UQAC
Université du Québec
à Chicoutimi

ANALYSE DES MÉTHODES HYBRIDES D'APPRENTISSAGE PROFOND POUR LA DÉTECTION DES CYBERATTAQUES DANS L'INTERNET DES OBJETS : UNE ÉTUDE COMPARATIVE DE RÉFÉRENCE

PAR DJIM MBATHIE

MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI EN VUE DE L'OBTENTION DU GRADE DE MAÎTRISE ÈS SCIENCES (M. SC.) EN INFORMATIQUE

QUÉBEC, CANADA

RÉSUMÉ

L'essor massif de l'Internet des Objets transforme les infrastructures numériques modernes et augmente de manière significative la surface d'attaque des réseaux. Les dispositifs connectés, souvent limités en ressources et peu sécurisés, deviennent des cibles privilégiées pour des attaques variées telles que les botnets, les attaques par déni de service ou les campagnes de reconnaissance. Dans ce contexte, la mise en place de mécanismes de détection fiables et adaptatifs constitue un enjeu essentiel pour garantir la fiabilité et l'intégrité des environnements IoT.

Ce mémoire propose une analyse comparative approfondie de plusieurs méthodes hybrides d'apprentissage profond appliquées à la détection de cyberattaques IoT. Trois modèles ont été conçus et évalués à partir du jeu de données CIC IoT 2023, soit les modèles CNN, CNN LSTM et Stacked CNN LSTM. Les données ont été prétraitées à l'aide d'un processus rigoureux comprenant le nettoyage, la normalisation, l'équilibrage et la transformation en séquences temporelles. Les performances obtenues ont été comparées à celles de deux méthodes classiques, Random Forest et Logistic Regression, afin de proposer une analyse expérimentale complète.

Les résultats montrent que les méthodes hybrides d'apprentissage profond exploitent efficacement la complémentarité entre les caractéristiques spatiales et temporelles, améliorant ainsi la détection de plusieurs cyberattaques complexes. Les approches classiques conservent toutefois un avantage notable dans les contextes où les motifs statistiques sont fortement marqués. L'analyse des matrices de confusion et des courbes ROC révèle également les difficultés persistantes liées aux attaques de reconnaissance et aux attaques par force brute qui imitent parfois les comportements légitimes.

Cette étude s'inscrit finalement dans une perspective large intégrant les enjeux de confidentialité, l'apprentissage distribué et l'explicabilité, dimensions essentielles à la conception de mécanismes de détection robustes et adaptés aux contraintes des environnements IoT.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

AI Artificial Intelligence (Intelligence artificielle)
AUC Area Under the Curve
CNN Convolutional Neural Network
CNN-LSTM Convolutional Neural Network – Long Short-Term Memory
CIC Canadian Institute for Cybersecurity
CIC-IoT-2023 Jeu de données IoT du Canadian Institute for Cybersecurity (2023)
DL Deep Learning (Apprentissage profond)
DoS Denial of Service
DDoS Distributed Denial of Service
FL Federated Learning (Apprentissage fédéré)
FN False Negative
FP False Positive
GAN Generative Adversarial Network
IDS Intrusion Detection System
IoT Internet of Things (Internet des Objets)
LIME Local Interpretable Model-agnostic Explanations
LSTM Long Short-Term Memory
ML Machine Learning (Apprentissage automatique)
ROC Receiver Operating Characteristic
RF Random Forest
SHAP SHapley Additive exPlanations
SVM Support Vector Machine
TN True Negative
TP True Positive
XAI Explainable Artificial Intelligence (Intelligence artificielle explicable)
XGBoost Extreme Gradient Boosting

DÉDICACE

À la mémoire de mon père, Ahmadou Mbathie, et de mes défuntes sœurs (Maimouna et Khoradia), dont l'amour, les valeurs et les sacrifices continuent de guider chacun de mes pas. J'espère que ce travail honore leur mémoire et reflète l'éducation qu'ils m'ont transmise.

