie sur la théorie de la résonance adaptative. Ce type d'apprentissage implique que des réseaux reçoivent des données non étiquetées et découvrent des modèles et des structures en leur sein pour atteindre un état final [20].

c. La fonction d'activation des neurones

Les transmissions d'activation lissent ou normalisent la donnée de sortie avant qu'elle ne soit transmise aux neurones suivants. Ces fonctions aident les réseaux de neurones à apprendre et à s'améliorer. Les neurones de chaque couche sont interconnectés aux neurones des autres couches par des liaisons appelées poids synaptiques notés Wi, j (fig.4) [20]. Chaque neurone

collecte les informations fournies par les neurones de la couche précédente avec lesquels elle est en relation, puis calcule son potentiel d'activation. Celui-ci est ensuite transformé par une fonction pour déterminer l'impulsion envoyée aux neurones de la couche suivante (potentiel de sortie), comme le montre la figure 3.

L'activation d'un neurone est donnée par la somme des potentiels de sortie de ses prédécesseurs, pondérée par les poids synaptiques. Ce potentiel d'activation est ensuite transformé par une fonction pour déterminer le potentiel de sortie [21].



y_i, désigne les sorties des prédécesseurs du neurone,
w_i, désigne les poids synaptiques.

Figure 4 : Architecture d'un réseau à couches 2. [21]

2.3.1 Réseaux neuronaux convolutifs

Pour comprendre les réseaux neuronaux convolutifs, il est essentiel de revenir aux réseaux neuronaux classiques, également connus sous le nom de réseaux de neurones de retour. Les architectures plus complexes qui ont suivi, tel que les CNN, ont été fondées sur les idées de neurones et de perceptrons [22].

En réalité, un neurone dans un réseau neuronal est une représentation mathématique basée sur la façon dont les neurones fonctionnent dans le corps. La fonction affine est la non-linéarité qu'un neurone applique à la somme pondérée de ses entrées. La représentation mathématique de cette opération est la suivante :

$$y = \sigma(\sum_{i=1}^{n} w_i x_i + b)$$

Où :

- y est la sortie du neurone
- σ est la fonction d'activation, généralement une fonction non linéaire telle que la sigmoïde, la tangente hyperbolique ReLU, etc.
- W_i sont les poids associés à chaque entrée X_i.
- b est le terme de biais

En apprenant les poids et les biais appropriés lors du processus d'entraînement, les neurones peuvent capturer des relations complexes au sein des données grâce à cette représentation mathématique. Le perceptron, introduit par Frank Rosenblatt, est un type de neurone à une seule couche qui sert de base aux réseaux neuronaux. Il reçoit plusieurs entrées binaires, les pondère, les somme et applique une fonction d'étape au résultat [23]. Le développement de perceptrons multicouches (MLP) ou de réseaux neuronaux entièrement connectés surmonte les limites des perceptrons à une seule couche dans leur capacité à apprendre des motifs complexes en introduisant plusieurs couches de neurones, permettant ainsi la création d'architectures profondes capables d'apprendre des représentations hiérarchiques de données [23].

2.3.2 Conclusion

Les réseaux de neurones artificiels, en particulier les réseaux convolutifs (CNN), ont transformé radicalement le secteur de l'intelligence artificielle en proposant des compétences inégalées pour traiter et examiner des informations complexes. Ces réseaux, en se basant sur le système nerveux humain, offrent une modélisation efficace de l'information grâce à des architectures multicouches qui captent et assimilent les attributs non linéaires des données.

2.4Contexte général et problématique de recherche

2.4.1Contexte général

L'objectif de cette étude est de proposer un modèle de détection d'intrusions dans les réseaux Vehicle-to-Grid (V2G), en particulier en s'appuyant sur les réseaux de neurones [24].

2.4.1.1 Détection des intrusions

Le système de détection d'intrusion (IDS) est une technique de cybersécurité utilisée pour détecter les intrusions et les attaques dans tout système de communication. Les véhicules connectés constituent un environnement de communication riche et critique qui nécessite une attention particulière en matière de cybersécurité. L'IDS est utilisé dans les véhicules connectés

en filtrant les données échangées entre les véhicules [25]. Parmi les exemples d'attaques qui interfèrent avec les véhicules connectés figurent les attaques par déni de service distribué (DDOS), les attaques de trou noir, les attaques Sybil et les attaques par synchronisation. L'attaque DDOS est considérée comme l'une des attaques les plus graves, car elle empêche l'utilisateur d'accéder aux services du réseau. Dans l'environnement des véhicules connectés, les attaques DOS peuvent manipuler les identités et diffuser de faux messages pour introduire un embouteillage dans le réseau ciblé [25].

La détection d'intrusion est une méthode pratiquée dans diverses industries depuis les débuts de la sécurité des réseaux, et elle est désormais appliquée aux cas d'utilisation automobile. Un IDS est couramment déployé en tant que dispositif physique dédié sur le réseau ou en tant que programme logiciel. De plus, il est utilisé comme une ligne de défense secondaire dans une architecture de sécurité réseau, souvent placé stratégiquement derrière un pare-feu pour détecter les attaques en cours [26]. Un système de détection d'intrusion (IDS) est une solution essentielle pour surveiller les réseaux et les systèmes afin de détecter des activités malveillantes ou des violations de politiques. En distinguant les menaces réelles des fausses alertes grâce à l'intégration avec des systèmes de gestion des informations et des événements de sécurité (SIEM), l'IDS améliore considérablement la sécurité informatique. Avec ses deux principales variantes :

– les systèmes de détection d'intrusion réseau (NIDS) pour le trafic réseau et les systèmes basés sur les hôtes (HIDS) pour les fichiers critiques des systèmes d'exploitation.

– l'IDS s'adapte aussi bien aux dispositifs individuels qu'aux réseaux étendus. En détectant les menaces à un stade précoce, ces systèmes jouent un rôle crucial dans la réduction des risques de violations de sécurité et dans le renforcement de la cybersécurité.

Les stations de recharge pour les véhicules électriques (VE) sont devenues populaires dans les villes intelligentes. Différents pays souhaitent adopter rapidement les stations de recharge pour les véhicules. Ces nouvelles stations de recharge utilisent l'Internet des objets (IoT) pour simplifier la vie et offrir aux opérateurs des véhicules électriques un meilleur contrôle. En tant que dispositif IoT, les véhicules électriques sont toujours connectés en ligne pour élargir les services destinés aux clients. Cela ouvre toutefois la porte à des cyberattaques sur l'écosystème des véhicules électriques. Les VE ne sont pas seulement affectées par ces attaques, mais elles impactent également l'infrastructure du réseau électrique et les clients de manière égale. Les clients, les stations de recharge pour véhicules électriques et le réseau

électrique composent l'écosystème des voitures électriques. Tous les composants de l'écosystème des voitures électriques sont vulnérables aux cyberattaques IoT [27]. Bien que l'expansion à long terme des VE nécessite un développement rapide des infrastructures, cela exige des stations de recharge fiables. Il convient de mentionner qu'un ensemble de protocoles régit la communication au sein de l'écosystème des voitures électriques. Le protocole Open Charge Point Protocol (OCPP) est un protocole de communication pour les stations de recharge pour véhicules électriques. Il permet la surveillance et la gestion à distance du processus de recharge, ainsi que la collecte des données d'utilisation [28].

2.4.2 Problématique de recherche

La cybersécurité des logiciels pour les véhicules électriques est un problème majeur, tout comme pour d'autres équipements électroniques. En 2015, deux pirates ont piraté à distance un Jeep Cherokee, et cette année, un groupe de pirates a piraté une Tesla Model S, suscitant des préoccupations quant à une attaque massive qui pourrait paralyser toute une ville. C'est pourquoi plusieurs constructeurs automobiles et prestataires de services ont fait appel à des experts, rémunérant des pirates informatiques pour trouver des failles dans leurs logiciels [29].

Il existe également une liste des meilleures stratégies de cybersécurité automobile. Il y a des problèmes liés à la sécurité des données et à la confidentialité dans ce domaine. À mesure que les autorités prennent conscience de ces problèmes, des règles provisoires commencent à être élaborées. Les systèmes autonomes nécessitent des tests approfondis en raison de la complexité des systèmes et du fait que chaque choix effectué par le logiciel affecte directement la vie humaine. La norme ISO 26262 établit un cadre pour les systèmes de guidage des véhicules qui tient compte de la sécurité fonctionnelle [29]. Le modèle V est utilisé dans l'industrie automobile depuis longtemps et est la norme ISO 26262 à l'échelle mondiale.

Tandis que la popularité des véhicules électriques augmente, les pirates informatiques persistent, cherchant des moyens d'exploiter le réseau de connexions numérique en expansion dont les véhicules électriques sont dépendants [30]. À mesure que ces véhicules se connectent davantage et dépendent des technologies numériques, ils deviennent de plus en plus vulnérables aux activités malveillantes, ce qui met en évidence l'importance cruciale de la vigilance en matière de cybersécurité, car des vies et la sécurité sont en jeu [31].

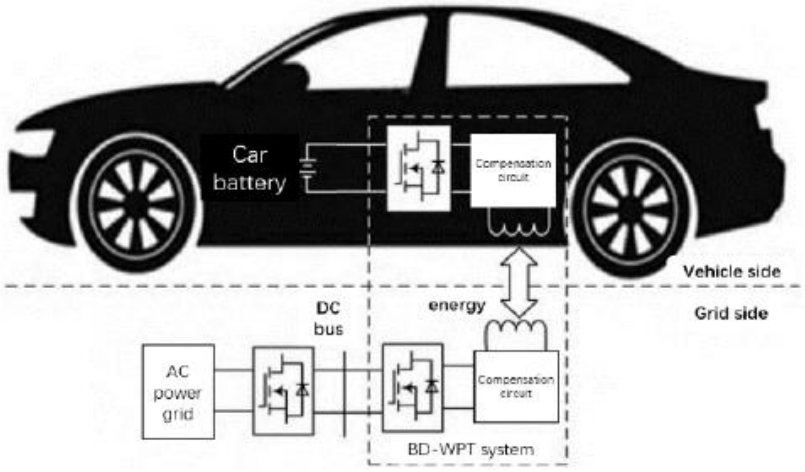


Figure 5 : Infrastructure de recharge des véhicules électriques. [31]

Source : Département de l'Énergie des États-Unis.

Les défis en matière de cybersécurité liés à l'infrastructure de recharge sont illustrés par la figure 5, car il existe de nombreuses occasions pour un acteur malveillant (menace) de manipuler un véhicule électrique en utilisant le réseau et les connecteurs. Il est légitime de dire que l'infrastructure de recharge des véhicules électriques souligne l'importance de protéger les systèmes électroniques des voitures, les réseaux de communication, les algorithmes de contrôle, les logiciels, les utilisateurs et les données sous-jacentes contre les attaques malveillantes, les dommages, l'accès non autorisé ou la manipulation [31].

Lorsqu'on examine l'infrastructure de recharge et la nature des communications, on peut se demander si des protocoles sécurisés sont utilisés. Par conséquent, il est essentiel de noter qu'il existe plusieurs façons de concevoir les points de recharge des VE pour minimiser les défis en matière de sécurité [31].

Par exemple, en utilisant des protocoles de connectivité simples tels que les modems cellulaires intégrés, ou en utilisant des passerelles industrielles dédiées desservant plusieurs points de recharge, des mesures peuvent être prises pour garantir que l'infrastructure de recharge soit plus sécurisée contre les cyberattaques. De plus, regrouper un grand nombre de points de recharge pour les acheminer via un seul emplacement augmente le risque d'un point de défaillance unique [30].

Il est important de comprendre que les problèmes majeurs sont les attaques, Il existe différents types d'attaques potentielles sur les stations de recharge pour véhicules électriques qui pourraient compromettre la sécurité et l'intégrité du processus de recharge, ainsi que la sécurité et la confidentialité des utilisateurs [32].

– Une attaque physique, dans laquelle un individu pourrait manipuler ou causer des dommages au matériel de la station de recharge, ce qui pourrait provoquer un incendie ou d'autres dangers pour la sécurité.

– Une autre forme d'attaque dans le réseau dans laquelle un individu pourrait obtenir accès non autorisé au réseau des véhicules électriques afin de manipuler ou perturber le processus de charge. Cela peut comprendre des actions telles que les attaques par déni de service (DoS) où l'attaquant surcharge le réseau de trafic pour empêcher les utilisateurs légitimes d'accéder à la station de recharge.

– Les attaques de type homme du milieu pour intercepter et modifier la communication entre la borne de recharge et le véhicule électrique. Cibler les systèmes de paiement dans les véhicules pour obtenir des informations sur les utilisateurs et des données financières, manipuler le système de facturation pour facturer aux utilisateurs plus que ce qu'ils ont utilisé [32].

D'autres types d'attaques concernent le protocole de communication utilisé entre deux véhicules électriques ou la communication Vehicle-to-Grid (V2G), où un attaquant pourrait intercepter ou manipuler la communication entre le VE et l'EVSE. Enfin, les véhicules électriques peuvent être un point d'entrée pour les attaquants souhaitant cibler l'infrastructure globale du réseau électrique [33].

Il est crucial de prendre en compte les vulnérabilités associées aux systèmes numériques et aux réseaux lorsqu'on se concentre sur la surface d'attaque d'une infrastructure de recharge. Les réseaux interconnectés, comme les applications mobiles et les plateformes en ligne, accroissent la vulnérabilité d'un véhicule électrique face à une attaque cybernétique [34]. Il existe différentes formes d'attaques potentielles, allant de l'accès non autorisé et des violations de données à l'insertion de logiciels malveillants ou de rançongiciels. Un acteur malveillant peut avoir différentes intentions, allant des motivations financières à la prise de contrôle de fonctions essentielles, compromettant ainsi les mécanismes de sécurité du véhicule ou manipulant des informations confidentielles [34].

2.5 Objectifs spécifiques de la recherche

L'objectif principal de ce mémoire est de concevoir un modèle de prédiction algorithmique de l'intelligence artificielle pour les intrusions dans le réseau (Vehicle to Grid) V2G afin de renforcer la sécurité des voitures électriques en anticipant les éventuelles menaces. Ce système de détection devra être en mesure de détecter les activités anormales et suspectes dans le réseau V2G en utilisant des algorithmes de détection d'attaques dans le réseau V2G. Par la suite, proposer les recommandations nécessaires pour assurer la sécurité des voitures électriques contre les attaques de cybersécurité. Nous présentons ci-dessous les objectifs détaillés de ce travail de recherche : :

- Assurer la sécurité des flux de communication entre les différentes entités afin d'assurer une intégration sécurisée dans le réseau.
- Proposer une méthode de détection d'anomalies basée avec les méthodes (Arbre de décision, Gradient Boosting, Régression Logistique, Forêt Aléatoire) afin de repérer les intrusions dans les réseaux V2G
- Identifier les facteurs de cyberattaques et proposer quelques recommandations.

2.6 Conclusion

Étant donné que le réseau véhiculaire V2G soit un dispositif, diverses ressources doivent participer à son fonctionnement. Chacune d'elles a pour objectif spécifique de recharger un véhicule électrique en minimisant les dangers potentiels. Cependant, son intégration soulève des défis importants en matière de cybersécurité à cause de la vulnérabilité croissante des véhicules électriques aux attaques malveillantes via les réseaux électriques.

Pour assurer une insertion sécurisée dans le réseau, il est crucial d'assurer la protection des flux de communication entre les entités engagées dans l'infrastructure de recharge.

Dans cette optique, nous avons proposé l'utilisation d'un moteur de détection d'anomalies basé sur un apprentissage automatisé pour repérer les intrusions potentielles dans les réseaux V2G. Cette approche vise à renforcer la sécurité des systèmes de recharge des véhicules électriques en détectant et en répondant aux tentatives d'intrusion ou d'attaques cybernétiques.

Finalement, l'objectif principal de notre travail de recherche est de contribuer à la sécurisation et à l'optimisation des infrastructures de recharge des véhicules électriques, afin de promouvoir leur adoption généralisée et leur intégration réussie dans les réseaux énergétiques.

Dans le chapitre suivant, nous présenterons une revue de littérature du réseau V2G et des méthodes de détection d'intrusions. Ce chapitre élaborera les solutions de détection contre les attaques dans le réseau V2G.

CHAPITRE 3 REVUE DE LITTÉRATURE

Cet état de l'art présente différents travaux de recherche sur la sécurité des réseaux véhiculaires V2G et la détection des intrusions. Des études approfondies ont été déployées par les chercheurs afin d'analyser et de suggérer des solutions pour garantir la sécurité des réseaux véhiculaires V2G et les attaques dans le système du réseau. Tout en examinant en détail les avancées scientifiques dans le domaine de la sécurité des réseaux véhiculaires V2G et les systèmes d'intrusion, nous mettrons l'accent sur une analyse critique des méthodes employées, des résultats obtenus et des limites des études antérieures.

3.1 Sécurité des réseaux V2G

La sécurité des réseaux véhiculaires V2G a attiré l'attention de plusieurs chercheurs qui ont effectué une contribution immense à la littérature scientifique.

Dans [35], les auteurs ont proposé un modèle de prévisions précises et de la mise en œuvre stratégique de la technologie V2G pour répondre aux besoins du réseau électrique. Ils soulignent l'importance de prendre en compte les aspects psychologiques et comportementaux des utilisateurs dans la conception des stratégies V2G et définissent les facteurs clés permettant de prévoir la demande de recharge des véhicules électriques. En outres, l'étude ne met pas l'accent sur les préoccupations en matière de confidentialité et de sécurité qui limitent également l'adoption de la technologie V2G, car les utilisateurs doivent partager des données détaillées avec les agrégateurs de véhicules électriques, s'exposant ainsi à des risques potentiels.

Les auteurs dans [36] ont proposé un schéma de protection de la vie privée pour les réseaux V2G basé sur le calcul en brouillard. Ce schéma s'applique à l'algorithme de chiffrement à clé publique sans certificat tout en utilisant des courbes elliptiques pour éviter l'opération de couplage bilinéaire. En présentant une architecture améliorer pour le réseau V2G à travers l'intégration de calcul en brouillard dans le processus de transmission des données de charge et de décharge à l'aide de la technologie d'agrégation de données, le schéma optimise l'efficacité du calcul. Bien que le schéma soit efficace, il présente certaines contraintes, notamment en termes de complexité et de la gestion des clés dans des environnements étendus.

