Université Paris 1 Panthéon-Sorbonne

MIAGE Sorbonne — Réseau des MIAGE de France

Mémoire de Master 1 MIAGE

Soutenu le : 2 Avril 2024

Anthony RODRIGUES

Développement de Techniques d'Intelligence Artificielle Explicables pour Améliorer la Détection des Intrusions dans les Réseaux de Véhicules VANET

Sous la direction de : Nourhène Ben Rabah

Année d'étude : 2023/2024

Jury : Daniel Diaz

Remerciements

Je tiens à consacrer cette section aux remerciements pour clore de manière appropriée le présent mémoire de recherche. Je souhaite exprimer ma profonde gratitude envers les membres du corps enseignant de la MIAGE de Paris 1 Panthéon Sorbonne. Leur expertise académique, leur passion pour la transmission du savoir et leur rigueur ont été les fondations sur lesquelles s'est construit mon apprentissage. Leur approche pédagogique a non seulement enrichi mes connaissances mais aussi aiguisé ma curiosité intellectuelle.

Un merci tout particulier à ma directrice de mémoire, Madame Nourhène Ben Rabah, pour sa patience remarquable et son désir contagieux de partager sa passion pour la recherche. Ses orientations précieuses et sa capacité à éveiller ma curiosité ont été le moteur de ce mémoire. Son accompagnement a été pour moi une réelle source d'inspiration.

Je souhaite également témoigner de ma gratitude envers mon tuteur en entreprise, Monsieur Emmanuel Huet, pour son accompagnement patient et attentif. Sa méthodologie et son expertise m'ont permis d'acquérir une rigueur professionnelle et une approche méthodique indispensables. Sa présence bienveillante a grandement contribué à mon épanouissement professionnel.

À ma famille et à mes proches, dont le soutien sans faille et l'encouragement constant ont été le pilier de ma persévérance. Leur confiance et leur amour m'ont donné la force nécessaire pour avancer et relever les défis.

Ce mémoire marque une étape clé de mon parcours en Master MIAGE, et il me tient à cœur de souligner l'importance de chaque contribution qui a rendu ce travail possible. Merci à toutes et à tous pour votre soutien, votre confiance et pour avoir été à mes côtés lors de cette aventure académique.

Table des matières

Abstract

This research explores the convergence of artificial intelligence (AI) and Vehicular Ad-Hoc Networks (VANETs), with a particular focus on the impact and potential of Explainable Artificial Intelligence (XAI) in improving intrusion detection. With the rise of VANETs, securing these networks becomes a major concern, posing complex challenges in terms of cybersecurity. Therefore, this study aims to analyze to what extent XAI techniques can help address these challenges by making the decision-making processes of AI systems not only more precise but also understandable and transparent to users.

To achieve this, we rely on a comprehensive analysis of various XAI methodologies, including LIME and SHAP, applied to the specific context of VANETs. Our goal is to determine how these techniques can facilitate the detection of malicious activities within networks while ensuring clear and accessible interpretation of decisions made by AI systems. Through a systematic exploration of the literature and the analysis of concrete cases, this research highlights the advantages and limitations of each approach, while examining their applicability in the dynamic and complex environment of VANETs.

Key words

Artificial Intelligence (IA) – Machine Learning (ML) — Explainability - Explainable AI (XAI) – Vehicular Ad-Hoc Network (VANET) - Intrusion Detection

Résumé

Cette recherche explore la convergence de l'intelligence artificielle (IA) et des réseaux de véhicules ad hoc (VANETs), avec un focus particulier sur l'impact et le potentiel de l'Intelligence Artificielle Explicable (XAI) dans l'amélioration de la détection des intrusions. Face à l'essor des VANETs, la sécurisation de ces réseaux devient un enjeu majeur, soulevant des défis complexes en termes de cybersécurité. Cette étude se propose donc d'analyser dans quelle mesure les techniques de XAI peuvent contribuer à relever ces défis, en rendant les processus décisionnels des systèmes d'IA non seulement plus précis mais également compréhensibles et transparents pour les utilisateurs.

Pour ce faire, nous avons recours à une analyse approfondie de différentes méthodologies de XAI, notamment LIME et SHAP, appliquées au contexte spécifique des VANETs. Notre objectif est de déterminer comment ces techniques peuvent faciliter la détection des activités malveillantes au sein des réseaux, tout en assurant une interprétation claire et accessible des décisions prises par les systèmes d'IA. À travers une exploration systématique de la littérature et l'analyse de cas concrets, cette recherche met en lumière les avantages et les limites de chaque approche, tout en examinant leur applicabilité dans l'environnement dynamique et complexe des VANETs.

Mots clés

Intelligence Artificielle (IA) – Apprentissage Automatique (ML) – Explicabilité - IA Explicable (XAI) – Réseau Ad-Hoc Véhiculaire (VANET) – Détection d’intrusion

Introduction

A. Contexte

L'histoire de la mobilité humaine est marquée par une quête incessante d'innovation, visant à rendre les moyens de transport toujours plus rapides, intelligents et efficaces. Dès le Moyen Âge, le cheval se présentait comme le symbole ultime de vitesse et de puissance, essentiel pour les longs voyages. Avec le temps, l'évolution technologique nous a menés des carrosses aux véhicules à vapeur de la Révolution industrielle, révolutionnant ainsi le transport de marchandises et de personnes.

Cette évolution s'est accélérée de manière spectaculaire au fil des siècles. La création de la première voiture à essence a ouvert la voie à l'industrialisation et à la démocratisation de l'automobile, touchant toutes les régions du monde. De nos jours, l'innovation technologique a transformé les véhicules en machines extrêmement sophistiquées, capables de se connecter entre elles pour former des réseaux de véhicules ad hoc, ou VANETs. Ces réseaux ouvrent de nouvelles possibilités pour améliorer la sécurité et la fiabilité des systèmes de transport.
L'avènement des véhicules autonomes représente une révolution dans le secteur, nécessitant une interconnexion pour une analyse et une décision optimale dans chaque situation. Ce système assure une sécurité renforcée grâce à l'usage de l'intelligence artificielle et du machine learning, marquant une avancée significative dans l'histoire de l'automobile. Cette évolution symbolise le début d'une nouvelle ère où la sécurité routière et la communication sont étroitement liées.

Toutefois, ces avancées technologiques ne sont pas sans défis, particulièrement en termes de cybersécurité [1]. Bien que la communication directe entre les véhicules améliore la sécurité et l'efficacité du trafic, elle introduit également de nouveaux risques, ouvrant la porte à d'éventuelles attaques malveillantes.

Face à ces défis, l'intelligence artificielle explicative (XAI) apparaît comme une solution prometteuse [2] [3]. Elle a le potentiel de détecter les intrusions tout en rendant les processus décisionnels transparents et compréhensibles, contribuant ainsi à une meilleure acceptation et confiance dans les technologies avancées de mobilité.

B. Définition de la problématique

Les réseaux de véhicules Ad-Hoc sont en pleine expansion, l'un des défis majeurs réside dans la sécurité et la détection des intrusions, essentielles pour garantir la fiabilité et la sûreté des communications entre véhicules. Cependant, la complexité et la variabilité des attaques nécessitent des solutions sophistiquées pour une identification précise et en temps réel des menaces. L'intelligence Artificielle (IA) mais plus précisément l'Intelligence Artificielle Explicable (XAI) [3], émerge comme une approche prometteuse pour apporter des méthodes capables de fournir non seulement des prédictions précises mais aussi des explications compréhensibles par tous sur le processus décisionnel.

Cependant, l'application de la XAI dans le contexte spécifique des réseaux de véhicules Ad-Hoc pose des questions liées à la dynamique et à l'échelle des réseaux, à la variabilité des données, et à la nécessité d'explications temps réel qui peuvent être intégrées de manière transparente dans les processus décisionnels. Ainsi, la question centrale de notre recherche est la suivante :

"Dans quelle mesure les techniques d'Intelligence Artificielle Explicables peuvent-elles améliorer la détection des intrusions dans les réseaux de véhicules Ad-Hoc ?"

L'objectif de ce travail est d'explorer cette question en évaluant l'efficacité de différentes techniques d'XAI. Par cette analyse, nous visons à contribuer à une meilleure compréhension de la façon dont la XAI peut être exploitée pour renforcer la sécurité dans les réseaux de véhicules Ad-Hoc, en fournissant des moyens pour une interprétation fiable et explicite des décisions prises par les systèmes d'IA.

Définition des termes clés

A. Background technique

1. Vers une mobilité intelligente : l'ère des VANETs

Un VANET, Vehicular Ad-Hoc Networks, est un réseau de véhicules sans fil qui communiquent entre eux. Ces VANET sont composés de véhicules intelligents équipés de capteurs, de processeurs et de nombreux équipements qui leur permettent d'être plus connectés et intelligents.

Les réseaux ad hoc véhiculaires (VANET) représentent une catégorie de réseaux de communication permettant des échanges d'informations entre des véhicules connectés et tout autre élément extérieur (V2X), incluant les communications véhicule-à-véhicule (V2V), véhicule-à-infrastructure (V2I), infrastructure-à-véhicule (I2V), infrastructure-à-infrastructure (I2I), ainsi que les communications intra-véhicule.

Un système VANET repose essentiellement sur deux composants majeurs [4]:

- **Autorité de Confiance (CA):** la CA est au cœur de l'authenticité de l'information dans le réseau car elle détient les identités réelles de tous les véhicules et est responsable de l'émission et de l'attribution des certificats ainsi que des pseudonymes de communication.
- **Unité en Bord de Route (Roadside Unit - RSU):** Les RSU sont installées en bord de route sous formes de feux de signalisation ou autres signalétiques. Leur rôle principal est d'assister l'Autorité de Confiance dans la gestion du trafic et des véhicules.
- **Unité Embarquée (On-Board Unit - OBU):** Les OBU sont des unités intégrées dans les véhicules intelligents, comprenant un ensemble de composants matériels et logiciels avancés tels que le GPS, le radar, des caméras, et divers capteurs. Elles facilitent la localisation, la réception, le traitement, le stockage, et l'envoi des données sur le réseau. Agissant comme des émetteurs-récepteurs, les OBU permettent la connexion du véhicule au réseau.

La RSU, en tant qu'équipement de communication situé en bord de route, fournit l'accès à Internet aux véhicules et facilite l'échange de données entre eux, jouant un rôle crucial dans le soutien de la gestion du trafic et des véhicules en collaboration avec l'Autorité de Confiance. L'OBU, unité de traitement et de communication mobile intégrée au véhicule, assure la communication avec d'autres véhicules ou avec les équipements d'infrastructure, en plus de prendre en charge des fonctions avancées comme la localisation et le traitement des données, grâce à son ensemble de technologies de pointe.

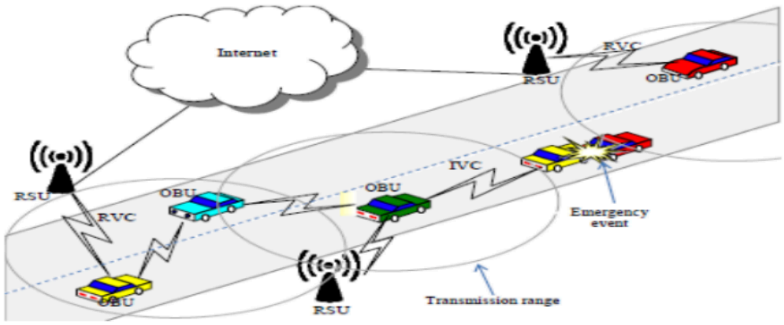

Figure 1 - Composants d'un réseau de véhicule ad-hoc

Un système VANET se compose de trois plans [4, 5] :

- Le plan véhiculaire
- Le plan RSU
- Le plan des services

Dans le plan véhiculaire, chaque véhicule est équipé d'une OBU permettant la communication V2V. Le plan RSU facilite les communications V2I, I2V, et I2I. Quant au plan des services, il permet le déploiement de différents types de services tels que la sécurité, l'infodivertissement, les paiements, l'accès à Internet, et les services basés sur le cloud.