À ma mère, Aissatou Camara, et à ma deuxième mère, Ya Ndeye Thioye, pour leur soutien indéfectible, leur patience, leurs prières et leur confiance, qui m'ont permis de persévérer malgré les épreuves.

À ma famille : mes sœurs (Ayda, Nguénar et Nogaye), mes frères (Modou, Cheikh et Macodou), mon neveu (Mamadou Lamine Thiaw) ainsi que mes nièces, pour leurs encouragements constants, leur affection et leur présence, qui ont été une source de force tout au long de mon parcours académique.

À mon directeur de recherche, le Professeur Fehmi Jaafar, pour sa confiance, son encadrement scientifique et son accompagnement tout au long de ce travail.

À toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce document, par leur soutien, leurs conseils ou leurs encouragements.

Qu'ALLAH leur accorde la santé, la prospérité et Sa miséricorde.

REMERCIEMENTS

Je souhaite exprimer, en premier lieu, ma profonde gratitude envers ALLAH pour la force, la patience et la santé qui m'ont accompagné tout au long de la réalisation de ce mémoire. Sa guidance m'a permis de persévérer dans les moments les plus exigeants de ce parcours.

J'adresse ensuite mes sincères remerciements à mon directeur de recherche, le professeur Fehmi Jaafar, pour la qualité remarquable de son encadrement. Son expertise, sa rigueur scientifique et ses conseils méthodologiques ont joué un rôle déterminant dans l'orientation et l'aboutissement de ce travail. Sa disponibilité, ses retours constructifs et sa vision académique ont constitué pour moi une source d'inspiration et un cadre d'apprentissage exceptionnel.

Je tiens également à remercier l'ensemble des professeurs rencontrés au cours de mon cheminement universitaire. Leurs enseignements, leurs encouragements et leur engagement pédagogique ont contribué à structurer ma compréhension des sciences informatiques et à nourrir ma curiosité intellectuelle.

Je souhaite également exprimer ma reconnaissance envers le personnel du Département d'informatique et de mathématiques (DIM) de l'UQAC, pour leur soutien administratif, leur disponibilité et le climat professionnel qu'ils contribuent à maintenir.

Je tiens également à exprimer ma profonde reconnaissance aux personnes qui m'ont accueilli et accompagné lors de mon arrivée au Canada. À Serigne Saliou Diouf, qui m'a chaleureusement accueilli à Montréal dès mon premier jour ; à Ramata Ndiaye et Samba Kébé, qui m'ont accueilli durant mes premiers jours à Chicoutimi ; à El Hadj Ahmad Barro Ndieguène et son épouse Ndeye Sarr Ndieguène, qui ont facilité mon intégration ; ainsi qu'à Abdou Aziz Mbodj pour ses précieux conseils. Je remercie également Ousmane Sine pour sa bienveillance et ses conseils, ainsi que son épouse. J'adresse également mes remerciements au professeur Souleymane Barry, président de la dahira tidjaniya de Chicoutimi, pour ses orientations et ses précieux conseils. Leur générosité et leur solidarité ont grandement facilité mon intégration et marqué positivement mon parcours.

Enfin, je remercie chaleureusement ma famille, mon guide spirituel, Serigne ASS Mouhamadou Ndieguène, mes amis et mes proches pour leur soutien indéfectible. Leur patience, leurs encouragements et leur présence bienveillante ont été essentiels pour maintenir l'équilibre nécessaire à la poursuite de mes études.

AVANT-PROPOS

Ce mémoire de maîtrise marque l'aboutissement d'un parcours académique à la fois exigeant et enrichissant, consacré à l'exploration de thématiques au cœur des enjeux actuels en informatique, notamment la cybersécurité et l'intelligence artificielle. À travers les enseignements suivis et les travaux de recherche réalisés, j'ai pu acquérir des compétences solides en développement d'applications, en sécurité des systèmes informatiques et, plus particulièrement, en intelligence artificielle appliquée à la détection des cybermenaces.