Dans [37], les auteurs ont proposé un système efficace d'analyse des performances qui démontrent que le schéma du système de cryptage de signature agrégée sans certificat (PCAS)

nécessite moins de temps de calcul que les schémas comparables. Lorsque des véhicules électriques de grande taille sont chargés et déchargés via des réseaux V2G, les données sont transmises de manière plus sûre et plus efficace grâce au schéma PCAS. Cela prouve que le système PCAS garantie la sécurité des réseaux véhiculaires V2G et assure la confidentialité. La limitation de cette étude est que le tiers de confiance est utilisé pour initialiser le système et distribuer des clés privées partielles dans le schéma PCAS, ce qui peut entraîner le problème de défaillance à point unique.

Dans l'étude [38], les auteurs ont conçu un schéma de connexion V2G léger qui garantit la confidentialité et la sécurité des informations échangées pendant les sessions de (dé)charge d'électricité, ce qui permet de résoudre le problème d'authentification des véhicules électriques. Cependant le schéma de connexion ne couvre pas la technique de sélection optimale pour les véhicules électriques durant les processus d'échange d'électricité dans le but d'assurer une sécurité totale du réseau V2G.

L'étude de **Saxena** [39] a proposé un schéma basé sur une technologie de couplage bilinéaire et des accumulateurs pour la vérification par lots, qui offre une plus grande efficacité du système, qui se défend contre diverses attaques de sécurité et qui maintient l'intracabilité, la confidentialité future et l'anonymat de l'identité. L'analyse des performances montre que par rapport aux solutions existantes, ce système entraîne beaucoup moins de frais de communication et de calcul dans les réseaux V2G domestiques et centralisés, tout en engendrant des frais comparables dans les réseaux V2G distribués de passage. Cependant, l'un des inconvénients majeurs de ces schémas et protocoles existants est qu'ils ne présentent pas de scénarios d'attaques de sécurité dans les réseaux V2G et que la plupart d'entre eux entraînent une surcharge importante.

L'étude de **Saxane et al** [40] a présenté une nouvelle architecture de sécurité pour soutenir le réseau V2G et présente des solutions de sécurité comme authentification anonyme, contrôle d'accès granulaire, confidentialité des informations, intégrité des messages, attestation à distance et un système de paiement. L'idée de cette étude est d'assurer des transactions sécurisées tout en protégeant la confidentialité des utilisateurs dans un environnement dynamique et distribué. Cependant, l'étude se limite uniquement à la communication des applications intelligentes des réseaux V2G sans prendre en compte la sécurité globale du réseau V2G.

Dans [41], les auteurs ont identifiées plusieurs contre-mesures. Celles-ci incluent le chiffrement de bout en bout pour protéger les données pendant leur transmission, les protocoles d'authentification pour la vérification des transactions, les systèmes de détection d'intrusions pour surveiller les activités suspectes, les mises à jour régulières du firmware, la limitation du taux de requêtes, les mesures de sécurité physique, l'horodatage et les numéros de séquence pour prévenir les attaques par rejeu, les vérifications de l'intégrité des données, le contrôle d'accès basé sur les rôles et la sensibilisation à la sécurité ainsi que la formation. L'étude est seulement basée sur l'identification des contre-mesures sans une proposition de schéma ou un modèle pour la sécurité des réseaux V2G des véhicules électriques.

Dans [42], les auteurs ont présenté un modèle pour soutenir l'exploitation sécurisée du réseau électrique, y compris la modulation de fréquence, la régulation de la tension, la capacité de réserve et l'intégration des énergies renouvelables. Par ailleurs, les normes actuelles et les projets de démonstration de l'application du V2G soutenant activement la sécurité du réseau électrique sont analysés. Enfin, des perspectives pour une intégration fiable et sécurisée du V2G dans le réseau électrique sont proposées. Cette évaluation multidimensionnelle n'aborde pas les limites des analyses unidimensionnelles des études existantes, offrant un cadre plus complet pour mieux comprendre le rôle multifacette des véhicules électriques dans le renforcement de la stabilité et de la sécurité du réseau.

Dans [43], les auteurs ont proposé un modèle d'apprentissage fédéré basé sur la blockchain pour la protection de la vie privée dans les interactions V2G. En déployant le modèle sur une blockchain, la décentralisation est assurée. Le chiffrement homomorphe est utilisé lors de l'entraînement local pour protéger la confidentialité temporelle des utilisateurs sans compromettre la précision de l'entraînement des modèles locaux. Lors de la phase d'agrégation des modèles, un protocole d'agrégation sécurisé et léger est utilisé pour masquer les véritables valeurs de gradient des modèles locaux, garantissant ainsi la robustesse et la tolérance aux pannes du modèle. Toutefois, cette thèse comporte certaines contraintes. À titre d'exemple, si un point de recharge (CP) télécharge un gradient de modèle perturbé, mais que le groupe de consensus C reçoit ce gradient après l'agrégation du modèle local en raison de la latence du réseau, le CP ne pourra pas prendre part au cycle d'apprentissage fédéré actuel.

La première démonstration d'une flotte de véhicules déchargeant pour soutenir la sécurité du système après une contingence de fréquence dans un réseau national a été réalisée dans [44]. Cette étude met en évidence le potentiel du concept véhicule-réseau (V2G), avec des véhicules

déchargeant dans les 6 secondes suivant l'événement de contingence, ainsi que ses limites, les véhicules reprenant la recharge avant que le système électrique ne se soit complètement rétabli. Les données issues de la réponse V2G révèlent des aspects pouvant être améliorés, notamment pour cette mise en œuvre spécifique du V2G et pour les paramètres du marché liés au contrôle de la fréquence.

Dans [45], les auteurs ont développé un algorithme de charge intelligente visant à réduire les coûts d'électricité d'un ménage en optimisant la charge et la décharge d'un véhicule électrique. La charge peut provenir du réseau électrique ou du système photovoltaïque du ménage, tandis que la décharge peut alimenter les appareils électroménagers de la maison ou être renvoyée au réseau, en utilisant des stratégies de véhicule à domicile (V2H) et de véhicule au réseau (V2G). L'algorithme a été testé dans MATLAB avec cinq scénarios différents illustrant ces possibilités et simulé en combinaison avec cinq profils d'utilisateurs distincts. Les résultats montrent que les pics de charge sont déplacés vers des périodes de moindre consommation et confirment que les coûts d'électricité peuvent être réduits sur une base quotidienne et annuelle. En revanche, l'étude ne se concentre pas sur les problèmes de sécurité liés au V2G pour les véhicules électriques, mais seulement sur le système de réseau électrique. En outre, la demeure examinée n'avait qu'un seul véhicule et n'utilisait l'électricité que pour les appareils électroménagers. L'étude d'un foyer avec plusieurs véhicules et l'analyse de l'interaction entre différents profils d'utilisateurs dans le même ménage permettraient d'intégrer un facteur d'interaction et de concurrence pour la charge entre les véhicules. Les résultats pourraient aussi être modifiés par l'étude de véhicules ayant des capacités de batterie différentes.

Dans [46], les auteurs ont analysé comment la technologie Vehicle-to-Grid (V2G) peut équilibrer le réseau et réguler les problèmes de qualité de l'énergie causés par la recharge des véhicules électriques (VE), garantissant ainsi un fonctionnement stable du système. La technologie V2G permet également aux VE de fonctionner comme des dispositifs de stockage d'énergie mobiles et favorise une distribution rationnelle de l'énergie grâce à des stations de charge bidirectionnelles. L'intégration des stations de charge avec le réseau, formant un réseau connecté de véhicules, est considérée comme la future direction du développement des VE. L'étude est limitée par la capacité de charge du réseau de distribution des mesures telles que l'ajustement opportun de la charge du réseau ou la mise à niveau de celui-ci sont nécessaires pour garantir un fonctionnement sûr du réseau de distribution.

3.2 Système de détection d’intrusion

Dans cette section, nous analyserons les recherches menées sur les systèmes de détection d’intrusion. Le but est d’examiner les méthodologies employées, d’identifier les forces et les faiblesses afin d’améliorer les performances de notre propre solution.

Dans [47], les auteurs ont proposé un système de détection d'intrusion (IDS) inspiré de la biologie basé sur l'apprentissage automatique pour prédire et atténuer les attaques sur le réseau V2G. Le modelé proposé a pour but de renforcer la sécurité des réseaux V2G en fournissant des solutions contre les attaques de type Homme du milieu (MitM) et Déni de service (DoS). Les simulations menées à l'aide du simulateur MiniV2G montrent que l'IDS proposé atteint une précision de détection de 98,93%, améliorant ainsi la fiabilité du réseau V2G pour les utilisateurs et offrant une meilleure protection pour les stations de recharge de véhicules électriques contre les attaques DoS et MitM. Malgré la performance du modèle qui a été testé sur un simulateur MiniV2G, il ne peut pas refléter totalement les conditions réelles du réseau. Il est optimisé pour les attaques MitM et DoS, mais pas pour d'autres menaces telles que les attaques internes ou les logiciels malveillants avancés.

Dans [48] les auteurs ont recommandé une technique de réseau de neurones profond (DNN) qui adopte les avancées récentes dans les études sur l'apprentissage profond, telles que l'initialisation des paramètres via l'entraînement non supervisé des réseaux de croyance profonds (DBN), améliorant ainsi la précision de détection. Les résultats expérimentaux démontrent que la technique proposée peut fournir une réponse en temps réel aux attaques avec un taux de détection significativement amélioré dans le bus de réseau de zone de contrôleur (CAN). Bien que les résultats performent avec un taux de détection élevé dans le bus CAN, il existe encore des défis importants en raison de la dépendance à des données de haute qualité et à la complexité des calculs.

Dans [49] les auteurs ont proposé une solution de planification sécurisée pour le système V2G, combinant l'apprentissage profond fédéré avec l'informatique en périphérie distribuée pour les opérations V2G. Dans ce cadre, chaque point de recharge est équipé d'un module informatique intelligent permettant de réaliser une planification en périphérie distribuée pour les véhicules électriques (VE) connectés. Cela garantit non seulement une efficacité dans le processus d'inférence, mais aussi la préservation de la confidentialité des utilisateurs de VE. Par conséquent, les limites de l’étude sont basées sur le problème d'acquisition précise des données

futures. En parallèle, des méthodes de regroupement basées sur l'espace et le temps sont appliquées pour améliorer la précision des prédictions.

Dans [50], les auteurs ont fourni une revue structurée sur les IDS intra-véhicule. L'approche employée consiste à catégoriser les travaux examinés en fonction de leur technique de détection et à analyser les caractéristiques utilisées, les méthodes de sélection des caractéristiques, les ensembles de données d'évaluation, les types d'attaques, les métriques de performance et les modèles de référence. Cependant, l'étude montre qu'aucun travail à ce jour n'a étudié comment le système devrait réagir aux intrusions. Cela est d'une importance cruciale, car des actions inappropriées peuvent entraîner des problèmes de sécurité.

Dans [51], les auteurs ont développé une nouvelle approche qui utilise l'apprentissage automatique pour détecter les attaques précoces. Les auteurs ont étudié l'efficacité de cette approche proposée en prenant en compte différentes résolutions temporelles des données de mesure. L'approche proposée a été testée sur un micro-réseau équipé de ressources énergétiques renouvelables ainsi que de véhicules électriques en mode V2G. Les résultats ont montré que la précision de détection augmente dans le cas de données de mesure à haute résolution temporelle par rapport à celles à basse résolution temporelle. L'étude montre également que l'approche proposée a réussi à détecter précocement les trois types de cyberattaques cyber, physiques et cyber-physiques) avec une précision moyenne de près de 98 %. Cependant, il est compréhensible que l'étude ne prenne pas en compte les attaques cyber-physiques. Ces attaques restent dormantes, mais lorsque le système électrique en subit les impacts, il est trop tard pour que le processus de détection soit efficace et permette des mesures d'atténuation.

Dans [52], les auteurs ont démontré que les prévisions basées sur la régression augmentent de manière significative les performances de détection pour les attaques affectant des rapports individuels pendant une session de charge. De plus, la méthode d'ensemble proposée, qui combine la classification basée sur des réseaux de neurones artificiels et la détection de nouveauté basée sur le facteur de nouveauté local, maintient un faible taux de fausses alertes tout en offrant de bonnes performances de détection pour les attaques connues ainsi qu'une généralisation aux attaques inconnues. L'étude soutient que la solution proposée peut contribuer positivement à la sécurité, à la résilience et à la fiabilité de la recharge des VE. Cependant, elle ne prend pas en compte l'usage futur du transfert d'énergie véhicule-vers-réseau (V2G), qui pourrait ouvrir de nouvelles voies d'attaques ainsi que des options pour la conception des IDS et l'ingénierie des caractéristiques, qui devraient être examinées.

Dans [53], les auteurs ont mis en évidence les défis auxquels sont confrontés les dispositifs de protection en cas de cyberattaques appliquées aux micro-réseaux avec des stations de recharge des VE. L'étude a démontré le cas critique d'un micro-réseau autonome et le dysfonctionnement des relais lorsque le micro-réseau est sous cyberattaque. Cependant, l'étude n'a pas proposé de méthode pour détecter ces cyberattaques.

L'étude présenté dans [54] a analysé la vulnérabilité et les risques liés à l'infrastructure de recharge des VE à travers différentes études de cas impliquant des attaques cyber-physiques. L'étude a conclu que la détection de ces attaques est importante pour prévenir les conséquences à la fois cybernétiques et physiques. Cependant, elle ne prend pas en compte l'usage futur du transfert d'énergie véhicule-vers-réseau (V2G), qui pourrait ouvrir de nouvelles voies d'attaques ainsi que des options pour la conception des IDS et l'ingénierie des caractéristiques, qui devraient être examinées.

Dans [55], les auteurs ont proposé un nouveau modèle IDS pour les réseaux intra-véhicule, appelé GIDS (GAN-based Intrusion Detection System), basé sur les réseaux antagonistes génératifs (Generative Adversarial Nets) et utilisant un modèle d'apprentissage profond. GIDS peut apprendre à détecter des attaques inconnues en utilisant uniquement des données normales. Le modèle GIDS basé sur GAN dépasse les limites des IDS traditionnels et engendre une solution de sécurité réseau CAN prometteuse, répondant aux exigences de précision et de détection des attaques inconnues tout en ouvrant la voie à des futures améliorations du système.

Dans [56], les auteurs ont étudié l'impact de la propagation de logiciels malveillants depuis des VE (véhicules électriques) infectés et des bornes de recharge (EVSE) vers le réseau électrique. L'étude a utilisé un modèle probabiliste pour estimer le nombre de bornes de recharge à isoler tout en soulignant la nécessité d'une approche permettant de détecter ces attaques avant leur propagation au réseau électrique.

Dans [57], les auteurs ont étudié l'impact des attaques de type botnet sur le système de distribution et l'utilisation de réseaux neuronaux pour détecter ces cyberattaques. Cependant, le modèle basé sur les réseaux neuronaux développé est sujet au *surapprentissage* en raison de sa complexité, ce qui peut limiter son applicabilité en raison de sa forte complexité computationnelle. De plus, la détection précoce de ces attaques, qui est cruciale, n'a pas été étudiée.

Dans [58], l'étude a examiné l'effet des attaques de type botnet impliquant des VE compromis et des stations de charge rapides (FSCs) sur la congestion du réseau de distribution et les

violations des limites de tension. L'étude a signalé une augmentation de la charge entraînant des congestions sur les lignes et des bus subissant des violations de sous-tension lorsque les VE étaient programmés pour se recharger simultanément à 7h00 du matin. L'étude a recommandé d'approfondir les recherches sur l'effet des attaques sur plusieurs stations de charge et la nécessité de développer des méthodes préventives pour détecter ces attaques. En l'absence de méthode pour détecter ces cyberattaques, les impacts en termes de surcharge seront observés sur le système de distribution, en particulier lorsque l'on considère les stations de charge rapide pour VE.

L'introduction de l'Internet des objets (IoT) dans les réseaux électriques les rend vulnérables aux cyberattaques. Dans [59], les auteurs ont utilisé la technique du *Power Fingerprinting*, qui repose sur le traitement du signal et la reconnaissance des motifs, pour détecter certaines activités malveillantes dans les systèmes de contrôle industriels des réseaux électriques. L'indice de corrélation est calculé pour identifier les échantillons modifiés et non modifiés. Les échantillons proches des traces de référence enregistrées présentent une valeur de corrélation plus élevée, tandis que les échantillons modifiés ont une valeur de corrélation plus faible. Cependant, aucune information n'est fournie dans l'étude concernant la précision de détection de la technique de Power Fingerprinting pour la détection des anomalies.

Dans [60], un algorithme de détection d'intrusion basé sur le comportement a été proposé pour détecter les anomalies dans une sous-station numérique basée sur la norme IEC 61850. Cependant, la détection de ces anomalies repose principalement sur les messages échangés, ce qui rend la détection précoce de ces anomalies impossibles.

Les auteurs dans [61] ont proposé une approche permettant de détecter les anomalies dans plusieurs sous-stations simultanément. Cependant, cette approche manque également d'un aspect de détection précoce et a montré de mauvaises performances contre des attaques inconnues.

Dans [62], les conséquences d'une cyberattaque ciblant un système SCADA par falsification du protocole de résolution d'adresse (ARP spoofing) ont été étudiées, mais sans fournir de technique de détection efficace.

Dans [63], les auteurs ont étudié l'efficacité de l'utilisation du logiciel Snort pour détecter différents types de cyberattaques sur les dispositifs électroniques intelligents (IEDs) basés sur la norme IEC 61850. Cependant, ce logiciel n'est efficace qu'après le début de l'attaque, rendant ainsi la détection précoce impossible.