Les VANET partagent certaines caractéristiques avec les réseaux ad hoc mobiles (MANET), comme la diffusion omnidirectionnelle, la portée de transmission courte, et la faible bande passante. Toutefois, les VANET présentent des caractéristiques spécifiques, notamment une topologie hautement dynamique due à la mobilité élevée des véhicules, ce qui entraîne des déconnexions fréquentes. De plus, les véhicules cibles peuvent être atteints en fonction de leur localisation géographique, et la propagation du signal est

influencée par l'environnement, comme les bâtiments ou les arbres. Contrairement aux MANET, les problématiques d'énergie, de stockage et de capacité de calcul sont moins critiques dans les VANET, bien que le traitement en temps réel de grandes quantités de données représente un défi majeur.

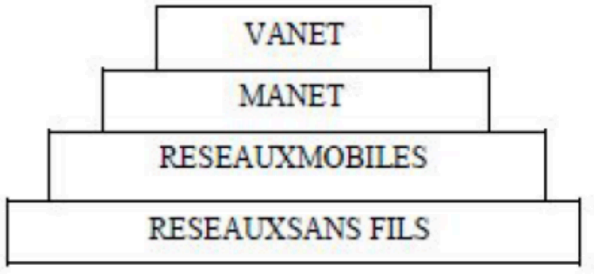

Figure 2 – Hiérarchie des réseaux sans fil

La diversité des schémas de communication et les caractéristiques inhérentes des communications sans fil rendent les VANET vulnérables à de nombreuses attaques et vulnérabilités de sécurité. Ces vulnérabilités peuvent affecter toutes les couches du réseau et tous les aspects de la communication.

2. Intrusions dans les VANETs

Les réseaux de véhicules ad hoc (VANETs) représentent une innovation cruciale pour améliorer la sécurité et l'efficience des systèmes de transport contemporains. Cependant, ils sont vulnérables à des attaques de sécurité significatives qui peuvent non seulement perturber l'opérationnalité des véhicules mais également compromettre la sécurité des passagers et des autres usagers de la route.

Caractéristiques	Attaque Sybil	DDOS	Attaque par trou noir	Spoofing	Attaque de rejeu	Attaque par révélation d'information	Attaque par perturbation de la route	Attaque par épuisement de l'énergie
Authenticité	X	X	X	X	X	X	X	X
Intégrité			X	X	X		X	
Confidentialité						X		
Disponibilité		X	X	X			X	X
Accessibilité		X	X	X			X	X

Tableau 1 - Caractéristiques de certaines attaques de sécurité VANET [5]

Parmi les nombreuses intrusions possibles, trois se distinguent par leur prévalence et leur potentiel de nuisance [6]: l'attaque Sybil, l'attaque par déni de service distribué (DDOS), et le Spoofing ou usurpation d'identité.

Attaque Sybil : Elle menace directement l'authenticité et l'accessibilité du réseau VANET. L'attaquant crée de multiples identités fictives pour influencer indûment le réseau, par exemple en envoyant des informations de trafic fausses ou exagérées, provoquant ainsi de la confusion et potentiellement des accidents.

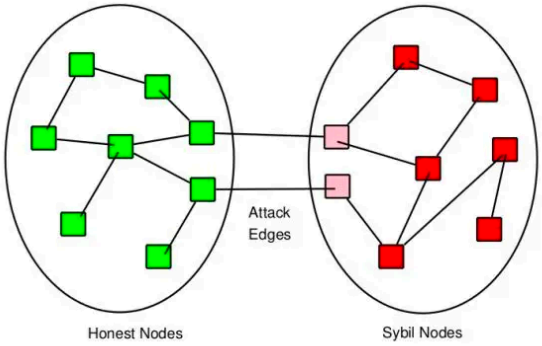

Figure 3 – Attaque Sybil par multiples instances frauduleuses

Attaque par déni de service (DDOS) : Cette attaque cible principalement la disponibilité et l'accessibilité du réseau en inondant le système avec un flot de requêtes illégitimes, ce qui peut paralyser la communication entre les véhicules et les infrastructures routières, et entraîner des retards critiques dans la transmission des informations de sécurité.

Spoofing (Usurpation d'identité) : Le Spoofing est un type d'attaque où l'attaquant se fait passer pour un autre véhicule ou réseau, envoyant de fausses informations qui pourraient induire en erreur les conducteurs et les systèmes automatisés, aboutissant à des décisions de conduite erronées ou à des actions inappropriées.

Les principes de sécurité dans les VANETs – authenticité, intégrité, confidentialité, disponibilité, et accessibilité – sont cruciaux pour maintenir l'efficacité et la fiabilité du réseau. L'authentification est indispensable pour confirmer l'identité des véhicules participants, assurant ainsi l'intégrité des données partagées. L'intégrité vérifie que les informations restent inaltérées, tandis que la confidentialité restreint l'accès aux données

aux entités autorisées. La disponibilité garantit l'accès constant aux services réseau pour la transmission d'informations cruciales en temps réel, et l'accessibilité permet une communication fluide entre les véhicules au sein du réseau.

Les conséquences d'une attaque sur les VANETs peuvent être catastrophiques, avec des informations de sécurité incorrectes ou retardées pouvant causer des accidents, gêner les interventions d'urgence, et semer le chaos sur les routes. Les implications sociétales sont considérables, englobant la protection des citoyens contre des menaces potentielles à leur sécurité individuelle et collective tout en préservant une circulation efficace et sûre. Par conséquent, la sécurisation des réseaux VANET constitue un champ de recherche et de développement primordial, visant à solidifier la confiance dans les technologies de transport futuristes.

3. Les Systèmes de Détection d'Intrusions dans les VANETs

Les systèmes de détection d'intrusion (IDS) constituent une composante essentielle pour la sécurité des réseaux de véhicules ad hoc (VANETs), en agissant comme des gardiens vigilants qui surveillent continuellement le réseau à la recherche d'activités suspectes ou malveillantes. Cette surveillance est particulièrement cruciale dans le contexte des VANETs, où la mobilité élevée et les changements fréquents de topologie affaiblissent l'efficacité des stratégies de sécurité conventionnelles.

Pour pallier ces limites, l'adoption de l'intelligence artificielle (IA) et de l'apprentissage automatique (ML) dans le développement des IDS s'est imposée comme une solution prometteuse [6] [7].

Les IDS basés sur l'IA/ML intègrent généralement un système de collecte de données exploitant des capteurs disséminés à travers le réseau, un moteur d'analyse traitant ces informations en temps réel, et un dispositif d'alerte informant les opérateurs des potentiels dangers détectés. L'avantage majeur de cette approche réside dans sa capacité à identifier des menaces auparavant inconnues, y compris les attaques dites de "jour zéro", grâce à l'aptitude des modèles de ML à apprendre et à s'adapter à de nouveaux schémas d'attaque continuellement.

Les avantages des IDS basés sur l'IA/ML pour la sécurité des VANETs incluent [6]:

- **Adaptabilité :** Ils s'ajustent aux comportements normaux du réseau et s'adaptent dynamiquement aux menaces émergentes, permettant la détection d'attaques inédites.
- **Capacité prédictive :** Ils anticipent les menaces potentielles avant leur matérialisation.
- **Efficience :** Capables de traiter d'importantes volumétries de données, ces systèmes identifient rapidement les activités suspectes.

L'utilisation de l'IA pour la prédiction de menaces comporte également des défis :

- **Complexité :** La mise en œuvre et la maintenance de ces systèmes peuvent s'avérer complexes.
- **Alertes inexactes :** Les risques de faux positifs ou négatifs persistent, comme pour tout système basé sur l'IA.
- **Manque de transparence :** Les décisions générées par les modèles de ML peuvent manquer de clarté pour les utilisateurs, soulignant l'importance de l'intégration de l'intelligence artificielle explicable (XAI) pour améliorer l'interprétabilité.

4. L'Intelligence Artificielle Explicable (XAI)

L'intégration de l'intelligence artificielle (IA) dans les systèmes de détection d'intrusions pour les réseaux de véhicules ad hoc (VANETs) a marqué un progrès significatif en termes d'efficacité et de précision. Cependant, la complexité croissante des modèles d'apprentissage profond (Deep Learning, DL), essentiels à cette évolution, introduit une opacité dans le processus décisionnel. Cette nature "boîte noire" [3] des modèles de DL pose un défi, particulièrement dans des contextes critiques comme la sécurité des VANETs, où comprendre le "pourquoi" et le "comment" des décisions prises par l'IA est crucial [7].

Face à ce défi, l'Intelligence Artificielle Explicable (Explainable AI, XAI) émerge comme une réponse prometteuse. La XAI vise à rendre les processus décisionnels des modèles d'IA transparents, permettant aux utilisateurs et aux parties prenantes de comprendre les

mécanismes sous-jacents à leurs prédictions ou décisions. Cette capacité est particulièrement pertinente dans le domaine des VANETs, où la détection précise et justifiable des intrusions est essentielle pour la prévention des risques et la gestion des incidents de sécurité.

Définition de l'explicabilité *:*
L'explicabilité dans le contexte de l'Intelligence Artificielle, dite « Explainable AI » (XAI), fait référence à la capacité d'un modèle à fournir des explications claires et compréhensibles sur ses décisions. Elle vise à rendre les décisions des modèles d'IA transparentes, compréhensibles et justifiables pour les utilisateurs et les parties prenantes concernées [3].

Protocole de recherche

1. Méthodologie de recherche : La revue systématique de littérature

La revue systématique est une méthode méthodique et rigoureuse visant à regrouper et à examiner de façon exhaustive toutes les recherches pertinentes concernant une question de recherche spécifique. Cette méthode cherche à réduire les biais et les erreurs en adoptant un processus méthodique et transparent.

Ce processus commence par la définition claire d'une question de recherche. Il se poursuit ensuite par une quête exhaustive et organisée des études appropriées, en utilisant des bases de données électroniques (exemple : Miage Scholar). Les recherches sélectionnées sont critiquées soigneusement pour évaluer leur qualité et leur applicabilité. Les données extraites de ces études sont par la suite synthétisées et analysées pour formuler des conclusions synthétiques qui répondent à la question de recherche posée initialement.

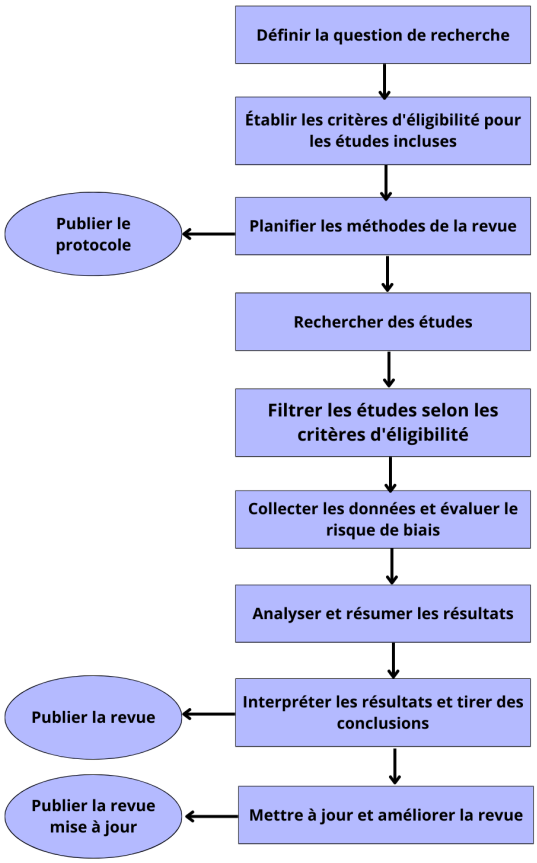

Figure 4 – Étapes de réalisation d'une revue systématique

Pour l'élaboration de ce mémoire, il a été décidé d'adopter l'approche de la Revue Systématique de Cartographie (Systematic Mapping Study, SMS). L'objectif de la SMS est de fournir une vue d'ensemble d'un champ de recherche spécifique.

Dans le contexte de ce mémoire, l'application d'une SMS est jugée la plus appropriée en raison de l'objectif fixé de répondre à la question de recherche suivante :

"Dans quelle mesure les techniques d'Intelligence Artificielle Explicables peuvent-elles améliorer la détection des intrusions dans les réseaux de véhicules Ad-Hoc ?"

Cela permettra d'obtenir une cartographie claire du domaine et de situer les recherches actuelles par rapport aux différentes catégories et dimensions de la question posée.

B. Questions et hypothèses de recherche

Cette étude vise à explorer la contribution des techniques d'Intelligence Artificielle Explicables (XAI) sur la détection d'intrusions dans les réseaux de véhicules ad hoc (VANETs). Elle se concentre sur la façon dont l'explicabilité peut améliorer la compréhension et la fiabilité des décisions prises par des systèmes d'IA. Ce mémoire s'intéresse à l'efficacité des méthodologies d'XAI, à leur influence sur la confiance et l'acceptation des utilisateurs des systèmes de sécurité des VANETs. Les questions spécifiques suivantes orienteront cette analyse systématique :

- *QR1 : Comment les techniques d'XAI peuvent-elles clarifier les décisions prises par des systèmes d'IA dans le contexte des VANETs ?*
- *QR2 : Quelles sont les méthodes et les modèles d'explicabilité mis en œuvre dans le domaine des VANETs pour améliorer la compréhensibilité des systèmes d'IA ?*
- *QR3 : Comment les chercheurs appréhendent-ils le défi du compromis entre la précision et l'explicabilité des modèles d'IA dans le contexte des VANETs ?*

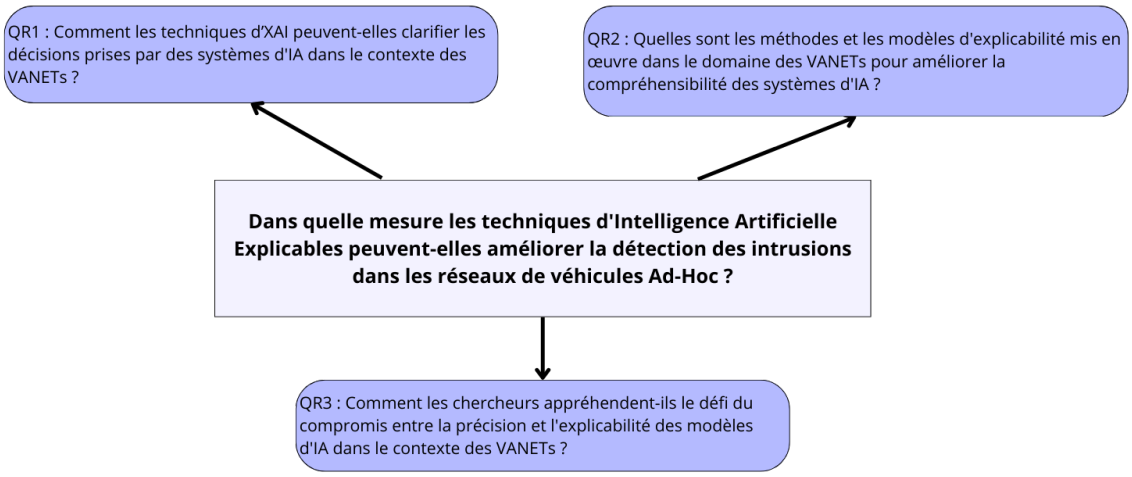

Figure 5 – Sous-questions émergeantes de la problématique principale

En me fondant sur mes connaissances actuelles et mes perspectives préliminaires relatives au domaine des réseaux de véhicules ad hoc (VANETs) et de l'intelligence artificielle explicative (XAI), j'ai formulé plusieurs hypothèses qui seront testées au fil de cette étude. Les résultats recueillis permettront de confirmer ou de réfuter ces suppositions initiales. Voici les hypothèses adaptées à mon sujet et aux questions de recherche définies :

- *H1: L'application de techniques d'XAI peut améliorer significativement la compréhension des décisions automatisées dans les systèmes de détection d'intrusions pour les VANETs.*
- *H2: Les chercheurs estiment que l'explicabilité des modèles d'IA n'est pas prioritaire et représente un frein à la recherche.*
- *H3 : Une méthode d'explicabilité domine dans les VANETs, éclipsant les autres par sa polyvalence et son adoption répandue.*

Ces hypothèses structureront le processus de recherche et d'analyse des données. Elles serviront de points de référence pour évaluer les contributions actuelles et potentielles de l'XAI dans le renforcement de la sécurité des VANETs. La méthodologie adoptée pour cette recherche sera présentée en détail pour démontrer la rigueur et la systématicité de l'approche choisie.

C. Conduite de la recherche

A. Méthodologie

Pour assurer une exploration méthodique des publications scientifiques pertinentes qui correspondent aux questions de recherche établies, il est primordial d'élaborer une stratégie de recherche qui soit à la fois rigoureuse, pertinente et impartiale. L'enjeu est de compiler une collection exhaustive d'articles scientifiques qui abordent les interrogations soulevées, démontrant ainsi l'importance d'une stratégie de recherche bien conçue.

À cet effet, les directives PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) ont été adoptées comme cadre structurant pour cette revue systématique. PRISMA fournit un ensemble de critères conçus pour orienter les chercheurs dans l'agrégation d'études pertinentes, en veillant à la transparence et à l'intégrité du processus de révision. Un diagramme de flux PRISMA sera employé pour illustrer visuellement la démarche de sélection des articles, détaillant le filtrage et le choix des études tout au long de la recherche.

Le diagramme suivant expose les différentes phases de ma stratégie de recherche documentaire, indiquant le volume d'articles retenus à chaque étape du processus.

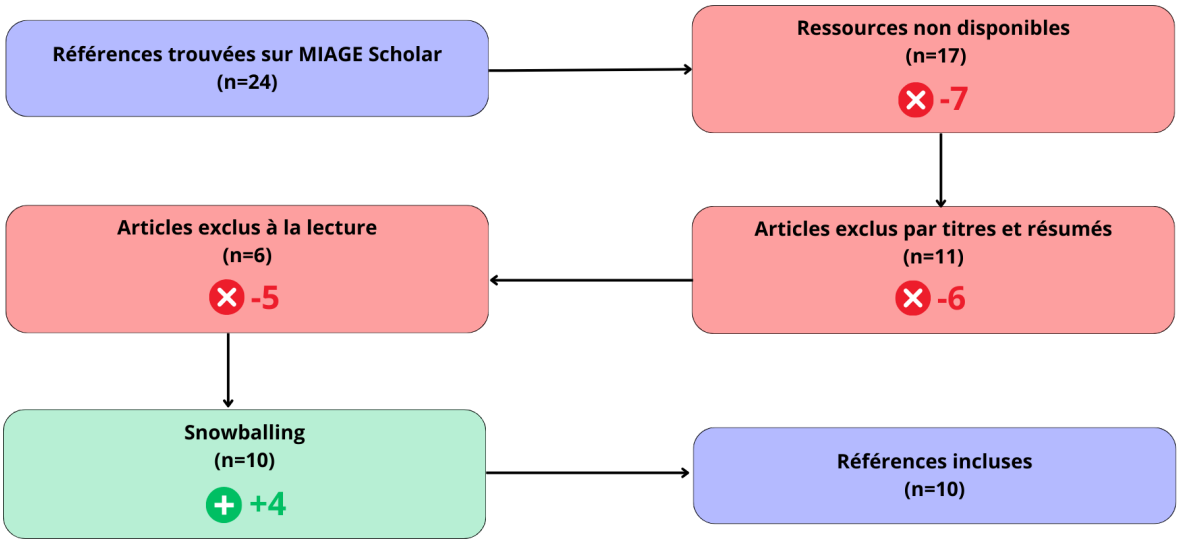

Figure 6 – Diagramme de flux PRISMA

La méthodologie PRISMA peut être synthétisée en plusieurs étapes :

- Identification des mots-clés pertinents pour la recherche afin de construire une requête efficace.
- Sélection, comprenant une première sélection suivie d'une sélection approfondie.
- Inclusion, visant à enrichir la liste des articles sélectionnés, notamment par la méthode de snowballing, qui consiste à explorer les références des articles déjà sélectionnés.

B. Requête de recherche

Pour approfondir la recherche d'articles pertinents concernant l'application des techniques d'Intelligence Artificielle Explicables (XAI) dans les réseaux de véhicules ad hoc (VANETs), une stratégie méthodique de recherche par mots-clés a été déployée. Cette démarche a commencé par l'élaboration d'un tableau exhaustif recensant tous les mots-clés pertinents et leurs synonymes, organisés de manière à optimiser la formulation des requêtes dans les moteurs de recherche académiques.

Vanet	XAI	machine learning	Intrusion detection
Autonomous vehicule	explainable artificial intelligence	ML	intrusion detection systems
vehicular ad hoc network	interpretation	AI	network intrusion detection
autonomous car ad hoc network	explanations	Artificial Intelligence	security
V2V	Interpretable Artificial Intelligence		IDS
	Explainable AI		malware

Tableau 2 - tableau mots-clés permettant d'affiner la recherche

Le tableau a été structuré selon une logique précise : chaque colonne regroupe des termes qui, lors de la recherche, sont combinés avec un opérateur logique "ET" (AND), tandis que les mots-clés et leurs synonymes listés sur une même ligne sont reliés par l'opérateur "OU" (OR). Cette disposition a pour but de créer une requête de recherche à la fois large et ciblée, permettant de capturer une gamme étendue d'articles tout en restant spécifique au sujet d'étude.

La requête finale, conçue à partir de cette matrice de mots-clés, se présente comme suit :

```
(TITLE-ABS-KEY("autonomous vehicle") OR TITLE-ABS-KEY("vanet") OR
TITLE-ABS-KEY("vehicular ad hoc network") OR TITLE-ABS-KEY("V2V")
OR TITLE-ABS-KEY("autonomous car ad hoc network"))
AND
(TITLE-ABS-KEY("explainable artificial intelligence") OR TITLE-ABS-
KEY("interpretation") OR TITLE-ABS-KEY("explanations") OR TITLE-
ABS-KEY("XAI") OR TITLE-ABS-KEY("Interpretable Artificial
Intelligence") OR TITLE-ABS-KEY("Explainable AI") OR TITLE-ABS-
KEY("interpretation"))
AND
(TITLE-ABS-KEY("machine learning") OR TITLE-ABS-KEY("ML") OR TITLE-
ABS-KEY("AI") OR TITLE-ABS-KEY("Artificial Intelligence"))
AND
(TITLE-ABS-KEY("intrusion detection") OR TITLE-ABS-KEY("intrusion
detection systems") OR TITLE-ABS-KEY("network intrusion detection")
OR TITLE-ABS-KEY("attacks") OR TITLE-ABS-KEY("security") OR TITLE-
ABS-KEY("malware") OR TITLE-ABS-KEY("IDS"))
```

Après cette sélection préliminaire, la technique de recherche en chaîne, également connue sous le nom de snowballing, a été appliquée. Cette technique consiste à utiliser les références des articles récupérés pour découvrir d'autres travaux pertinents, favorisant ainsi l'élargissement progressif du corpus documentaire.