Les auteurs dans [64] ont développé un modèle pour répondre aux conditions émergentes et futures les plus perturbatrices pour la sécurité des dispositifs embarqués des bornes de recharge pour véhicules électriques (VE) et de leurs réseaux V2G. L'innovation de cette recherche réside dans la prise en compte des menaces hybrides en cybersécurité affectant les intérêts des propriétaires de systèmes, des opérateurs et des utilisateurs, en répondant à des préférences basées sur des scénarios pour les technologies en rapide évolution dans le cadre du réseau V2G. En tant que méthodologie basée sur des scénarios, cette étude a identifié les scénarios les moins et les plus perturbateurs uniquement en fonction des scénarios particuliers définis, basés sur les sources et les données disponibles au moment de l'étude. L'accès limité à des données et documents supplémentaires ainsi que l'engagement des parties prenantes dans l'étude sont d'autres limitations à mentionner.

L'étude [65] a proposé un schéma d'authentification anonyme de groupe pour les communications V2G. Ce schéma permet l'intégration et la révocation dynamiques des VE et réduit considérablement les charges liées à la révocation des VE. L'analyse théorique montre que ce schéma peut garantir la confidentialité de l'identité des utilisateurs de VE et la sécurité des transmissions de données pendant les processus de charge et de décharge. Par ailleurs, ce schéma résout le problème de gestion complexe de la révocation dans le cadre de la charge et de la décharge, Cependant c'est une contrainte qui persiste dans l'analyse.

Dans cette étude [66], les auteurs ont testé une méthode non supervisée (arbre de décision) de détection d'intrusions sur plusieurs ensembles de données, notamment WNS-DS (84,05 % de précision), KDDCup (60,63 %), UNSW_NB15 (73,62 %) et ISCX (74,83 %). Les résultats montrent que cette méthode est généralisée et robuste pour détecter les intrusions. Elle peut être déployée en temps réel pour surveiller le trafic CAN des véhicules et alerter de manière proactive en cas d'attaques. Cependant, des faiblesses subsistent : les performances varient considérablement selon les ensembles de données, avec une précision particulièrement faible sur KDDCup (60,63 %), ce qui remet en question l'efficacité de la méthode dans certains scénarios spécifiques. De plus, aucune comparaison approfondie avec d'autres approches de pointe n'a été fournie pour justifier la supériorité du modèle.

Dans [67], les auteurs ont proposé un système de détection d'intrusion basé sur l'apprentissage profond pour identifier les attaques par déni de service (DoS) dans les stations de recharge de véhicules électriques (EVCS). Ils ont utilisé des algorithmes de réseaux neuronaux profonds (DNN) et de mémoire à long court terme (LSTM) pour détecter et classer ces attaques. Les

résultats montrent que les deux approches atteignent une précision supérieure à 99 %, le LSTM surpassant le DNN en termes de précision, rappel et F-mesure. Certaines faiblesses ont été identifiées : les résultats, très optimistes (précision de 99 %), pourraient indiquer un *surapprentissage*, limitant la généralisation à d'autres contextes. En outre, l'étude ne teste pas la robustesse face à des attaques inconnues ou des variations dans les données, ce qui pose des questions sur l'efficacité du modèle dans des scénarios réels.

Dans [68], les auteurs ont proposé un système hybride de détection d'intrusion combinant une prévision des sessions de recharge basée sur la régression et une détection d'anomalies. Cette approche intègre des modèles de classification et de détection de nouveauté, ainsi qu'une méthode d'ensemble qui combine des réseaux neuronaux et un facteur de localité d'anomalie (LOF). Les résultats montrent une amélioration significative des performances pour la détection des attaques connues et inconnues, tout en maintenant un faible taux de fausses alertes. Des faiblesses subsistent : la complexité du modèle hybride pourrait compliquer son déploiement en temps réel dans des environnements à ressources limitées. De plus, l'étude n'aborde pas en détail l'efficacité du système face à des attaques très sophistiquées ou à des scénarios rares, ce qui limite l'évaluation complète de sa robustesse.

Dans [69], il est proposé un système de détection d'intrusion pour prédire les attaques dans les réseaux V2G utilisant un classificateur basé sur le comportement des cafards pour renforcer la sécurité contre les attaques par homme du milieu (MitM) et par déni de service (DoS). Les résultats de simulation avec MiniV2G montrent une précision de 98,93 %, ce qui renforce la fiabilité des réseaux V2G et améliore la protection des stations de recharge de véhicules électriques (EVCS). Des faiblesses sont à noter : le modèle repose uniquement sur des simulations (MiniV2G), ce qui peut ne pas refléter pleinement les défis des déploiements réels. De plus, la méthode ne prend pas en compte d'autres vecteurs d'attaques dans les réseaux V2G, ce qui limite son applicabilité à des scénarios spécifiques de sécurité.

Dans [70], les auteurs proposent une approche antagoniste hiérarchique utilisant DRL (Deep Reinforcement Learning) HADRL (Hierarchical Antagonistic Deep Reinforcement Learning), qui détecte efficacement les cyberattaques furtives sur les bornes de recharge pour VE. Cette méthode est divise en deux : La première approche est un schéma qui exploite DRL pour développer des méthodes d'attaque avancées et furtives pour stimuler les attaques sophistiquées capable de contourner les systèmes de détection d'intrusion et le deuxième, un schéma basé sur DRL profond pour les bornes de recharge des VE visant à détecter et à contrer ces attaques

sophistiquées. Le deuxième schéma est constitué d'ensembles de données générés à partir du premier schéma, ce qui crée un système de détection d'intrusion solide et efficace. Les résultats montrent que le cadre HADRL obtient des résultats supérieurs par rapport aux approches existantes. Il détecte avec précision les véhicules trompeurs tout en maintenant un faible de taux de fausse alarme même face à des attaques inédites. Cette avancé contribue à rendre les infrastructures de recharge plus sûres et résilientes contre les cyberattaques. Malgré la performance du cadre DRL, ils ne fournissent pas de détails précis sur les métriques d'évolution (précision, rappel taux de faux positif).

Dans [71] les auteurs ont proposé différents algorithmes de classification d'apprentissage automatique pour détecter les attaques par déni de service distribué (DDoS) dans l'environnement réseau EVCS en utilisant l'ensemble de données typiques de l'Internet des objet (IoT) obtenu à partir du trafic réel. Deux classificateurs classiques employés dans l'apprentissage automatique : le classificateur d'arbre de décision et le classificateur filtré. Selon les résultats obtenus en termes d'exactitude, de précision, de rappel, de score F-1 et de temps de modélisation, le classificateur filtré se distingue comme la solution la plus efficace pour détecter les attaques DDoS dans l'environnement IoT avec une précision de 99,99 % contre 99,97 %. Cela contribue à améliorer la stabilité du système EVCS et à réduire considérablement le nombre de cyberattaques qui pourraient perturber les activités de la vie quotidienne associées à l'écosystème EVCS. Malgré l'utilisation de méthode d'apprentissage automatique, les modèles IDS se révèlent généralement efficaces face aux attaques identifiées, ils peuvent être moins efficaces dans la détection d'attaques *Zeroday* ou de variantes sophistiquées. La méthode suggérée pourrait exiger une réorganisation régulière pour s'adapter à des nouvelles menaces. Les attaques par déni de service (DoS) dans l'EVCS peuvent effectivement se produire grâce à la couche de communication ouverte du système IoT.

Cependant dans l'étude [72], les auteurs ont développé un système de détection d'intrusions basé sur un apprentissage profond, en combinant les techniques d'apprentissage du réseau de neurones profond DNN et de la mémoire à long terme (LSTM) pour détecter et classer ces attaques. L'efficacité des deux méthodes a atteint 99 % de réussite, L'exactitude, la précision, le rappel et les mesures de la méthode LSTM dépassent celles de la méthode DNN. Ils se sont concentrés spécifiquement sur la cyberattaque par le biais du déni de service distribué (DDoS). Même si la détection des attaques DoS favorise l'accessibilité du système, les conséquences sur la protection de la vie privée et l'intégrité des informations ne sont pas abordées, ce qui rend d'autres aspects de la cybersécurité moins pris en compte.

Dans [73], les auteurs ont proposé un modèle d'apprentissage profond pour la détection d'intrusion à l'aide de réseaux neuronaux récurrents (RNN-IDS). Le modèle RNN-IDS se distingue par son efficacité de modélisation, ainsi que sa précision dans la classification binaire et multi-classes. Les auteurs ont analysé les performances du modèle pour la classification binaire ainsi que pour la classification multi-classe en évaluant également l'impact du nombre de neurones et des différents taux d'apprentissage sur l'efficacité du modèle suggéré, tout en utilisant l'ensemble de données NSL-KDD et l'ensemble de formation et de test pour évaluer leur performance en matière de détection dans la classification binaire et multi-classe, en les comparant aux méthodes d'apprentissage automatique. Dans la classification binaire la performance a été comparée à celles du J48, du réseau de neurones artificiel (ANN), la classification bayésienne naïf, foret aléatoire, d'un perceptron multicouche et d'une machine à vecteurs de support (SVM). Les résultats montrent que le RNN-IDS offre une précision supérieure et se révèle particulièrement adapté pour la classification, il surpasse les approches classiques tout en proposant une nouvelle méthode pour améliorer la détection d'intrusion. Malgré la performance du modèle, il manque de données claires ou de comparaisons quantitatives, en plus il n'y a aucun détail sur les métriques utilisées pour évaluer la performance (précision, rappel, score F1).

La revue complète des études existantes sur les réseaux Vehicle-to-Grid (V2G) et les mécanismes de sécurité associés mettent en lumière des lacunes de recherche significatives qui méritent d'être explorées davantage. Malgré les avancées dans la sécurisation des réseaux V2G, la plupart des études se concentrent sur des aspects spécifiques tels que les systèmes de détection d'intrusion (IDS), les techniques de cryptage ou les performances du système. Cependant, il manque des cadres holistiques intégrant ces mesures pour offrir une sécurité complète à l'ensemble de l'écosystème V2G. Les modèles et protocoles existants ne parviennent souvent pas à aborder les capacités de réponse en temps réel, les scénarios de menaces multidimensionnels et les mécanismes de sécurité adaptatifs qui évoluent avec les avancées technologiques. De plus, bien que des études comme celles de [47] et [68] mettent l'accent sur la détection des intrusions et des anomalies, elles n'explorent pas leur intégration avec des techniques de mitigation proactive, notamment dans des systèmes décentralisés comme ceux reposant sur la blockchain ou l'apprentissage fédéré. Cela crée une lacune critique dans l'assurance de la résilience face aux attaques cyber-physiques coordonnées sur les réseaux V2G.

Une autre lacune importante réside dans la prise en compte des préoccupations centrées sur l'utilisateur, telles que la confidentialité et l'adoption comportementale des technologies V2G. Des études comme celles de [35] soulignent l'importance des facteurs psychologiques et comportementaux, mais ne relient pas ces éléments à des solutions techniques garantissant la confiance des utilisateurs et la sécurité du système. De plus, bien que les avancées en apprentissage fédéré et en cryptage homomorphe, telles que proposées dans [37] offrent des perspectives prometteuses pour des solutions préservant la confidentialité, les problèmes de scalabilité et de latence inhérente à ces technologies restent non résolus. Les recherches futures doivent se concentrer sur la conception de systèmes de sécurité V2G robustes, évolutifs et conviviaux, équilibrant efficacité technique, adoption par les utilisateurs et rentabilité. En comblant ces lacunes, les chercheurs peuvent contribuer au développement de réseaux V2G résilients, assurant une intégration fluide avec les systèmes d'énergie renouvelable et les infrastructures de réseaux intelligents plus larges.

4. Conclusion

La revue de la littérature présentée met en évidence des avancées significatives et des défis persistants dans le domaine de détection des intrusions et la sécurité de réseaux V2G. Le prochain chapitre abordera la méthodologie utilisée dans notre étude, en détaillant les éléments nécessaires pour répondre aux questions de recherche et aux objectifs, en se basant sur plusieurs aspects de l'apprentissage automatique et en utilisant des algorithmes et des données afin d’arriver à de nouveaux résultats.

CHAPITRE 4 MÉTHODOLOGIE

4. Introduction

Dans cette partie, nous exposons la méthode utilisée dans notre étude. L'approche utilisée dans cette étude associe l'intelligence artificielle, l'apprentissage automatique, des algorithmes et des méthodes de collecte de données rigoureuses afin de fournir une analyse exhaustive et fiable. Cette méthodologie mettra en œuvre la détection des méthodes courantes de cyberattaques sur les réseaux véhiculaires V2G, l'analyse descriptive, la matrice de corrélation et les algorithmes de l'apprentissage automatisé.

4.1 Méthodes utilisées pour l'analyse

La méthodologie de cette étude présente les éléments d'analyse qui seront utilisés, notamment les statistiques descriptives, ainsi que plusieurs algorithmes d'apprentissage automatique tels que la classification, la régression logistique, les arbres de décision, le gradient boosting et la forêt aléatoire.

4.2 L'apprentissage machine et l'intelligence artificielle

Cette partie de la méthodologie est la plus importante, avec les données collectées des attaques et les facteurs qui peuvent influencer les attaques sur le réseau véhiculaire V2G. Cette section de la méthodologie élaborera comment nous allons utiliser les méthodes utiles de l'apprentissage machine dans le but de trouver une solution. En général, L'apprentissage machine signifie que la performance d'un programme informatique s'améliore avec l'expérience par rapport à une classe de tâches et de mesures de performance [74]. En tant que tel, il vise à automatiser la tâche de construction de modèles analytiques pour effectuer des tâches cognitives telles que la détection d'objets ou la traduction automatique. Cependant, la méthodologie prise en considération dans cette étude est l'apprentissage supervisé par classification.

Donc, il est important de comprendre que, la méthodologie commence avec l'explication de l'analyse descriptive, la matrice de corrélation et les algorithmes de l'apprentissage automatisé tels que : la régression logistique, le classificateur de l'arbre de décision, le classificateur par gradient boosting et le classificateur par foret aléatoire.

4.2.1 Les statistiques descriptives

Les statistiques descriptives sont des coefficients informatifs succincts qui permettent de résumer un ensemble de données, qu'il représente une population entière ou un simple échantillon de celle-ci. Ces statistiques se divisent en deux catégories principales : les mesures de tendance centrale et les mesures de dispersion (ou de variabilité) [75]. Les mesures de tendance centrale comprennent la moyenne, la médiane et le mode, tandis que les mesures de dispersion englobent l'écart type, la variance, les valeurs minimales et maximales, ainsi que les indicateurs de kurtosis (aplatissement) et d'asymétrie (skewness) [75].

4.2.2 Heatmap (la matrice de corrélation)

C'est une carte thermique de corrélation et un outil graphique qui affiche la corrélation entre plusieurs variables sous la forme d'une matrice codée par des couleurs. C'est un peu comme un tableau coloré qui nous montre à quel point différentes variables sont liées entre elles. Chaque variable est représentée par une ligne et une colonne dans une carte thermique de corrélation, et les cellules représentent la relation entre elles. La teinte de chaque cellule témoigne de la puissance et de la direction de la corrélation, les teintes plus sombres étant celles qui sont plus fortes [76].

4.3 Les algorithmes de classification

Les techniques d'apprentissage automatique appelées algorithmes de classification permettent de prédire la catégorie ou la classe d'une donnée en se basant sur ses caractéristiques. Ces algorithmes acquièrent des connaissances en utilisant un ensemble de données étiquetées (données avec des classes ou des catégories déjà définies) afin de donner une étiquette ou une classe aux nouvelles données non étiquetées. Les algorithmes de l'apprentissage automatisé sont tels que : la régression logistique, le classificateur de l'arbre de décision, le classificateur par gradient boosting et le classificateur par foret aléatoire.

4.3.1. Algorithme de la régression logistique

L'algorithme de régression logistique est couramment employé dans des tâches de classification binaire, comme déterminer si un courriel est un spam ou non, ou pour diagnostiquer des maladies en vérifiant la présence ou non de conditions spécifiques basée sur les résultats d'examens médicaux [77].

Dans le domaine de la régression logistique, on utilise la fonction sigmoïde pour modéliser la

probabilité qu'une observation soit classée (par exemple, 1 pour « oui » ou 0 pour « non »). En français, on peut exprimer la formule de la manière suivante :

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

Explications :

- P(y=1|X) : Probabilité que la variable dépendante y prenne la valeur 1, donnée les variables explicatives X.
- β_0: Interception (ou biais).
- β_1, β_2…. β_n: Coefficients de régression associés aux variables explicatives X_1, X_2…X_n.
- X_1, X_2, …, X_n : Variables explicatives (ou caractéristiques) du modèle.
- e: Base du logarithme naturel (environ 2,718).

Cette formule utilise une transformation logistique (fonction sigmoïde) pour s'assurer que la sortie soit une probabilité entre 0 et 1. Si P(y=1|X) dépasse un seuil donné (généralement 0,5), l'observation est classée dans la catégorie y=1. Dans le cas contraire, elle est classée comme y=0.

4.3.2 Algorithme du classificateur de l'arbre de décision

Un arbre de décision est une méthode de division récurrente employée pour faciliter la prise de décisions. Il utilise un modèle arborescent pour illustrer les décisions et leurs effets potentiels, comprenant les conséquences d'événements imprévus, le coût des ressources et la valeur [78]. La formule de l'algorithme du classificateur de l'arbre de décision est la suivante avec l'indice de Gini:

L'indice de Gini, une autre mesure d'impureté, est calculé comme suit :

$$Gini(S) = 1 - \sum_{i=1}^{c} p_i^2$$

Explications :

- Gini(S) : Impureté du nœud S.

- P_i : Proportion des observations de la classe i dans le nœud S.

4.3.3 Algorithme du classificateur du gradient boosting

Le boosting gradient est une méthode d'apprentissage automatique fondée sur le concept de boosting dans un espace fonctionnel, où la cible est des pseudo-résidus et non des résidus comme dans le boosting classique. Il génère un modèle prédictif en utilisant une série de faibles modèles de prédiction (comme des arbres décisionnels).