C. Critères de sélection des ressources

Après la recherche primaire, il a fallu sélectionner les articles en se basant sur des critères incluant des éléments d'inclusion et d'exclusion, principalement liés aux questions de recherche. Voici un tableau des critères utilisés :

+ Critères d'inclusion +	- Critères d'exclusion -
Article traitant de l'IA Explicable appliqué aux VANETs	Article n'est pas rédigé en Français ou en Anglais.
Article traitant des techniques de détection des intrusions dans les VANETs	Article ne traitant aucune fois d'explicabilité

	Article inaccessible ou introuvable
	Article n’est pas un article scientifique

Tableau 3 - tableau des critères de sélection des articles

Critères d’inclusion

Les articles sélectionnés devaient se focaliser principalement sur l'application de l'intelligence artificielle explicative (XAI) dans le domaine des réseaux de véhicules ad hoc (VANETs). Ces travaux pouvaient explorer divers angles relatifs à cette thématique, qu'ils proposent des innovations techniques, analysent des défis opérationnels ou considèrent les implications de l'adoption de l'XAI pour la sécurité des VANETs. Vu la nature contemporaine de cette recherche, aucune limite de date n'a été fixée pour la sélection des articles ; le document le plus ancien intégré dans cette étude datait de 2018. De plus, il était essentiel que tous les articles retenus soient disponibles en ligne pour consultation.

Critères d’exclusion

Les articles devaient impérativement traiter de l'explicabilité de l'IA dans les VANETs. Ainsi, les documents abordant des techniques d'IA sans se pencher sur l'aspect explicatif étaient exclus. De même, les études présentant des approches d'explicabilité hors du cadre des VANETs n'étaient pas prises en compte. Enfin, les travaux non disponibles en anglais étaient également écartés.

Extraction et analyse des données

A. Analyse quantitative

Dans cette étape de l'analyse, une évaluation quantitative unique des articles sélectionnés a été réalisée. L'objectif était de dresser un aperçu global des recherches compilées et de contribuer aux questions de recherche définies. Les articles ont été triés par année de publication pour identifier la trajectoire de recherche et l'intérêt croissant pour notre sujet.

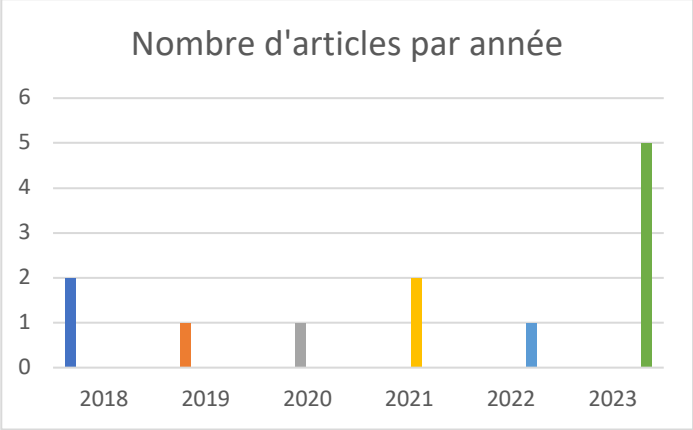

Figure 7 - Histogramme de classification des articles par année

Une tendance marquée se dégage à partir de l'année 2023, où l'on constate une hausse significative des recherches. Cette augmentation témoigne de la montée en popularité du sujet de l'explicabilité ces dernières années, stimulant ainsi un intérêt accru dans la communauté scientifique et suscitant de nombreuses études.

2. Analyse qualitative

1. XAI : un outil pour mieux comprendre les décisions dans le domaine des VANETS

Dans un monde de plus en plus numérisé, les systèmes d'intelligence artificielle (IA) jouent un rôle crucial dans de nombreux secteurs, notamment dans la gestion des réseaux ad hoc véhiculaires (VANETs). Ces réseaux, essentiels pour la sécurité et l'efficacité du trafic routier, reposent sur des décisions automatisées prises par des systèmes d'IA pour fonctionner correctement [4] [6]. Cependant, malgré leur importance croissante, la complexité de ces systèmes rend souvent leurs décisions difficiles à comprendre pour les

humains. Dans ce contexte, l'émergence des techniques d'Intelligence Artificielle Explicable (XAI) représente une avancée prometteuse pour rendre les décisions de l'IA plus transparentes et compréhensibles. Cette nécessité d'explication nous amène à nous interroger :

QR1 : Comment les techniques d'XAI peuvent-elles clarifier les décisions prises par des systèmes d'IA dans le contexte des VANETs ?

Approche XAI

Les progrès récents dans le domaine de l'XAI ont révélé deux approches principales pour rendre les décisions des modèles d'IA plus transparentes et compréhensibles : les modèles transparents (ou ante-hoc) et les modèles post-hoc [3].

Ces deux méthodes visent à améliorer la compréhension et l'interprétation des décisions générées par les systèmes d'IA, surtout dans des domaines où les enjeux sont importants, comme la sécurité routière dans les VANETs.

- Les *modèles ante-hoc* sont conçus dès le départ pour être transparents et compréhensibles. Ils sont construits avec des structures ou des règles explicites qui illustrent clairement le processus de prise de décision. Des exemples de ces modèles incluent la régression linéaire et les arbres de décision. Ces modèles sont particulièrement utiles dans les VANETs pour des applications où la justification des décisions est aussi cruciale que la décision elle-même [7].

- Les *modèles post-hoc*, en revanche, sont des techniques appliquées à des modèles d'IA existants, souvent complexes, pour expliquer a posteriori leurs décisions. Plutôt que de dévoiler le fonctionnement interne du modèle, ces approches se concentrent sur l'analyse des comportements du modèle pour justifier ses décisions. Des outils comme LIME (Local Interpretable Model-Agnostic Explanations) ou SHAP (SHapley Additive exPlanations) permettent de lier les entrées spécifiques aux sorties du modèle [1] [2], montrant l'impact des caractéristiques sur la décision finale [1]. Dans les VANETs, cette approche est particulièrement précieuse pour comprendre le comportement des systèmes basés sur des réseaux de neurones profonds, qui sont utilisés pour des tâches complexes comme la prédiction des intrusions dans un réseau VANET.

Type d'explication

L'intelligence artificielle explicative (XAI) émerge comme un domaine crucial pour rendre les processus de décision des IA compréhensibles pour les humains. Pour atteindre cet objectif, différentes méthodes d'explication sont utilisées. Voici les principales catégories de ces méthodes [1] [3] :

- *Explications sémantiques* - L'articulation en langage naturel fourni des explications textuelles/visuelles qui détaillent le processus décisionnel basé sur les données d'entrée et de sortie. Un modèle décisionnel explicite basé sur cette approche a été mis en œuvre, qui génère des justifications sémantiques pour chaque décision permettant une meilleure interprétation et confiance au système.
- *Visualisations* - Les techniques de visualisation ont prouvé leur utilité en illustrant les caractéristiques ou paramètres influents sur la décision finale d'un modèle.
- *Explications locales* - Ces explications ciblent des instances spécifiques, offrant une justification détaillée pour une prédiction donnée, ce qui est crucial dans des situations où chaque décision peut avoir des conséquences importantes.
- *Explications basées sur des exemples* - Trouver des instances de formation qui se rapprochent le plus de l'élément d'entrée en question permet de créer des parallèles qui facilitent l'explication des décisions prises par le modèle.

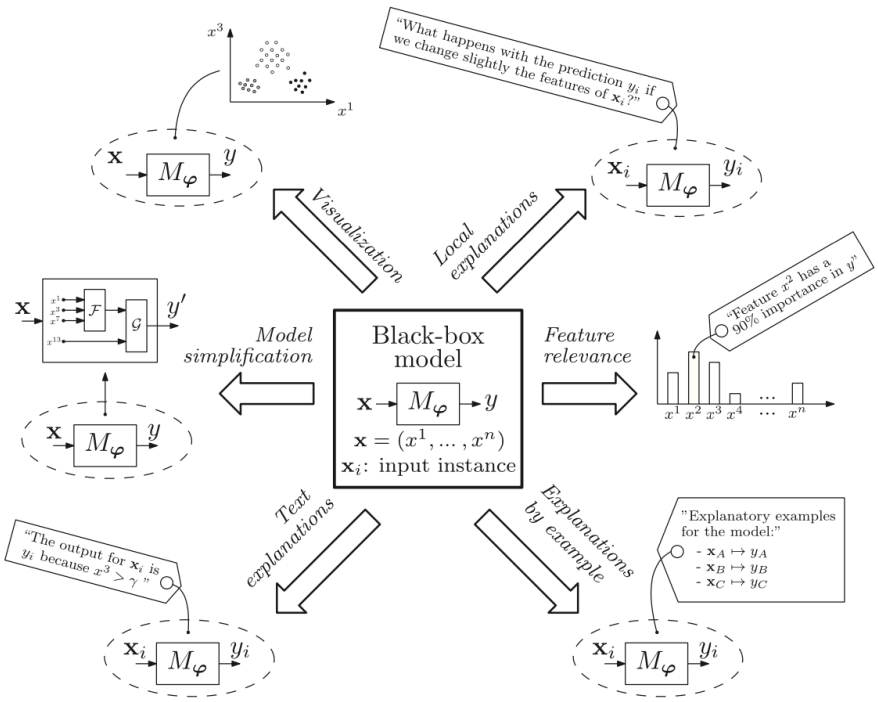

Figure 8 – Les différentes approches d'explicabilité post hoc [3]

Niveau d'explication

Dans le domaine de l'intelligence artificielle explicative (XAI), la transparence des modèles joue un rôle crucial. Cette transparence est souvent abordée selon trois niveaux dans la littérature [3] :

- *Simulabilité* - Un modèle est dit simulable lorsqu'un humain peut le comprendre et le raisonner complètement, ce qui fait de la complexité une considération majeure. Pour qu'un modèle décomposable soit simulable, il doit être suffisamment autonome pour qu'un humain puisse réfléchir à son fonctionnement global.
- *Décomposabilité* - Cela se réfère à la capacité d'expliquer chaque composant d'un modèle (entrées, paramètres, et calculs). Il est synonyme d'intelligibilité. Pour qu'un modèle soit décomposable, chaque entrée doit être interprétable, ce qui signifie que les caractéristiques complexes ou lourdes ne répondent pas à cette exigence. Pour qu'un modèle transparent sur le plan algorithmique devienne décomposable, chaque partie du modèle doit être compréhensible par un humain sans outils supplémentaires.
- *Transparence algorithmique* - Ce concept concerne la capacité des utilisateurs à comprendre le cheminement suivi par le modèle pour produire un résultat à partir des données d'entrée. Cela implique, par exemple, qu'un modèle linéaire est transparent parce que sa surface d'erreur est logique et permet de prévoir comment le modèle réagira face à toutes les situations possibles. En revanche, cela n'est pas possible avec des architectures profondes où le paysage des pertes peut être opaque et la solution doit être approximée. Pour qu'un modèle soit pleinement exploré par analyse mathématique, il doit être entièrement accessible à cette analyse.