Le modèle de gradient boosting à l'étape t est donné par :

$$F_t(x) = F_{t-1}(x) + \nu \cdot h_t(x)$$

Explication :

- $F_t(x)$: Modèle à l'étape t.
- $F_{t-1}(x)$: Modèle à l'étape précédente t−1.
- V : Taux d'apprentissage (learning rate), un facteur d'échelle pour ajuster la contribution de $h_t(x)$.
- $h_t(x)$: Nouveau modèle faible (par exemple, un arbre de décision) ajusté sur les pseudo-résidus.

4.3.4 Algorithme du classificateur par foret l'aléatoire

Les forêts aléatoires, également appelées forêts d'arbres de décision aléatoires, sont une technique d'apprentissage ensembliste employée dans le domaine de la classification, de la régression et d'autres domaines [79]. Cette technique opère en créant une multitude d'arbres de décision au cours du processus d'entraînement [79].

Pour les opérations de catégorisation, le sort aléatoire de la forêt correspond à la catégorie sélectionnée par la plupart des arbres. Pour les opérations de régression, la sortie correspond à la moyenne des prévisions effectuées par les arbres.

Pour une observation x, la classe prédite par la forêt aléatoire est donnée par :

$$\hat{y} = \arg\max_{c \subset C} \sum_{i=1}^{T} 1(h_i(x) = c)$$

Explications :

- $\hat{y}$: Classe prédite pour x.
- C : Ensemble des classes possibles.
- T : Nombre total d'arbres dans la forêt.
- $h_i(x)$: Prédiction de l'arbre i pour l'observation x.
- $1(\cdot)$: Fonction indicatrice, qui vaut 1 si l'expression est vraie et 0 sinon.

La classe c ayant le plus grand nombre de votes parmi les T arbres est sélectionnée.

4.4 Logiciels utilisés et données

Les données utilisées sont des données quantitatives relatives aux attaques sur les réseaux véhiculaires V2G. Ces données sont exploitées afin d'analyser les menaces potentielles et d'évaluer leur impact à l'aide d'algorithmes de classification adaptés. L'analyse des données quantitatives est un processus qui consiste à examiner, organiser et interpréter des données numériques ou mesurables afin d'identifier des tendances, des relations ou des schémas. Cette analyse repose souvent sur des méthodes statistiques, des calculs mathématiques ou des techniques algorithmiques pour synthétiser l'information et en tirer des conclusions.

Dans le but de bien mener à l'analyse, il y a deux logiciels qui ont été utilisés qui sont *Anaconda* pour déclencher le deuxième logiciel qui est *Python* qui permettra à faire une analyse détaillée cette étude de recherche. Le logiciel Python est un programme ou une application développée à l'aide du langage de programmation Python. Python est un langage de haut niveau, interprété, orienté objet, et facile à apprendre, souvent utilisé pour le développement d'applications web, scientifiques, d'analyse de données, d'intelligence artificielle et d'apprentissage automatique.

4.5 Conclusion

La modernisation et l'optimisation des systèmes de véhicules électriques sont essentielles grâce à l'intelligence artificielle (IA) et à l'apprentissage machine (ML), en particulier en intégrant la technologie Vehicle-to-Grid (V2G) avec des systèmes des véhicules électriques. En résumé, l'utilisation de l'intelligence artificielle et de l'apprentissage automatique dans le domaine des véhicules électriques et de la technologie V2G offre un immense potentiel pour améliorer la sécurité et l'efficacité énergétique. L'importance de l'utilisation de ces technologies avancées

pour faire face aux défis actuels et futurs dans le domaine de la mobilité électrique est mise en évidence par les résultats de cette étude.

Le chapitre suivant se concentrera sur l'analyse des données de cette étude de recherche à l'aide des techniques du Machine Learning et des algorithmes énumérés dans le chapitre 4 afin d'atteindre les objectifs de recherche que nous avons défini.

CHAPITRE 5 ANALYSE DES DONNÉES ET RÉSULTATS DE L'APPRENTISSAGE SUPERVISÉ

5.0 Introduction

L'analyse des données sur les attaques potentielles auxquelles les voitures électriques peuvent être exposées, ainsi que l'utilisation de l'apprentissage machine et des algorithmes pour faire face à ces défis, sont examinées en détail dans cette section. Ces voitures sont aussi exposées à différentes attaques qui peuvent mettre en péril leur fonctionnement et leur sécurité. La compréhension des tendances et des anomalies est facilitée par l'analyse descriptive des données, tandis que l'apprentissage automatique fournit des outils performants pour anticiper et prévenir les attaques, ainsi que pour améliorer les méthodes d'authentification.

5.1 Méthodes de cyber-attaques dans les voitures électriques

Les véhicules électriques, étant de plus en plus complexes et dépendants de la technologie, sont vulnérables à de nombreuses attaques potentiellement destructrices. Chaque vecteur d'attaque comporte des risques spécifiques, qu'il s'agisse d'une compromission physique ou de l'exploitation de failles logicielles en passant par des attaques sur les systèmes de communication et de recharge.

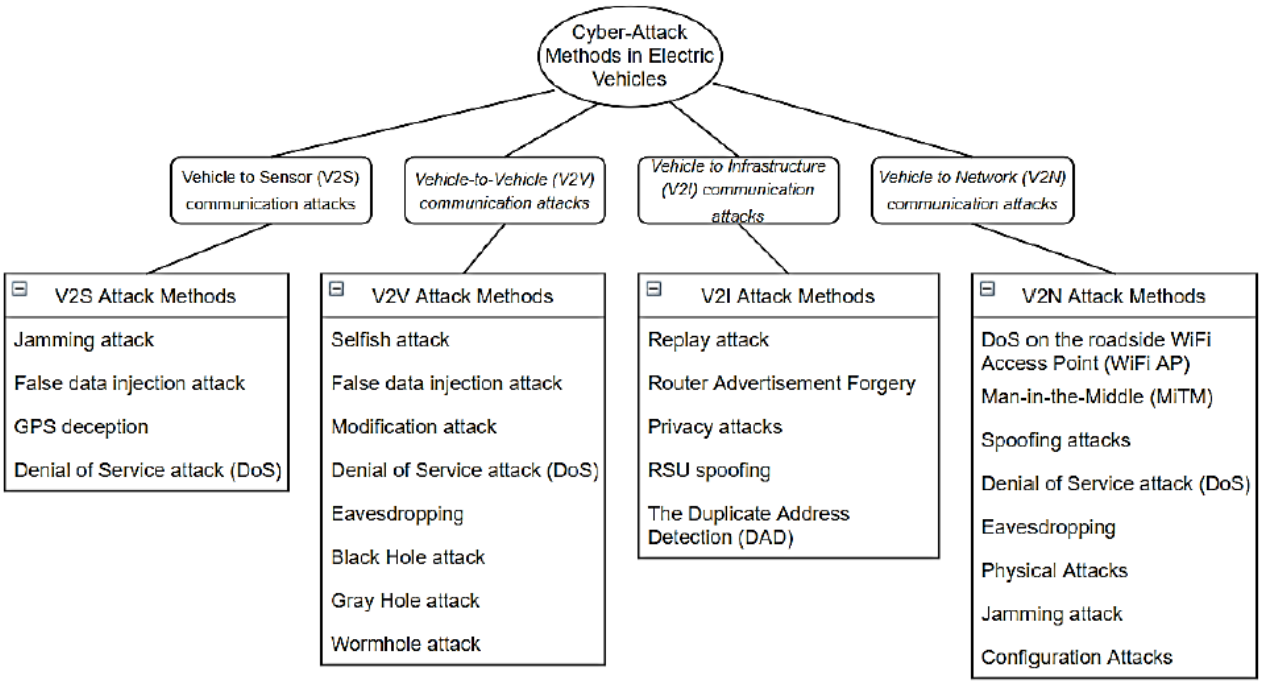


Figure 6 : Méthodes de cyber attaques dans les voitures électriques. Source : (Ozarpa et al, 2021).

Les méthodes de cyber-attaques sur les véhicules électriques désignent les différentes techniques utilisées par des individus malveillants pour compromettre la sécurité des systèmes informatiques et électroniques présents dans les véhicules électriques. Ces attaques visent à perturber le fonctionnement normal du véhicule, compromettre la sécurité des passagers ou accéder à des données sensibles. Voici quelques-unes des méthodes couramment utilisées :

a. **Piratage du système de divertissement et de navigation :** Les pirates peuvent exploiter les vulnérabilités des systèmes de divertissement et de navigation pour accéder au réseau informatique du véhicule et prendre le contrôle de certaines fonctions, telles que la climatisation, la radio ou les paramètres de conduite.
b. **Attaques par déni de service (DDoS) :** Les attaques par déni de service visent à saturer les serveurs ou les systèmes de communication du véhicule avec un trafic malveillant, ce qui peut entraîner une interruption du service ou des dysfonctionnements du système.
c. **Piratage des systèmes de contrôle :** Les pirates peuvent tenter de pirater les systèmes de contrôle électronique du véhicule, tels que le système de freinage ou de direction assistée, afin de compromettre la sécurité des passagers et provoquer des accidents.

d. **Vol de données personnelles :** Les données personnelles stockées dans les systèmes informatiques des véhicules électriques, tels que les informations de navigation, les contacts téléphoniques ou les données de localisation, peuvent être ciblées par des pirates dans le but de les voler ou de les exploiter à des fins malveillantes.
e. **Ingénierie sociale :** Les pirates peuvent utiliser des techniques d'ingénierie sociale pour tromper les utilisateurs et leur faire divulguer des informations sensibles, telles que des identifiants de connexion ou des mots de passe, ce qui leur permet d'accéder aux systèmes informatiques du véhicule.

Il est essentiel pour les constructeurs automobiles et les utilisateurs de véhicules électriques de prendre des mesures de sécurité appropriées pour protéger les systèmes informatiques et électroniques des véhicules contre les cyber-attaques. Cela peut inclure l'utilisation de logiciels de sécurité, de pare-feu, de systèmes de détection d'intrusion, ainsi que la mise à jour régulière des logiciels pour combler les failles de sécurité connues. De plus, la sensibilisation des utilisateurs aux risques de cyber-attaques et aux bonnes pratiques en matière de sécurité informatique est également essentielle pour prévenir les attaques et protéger les données des utilisateurs.

5.2 Analyse descriptive

Les statistiques descriptives résument l'échantillon étudié sans tirer de conclusions fondées sur la théorie des probabilités. Bien que l'étude vise principalement des statistiques inférentielles, les statistiques descriptives sont toujours employées pour fournir un résumé global. Les statistiques descriptives sont utilisées lorsque nous décrivons la population en utilisant des outils tels que les tableaux de distribution de fréquence, les pourcentages et d'autres mesures de tendance centrale comme la moyenne, par exemple.
Cependant, cette section de l'analyse met en évidence les composants de l'analyse descriptive qui sont la moyenne, l'écart type, le minimum, le maximum et le pourcentage en termes de 25% à 50%. Mais avant tout, il est important de donner la description des variables de cette analyse qui est ci-dessous :

Tableau 1: Description des variables.

Feature	Description	Type de Donnée
Flow Duration	Durée du flux	int64
Bwd Pkts/s	Nombre de paquets en retour par seconde	float64
Fwd Pkts/s	Nombre de paquets en avant par seconde	float64
Flow IAT Min	Intervalle minimum entre deux flux	float64
Flow IAT Mean	Intervalle moyen entre deux flux	float64
Flow IAT Max	Intervalle maximum entre deux flux	float64
Flow Pkts/s	Nombre de paquets transférés par seconde	float64
Flow Total	Nombre total de paquets	int64
Flow IAT Std	Écart type de l'intervalle entre les flux	float64
Flow IAT Tot	Intervalle temporel total entre les flux	float64
Fwd IAT Mean	Intervalle moyen entre paquets en avant	float64
Fwd IAT Std	Écart type de l'intervalle en avant	float64
Fwd IAT Max	Intervalle maximum entre paquets en avant	float64
Fwd IAT Min	Intervalle minimum entre paquets en avant	float64
Subflow Fwd Pkts	Paquets sous-flux en avant	int64
Active Mean	Durée moyenne d'activité	float64
Active Std	Écart type de la durée d'activité	float64
Active Max	Durée maximale d'activité	float64
Active Min	Durée minimale d'activité	float64
Idle Mean	Durée moyenne d'inactivité	float64
Idle Std	Écart type de la durée d'inactivité	float64
Idle Max	Durée maximale d'inactivité	float64
Idle Min	Durée minimale d'inactivité	float64
ATT	Variable cible indiquant si une attaque a eu lieu (1) ou non (0)	int64

Après avoir obtenu les informations de la base de données des intrusions d'attaques, la prochaine étape est le nettoyage de cette base de données. Pour la basse de données, il y a eu une analyse descriptive pour chaque variable.

- Ensuite, une colonne "ATT" (attaque) est créée dans chaque fichier CSV correspondant aux bases de données partielles nettoyées.
- Dans le fichier CSV correspondant aux données relatives au scénario sans attaque : toutes les valeurs de la variable ATT sont à "0".
- Dans le fichier CSV correspondant aux données relatives au scénario avec une attaque de type homme du milieu : toutes les valeurs de la variable ATT sont à "1". La variable ATT nous permet d'étiqueter nos données et de faire la différence entre les données avec et sans attaque.

Enfin, après l'ajout de la variable ATT, les trois bases de données partielles nettoyées ont été fusionnées, ce qui donne l'obtention de la base de données finale.

Ci-dessous (Tableau 2) est présentée l'analyse descriptive de cette étude de recherche.

Tableau 2:Analyse descriptive.

	count	mean	std	min	25%	50%
Flow Duration	283.00	1730636.83	3270323.15	-75.00	333846.50	569114.00
Bwd Pkts/s	283.00	9867.98	68498.27	0.00	0.38	1.61
Fwd Pkts/s	283.00	9924.80	68492.37	0.00	1.58	3.17
Flow IAT Min	283.00	180615.32	817211.00	-254410.00	3.00	6048.00
Flow IAT Mean	283.00	488155.01	1184595.57	-75.00	6792.55	295870.00
Flow Pkts/s	283.00	-19627.25	221859.21	-1000000.00	2.51	5.02
Flow IAT Max	283.00	1430153.91	3107989.39	-75.00	90405.00	350874.00
Tot Fwd Pkts	283.00	22.00	47.27	1.00	1.00	3.00
Flow IAT Std	283.00	543695.44	1621858.73	0.00	0.00	3968.81
Fwd IAT Tot	283.00	1503934.65	3251078.27	0.00	0.00	254505.00
Fwd IAT Mean	283.00	405785.42	1360455.96	0.00	0.00	5306.69
Fwd IAT Std	283.00	619933.95	1958352.16	0.00	0.00	2600.86
Fwd IAT Max	283.00	1245627.91	3075196.50	0.00	0.00	29705.00
Fwd IAT Min	283.00	4102.35	84144.32	-254410.00	0.00	0.00
Subflow Fwd Pkts	283.00	22.00	47.27	1.00	1.00	3.00
Active Mean	283.00	20213.62	124242.29	0.00	0.00	0.00
Active Std	283.00	0.00	0.00	0.00	0.00	0.00
Active Max	283.00	20213.62	124242.29	0.00	0.00	0.00
Active Min	283.00	20213.62	124242.29	0.00	0.00	0.00
Idle Mean	283.00	467431.43	2837041.59	0.00	0.00	0.00
Idle Std	283.00	0.00	0.00	0.00	0.00	0.00
Idle Max	283.00	467431.43	2837041.59	0.00	0.00	0.00
Idle Min	283.00	467431.43	2837041.59	0.00	0.00	0.00
ATT	283.00	0.34	0.47	0.00	0.00	0.00

Sur le tableau ci-dessus nous pouvons voire les informations de l'analyse descriptive de chaque variable en fonction de la moyenne, l'écart-type, le maximum et le minimum et le pourcentage. La suite de cette analyse descriptive est présentée dans le tableau 3 ci-dessous.

Tableau 3: Analyse descriptive 2.

	75%	max
Flow Duration	1176512.50	25945636.00
Bwd Pkts/s	2.17	1000000.00
Fwd Pkts/s	5.56	1000000.00
Flow IAT Min	296846.50	13327440.00
Flow IAT Mean	539173.00	13327440.00
Flow Pkts/s	7.86	2000000.00
Flow IAT Max	898613.50	25945617.00
Tot Fwd Pkts	7.00	205.00
Flow IAT Std	463793.70	14979703.47
Fwd IAT Tot	1140583.00	25945578.00
Fwd IAT Mean	325633.38	12972789.00
Fwd IAT Std	549429.44	18346349.30
Fwd IAT Max	878362.50	25945617.00
Fwd IAT Min	3598.00	329261.00
Subflow Fwd Pkts	7.00	205.00
Active Mean	0.00	1143768.00
Active Std	0.00	0.00
Active Max	0.00	1143768.00
Active Min	0.00	1143768.00
Idle Mean	0.00	25945617.00
Idle Std	0.00	0.00
Idle Max	0.00	25945617.00
Idle Min	0.00	25945617.00
ATT	1.00	1.00

Après avoir examiné les statistiques descriptives de cette analyse, nous avons commencé à nettoyer les données en fonction des variables que nous avons actuellement afin d'obtenir une précision adéquate dans le processus de nettoyage. En analysant les données, nous avons constaté que 13 lignes étaient dupliquées dans la base de données, ce qui nous a poussés à préserver les données restantes, soit 270 lignes de données qui pourraient être utilisées pour l'analyse des données.

Par la suite, nous avons étudié les données manquantes et les anomalies présentes dans notre jeu de données. Selon le cas, les valeurs manquantes ont été traitées en imputant ou en supprimant les lignes concernées. Les anomalies ont été repérées et analysées afin de déterminer s'il s'agissait d'erreurs de saisie ou de valeurs extrêmement pertinentes. Nous avons

pu améliorer la qualité et la fiabilité de notre jeu de données grâce à ce processus de nettoyage, ce qui nous a assuré des résultats d'analyse plus précis et significatifs.

Finalement, nous avons standardisé et étendu les variables requises afin de garantir qu'elles soient à une échelle similaire, ce qui revêt une importance capitale pour certains algorithmes de modélisation. Il est crucial de réaliser une préparation minutieuse des données afin d'obtenir des modèles prédictifs solides et des informations fiables.