L'efficacité de ces niveaux de transparence, de la simulabilité à la transparence algorithmique, est essentielle pour permettre aux utilisateurs des VANETs de comprendre et de se fier aux systèmes d'IA. Cela encourage non seulement l'adoption de ces technologies avancées mais ouvre également la voie à de nouvelles recherches pour améliorer l'intégrabilité et la facilité d'utilisation des modèles d'IA dans les applications de transport intelligent.

QR1 : Comment les techniques d'XAI peuvent-elles clarifier les décisions prises par des systèmes d'IA dans le contexte des VANETs?

RR1 : Nous avons examiné comment l'explicabilité de l'intelligence artificielle peut avoir un impact sur la transparence des décisions automatisées. Il est clairement apparu que les méthodes d'XAI offrent une clé précieuse pour décrypter les opérations complexes des modèles d'IA employés dans la gestion et la sécurité des VANETs.

Les techniques d'explicabilité, en fournissant des explications sémantiques, visuelles, locales, ou encore basées sur des exemples, jouent un rôle crucial pour lutter contre le phénomène de « boite noire ». Ces techniques permettent de d'interpréter de manière plus intuitive les décisions critiques prises par les systèmes d'IA, ce qui est essentiel dans un domaine où chaque décision peut avoir des conséquences immédiates sur la sécurité des usagers.

En déployant des modèles transparents dès la conception (*ante-hoc*) ou en appliquant des techniques explicatives après coup (*post-hoc*) à des systèmes complexes, l'XAI contribue significativement à augmenter la confiance et la compréhension dans les technologies automatisées des VANETs.

Cette **RR1** confirme notre première hypothèse :

- *H1: L'application de techniques d'XAI peut améliorer significativement la compréhension des décisions automatisées dans les systèmes de détection d'intrusions pour les VANETs.*

Cette hypothèse se vérifie, mettant en lumière l'influence bénéfique des explications fournies par l'XAI sur la transparence des décisions prises par les systèmes d'IA. Il est important de noter que cette amélioration de la compréhension découle non seulement des explications textuelles mais aussi d'autres formes d'explications, telles que les visualisations, les analyses locales et les comparaisons par exemple, enrichissant la perception et l'interprétation des processus décisionnels complexes au sein des VANETs.

2. Méthodes d'XAI et cas d'application dans le domaine des VANETS

Nous venons de détailler les approches et les méthodologies d'explicabitlité généralement adoptées dans les systèmes VANET, ainsi que les différentes formes d'explications que ces méthodes peuvent fournir. Passons maintenant à la catégorisation des diverses méthodes d'explicabilité spécifiquement mises en œuvre dans les VANETs, en les alignant avec le cadre conceptuel que nous avons établi. Cela permettra de discerner comment chaque méthode contribue à la clarté et à la fiabilité des systèmes d'IA au sein de cet environnement réseau particulièrement complexe et dynamique.

Nous allons tenter cette catégorisation des méthodes d'explicabilité à travers la question de recherche suivante :

QR2 : Quelles sont les méthodes et les modèles d'explicabilité mis en œuvre dans le domaine des VANETs pour améliorer la compréhensibilité des systèmes d'IA ?

LIME

LIME est une technique d'explicabilité qui permet de décomposer les prédictions des modèles de machine learning complexes, en les rendant intelligibles à un niveau local.

LIME repose sur l'idée que bien que le modèle global puisse être complexe et opaque, il devrait être possible d'approximer et d'expliquer ses prédictions de manière fidèle localement, autour de la prédiction d'intérêt. Cette approximation est réalisée en utilisant des modèles linéaires simples, qui sont intrinsèquement plus interprétables [2].

Pour expliquer la prédiction d'une instance donnée, LIME génère d'abord un ensemble de nouveaux échantillons en perturbant l'instance originale et en observant les prédictions du modèle sur ces nouvelles données. Ensuite, LIME sélectionne un sous-ensemble de ces échantillons qui sont proches de l'instance originale et ajuste un modèle linéaire simple, comme la régression linéaire, qui sert de modèle explicatif local [1]. Les coefficients du modèle linéaire indiquent l'importance de chaque caractéristique par rapport à la prédiction.

Avantages :

- *Interprétabilité locale* : LIME fournit des explications faciles à comprendre pour des prédictions individuelles, ce qui est précieux pour les utilisateurs finaux qui doivent comprendre le raisonnement derrière des décisions spécifiques [3].
- *Modèle-agnostique* : Il peut être appliqué à n'importe quel modèle de machine learning, rendant cette technique très flexible.
- *Personnalisation des explications* : Les utilisateurs peuvent choisir les caractéristiques qu'ils souhaitent explorer, ce qui permet de focaliser l'explication sur les aspects les plus pertinents pour eux.

Inconvénients :

- *Approximation locale* : L'explication ne représente qu'une approximation locale et pourrait ne pas refléter le comportement global du modèle.
- *Choix des caractéristiques et des paramètres* : Les explications peuvent varier en fonction des caractéristiques sélectionnées et des paramètres du modèle linéaire.
- *Absence de considération des interactions* : LIME ne tient pas compte des interactions complexes entre caractéristiques, ce qui peut limiter la précision de l'explication [1] [2].

Dans l'étude [1], LIME a été utilisé pour interpréter les prédictions individuelles faites par l'IDS concernant le trafic réseau normal ou malveillant.

L'utilisation de LIME, via l'outil Lime Tabular Explainer, nous a permis de décortiquer les prédictions d'un modèle complexe et de les présenter de manière intelligible, appliquée spécifiquement à des données structurées sous forme de tableau. Ces données, regroupées dans l'ensemble Edge-IIoT, comportent des informations sur le trafic normal et des activités malveillantes, essentielles pour l'entraînement et l'évaluation de l'efficacité des IDS dans les VANETs.

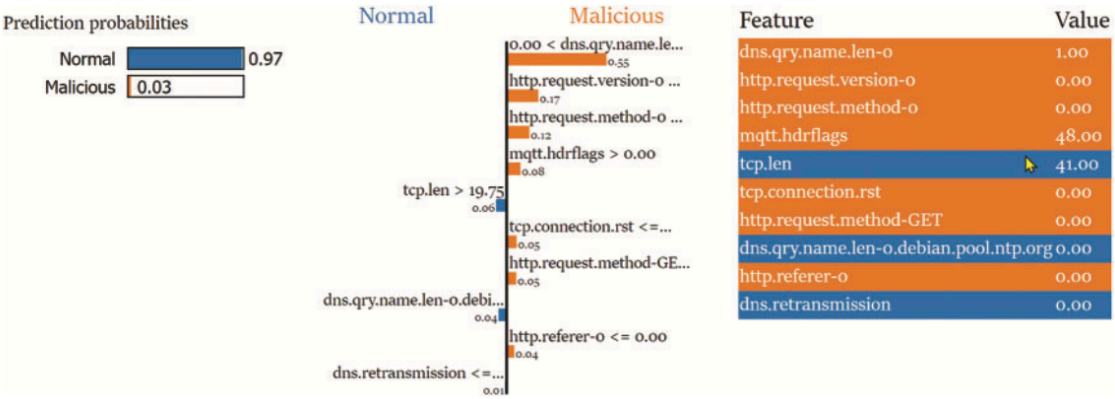

Figure 9 – Analyse du résultat de LIME pour une classification de type normale [1]

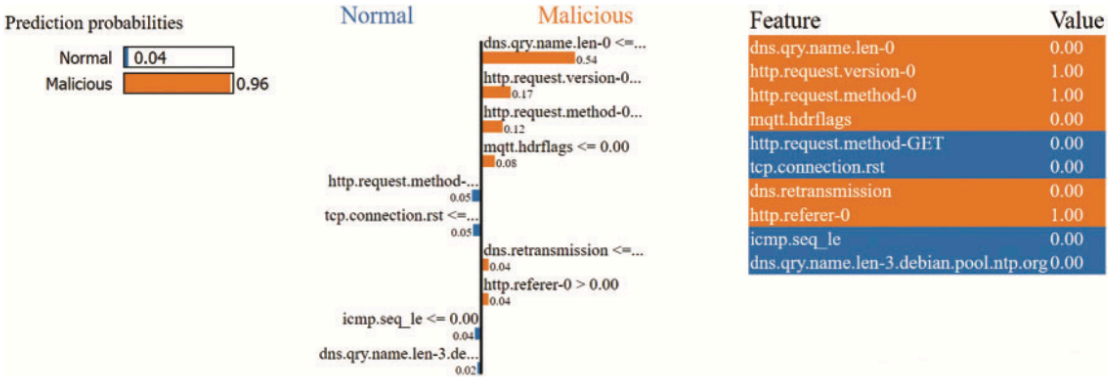

Figure 10 – Analyse du résultat de LIME pour une classification de type malveillante [1]

Pour garantir une analyse cohérente et la reproductibilité des résultats, une partie des données a été désignée pour l'entraînement du modèle, avec un paramètre d'état aléatoire fixé. Les figures 9 et 10 illustrent l'application de LIME à des instances appartenant aux classes "Normal" et "Malveillant", mettant en lumière l'importance de certaines caractéristiques selon la classe prédite. Ces insights révèlent que les caractéristiques comme les noms courts de requête DNS, les versions et méthodes de requête HTTP basses, et les indicateurs d'en-tête MQTT sont indicatifs d'un trafic normal, tandis que les attaques se caractérisent par des noms de requête DNS inexistants ou très courts, et l'absence d'indicateurs d'en-tête MQTT.

Le dataset Edge-IIoT, initialement conçu pour les environnements IIoT mais pertinent pour cette étude sur les VANETs, comprend plus de vingt millions d'enregistrements, dont environ la moitié représente du trafic malveillant. Cette richesse de données permet d'adresser la diversité des menaces auxquelles les réseaux VANETs sont confrontés, avec des attaques allant du déni de service (DoS/DDoS) au scan de ports et à l'injection SQL. La classification binaire adoptée, distinguant le trafic "normal" du "malveillant", s'appuie sur un échantillon uniforme des enregistrements de chaque type d'attaque pour équilibrer le dataset, facilitant ainsi un entraînement et un test précis des IDS.

SHAP

SHAP (SHapley Additive exPlanations) est une méthode avancée d'explicabilité qui attribue une importance à chaque caractéristique d'un modèle pour une prédiction donnée, en se basant sur les valeurs de Shapley, issues de la théorie des jeux. Le but est de comprendre la contribution de chaque caractéristique à la prédiction finale d'un modèle, quelle que soit la complexité de ce dernier.

Dans la théorie, les valeurs de Shapley, sur lesquelles s'appuie SHAP, offrent une méthode équitable pour distribuer le "paiement" (dans notre cas, l'impact sur la prédiction) parmi les "joueurs" (les caractéristiques d'un modèle). Dans la théorie des jeux, cela revient à déterminer la juste valeur de la contribution de chaque joueur à la victoire de l'équipe [1].

SHAP évalue l'impact de chaque caractéristique en prenant en compte toutes les combinaisons possibles de caractéristiques et en calculant la prédiction que le modèle aurait faite avec ou sans chaque caractéristique. Pour une prédiction donnée, SHAP compare la sortie du modèle complet avec celle d'un modèle où une caractéristique est "absente". L'absence d'une caractéristique est simulée en intégrant les valeurs de cette caractéristique dans un ensemble de référence tiré des données d'entraînement.

SHAP calcule ensuite la différence moyenne de la prédiction causée par l'ajout de la caractéristique au modèle. Cela se fait en considérant toutes les permutations possibles des caractéristiques et en moyennant l'effet de l'ajout d'une caractéristique sur la prédiction. Cela donne un score SHAP pour chaque caractéristique pour une instance donnée, reflétant sa contribution positive ou négative à la prédiction finale [2].