5.3 Analyse exploratoire des données d'attaques

L'analyse exploratoire des données (EDA) en statistiques est l'étude d'ensembles de données afin de mettre en lumière leurs principales caractéristiques, généralement en utilisant des graphiques statistiques et d'autres techniques de visualisation. Si l'on peut utiliser un modèle statistique, l'EDA se focalise sur la recherche d'insights à partir des données, au-delà de ce que la modélisation formelle peut offrir, en opposition aux tests d'hypothèses classiques.

Les codes de l'analyse exploratoire des données de cette étude sont présentés ci-dessous (fig.9).

```
values = list(df_cleaned["ATT"].value_counts())
mylabels = ["NO Attack", "Attack"]

plt.figure(figsize=(6,6))
plt.pie(values, labels=mylabels, colors=['#c1f55b', '#f56f5b'], autopct="%1.
  ↪1f%%", startangle=45)
centre_circle = plt.Circle((0,0),0.70,fc='white')
fig = plt.gcf()
fig.gca().add_artist(centre_circle)

plt.title('Distribution of ATT Column', fontsize=14)
plt.axis('equal')
plt.show()
```

Figure 7 : codes de l'analyse exploratoire des données.

Selon les résultats de l'analyse exploratoire de cette étude qui est présentée à la figure 8 ci-dessous, nous pouvons voir la classification des données avec attaque et sans attaque. Avec attaque, c'est 30.4% des données et sans attaque, c'est 69.6%.

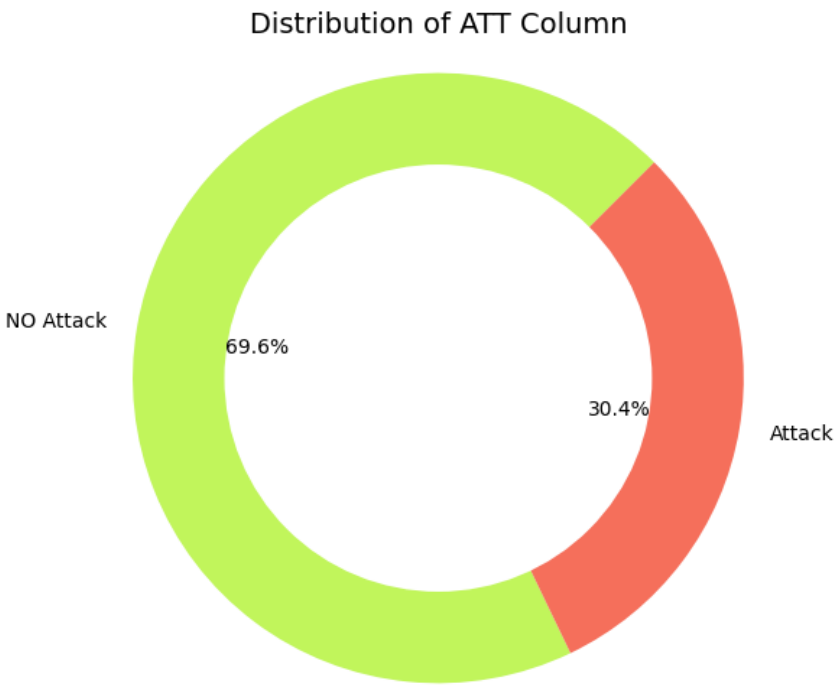


Figure 8 : la distribution de la variable ATT.

En analysant la variable ATT, qui symbolise l'attaque, nous pouvons identifier une matrice de corrélation afin de repérer les variables qui présentent une corrélation positive avec ATT. Par la suite, ces variables peuvent être perçues comme des éléments influençant la variable ATT.

Nous commençons par analyser les coefficients de corrélation entre ATT et les autres variables du jeu de données. Lorsque la valeur de la variable augmente, la valeur d'ATT tend également à augmenter, ce qui indique une corrélation positive. Les indicateurs clés ou les contributeurs aux performances d'attaque peuvent être des variables avec une corrélation significative et positive.

Par la suite, nous examinons ces variables afin de saisir leur relation avec ATT. Cela peut nécessiter des représentations visuelles supplémentaires, comme des nuages de points ou des diagrammes de dispersion, afin d'observer l'interaction entre les valeurs de ces variables et ATT.

5.4 Heatmap (Matrice de corrélation)

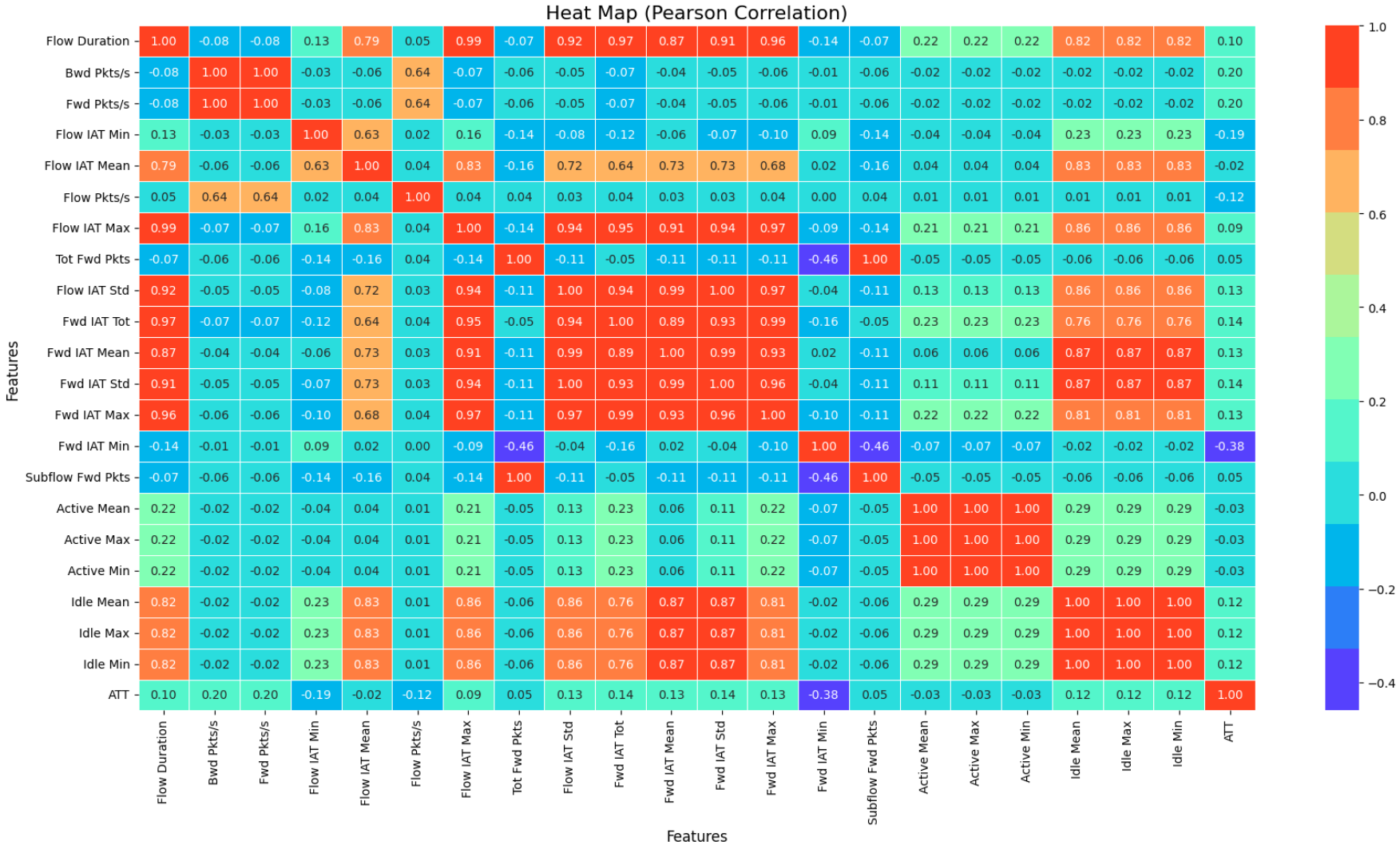
Il s'agit d'un instrument visuel qui présente la relation entre plusieurs variables sous la forme d'une matrice coloriée. C'est un tableau de couleurs qui nous montre combien les variables sont interdépendantes.

Grâce à une *Heatmap* de corrélation (Tableau 4), il est facile de repérer quelles variables présentent des relations fortes, positives ou négatives, et lesquelles sont faiblement ou pas du tout liées. Les teintes vives, fréquemment dans des teintes rouge ou bleue, témoignent de corrélations élevées (positives ou négatives), tandis que les teintes plus claires révèlent des corrélations faibles ou inexistantes.

Chaque carré illustre la relation entre les variables sur chaque axe pour l'interprétation. La relation oscille entre -1 et +1. Il n'y a pas de tendance linéaire entre les deux variables si les valeurs sont proches de zéro. Les variables sont plus positivement corrélées lorsque la corrélation est proche de 1, c'est-à-dire que lorsque l'une augmente, l'autre augmente également, et plus la corrélation est proche de 1, plus cette relation est forte. On observe une corrélation similaire à -1 : au lieu d'augmenter les deux variables, l'une diminue et l'autre augmente.

Par conséquent, une corrélation de +1 correspond à une relation linéaire parfaite positive, où toutes les variations de l'une des variables sont parfaitement reflétées par l'autre. De la même manière, une corrélation de -1 correspond à une relation linéaire parfaite négative, où une variation de l'une des variables est exactement égale à une diminution de l'autre. Les valeurs intermédiaires indiquent des relations linéaires variables, avec des valeurs plus éloignées de zéro qui témoignent de relations plus intenses.

Tableau 4:Corrélation Heat Map.



Comme on peut le voir sur la matrice de corrélation, les variables telles que Flow Duration, Fwd Pkts/s, Flow IAT Mean, Fwd IAT Tot ont une corrélation significative avec ATT, pour cela, nous allons les prendre comme des variables ciblées.

```
# features and target variable
X = data[['Flow Duration', 'Fwd Pkts/s', 'Flow IAT Mean', 'Fwd IAT Tot']]
y = data['ATT']
```

Figure 9 : Code de sélection des variables.

5.5 Machine Learning (Fonction du modèle de classification)

Pour cette étude, nous allons utiliser quatre algorithmes d'analyse : la régression logistique, le classificateur d'arbres de décision, le classificateur de gradient boosting et le classificateur de forêt aléatoire.

On utilise la régression logistique afin de représenter la probabilité d'une classe ou d'un événement spécifique. Elle est particulièrement bénéfique lorsqu'il s'agit d'une variable binaire.

Le classificateur d'arbres de décision crée un modèle de décisions en utilisant un arbre, où chaque nœud interne représente une "question" concernant un attribut, chaque branche représente le résultat de cette question, et chaque feuille représente une classe cible.

Le classificateur de boosting de Gradient combine différents modèles faibles, habituellement des arbres de décision, afin de générer un modèle puissant. Chaque modèle suivant rectifie les erreurs des modèles précédents, ce qui améliore les performances de Prédiction.

5.5.1 Régression logistique

Les problèmes de classification binaire et linéaire peuvent être résolus de manière plus simple et plus efficace en utilisant la régression logistique. Il s'agit d'un modèle de classification extrêmement simple à mettre en place et qui offre de très bonnes performances avec des classes séparables de manière linéaire. Il s'agit d'un algorithme couramment utilisé dans le domaine de la classification.

```
list_of_all_ml_metrics = []
# train logisitic regression
list_of_all_ml_metrics.append(build_ml_model(LogisticRegression(), 'Logistic␣
 ↪Regression', X_train_scaled, y_train, X_test_scaled, y_test)
)
```

Figure 10 : Code de la régression linéaire.

À partir de l'analyse de la régression logistique, nous présentons une explication de la validation croisée du modèle et une prédiction du modèle (fig. 11).

5.5.1.1 Validation et test de données

```
================================================================================
              Logistic Regression Model Cross-Validation (CV=5)
================================================================================
Accuracy: 0.704 ± 0.051
Precision_macro: 0.687 ± 0.119
Recall_macro: 0.593 ± 0.058
F1_macro: 0.583 ± 0.076
================================================================================
              Logistic Regression Model Evaluation on Test Data
================================================================================
Accuracy:       0.667
Precision:      0.593
Recall:         0.539
F1-score:       0.513
AUC:            0.539
================================================================================
```

Figure 11 : Validation du modèle de la régression logistique.

Pour la régression logistique, il est important d'expliquer la validation croisée du modèle et l'évaluation du modèle sur les données de test. On a déjà discuté de la précision du modèle, de la précision (précision), du rappel (recall) et du score F1.

Du point de vue du modèle basé sur la validation croisée, nous pouvons voir que la précision (accuracy) est de 0,704, ce qui est une bonne précision. En termes de précision macro (précision macro), elle est de 0,687. Le rappel macro (recall macro) est de 0,593 et le F1_macro est de 0,583. Pour rappel, cela signifie qu'il y a un pourcentage de faux négatifs de 59,3%.

Concernant l'évaluation du modèle de régression logistique sur les données de test, il est possible de constater que la précision du modèle est de 0,667, la précision est de 0,593 et le rappel est de 0,539. En ce qui concerne les vrais positifs et les faux négatifs du modèle, cela indique qu'il existe une probabilité de 53,9 % d'avoir des faux négatifs. En ce qui concerne l'interprétation pour équilibrer la précision et le rappel, le score F1 s'élève à 0,513.

En conclusion, on peut dire que la régression logistique a produit des résultats acceptables, mais il y a des aspects à améliorer pour optimiser ses performances.

- Validation croisée : Avec une précision de 70,4%, le modèle montre une bonne capacité de prédiction globale. Cependant, le rappel de 59,3% indique que le modèle échoue encore à détecter une proportion significative des instances positives. La précision de

68,7% et le score F1 de 58,3% révèlent un équilibre modéré entre la précision et le rappel, mais la performance peut être optimisée davantage.

- Données de test : Sur les données de test, la précision tombe à 66,7%, ce qui est attendu, car les modèles ont souvent des performances légèrement inférieures sur des données non vues. Le rappel de 53,9% et le score F1 de 51,3% confirment la difficulté du modèle à généraliser ses prédictions, montrant un besoin d'amélioration pour mieux identifier les instances positives tout en maintenant un équilibre entre précision et rappel.

5.5.1.2 Matrice de confusion

Le résumé des prédictions sous forme matricielle est représenté par une matrice de confusion (Fig. 12). Elle met en évidence le nombre de prédictions correctes et incorrectes par catégorie. Elle permet de saisir les classes confondues par le modèle avec d'autres classes.

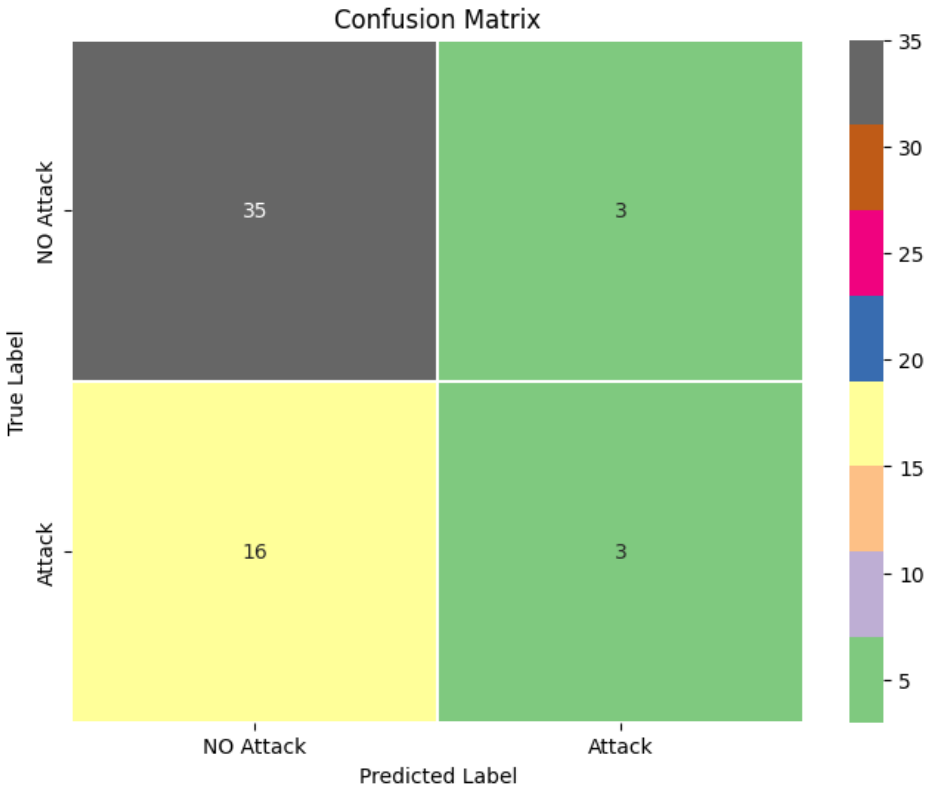


Figure 12 : Matrice de confusion- Régression logistique.

La matrice de confusion combine les éléments, vrai positif, faux positif, faux négatif et vrai négatif, pour prédire la variable attaque (ATT) (Tableau 5).

Tableau 5: Tableau d'interprétation de la matrice de confusion.

	Prédiction Positive	Prédiction Négative
Classe Positive (Vrai)	Vrai Positif (TP)	Faux Négatif (FN)
Classe Négative (Faux)	Faux Positif (FP)	Vrai Négatif (TN)

En se basant sur la prédiction du modèle avec la matrice de confusion, il est démontré qu'il y a 35 vrais positifs, 16 faux positifs, 3 vrais négatifs et 3 faux négatifs qui sont prédits par les variables Flow Duration, Fwd Pkts/s, Flow IAT Mean, Fwd IAT Tot.

Ensuite, la courbe de régression logistique figure 13 ci-dessous montre exactement comment la prédiction est liée à la conjoncture aléatoire de l'étude

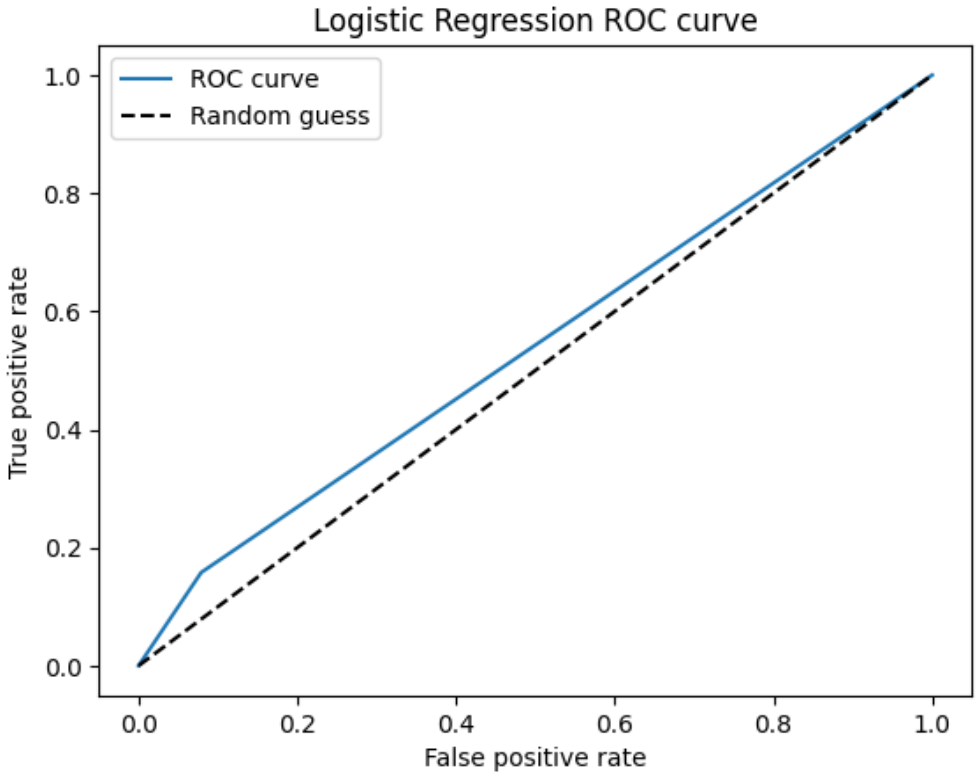


Figure 13 : La courbe de la régression logistique.

5.5.2 Classificateur de l'arbre de décision

Un classificateur basé sur un arbre de décision est un modèle de prédiction employé dans le domaine de l'apprentissage automatique qui représente les observations sur les cibles

prédictives. Il adopte une structure arborescente où chaque nœud interne correspond à un test sur une caractéristique (par exemple, si une variable est supérieure ou inférieure à une valeur donnée), chaque branche correspond au résultat du test, et chaque feuille correspond à une étiquette de classe (la décision prise après avoir examiné toutes les caractéristiques).

5.5.2.1 Validation et test de données

```
================================================================================
            Decision Tree Classifier Model Cross-Validation (CV=5)
================================================================================
Accuracy: 0.978 ± 0.020
Precision_macro: 0.974 ± 0.023
Recall_macro: 0.977 ± 0.023
F1_macro: 0.975 ± 0.022
================================================================================
            Decision Tree Classifier Model Evaluation on Test Data
================================================================================
Accuracy:      0.930
Precision:     0.921
Recall:        0.921
F1-score:      0.921
AUC:           0.921
================================================================================
```

Figure 14 : Validation du modèle de l'arbre de décision.

Pour le classificateur par arbre de décision, lors de l'évaluation en validation croisée du modèle, il est important de mentionner que le modèle présente une forte précision avec une valeur de 0,978, une précision_macro de 0,974, un rappel_macro de 0,977 et un score F1_macro de 0,975.

En ce qui concerne l'évaluation du modèle sur les données de test, l'exactitude est de 0,930, la précision est de 0,921, le rappel est de 0,921 et le score F1 est de 0,921 pour la prédiction du modèle.

Un classificateur par arbre de décision est un modèle de machine learning qui divise récursivement les données en sous-groupes homogènes en fonction des caractéristiques des données. Chaque division se fait de manière à maximiser la pureté des sous-groupes résultants en termes de classes cibles. Les nœuds de l'arbre correspondent à des tests sur les caractéristiques des données, les branches représentent les résultats de ces tests, et les feuilles de l'arbre contiennent les prédictions finales. Ce modèle est souvent utilisé pour sa capacité à capturer des relations non linéaires entre les caractéristiques et les classes cibles, tout en étant relativement interprétable.

5.5.2.2 Matrice de confusion

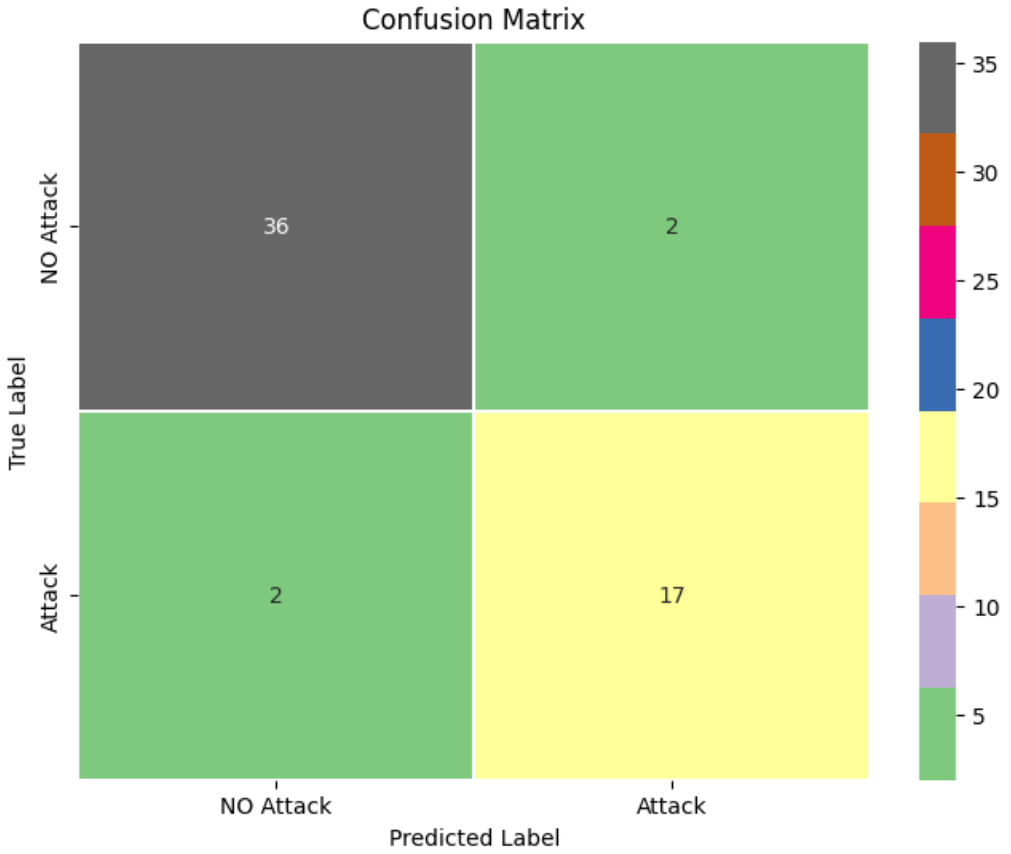


Figure 15 : Matrice de confusion-arbre de décision.

Il est essentiel de consulter le tableau ci-dessous afin d'interpréter la matrice de confusion, qui combine les éléments de vrai positif, faux positif, faux négatif et vrai négatif pour prédire la variable attaque (ATT).

Tableau 6:Tableau d'interprétation de la matrice de confusion 2.

	Prédiction Positive	Prédiction Négative
Classe Positive (Vrai)	Vrai Positif (TP)	Faux Négatif (FN)
Classe Négative (Faux)	Faux Positif (FP)	Vrai Négatif (TN)

Du point de vue de la matrice de confusion, il est important de savoir que le modèle de la matrice de confusion pourrait avoir 36 vrais positifs, 2 faux positifs, 2 faux négatifs et 17 vrais négatifs prédits par les variables suivantes : Durée du flux, Paquets/s en aval, moyenne des temps inter-arrivées de flux, et Total des temps inter-arrivées en avant.

Par la suite, on peut expliquer la figure 16 de l'arbre de décision en utilisant le graphique ci-dessous, qui illustre comment la courbe aléatoire se rapproche de la courbe ROC.

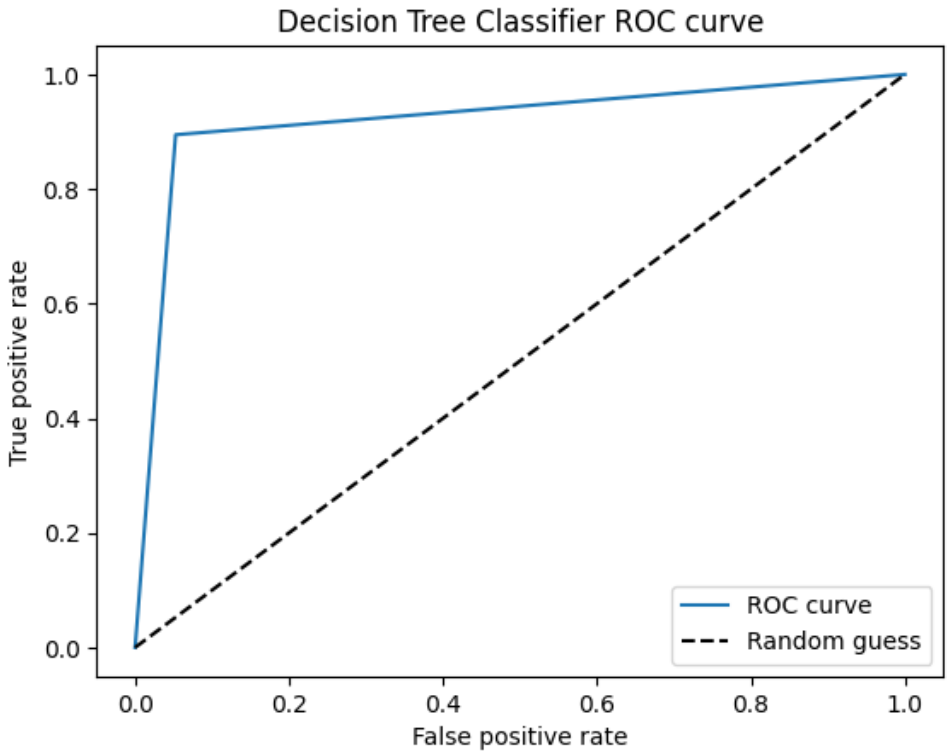


Figure 16 : La courbe de l'arbre de décision.

5.5.3 Classificateur par gradient boosting

Le classificateur par gradient boosting est une méthode d'ensemble où plusieurs modèles de faible performance, généralement des arbres de décision simples, sont combinés de manière séquentielle pour améliorer les prédictions. À chaque étape, un nouvel arbre est construit pour corriger les erreurs résiduelles des prédictions précédentes. Cela permet au modèle global de s'adapter progressivement aux données en se concentrant sur les exemples mal prédits.

5.5.3.1 Validation et test de données

```
================================================================================
          Gradient Boosting Classifier Model Cross-Validation (CV=5)
================================================================================
Accuracy: 0.978 ± 0.020
Precision_macro: 0.977 ± 0.023
Recall_macro: 0.974 ± 0.023
F1_macro: 0.975 ± 0.022
================================================================================
          Gradient Boosting Classifier Model Evaluation on Test Data
================================================================================
Accuracy:      0.965
Precision:     0.961
Recall:        0.961
F1-score:      0.961
AUC:           0.961
================================================================================
```

Figure 17 : Validation du modèle gradient boosting.

Il est important de savoir qu'en termes d'exactitude pour évaluer l'efficacité globale du modèle, on peut comprendre que l'exactitude est de 0,978, la **précision-macro** est de 0,977, le **rappel-macro** est de 0,974 et le score **F1_macro** est de 0,975, du point de vue de la validation croisée du modèle.

Cependant, pour interpréter l'évaluation du modèle sur les données de test, il est crucial de noter que son exactitude est de 0,965. La précision est de 0,961, ce qui indique que la précision des prédictions positives atteint presque 96,1%. Le rappel a une valeur de 0,961, ce qui signifie que le modèle est capable d'identifier la majorité des instances pertinentes parmi celles qui sont réellement positives. Le score F1 est également de 0,961, ce qui reflète la capacité du modèle à équilibrer précision et rappel, fournissant ainsi une mesure globale de son utilité prédictive.

En somme, ces méthodes d'évaluation offrent la possibilité de mesurer la performance du modèle à la fois lors de la validation croisée et sur des données de test différentes, offrant ainsi une vision approfondie de ses capacités prédictives dans divers contextes.

5.5.3.2 Matrice de confusion

D'après la matrice de confusion, il convient de souligner que cette partie de l'analyse explique principalement la façon dont la matrice est comprise en termes de prédiction des attaques (ATT) en se basant sur certaines variables liées à ATT. Le tableau suivant est employé afin d'analyser la matrice de confusion.

Tableau 7:Tableau d'interprétation de la matrice de confusion 3.

	Prédiction Positive	Prédiction Négative
Classe Positive (Vrai)	Vrai Positif (TP)	Faux Négatif (FN)
Classe Négative (Faux)	Faux Positif (FP)	Vrai Négatif (TN)

En interprétant la figure ci-dessous, on peut comprendre que pour la prédiction d'ATT, il y a eu 37 vrais positifs, 1 faux négatif, 1 faux positif et 18 vrais négatifs pour le modèle prédit par les variables suivantes : Durée du flux, Paquets/s en aval, moyenne des temps inter-arrivées de flux, et Total des temps inter-arrivées en avant.

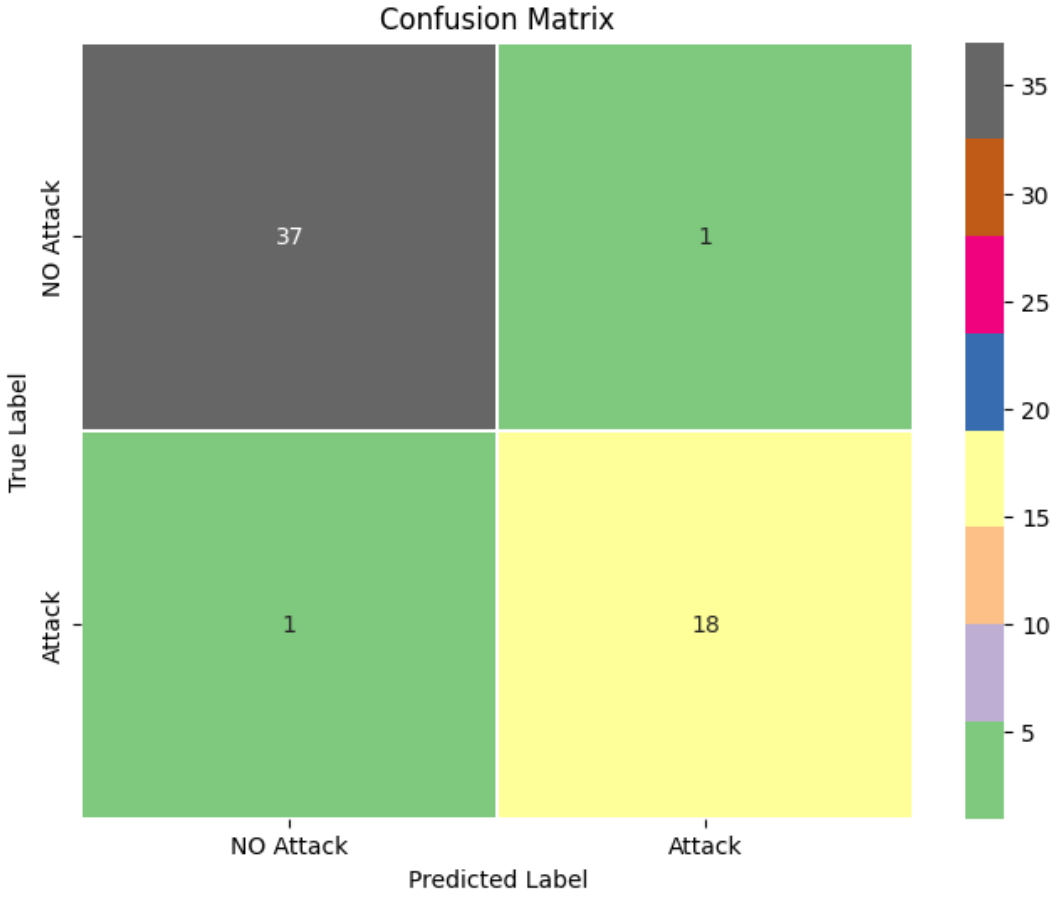


Figure 18 : Matrice de confusion du gradient boosting.

Ainsi, en ce qui concerne la compréhension de la courbe ROC, nous nous interrogeons sur la proximité de la courbe ROC pour la prédiction, ce qui n'est pas le cas pour le graphique ci-dessous. Donc, la prédiction d'ATT est bonne.

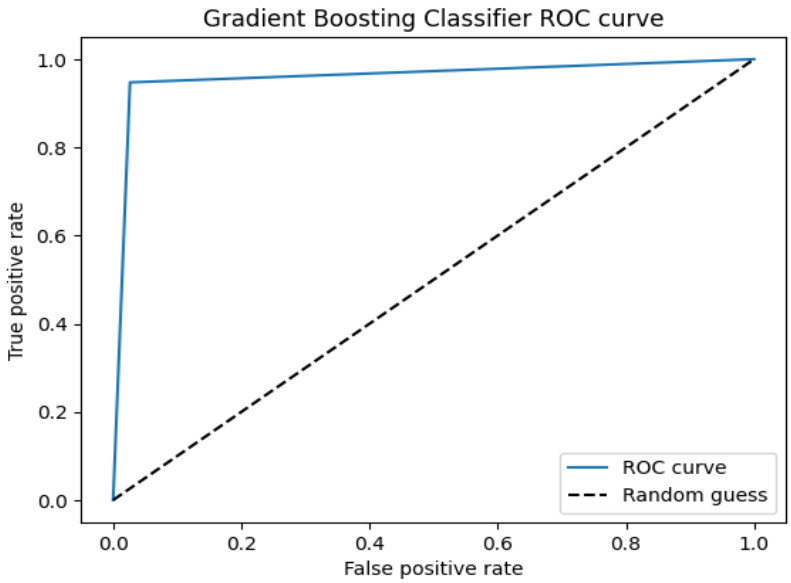


Figure 19 : Courbe du gradient boosting.