Avantages :

- *Interprétabilité globale et locale* : SHAP permet non seulement de comprendre l'importance globale des caractéristiques sur l'ensemble du modèle, mais aussi l'impact spécifique des caractéristiques sur les prédictions individuelles.
- *Compatibilité* : SHAP est compatible avec de nombreux types de modèles de machine learning, y compris les modèles complexes comme les réseaux de neurones profonds.
- *Précision* : Les valeurs de Shapley garantissent une répartition équitable et précise de l'importance attribuée à chaque caractéristique.

Limitations de SHAP :

- *Complexité de calcul* : SHAP peut être gourmand en ressources informatiques, surtout lorsqu'il s'agit de gros volumes de données.

Nous allons maintenant nous pencher sur l'utilisation de SHAP (SHapley Additive exPlanations) sur le même jeu de données que celui utilisé pour LIME [1], permettant ainsi une analyse comparative approfondie entre ces deux méthodes d'explicabilité. Cette démarche offre une opportunité unique de comparer l'efficacité et la pertinence de chaque approche dans le contexte spécifique des systèmes de détection d'intrusions pour les réseaux ad hoc (VANETs). En utilisant le dataset Edge-IIoT, qui inclut à la fois des données sur le trafic normal et des activités malveillantes, nous pouvons évaluer comment LIME et SHAP contribuent à notre compréhension des modèles d'apprentissage automatique utilisés pour identifier les menaces potentielles au sein des VANETs. Cette comparaison nous permet non seulement de discerner les avantages uniques de chaque méthode mais aussi de mettre en évidence leur complémentarité potentielle dans l'amélioration de la transparence et de la fiabilité des systèmes IDS.

Le plot résumé de SHAP classe les caractéristiques sur l'axe vertical, la caractéristique la plus importante étant située en haut et la moins importante en bas. Les valeurs horizontales représentent la valeur SHAP de chaque caractéristique, indiquant ainsi la contribution de cette caractéristique à la prédiction pour chaque point de données. Les

valeurs positives signifient que la caractéristique augmente la prédiction, tandis que les valeurs négatives indiquent qu'elle diminue la prédiction. La longueur de la barre indique l'ampleur et la direction de l'effet de la caractéristique sur la sortie du modèle. Les couleurs rouge et bleu indiquent respectivement les valeurs SHAP augmentées et diminuées.

Les figures *11 et 12* montrent les résultats du plot résumé de SHAP du point de vue des classes "Normal" et "Malveillant". La caractéristique "dns.qry.name.len-0" s'avère avoir le plus grand impact sur la prédiction. Des valeurs SHAP positives suggèrent qu'une valeur élevée de cette caractéristique est associée à des prédictions positives pour la classe "Normal". Inversement, une valeur faible de "dns.qry.name.len-0" entraîne une valeur SHAP plus élevée pour la classe "Malveillant", indiquant que cette caractéristique est un indicateur de la classe Malveillant".

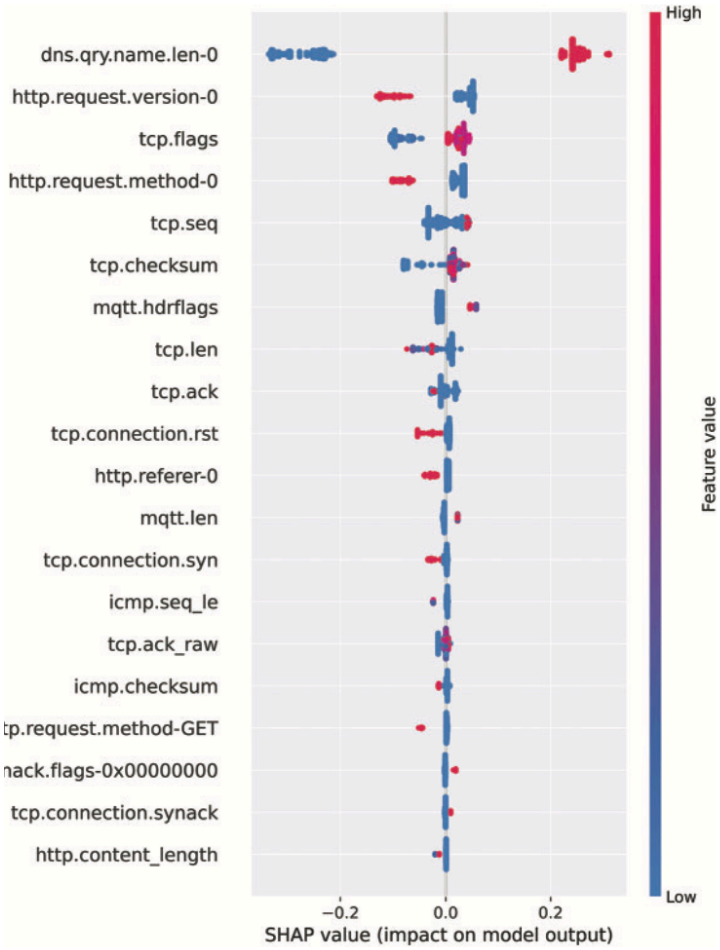

Figure 11 – Analyse du résultat de SHAP pour une classification de type normal [1]

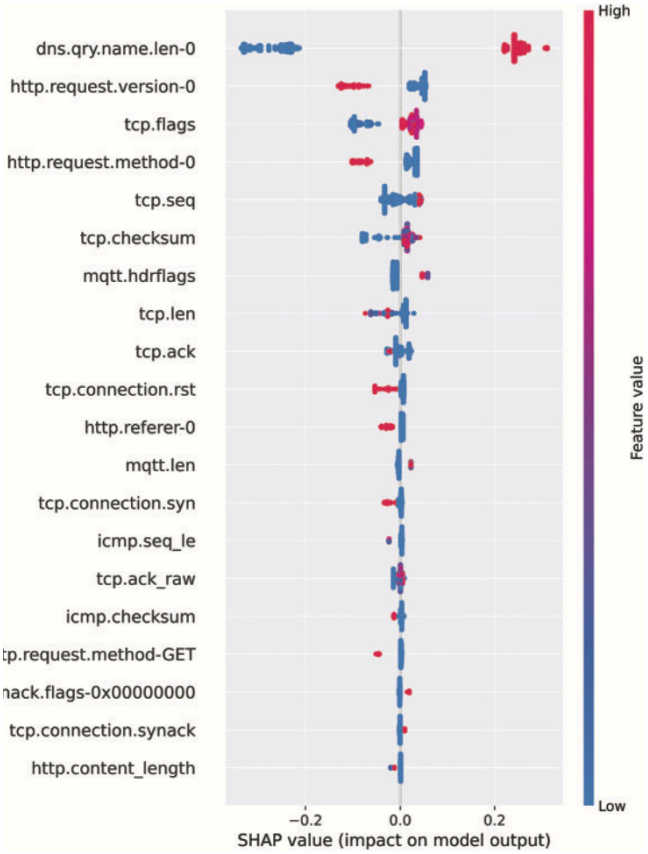

Figure 12 – Analyse du résultat de SHAP pour une classification de type malveillante [1]

La figure *13* combine ces informations en un seul graphique, montrant la moyenne de l'impact (valeur SHAP moyenne absolue) sur la prise de décision pour une caractéristique spécifique. Cela révèle combien une caractéristique spécifique peut changer la prédiction du modèle. La caractéristique "dns.qry.name.len-0" se distingue comme étant la plus influente dans la différenciation entre les données d'attaque et normales.

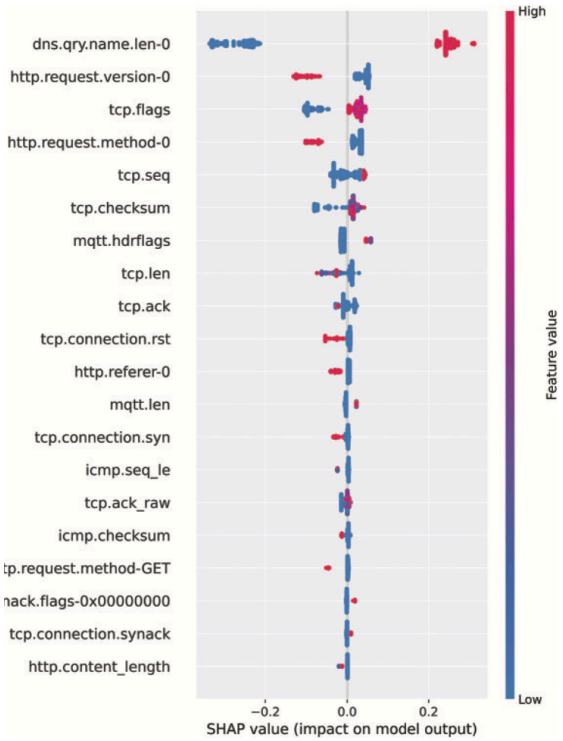

Figure 13 – Analyse du résultat de SHAP pour une classification de type malveillante [1]

L'utilisation de SHAP sur le même ensemble de données que LIME permet non seulement une analyse comparative directe mais aussi souligne l'importance de la synergie entre ces méthodes pour améliorer la compréhension et la transparence des modèles de détection d'intrusions basés sur l'apprentissage automatique. Cet angle d'approche révèle l'utilité indéniable de juxtaposer les insights obtenus par chacune de ces techniques pour forger un système de détection robuste et fiable. Bien que LIME et SHAP partagent des points communs, notamment l'importance accordée à certaines caractéristiques comme "dns.qry.name.len-0", ils offrent également des perspectives distinctes sur la manière dont ces caractéristiques influencent les prédictions. SHAP, avec son analyse basée sur les valeurs de Shapley, fournit une compréhension globale et locale de l'impact des caractéristiques, enrichissant ainsi notre capacité à évaluer l'efficacité du modèle IDS. Les résultats obtenus via SHAP complètent ceux de LIME, en mettant en lumière l'importance des protocoles TCP et HTTP et en soulignant l'éventuelle pertinence des caractéristiques liées au protocole MQTT dans la détection des intrusions.

Nous allons maintenant découvrir les arbres de décision (Decision Tree) et Random Forest. Ils ne sont pas des méthodes d'XAI (Intelligence Artificielle Explicable) au sens traditionnel, car ils sont plutôt des algorithmes de machine learning utilisés pour construire des modèles prédictifs. Cependant, ils sont souvent considérés dans les discussions sur l'explicabilité en raison de leur nature intrinsèquement plus interprétable comparée à des modèles plus complexes comme les réseaux de neurones profonds. [1]

Arbre de Décision (Decision Tree Model)

Les arbres de décision sont une méthode d'apprentissage supervisé largement utilisée pour la classification et la régression. Dans le contexte de la détection des intrusions dans les réseaux VANET, les arbres de décision peuvent être particulièrement utiles pour leur capacité à offrir des modèles prédictifs clairs et facilement interprétables [7].

Les arbres de décision modélisent les décisions et leurs conséquences potentielles, y compris les résultats des événements aléatoires, les coûts des ressources et l'utilité. Ils sont constitués de nœuds représentant les décisions à prendre ou les événements aléatoires et de branches qui relient ces nœuds, représentant les choix possibles ou les résultats des événements.

Fonctionnement :

- *Construction de l'arbre* : À partir d'un ensemble de données d'entraînement, l'arbre est construit en divisant les données en sous-ensembles basés sur les caractéristiques qui apportent le plus d'information concernant la variable cible. Ce processus est répété récursivement pour chaque sous-ensemble jusqu'à ce qu'un critère d'arrêt soit atteint, comme un nombre minimal de points dans un nœud ou une pureté maximale.
- *Prédiction* : Pour classer une nouvelle instance, on suit les décisions prises à chaque nœud, de la racine jusqu'à une feuille, qui indique la prédiction finale de l'arbre.

Avantages :

- *Interprétabilité* : Les règles de décision simples et séquentielles permettent aux utilisateurs de comprendre facilement pourquoi et comment une décision a été prise.

- *Flexibilité* : Peut être utilisé pour des problèmes de classification et de régression.
- *Pas de normalisation nécessaire* : Les arbres de décision ne nécessitent pas la normalisation des caractéristiques.

Inconvénients :

- *Surapprentissage* : Sans élagage ou avec une profondeur de l'arbre trop importante, ils peuvent mémoriser les données d'entraînement, réduisant leur capacité à généraliser à de nouvelles données.
- *Sensibilité aux variations* : Des petites variations dans les données d'entraînement peuvent aboutir à des arbres très différents.

Les arbres de décision sont particulièrement efficaces pour identifier et classer les différents types d'attaques dans les VANETs [7]. Leur structure transparente permet aux experts en sécurité de comprendre les critères selon lesquels les activités réseau sont jugées normales ou malveillantes, facilitant l'ajustement et l'amélioration des stratégies de détection.

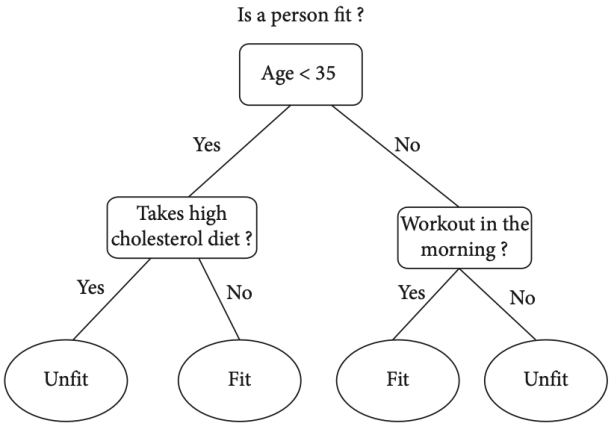

Figure 14 – Exemple d'un arbre de décision [7]

Random Forest

La méthode Random Forest est une technique d'apprentissage ensembliste pour la classification, la régression et d'autres tâches, qui opère en construisant une multitude d'arbres de décision au moment de l'entraînement. Pour les prédictions, le résultat de la majorité des arbres est pris pour les tâches de classification, et la moyenne pour la régression.

Random Forest est basé sur le principe que l'union d'un grand nombre de modèles (arbres de décision, dans ce cas [4]) moins précis ou légèrement biaisés peut conduire à un modèle plus robuste et précis en réduisant la variance et le biais. Chaque arbre de décision est construit à partir d'un échantillon aléatoire des données avec remplacement (bootstrap sample), et lors de la division d'un nœud, une sélection aléatoire des caractéristiques est considérée. Ce processus d'ajout de l'aléatoire aide à décorréléer les arbres et améliore la performance globale du modèle.

Random Forest est particulièrement adapté à la détection des intrusions dans les VANETs en raison de sa capacité à gérer des ensembles de données complexes et hétérogènes. Il peut efficacement classifier les comportements du réseau comme normaux ou malveillants en tirant parti de ses multiples arbres pour capturer diverses signatures d'attaque [1], tout en étant résistant au bruit et aux anomalies dans les données.

Avantages :

- *Robustesse aux Données hétérogènes*: Excellente capacité à traiter des données complexes et variées typiques des VANETs, offrant des prédictions fiables même en présence de bruit.
- *Réduction de la Variance et du biais*: L'agrégation des prédictions de nombreux arbres réduit le risque de surapprentissage et améliore la généralisabilité du modèle.
- *Interprétabilité relative*: Bien que moins interprétable qu'un seul arbre de décision, Random Forest permet d'évaluer l'importance des caractéristiques, fournissant ainsi des insights sur les facteurs influençant les comportements malveillants.

Inconvénients :

- *Complexité computationnelle*: L'entraînement de multiples arbres exige davantage de ressources computationnelles, ce qui peut être un défi pour les systèmes en temps réel des VANETs.
- *Interprétabilité limitée*: Par rapport aux modèles plus simples, l'ensemble complexe des arbres rend l'explication des prédictions spécifiques moins directe, ce qui peut compliquer la compréhension des raisons sous-jacentes à une décision.

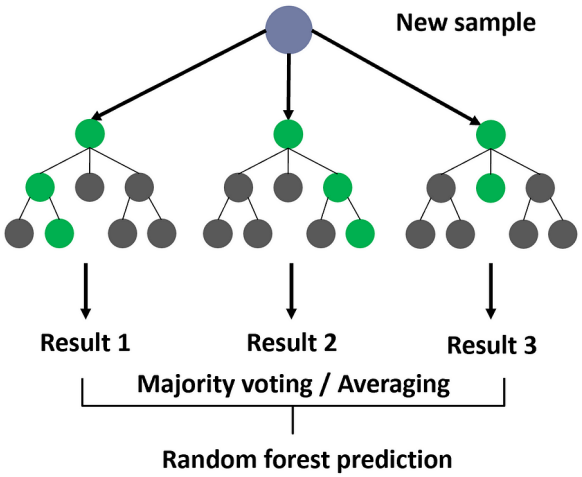

Figure 15 – Schéma de fonctionnement global de l'algorithme Random Forest

Méthode	Description	Approche	Relation au modèle	Type d'explication
LIME	Crée des modèles de substitution locaux pour expliquer des prédictions spécifiques.	Post-hoc	Agnostique	Locale, Sémantique, Visualisation
SHAP	Attribue des contributions individuelles à chaque caractéristique d'entrée pour expliquer les décisions du modèle.	Post-hoc	Agnostique	Locale (ou non), Sémantique, Visualisation
Decision Tree	Utilise une structure arborescente pour modéliser des séquences de règles décisionnelles.	Ante-hoc	Spécifique au modèle	Globale, Sémantique, Visualisation par arbre
Random Forest	Construit un ensemble d'arbres de décision pour améliorer la précision et contrôler le surapprentissage.	Ante-hoc	Spécifique au modèle mais considère diversité des arbres	Globale (à travers l'importance des caractéristiques), Visualisation par importance des caractéristiques

Tableau 4 - Tableau récapitulatif des méthodes et Modèles XAI répertoriés

QR2 : Quelles sont les méthodes et les modèles d'explicabilité mis en œuvre dans le domaine des VANETs pour améliorer la compréhensibilité des systèmes d'IA ?

RR2 : De nombreuses méthodes et classificateurs d'Intelligence Artificielle Explicable (XAI) sont activement explorés et implémentés dans le domaine des réseaux ad hoc de véhicules (VANET) pour améliorer la compréhensibilité des systèmes d'IA.

Pour ce qui est des méthodes *post-hoc*, nous avons identifié LIME et SHAP, qui fournissent des explications principalement pour des modèles prédictifs complexes. Ces méthodes sont particulièrement utiles dans les VANET, où la performance prédictive est critique et les modèles peuvent être très opaques.
Nous avons également identifié des méthodes *ante-hoc* telles que les arbres de décision (Decision Trees) et les forêts aléatoires (Random Forests), qui sont intrinsèquement plus transparents et fournissent une explication globale grâce à leur structure interne et à la visualisation de l'importance des caractéristiques.

Ces méthodes ante-hoc permettent de modéliser des séquences de règles décisionnelles qui sont directement déchiffrables, offrant une transparence dès la phase de conception du modèle.

En plus de ces méthodes, il est pertinent de noter l'émergence de techniques hybrides et d'approches ensemblistes qui combinent les forces des approches *ante-hoc* et *post-hoc* pour fournir à la fois une haute performance prédictive et une explicabilité robuste. L'objectif ultime reste d'équiper les VANETs avec des systèmes d'IA qui ne sont pas seulement performants, mais aussi suffisamment explicables pour permettre aux opérateurs humains et aux parties prenantes de comprendre et de faire confiance aux décisions prises par l'IA. Cette évolution est cruciale pour l'adoption généralisée de l'IA dans les systèmes de transport intelligents et pour garantir leur sûreté et leur fiabilité à long terme.

Cette **RR2** infirme notre troisième hypothèse :

- *H3 : Une méthode d'explicabilité domine dans les VANETs, éclipsant les autres par sa polyvalence et son adoption répandue.*

La réponse RR2 montre clairement qu'aucune méthode unique d'explicabilité ne domine le champ des VANETs. Au lieu de cela, une combinaison de méthodes est privilégiée pour répondre aux divers besoins d'explicabilité et de performance. Cela souligne l'importance de l'adaptabilité et de la personnalisation dans le choix des méthodes d'explicabilité, ainsi que la tendance vers des approches hybrides pour maximiser à la fois la compréhension humaine et l'efficacité algorithmique.

3. Explicabilité et Efficacité : Un Dilemme ?

Dans le cadre de notre, nous avons identifié plusieurs méthodes d'XAI pertinentes. Ces méthodes permettent d'élucider les décisions algorithmiques prises par des systèmes d'IA potentiellement opaques, qu'ils soient spécifiquement conçus pour être explicables (*approche ante-hoc*) ou qu'ils nécessitent des techniques explicatives après leur développement (*approche post-hoc*).

Dans le domaine des VANET, où la détection rapide et précise des intrusions est primordiale pour la sécurité et la fiabilité des communications, il est crucial d'utiliser des modèles d'IA très performants. Cependant, la complexité de ces modèles performants peut les rendre moins transparents. L'enjeu de l'XAI est donc de maintenir la haute performance de ces systèmes tout en les rendant compréhensibles, ce qui est essentiel pour établir la confiance des utilisateurs et assurer une réponse adéquate lorsqu'une intrusion est détectée.

Toutefois, la question de savoir si l'interprétabilité doit l'emporter sur la performance est l'objet de discussions au sein de la communauté scientifique, en particulier dans le contexte des VANETs. Certains soutiennent qu'un équilibre doit être trouvé entre la transparence et la performance des systèmes d'IA, en privilégiant des modèles plus interprétables qui peuvent renforcer la responsabilisation et la compréhension des décisions de sécurité critiques. D'autres, cependant, argumentent que la priorité devrait être donnée à la performance des modèles, même si cela signifie accepter un certain degré d'opacité.

Nous allons tenter de répondre à cette interrogation à travers la question de recherche suivante :

QR3 : Comment les chercheurs appréhendent-ils le défi du compromis entre la précision et l'explicabilité des modèles d'IA dans le contexte des VANETs ?

La transparence, nécessaire à la confiance, est mise en avant comme un prérequis dans l'acceptation et la commercialisation des systèmes autonomes tels que les VANETs [8]. L'argument principal est que comprendre le fonctionnement d'un système d'IA, comme

celui des véhicules autonomes, accroît la confiance des utilisateurs [3] et est déterminant pour le taux d'adoption de cette technologie.

En matière de systèmes de détection d'intrusion, un équilibre doit être trouvé entre la performance élevée et la capacité d'explication des décisions prises par l'IA. L'utilisation de modèles d'arbres de décision [2], reconnus pour leur interprétabilité, permet une meilleure compréhension des caractéristiques des attaques malveillantes, renforçant ainsi la confiance dans ces systèmes. Néanmoins, cette recherche reconnaît le compromis entre la complexité et la performance des modèles d'IA et leur transparence et responsabilité.

L'utilisation de modèles d'apprentissage profond, bien que fournissant une précision prédictive significative dans les VANETs, opère souvent comme des "boîtes noires" [7]. Ce manque de transparence dans les processus décisionnels suscite des inquiétudes concernant la gestion de la confiance en cybersécurité, soulignant que l'efficacité des modèles ne devrait pas se faire au détriment de la confiance dans des systèmes critiques.