5.5.4 Classificateur par forêt aléatoire

Leo Breiman et Adele Cutler ont développé un algorithme d'apprentissage automatique connu sous le nom de forêt aléatoire, qui combine les résultats de plusieurs arbres de décision pour obtenir un résultat unique. Son adoption est due à sa simplicité d'utilisation et à sa souplesse, car il aborde à la fois les problèmes de classification et de régression.

Du point de vue du classificateur de forêt aléatoire, il est important de mettre en place le modèle qui peut être interprété.

5.5.4.1 Validation et test de données

```
================================================================================
             Random Forest Classifier Model Cross-Validation (CV=5)
================================================================================
Accuracy: 0.973 ± 0.017
Precision_macro: 0.974 ± 0.020
Recall_macro: 0.967 ± 0.018
F1_macro: 0.970 ± 0.019
================================================================================
             Random Forest Classifier Model Evaluation on Test Data
================================================================================
Accuracy:       0.947
Precision:      0.947
Recall:         0.934
F1-score:       0.940
AUC:            0.934
================================================================================
```

Figure 20 : Validation du modèle foret aléatoire.

En examinant la validation croisée du modèle de classificateur, il est démontré que l'exactitude, qui reflète l'efficacité globale du modèle, montre une exactitude de 0,973, une précision_macro de 0,974, un rappel_macro de 0,967 et un score F1_macro de 0,970.

L'exactitude du modèle est de 0,947. Pour la précision, cela indique que la précision des prédictions positives est de 0,947, soit 94,7%. Le rappel est de 0,934, ce qui signifie que le modèle est capable d'identifier la plupart des instances pertinentes, en tenant compte du coût des erreurs de classification (faux positifs et faux négatifs). Pour le score F1, la prédiction est de 0,940, ce qui est utile et bon en termes d'équilibre entre la précision et le rappel.

En résumé, ces mesures d'évaluation permettent de quantifier la performance du modèle de différentes manières :

- L'exactitude (Accuracy) mesure la proportion totale de prédictions correctes.
- La précision (Precision) mesure la proportion des vraies prédictions positives parmi toutes les prédictions positives, ce qui est crucial lorsque le coût des faux positifs est élevé.
- Le rappel (Recall) mesure la proportion des vraies prédictions positives parmi toutes les instances réellement positives, important lorsque le coût des faux négatifs est élevé.
- Le score F1 est la moyenne harmonique de la précision et du rappel, fournissant une mesure équilibrée qui prend en compte à la fois les faux positifs et les faux négatifs.

Ces mesures, obtenues à la fois par validation croisée et par évaluation sur des données de test, fournissent une compréhension détaillée des capacités prédictives du modèle dans différents contextes et aident à évaluer sa robustesse et sa fiabilité.

En examinant la validation croisée du modèle de classificateur, il est démontré que la précision, qui montre l'efficacité globale du modèle, indique une précision de 0,973, une précision_macro de 0,974, un rappel_macro de 0,967 et un F1_macro de 0,970. La précision du modèle est de 0,947. Pour la précision, cela indique que l'exactitude des prédictions positives est de 0,947, soit 94,7 %. Un rappel de 0,934 est basé sur l'identification des instances pertinentes et le coût des erreurs positives et négatives. Pour le score F1, la prédiction est de 0,940, ce qui est utile et bon en termes d'équilibre entre précision et rappel.

5.5.4.2 Matrice de confusion

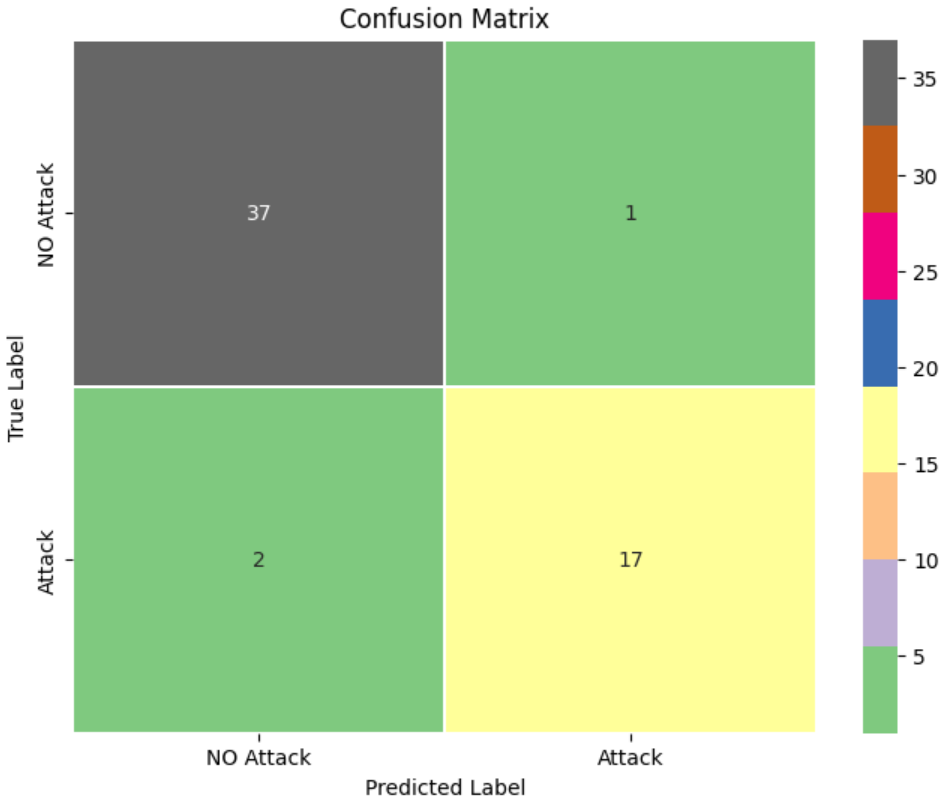


Figure 21 : : Matrice de confusion du modèle forêt aléatoire.

En analysant la matrice de confusion, il est possible de constater que le tableau ci-dessus présente 37 affirmations positives pour l'absence d'attaque, 2 affirmations négatives pour l'attaque, 1 affirmation négative pour l'absence d'attaque et 17 affirmations négatives pour l'attaque prédiquées par les variables suivantes : Flow Duration, Fwd Pkts/s, Flow IAT Mean, Fwd IAT Tot.

Tableau 8:Tableau d'interprétation de la matrice de confusion 4.

	Prédiction Positive	Prédiction Négative
Classe Positive (Vrai)	Vrai Positif (TP)	Faux Négatif (FN)
Classe Négative (Faux)	Faux Positif (FP)	Vrai Négatif (TN)

En outre, la courbe ROC du classificateur par forêt aléatoire repose sur la comparaison entre la prédiction aléatoire du modèle et le modèle de classificateur par forêt aléatoire lors des tests et de la validation en ce qui concerne la précision et la prédiction. Toutefois, cela ne

s'applique pas au classificateur par forêt aléatoire de cette étude, comme on le peut observer comme suit :

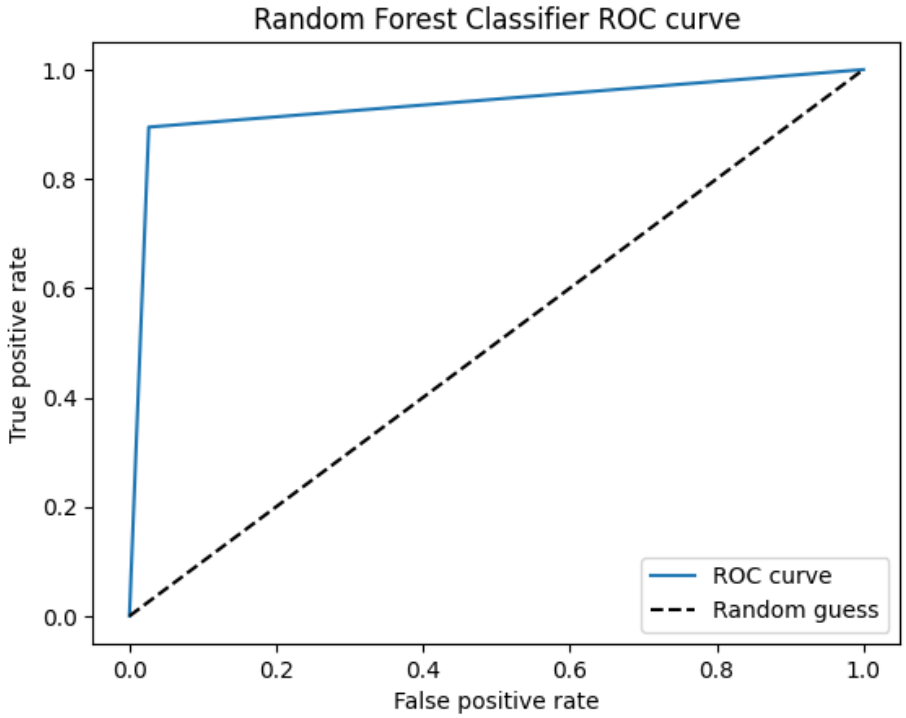


Figure 22 : Courbe ROC du modèle forêt aléatoire.

5.5.5 Comparaison des modèles algorithmiques

Les performances de la régression logistique sont modestes, avec une précision et un score F1 plutôt faibles par rapport aux autres modèles. Les performances du classificateur par arbre de décision sont excellentes, avec des scores élevés en précision, rappel et score F1, ainsi qu'une validation croisée très solide. La performance du classificateur par gradient boosting est remarquable, avec des scores très élevés sur tous les critères, et une validation croisée confirmant sa solidité. Les performances du classificateur par forêt aléatoire sont également excellentes, avec une précision légèrement inférieure à celle du classificateur par gradient boosting, mais il reste une option solide avec une validation croisée élevée. Le résumé des résultats est ci-dessous :

Tableau 9:Comparaison des algorithmes.

Model Name	Accuracy Score	Precision	Recall	F1-score	CV Accuracy	CV Precision	CV Recall	CV F1-score
Logistic Regression	0.67	0.59	0.54	0.51	0.70	0.69	0.59	0.58
Decision Tree Classifier	0.93	0.92	0.92	0.92	0.98	0.97	0.98	0.98
Gradient Boosting Classifier	0.96	0.96	0.96	0.96	0.98	0.98	0.97	0.98
Random Forest Classifier	0.95	0.95	0.93	0.94	0.97	0.97	0.97	0.97

Le classificateur par gradient boosting et le classificateur par forêt aléatoire sont les modèles les plus efficaces parmi les quatre évalués, avec des scores de précision, de rappel et de score F1 très élevés. Les performances du classificateur par arbre de décision sont également très élevées. Malgré son efficacité dans certains contextes, la régression logistique présente des performances moindres par rapport aux autres modèles dans cette évaluation.

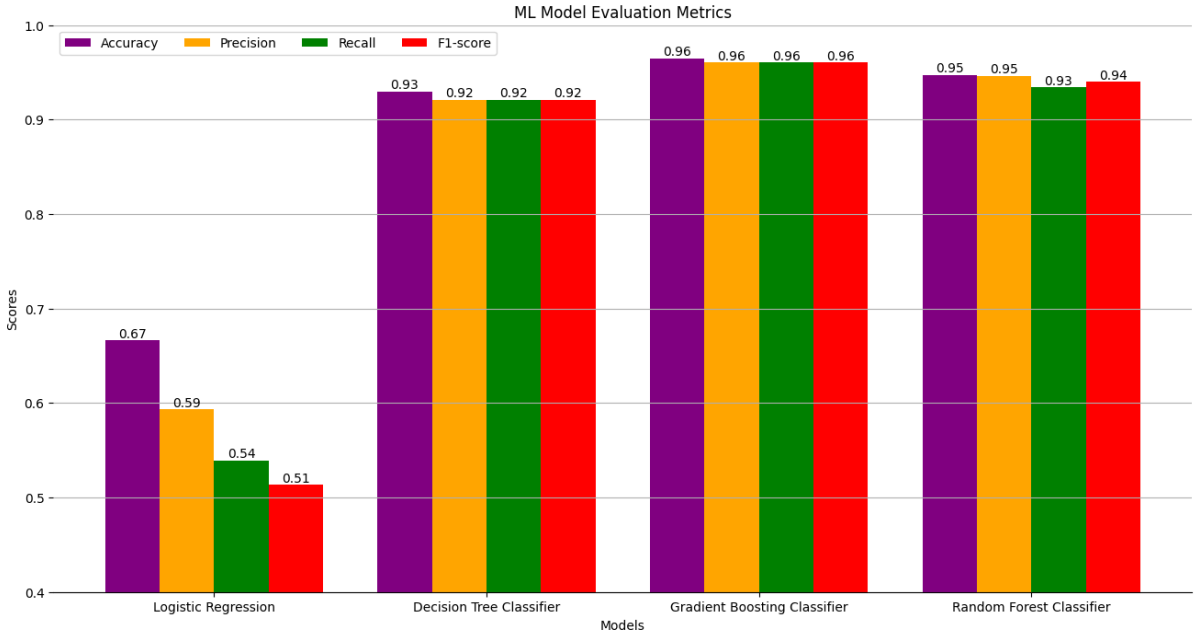


Figure 23 : Comparaison des algorithmes.

5.6 Conclusion

L'étude des données concernant l'analyse et les risques d'attaques contre les véhicules électriques met en évidence des obstacles importants à surmonter. La sécurité des véhicules repose sur l'utilisation de méthodes d'authentification solides, tandis que l'analyse descriptive des données permet de repérer les anomalies et les tendances. Les algorithmes avancés de l'apprentissage machine sont un outil précieux pour renforcer ces méthodes d'authentification

et éviter les attaques. En associant ces méthodes, il est envisageable de concevoir des systèmes plus sécurisés et plus fiables pour les véhicules électriques, garantissant ainsi une protection accrue contre les éventuels dangers.

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

6.1 Conclusion générale

Cette étude fournit une analyse non exhaustive des véhicules électriques en intégrant l'infrastructure des réseaux de neurones à l'aide d'algorithmes d'intelligence artificielle. L'examen de l'histoire des véhicules électriques a permis de mettre en lumière les progrès réalisés et les défis surmontés au fil des années. En abordant les réseaux de neurones, l'étude a fourni une vue d'ensemble des concepts fondamentaux et a identifié les problèmes actuels, tels que le surapprentissage et les vulnérabilités face aux cyberattaques. Nous avons présenté à travers une la revue de la littérature les études sur le V2G et spécialement la sécurité de ce réseau. Ensuite, nous avons donné une explication de la méthodologie qui a été établie avec des algorithmes d'apprentissage automatique tels que la régression logistique, la classification par arbre de décision, le gradient boosting et la forêt aléatoire, ensuite, nous avons présenté une explication comparative de ces méthodes. Enfin, nous avons présenté une comparaison des résultats obtenus afin que nous puissions comprendre la précision, le rappel et le score F1 des attaques sur les véhicules électriques V2G.

La méthodologie de recherche a été conçue en utilisant divers algorithmes d'apprentissage automatique, chacun apportant des perspectives uniques pour la détection d'anomalies et la prévention des intrusions dans les réseaux V2G. Les algorithmes de régression logistique, de classification par arbre de décision, de gradient boosting et de forêt aléatoire ont été évalués et comparés pour leur efficacité dans la détection des cyberattaques. Les résultats de l'étude ont montré des niveaux variés de précision, de rappel et de score F1, soulignant l'importance de choisir le bon algorithme en fonction du contexte spécifique des véhicules électriques V2G.

En conclusion, cette recherche a fourni des recommandations pratiques pour améliorer la sécurité et la résilience des systèmes V2G. Les recommandations issues de cette recherche visent à renforcer la détection des anomalies et à prévenir les cyberattaques potentielles, assurant ainsi un avenir plus sûr et plus fiable pour les véhicules électriques connectés.

6.2 Recommandations et travaux futurs

Pour assurer la sécurité des véhicules électriques (VE) dans un environnement connecté tel que le réseau V2G, plusieurs mesures essentielles sont à mettre en œuvre. Il est primordial de renforcer les mécanismes d'authentification et de chiffrement afin de garantir la confidentialité et l'intégrité des communications entre les véhicules, les stations de recharge et les infrastructures centrales. La mise en place de l'authentification multi-facteurs et de protocoles de chiffrement avancés permet de protéger efficacement les accès aux systèmes critiques.

Par ailleurs, il est recommandé de développer et de déployer des systèmes de détection d'intrusion (IDS) intelligents, reposant sur l'apprentissage automatique et l'intelligence artificielle, pour surveiller en temps réel les communications et repérer des attaques comme le déni de service (DoS), le spoofing ou les botnets. Ces systèmes doivent être distribués à plusieurs niveaux du réseau (VE, stations de recharge, agrégateurs) pour une couverture étendue.

L'adoption de mécanismes de détection hybrides, combinant l'analyse logicielle et l'évaluation des impacts physiques (tension, stabilité de la charge), est également essentielle pour une détection précoce et précise des menaces.

Enfin, il convient de maintenir une surveillance continue, de mettre à jour régulièrement les logiciels pour corriger les vulnérabilités, et de segmenter le réseau des VE afin de limiter la propagation des attaques et faciliter la gestion des risques. Ces actions coordonnées sont indispensables pour garantir la résilience et la fiabilité des véhicules électriques dans un contexte de cybersécurité en constante évolution

BIBLIOGRAPHIE

[1] R. S. S. &. A. R. Gupta, "Electric vehicles and climate change mitigation: A review of Vehicle-to-Grid systems," *Energy Policy and Sustainability Journal,* vol. 8, no. 3, p. 89–103, 2021.