L'importance de l'interprétation des caractéristiques est soulignée, indiquant que toutes les fonctionnalités d'un ensemble de données ne contribuent pas également aux prédictions d'un modèle. Comprendre les caractéristiques qui dirigent les décisions du modèle peut améliorer la confiance et la fiabilité, surtout dans des applications sensibles à la sécurité comme la détection d'intrusion dans les VANETs.

L'efficacité des arbres de décision [2] pour renforcer l'explicabilité est mise en évidence en raison de leur simplicité et de leur nature basée sur des règles, fournissant ainsi des explications compréhensibles aux opérateurs humains. Cela contraste avec des modèles plus complexes qui ne fournissent pas le même niveau de transparence.

QR3 : Comment les chercheurs appréhendent-ils le défi du compromis entre la précision et l'explicabilité des modèles d'IA dans le contexte des VANETs ?
RR3 : La recherche indique que, bien qu'il y ait une reconnaissance claire de l'importance de l'exactitude et de l'explicabilité dans les systèmes d'IA pour les VANETs, la communauté scientifique n'a pas encore atteint un consensus sur la façon d'équilibrer ces deux aspects cruciaux. Cela se reflète dans les efforts de recherche continus pour développer des modèles d'IA qui sont à la fois performants et

interprétables, ainsi que dans les opinions diverses sur la nécessité de l'explicabilité dans des applications à enjeux élevés comme la détection d'intrusion.

L'explicabilité n'est pas seulement une question de facilité de compréhension des modèles d'IA mais aussi une question d'importance des caractéristiques. Comprendre quelles caractéristiques influencent le plus la prise de décision permet de renforcer la confiance dans les prédictions du modèle, en particulier dans des applications sensibles à la sécurité comme la détection d'intrusion dans les VANETs.

Cette **RR3** infirme notre deuxième hypothèse :

- *H2: Les chercheurs estiment que l'explicabilité des modèles d'IA n'est pas prioritaire et représente un frein à la recherche.*

La réponse **RR3** montre clairement qu'un débat persiste sur la balance entre explicabilité et performance. D'une part, la nécessité de systèmes de détection précis dans les VANETs est indéniable. D'autre part, la complexité de ces systèmes s'accompagne souvent d'un manque de transparence et de confiance. Les papiers examinés suggèrent que l'explicabilité est un champ d'intérêt croissant, mais qu'il n'y a pas de consensus définitif sur la mesure dans laquelle elle devrait être privilégiée par rapport à la performance dans le développement de systèmes d'IA pour les VANETs .

Discussion

Ce travail de recherche a permis de mettre en lumière les enjeux de l'Intelligence Artificielle Explicable (XAI) dans le processus de détection des intrusions sur les réseaux de véhicules VANET. À travers notre étude, nous avons visé à éclaircir diverses interrogations initiales, permettant ainsi de confirmer ou d'invalider les hypothèses préliminaires de notre projet.

Chaque question de recherche a été détaillée et synthétisée dans des sections distinctes de notre document, illustrant notre progression méthodique dans la quête de réponses à la problématique principale. Voici donc un résumé des principales découvertes et du parcours accompli.

Au début, notre attention s'est portée sur l'impact potentiel des approches d'XAI pour démêler les décisions algorithmiques au cœur des VANETs. Il est ressorti que l'XAI, par ses multiples techniques explicatives qu'elles soient sémantiques, visuelles, spécifiques ou basées sur des exemples constitue une clé essentielle pour percer l'opacité des modèles d'IA dans la gestion sécuritaire des VANETs.

Ces approches d'explicabilité, en facilitant une interprétation intuitive des jugements critiques émis par les IA, se révèlent indispensables dans un domaine où chaque décision influe directement sur la sûreté des utilisateurs. L'adoption de modèles transparents dès leur conception ou l'intégration a posteriori de techniques explicatives à des architectures complexes renforce considérablement la confiance et la compréhension des technologies automatisées au sein des VANETs.

Par la suite, notre exploration s'est étendue aux méthodologies et aux modèles d'XAI appliqués spécifiquement aux VANETs pour renforcer la clarté des systèmes d'IA. Nous avons mis en lumière l'utilisation de méthodes *post-hoc* telles que LIME et SHAP, qui se démarquent dans l'explication des modèles prédictifs complexes, ainsi que l'adoption de stratégies *ante-hoc* comme les arbres de décision et les forêts aléatoires, qui par leur nature offrent une transparence intrinsèque et une compréhension globale grâce à leur structure interne.

Ces approches *ante-hoc* facilitent l'élaboration de modèles basés sur des séquences de règles clairement interprétables, assurant une transparence dès la phase de conception.

Outre ces méthodes, l'avènement de techniques hybrides et d'approches combinatoires promet de marier efficacement la performance prédictive élevée et une explicabilité approfondie, essentielle pour l'adoption et la confiance accrues dans les systèmes d'IA au sein des infrastructures de transport intelligentes.

La dernière phase de cette étude nous a amenés à approfondir la dualité entre la précision des modèles d'Intelligence Artificielle (IA) et leur capacité à être expliqués dans le cadre des Réseaux Ad hoc Véhiculaires (VANETs). L'objectif était de démêler le noeud gordien liant la nécessité d'une haute exactitude dans la détection des intrusions à l'impératif d'une transparence qui facilite la compréhension et la confiance des utilisateurs dans ces systèmes critiques.

Le débat sur la prépondérance de l'explicabilité par rapport à la performance des modèles d'IA dans les VANETs a révélé une diversité d'approches et de perspectives au sein de la communauté scientifique. D'un côté, l'urgence de détecter efficacement toute intrusion malveillante pousse vers l'emploi de modèles hautement sophistiqués, capables de prédictions précises mais souvent opaques. De l'autre, la nécessité de rendre ces systèmes compréhensibles et fiables aux yeux des opérateurs humains souligne l'importance cruciale de l'explicabilité.

Dans notre exploration, nous avons constaté que les techniques d'Intelligence Artificielle Explicable (XAI), notamment LIME et SHAP pour l'approche post-hoc, ainsi que les arbres de décision et les forêts aléatoires pour l'approche ante-hoc, présentent des avantages significatifs pour démystifier les décisions prises par les IA dans les VANETs. Ces méthodes, en rendant les modèles plus transparents, contribuent à une meilleure appréhension des mécanismes sous-jacents aux prédictions, renforçant ainsi la confiance dans ces systèmes.

L'enjeu du compromis entre précision et explicabilité a été particulièrement palpable dans nos discussions avec les chercheurs et à travers l'analyse de la littérature existante. Il est apparu que, loin d'être un choix binaire, la recherche d'un équilibre optimal entre ces deux aspects est une quête continue. La tendance émergente vise à intégrer des méthodes hybrides et des approches ensemblistes, qui promettent de concilier la robustesse prédictive avec une explicabilité satisfaisante. Cette approche hybride, alliant les avantages des méthodes ante-hoc et post-hoc, est envisagée comme une voie prometteuse

pour développer des systèmes d'IA à la fois fiables et compréhensibles dans le contexte des VANETs.

De plus, ces recherches nous ont donné la possibilité de vérifier si nos hypothèses étaient valides, comme le montre le tableau récapitulatif ci-dessous :

Hypothèse	Validée	Réfutée	Nuancée
H1 :L'application de techniques d'XAI peut améliorer significativement la compréhension des décisions automatisées dans les systèmes de détection d'intrusions pour les VANETs	X		
H2: Les chercheurs estiment que l'explicabilité des modèles d'IA n'est pas prioritaire et représente un frein à la recherche.			X
H3 : Une méthode d'explicabilité domine dans les VANETs, éclipsant les autres par sa polyvalence et son adoption répandue.		X	

Tableau 5 - Tableau récapitulatif des hypothèses

Conclusion

En abordant les méandres de l'intelligence artificielle (IA) appliquée aux réseaux de véhicules Ad-Hoc (VANETs), cette recherche a démontré l'importante et la pertinence d'approches explicatives face aux enjeux de cybersécurité émergents. La fusion entre l'IA et les VANETs, vecteur d'innovation, requiert une vigilance accrue contre les menaces numériques en perpétuelle évolution. À travers une analyse de méthodes telles que LIME et SHAP, nous avons examiné la contribution significative de l'XAI à la détection des intrusions, révélant ainsi l'efficacité et les limites intrinsèques à ces techniques.

Les conclusions de notre étude éclairent un double aspect fondamental : d'une part, l'impératif de précision dans la détection des activités malveillantes au sein des VANETs, et d'autre part, la nécessité d'une transparence totale pour l'utilisateur final. Les avancées permises par l'XAI, bien que prometteuses, doivent être envisagées avec prudence et discernement, en tenant compte des contextes d'utilisation spécifiques et des exigences en matière de fiabilité et d'intelligibilité des processus décisionnels automatisés.

Face aux défis soulevés par la complexité des modèles d'IA, la recherche a mis en lumière l'importance cruciale d'une approche explicative pour le développement futur des systèmes de transport intelligent. L'XAI n'est pas une simple amélioration technique, elle incarne un virage stratégique essentiel à l'instauration d'une confiance durable entre les technologies avancées et la société.

Bibliographie

[1] F. HASSAN, J. YU et Z. S. SYED, «Achieving Model Explainability for Intrusion Detection in VANETs with LIME.,» *PeerJ Computer Science,* vol. 9, 2023.

[2] M. RAHMAN, R. K. NAVID et M. M. HOSSAIN BHUY, «Exploring the intersection of machine learning and explainable artificial intelligence: An analysis and validation of ML models through XAI for intrusion detection,» *Brac University,* 2023.

[3] A. B. ARRIETA, N. DÍAZ-RODRÍGUEZ et J. DEL SER, «Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,» *Information fusion,* vol. 58, 2020.

[4] K. RASHID, Y. SAEED et A. ALI, «An adaptive real-time malicious node detection framework using machine learning in vehicular ad-hoc networks (vanets),» *Sensors,* vol. 23, 2023.

[5] N. BEN RABAH et H. IDOUDI, «A machine learning framework for intrusion detection in VANET communications,» *Springer International Publishing,* 2022.

[6] M. GOPALAKRISHNAN et U. ELANGOVAN, «Intelligent Communication Technologies and Virtual Mobile Networks,» *Springer International Publishing,* 2020.

[7] B. MAHBOOBA, M. TIMILSINA et R. SAHAL, «Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model.,» *Complexity,* 2021.

[8] T. ZHANG et Q. ZHU, «Distributed privacy-preserving collaborative intrusion detection systems for VANETs,» *IEEE Transactions on Signal and Information Processing over Networks,* vol. 4, 2018.

[9] S. AZIZ, M. T. FAIZ et A. M. ADENIYI, «Anomaly detection in the internet of vehicular networks using explainable neural networks (xNN),» *Mathematics,* 2022.

[10] L. M. CYSNEIROS, M. RAFFI et J. C. S. DO PRADO LEITE, «Software transparency as a key requirement for self-driving cars.,» *2018 IEEE 26th international requirements engineering conference (RE),* 2018.

Table des figures

Table des tableaux

Lexique

IA : Intelligence Artificielle
ML : MachinE Learning
DL : Deep Learning
XAI : Intelligence Artificielle Explicable
SMS : Revue Systématique de Cartographie
PRISMA : Preferred Reporting Items for Systematic Reviews and Meta-Analyses
LIME : Local Interpretable Model-agnostic Explanations
SHAP : SHapley Additive exPlanations
VANET : Vehicular Ad-Hoc Network