[2] P. F. T. &. J. K. Dik, "Smart charging and energy flexibility in Vehicle-to-Grid (V2G) systems," *Energy Systems Integration Journal,* vol. 14, no. 1, p. 45–59, 2022.

[3] M. R. L. &. B. F. Cozzi, "Smart charging and Vehicle-to-Grid technology: An analysis of cost-effective solutions," *International Journal of Energy Economics and Policy,* vol. 10, no. 4, p. 67–74, 2020.

[4] Z. L. M. &. Z. Y. Shang, "Exploring the energy storage and bidirectional flow potential of V2G-enabled electric vehicles," *Applied Energy,* vol. 309, p. 118–134., 2022.

[5] R. S. A. &. K. R. Boodoo, "Integration of Vehicle-to-Grid (V2G) technology with renewable energy for sustainable energy systems," *Journal of Renewable Energy Research,* vol. 12, no. 3, p. 123–145, 2024.

[6] C. Strom, "The future of transportation: Electric vehicles and their impact on sustainable development," *Journal of Modern Transportation Studies,* vol. 17, no. 1, p. 56–74, 2024.

[7] M. &. R. R. Devanandanan, "Addressing cybersecurity vulnerabilities in electric vehicles: A focus on Vehicle-to-Grid systems," *Cybersecurity and Energy,* vol. 15, no. 2, p. 201–217, 2024.

[8] P. J. N. &. P. R. Malimage, "Cybersecurity challenges in Vehicle-to-Grid systems: Emerging threats and mitigation strategies," *Cybersecurity Advances,* vol. 18, no. 4, p. 123–137, 2023.

[9] P. A. R. &. M. V. Hamdare, "Securing the future of smart transportation: Cybersecurity measures for electric vehicles," *Transportation Security Journal,* vol. 9, no. 1, p. 34–48, 2023.

[10] P. &. L. T. Lulia, "The Evolution of Electric Vehicles: A Historical Overview," *Journal of Transport History,* vol. 43, no. 1, p. 25–42, 2022.

[11] R. Willings, "How Policies Shaped the Electric Vehicle Revolution," *Journal of Energy and Environmental Policy,* vol. 5, no. 2, p. 89–104, 2017.

[12] B. Schmidt, "The Role of Vehicle-to-Grid (V2G) in the Energy Transition," *Energy & Environment Journal,* vol. 31, no. 8, p. 1447–1463, 2020.

[13] C. A. F. D. &. R. R. Robledo, "Vehicle-to-grid integration: Benefits and technical challenges," *Energy Policy Journal,* vol. 38, no. 5, p. 753–762, 2018.

[14] S. &. G. M. A. Liasi, "Vehicle-to-grid (V2G): Perspectives, challenges, and key enablers for sustainable energy integration," *Energy Conversion and Management,* vol. 151, p. 473–484, 2017.

[15] K. &. Q. X. Liu, "Emerging Trends in Vehicle-to-Grid (V2G) Technology," *IEEE Transactions on Smart Grid,* vol. 3, no. 2, p. 1205–1213, 2012.

[16] S. Masood, "Foundations of Neural Networks and Their Applications in Artificial Intelligence," Springer, 2023.

[17] D. W. J. &. P. R. Hollings, "Deep Learning Essentials: A Comprehensive Overview of Neural Networks," Cambridge University Press, 2023.

[18] S. K. M. A. M. R. &. K. G. Dargan, "A survey of deep learning and its applications: A new paradigm to machine learning," *Archives of Computational Methods in Engineering,* vol. 27, no. 4, p. 1071–1092, 2020.

[19] S. K. T. &. R. A. Ahmad, "Supervised and Unsupervised Learning: A Detailed Comparison and Applications," *Journal of Artificial Intelligence Research,* vol. 45, no. 2, p. 178–195, 2023.

[20] J. L. Y. &. Z. S. Dong, "Unsupervised Learning and Its Applications in Data Analytics," *Proceedings of the IEEE International Conference on Big Data,* vol. 15, no. 3, p. 341–348, 2021.

[21] S. &. A. A. Nawaz, "Functions of Activation in Artificial Neural Networks: Recent Developments," *Journal of Computer Science and Applications,* vol. 49, no. 4, p. 301–320, 2023.

[22] R. M. P. &. K. R. Ramambason, "Convolutional Neural Networks: Theory, Design, and Application," Elsevier Academic Press, 2022.

[23] C. I. W. G. A. &. M. S. Nwankpa, "Activation functions: Comparison of trends in practice and research for deep learning," *Neural Processing Letters,* vol. 53, no. 3, p. 1803–1828, 2021.

[24] S. &. A. A. Sarp, "Big Data and Neural Network Applications in Electric Vehicles: A Review," *Renewable Energy and Smart Systems,* vol. 14, no. 8, pp. 102-115, 2020.

[25] X. L. J. &. Z. Y. Wei, "Intrusion Detection Systems in Connected Vehicles: Techniques and Challenges," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 5, pp. 3051-3060, 2022.

[26] S. &. R. V. Kabilan, "Unsupervised intrusion detection on intrusion detection datasets for V2G systems," *Journal of Artificial Intelligence in Cybersecurity,* vol. 5, no. 2, pp. 199-210, 2024.

[27] T. Yoshoka, "IoT-Based Electric Vehicle Charging Stations and Their Cybersecurity Challenges," *Smart City Journal,* vol. 10, no. 3, pp. 55-67, 2022.

[28] A. K. M. &. B. N. El Houda, "Exploration of Open Charge Point Protocol (OCPP) and Its Role in Electric Vehicle Charging Systems," *International Journal of IoT Systems,* vol. 1, no. 12-20, p. 11, 2022.

[29] P. Kovacs, "ISO 26262 and Automotive Cybersecurity: A Focus on Safety in Electric Vehicles," *International Journal of Automotive Safety,* vol. 17, no. 6, pp. 34-49, 2023.

[30] A. Dhingra, "Cybersecurity risks in electric vehicle networks: A growing concern," *Network Security Insights,* vol. 12, no. 4, p. 101–118, 2023.

[31] J. Banda, "Cybersecurity challenges and solutions for EV infrastructure," *Energy Systems Research Journal,* vol. 15, no. 2, pp. 115-129, 2022.

[32] S. &. S. K. Suriya, "Analyzing threats in EV charging networks," *Journal of IoT Security,* vol. 10, no. 5, p. 98–112, 2022.

[33] M. Z. N. &. W. K. Elkashlan, "Cybersecurity in Electric Vehicles and Vehicle-to-Grid Communication: Emerging Threats and Countermeasures," *IEEE Vehicular Technology Magazine,* vol. 18, no. 2, pp. 50-59, 2023.

[34] P. Riley, "Digital Vulnerabilities in Electric Vehicle Ecosystems: Emerging Cyber Threats and Solutions," *Cybersecurity & IoT Journal,* vol. 9, no. 5, pp. 92-102, 2023.

[35] C. &. I. G. Comi, "Strategic implementation and accurate forecasting of V2G technology," *Electric Vehicles Journal,* vol. 36, no. 1, p. 77–92, 2024.

[36] B. L. J. &. Z. S. Wang, "Certificateless Privacy Protection Scheme based on Fog Computing in V2G Network," *Frontiers in Science and Engineering,* vol. 1, no. 2, p. 200–215, 2021.

[37] T. L. Z. &. Z. H. Wang, "Efficient cryptographic solutions for secure V2G communications," *IEEE Transactions on Vehicular Technology,* vol. 73, no. 4, p. 1201–1212, 2024.

[38] K. H. S. &. A.-F. A. Abdallah, "A lightweight V2G connection scheme ensuring confidentiality and security during electricity exchange sessions," *Journal of Computer Security,* vol. 24, no. 3, p. 243–263, 2016.

[39] A. (. Saxena, "A bilinear pairing-based scheme for V2G security," *Information Security Journal,* vol. 19, no. 6, p. 145–156, 2016.

[40] A. S. M. &. G. R. Saxena, "Secure architecture for V2G networks: Privacy-preserving applications," *Smart Grid Security Journal,* vol. 12, no. 3, p. 67–84, 2016.

[41] R. &. I. S. Novak, "Countermeasures against cybersecurity threats in V2G networks," *Journal of Cybersecurity,* vol. 25, no. 4, p. 299–314, 2020.

[42] J. C. L. &. W. X. Yang, "Comprehensive V2G models for secure power grid operations," *Renewable Energy Journal,* vol. 32, no. 9, p. 200–219, 2024.

[43] L. C. R. &. Z. F. Yan, "Federated learning with blockchain for privacy protection in V2G," *IEEE Transactions on Smart Grid,* vol. 15, no. 2, p. 530–543, 2024.

[44] M. T. B. &. W. J. Stumberg, "Demonstrating V2G for frequency contingency in national power grids," *Energy Research Letters,* vol. 45, no. 1, p. 103–118, 2024.

[45] A. Timm, "Smart charging algorithms for household electricity optimization," *Energy Efficiency Journal,* vol. 18, no. 7, p. 55–73, 2023.

[46] M. Cao, "Balancing the grid with V2G technology: Energy quality control and grid stability," *Energy Reports,* vol. 9, pp. 289-302, 2023.

[47] K. E. K. H. &. M. A. Mekkaoui, "Biology-inspired intrusion detection system (IDS) based on machine learning for predicting and mitigating attacks on V2G networks," *International Journal of Computer Science and Information Security,* vol. 21, no. 8, pp. 101-115, 2023.

[48] H. &. K. H. Kang, "A deep neural network-based intrusion detection approach for in-vehicle networks," *Journal of Electrical Engineering & Technology,* vol. 11, no. 3, pp. 824-834, 2016.

[49] Y. Z. X. &. Z. Z. Shang, "Federated deep learning and distributed edge computing for secure planning in V2G systems," *IEEE Access,* vol. 12, pp. 56442-56453, 2024.

[50] I. A.-D. M. &. A.-N. A. Al-Jarrah, "Intrusion detection systems for in-vehicle networks: A review.," *Journal of Cyber Security Technology,* vol. 3, no. 1, pp. 38-59, 2019.

[51] Z. &. M. W. Warraich, " Evaluating the effectiveness of intrusion detection in V2G networks with varying time resolutions," *IEEE Transactions on Vehicular Technology,* vol. 72, no. 4, pp. 2051-2062, 2023.

[52] R. B. L. &. R. B. Kern, "Enhancing intrusion detection in V2G systems using regression and novelty detection methods," *Cybersecurity and Digital Trust Journal,* vol. 6, no. 2, pp. 103-115, 2023.

[53] F. Richert, "Challenges in protecting microgrids from cyberattacks with EV charging stations," *IEEE Transactions on Smart Grid,* vol. 12, no. 5, pp. 4167-4176, 2021.

[54] T. &. H. H. Nguyen, "Vulnerabilities and risks in electric vehicle charging infrastructure: A study of cyber-physical attacks," *Journal of Electric Power Systems Research,* vol. 181, p. 106184, 2020.

[55] Y. K. B. &. C. M. Seo, "GIDS: A GAN-based intrusion detection system for in-vehicle networks," *Journal of Communications and Networks,* vol. 20, no. 3, pp. 262-272, 2018.

[56] W. Z. C. &. X. X. Wang, "Malware propagation from EVs and charging stations to the electrical grid: A probabilistic approach," *International Journal of Electrical Power & Energy Systems,* vol. 68, pp. 246-255, 2015.

[57] L. &. F. S. Falk, " Impact of botnet attacks on power grid systems: Neural network-based detection," *IEEE Transactions on Industrial Informatics,* vol. 8, no. 5, pp. 1089-1097, 2012.

[58] S. L. J. &. S. H. Hong, "Impact of botnet attacks on distribution systems with EVs and fast charging stations," *IEEE Transactions on Power Delivery,* vol. 29, no. 1, pp. 220-228, 2024.

[59] T. &. G. R. Reed, "Power fingerprinting for detection of malicious activities in industrial control systems," *International Journal of Critical Infrastructure Protection,* vol. 5, no. 4, pp. 244-255, 2012.

[60] J. L. S. &. K. M. Kwon, "Behavior-based intrusion detection algorithm for IEC 61850-based substations," *Journal of Electrical Engineering & Technology,* vol. 10, no. 3, pp. 938-948, 2015.

[61] J. L. C.-C. &. G. M. Hong, "Integrated anomaly detection for cyber security of the substations," *IEEE Transactions on Smart Grid,* vol. 5, no. 4, pp. 1643-1653, 2014.

[62] S. &. Z. T. Yang, "Cyberattack consequences on SCADA systems via ARP spoofing protocol falsification," *Cybersecurity Technology & Applications,* vol. 3, no. 6, pp. 45-56, 2012.

[63] U. S. J. S. T. S. &. B. R. Premaratne, "An intrusion detection system for IEC61850 automated substations," *IEEE Transactions on Power Delivery,* vol. 25, no. 4, pp. 2376-2383, 2010.

[64] A. &. M. A. Moghadasi, "Responding to emerging security threats in EV charging and V2G networks," *International Journal of Cyber Security and Digital Forensics,* vol. 11, no. 2, pp. 142-157, 2023.

[65] Y. Sun, "Group anonymous authentication scheme for V2G communications," *IEEE Transactions on Vehicular Technology,* vol. 66, no. 4, pp. 2708-2718, 2017.

[66] A. P. R. &. Y. T. Kabilan, "Intrusion Detection Systems: Role in Automotive Cybersecurity," *Advances in Cybersecurity,* vol. 20, no. 4, pp. 78-90, 2024.

[67] R. &. H. A. Basnet, " Deep learning-based intrusion detection system for detecting DoS attacks in EV charging stations," *Proceedings of the SPIE: Applications of Artificial Intelligence,* vol. 2020, pp. 1-9, 2020.

[68] R. B. L. &. R. B. Kern, "Enhancing intrusion detection in V2G systems using regression and novelty detection methods," *Cybersecurity and Digital Trust Journal,* vol. 6, no. 2, pp. 103-115, 2023.

[69] D. Kheireddine, "Cockroach behavior-based intrusion detection for V2G networks," *Energy & Security Journal,* vol. 13, no. 1, pp. 34-47, 2024.

[70] A. &. Z. F. Al-Mehdhar, "HADRL: Hierarchical antagonistic deep reinforcement learning for detecting covert cyberattacks on EV charging stations," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 34, no. 2, pp. 457-469, 2023.

[71] A. &. Z. X. ElKashlan, "Machine learning classification algorithms for DDoS attack detection in EVCS network environments," *Journal of Cybersecurity,* vol. 24, no. 7, pp. 15-27, 2020.

[72] R. &. H. A. Basnet, "Multi-class classification of network flows for intrusion detection using deep learning techniques," in *Proceedings of the 18th International Conference on Machine Learning and Applications (ICMLA)*, 1-8, 2019.

[73] Z. L. Y. &. Z. Y. Yin, "RNN-based intrusion detection system for V2G networks," *International Journal of Cyber Security and Digital Forensics,* vol. 11, no. 3, pp. 123-135, 2022.

[74] M. I. &. M. T. M. Jordan, "Machine learning: Trends, perspectives, and prospects," *Science,* vol. 349, no. 6245, pp. 255-260, 2015.

[75] G. Andrew, "Statistics: A comprehensive guide," Springer, 2017.

[76] S. L. X. &. C. L. Simon, "Data visualization with correlation heatmaps," *International Journal of Data Science,* vol. 3, no. 2, pp. 45-56, 2014.

[77] D. &. D. P. Palei, "Logistic regression in machine learning: A survey of applications," *International Journal of Computer Applications,* vol. 16, no. 4, pp. 12-19, 2009.

[78] L. Y. D. &. P. J. C. Deng, "A comprehensive review of deep learning techniques for speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 19, no. 1, pp. 1-14, 2011.

[79] R. Anita, "Random forests: Methods and applications in machine learning," Cambridge University Press, 2007.

[80] D. W. J. &. P. R. Hollings, "Deep Learning Essentials: A Comprehensive Overview of Neural Networks.," Cambridge University Press., 2023.

[81] J. L. Y. &. Z. S. Dong, "Unsupervised Learning and Its Applications in Data Analytics," *Proceedings of the IEEE International Conference on Big Data,* vol. 15, no. 3, p. 341–348, 2021.

[82] C. I. W. G. A. &. M. S. Nwankpa, "Activation functions: Comparison of trends in practice and research for deep learning," *Neural Processing Letters,* vol. 53, no. 3, p. 1803–1828, 2021.

[83] T. Yoshoka, "IoT-Based Electric Vehicle Charging Stations and Their Cybersecurity Challenges," *Smart City Journal,* vol. 10, no. 3, pp. 55-67, 2022.

[84] R. B. L. &. R. B. Kern, "Enhancing intrusion detection in V2G systems using regression and novelty detection methods," *Cybersecurity and Digital Trust Journal,* vol. 6, no. 2, pp. 103-115, 2023.

[85] M. &. P. E. Giannelos, " F-factor methodology for quantifying the contribution of V2G to supply security," *IEEE Transactions on Smart Grid,* vol. 15, no. 1, pp. 374-384, 2024.

[86] E. &. S. W. Kang, "Cyber-attacks on electric vehicle batteries via smartphones: A review of risks and preventive measures," *Journal of Cybersecurity and Privacy,* vol. 3, no. 2, pp. 45-58, 2019.

[87] X. L. Y. Z. W. &. C. Z. Han, "Experimental verification system for secure EV charging: Hardware and software solutions," *Journal of Energy Systems,* vol. 39, no. 2, p. 125–138, 2024.

[88] X. L. Y. Z. W. &. C. Z. Han, "Experimental verification system for secure EV charging: Hardware and software solutions," *Journal of Energy Systems,* vol. 39, no. 2, p. 125–138, 2024.

[89] J. L. M. Z. H. &. X. T. Wang, "Privacy protection scheme for V2G networks using fog computing and certificateless public key encryption," *International Journal of Smart Grid Technologies,* vol. 12, no. 1, p. 54–72, 2024.