Ce travail s'inscrit dans le cadre de l'analyse des méthodes hybrides d'apprentissage profond pour la détection des cyberattaques dans l'Internet des Objets (IoT). L'objectif principal de ce mémoire est d'évaluer, de manière comparative et rigoureuse, l'apport des architectures hybrides combinant apprentissage profond et méthodes classiques de machine learning pour la détection multi-classes des cyberattaques dans des environnements IoT complexes. Il vise également à mettre en évidence les limites et les compromis liés à la complexité des modèles, à leur comportement de classification et à leur capacité de généralisation.

Dans un contexte où les systèmes basés sur l'intelligence artificielle sont de plus en plus utilisés pour la cybersécurité, ce travail accorde une attention particulière à la compréhension du comportement des modèles et à la nécessité d'évaluations transparentes et reproductibles. L'étude s'appuie sur des jeux de données publics et sur une méthodologie expérimentale unifiée, afin de fournir une analyse de référence utile tant pour la recherche académique que pour les praticiens du domaine.

Mon parcours de recherche a été jalonné de défis, de périodes de stress et d'obstacles techniques, mais également de découvertes enrichissantes et de résultats significatifs. Chaque chapitre de ce mémoire reflète les nombreuses heures consacrées à la collecte des données, à l'expérimentation, à l'analyse des résultats et à la réflexion critique, dans un effort constant pour comprendre les mécanismes sous-jacents à la détection des cyberattaques dans l'Internet des Objets.

Je prends l'entière responsabilité des propos, des analyses et des conclusions présentés dans ce document. Conscient toutefois que tout travail scientifique peut être amélioré, je resterai reconnaissant pour toute remarque, critique ou suggestion susceptible de contribuer à l'enrichissement et à l'approfondissement de cette étude.

CHAPITRE I
INTRODUCTION GÉNÉRALE

L'Internet des Objets (IoT) constitue aujourd'hui l'un des piliers majeurs de la transformation numérique mondiale. L'intégration massive de capteurs intelligents, d'actionneurs, de dispositifs connectés et d'applications distribuées a favorisé l'émergence de nouveaux environnements opérationnels dans les secteurs de la santé, de l'industrie, de la logistique, de l'agriculture et des infrastructures critiques. Cette croissance rapide s'accompagne d'une augmentation significative de la complexité des flux réseau et des surfaces d'attaque, rendant les systèmes IoT particulièrement vulnérables Hou *et al.* (2024); Lee *et al.* (2023).

Malgré leur importance stratégique, les dispositifs IoT demeurent fortement contraints sur les plans computationnel, énergétique et logiciel, et reposent souvent sur des mécanismes de communication hétérogènes ou insuffisamment sécurisés. Les travaux récents montrent que ces limitations exposent les réseaux IoT à diverses familles de cyberattaques, incluant les campagnes de déni de service distribué, les intrusions furtives, les attaques par reconnaissance et les botnets massifs Antonakakis *et al.* (2017); Falliere *et al.* (2012); Mohurle & Patil (2017). Dans ce contexte, la capacité à identifier et analyser efficacement ces comportements malveillants constitue un enjeu central pour assurer la résilience et la fiabilité des environnements connectés.

Pour répondre à ces menaces, de nombreuses approches de classification et d'analyse comportementale ont été développées. Les méthodes classiques d'apprentissage automatique, telles que Random Forest, Support Vector Machines ou XGBoost, ont démontré une efficacité notable dans certains scénarios, notamment lorsque les données sont bien struc-

turées et correctement prétraitées Owusu *et al.* (2024). Toutefois, ces approches reposent largement sur des caractéristiques extraites manuellement et peinent à capturer la nature séquentielle et dynamique des attaques modernes.

Les modèles d'apprentissage profond offrent une alternative prometteuse pour analyser des flux réseau de grande dimension. Les modèles convolutionnels (CNN), récurrents (LSTM) et les combinaisons hybrides capables de traiter simultanément les relations spatiales et temporelles se révèlent particulièrement performants pour distinguer des comportements malveillants subtils ou évolutifs Xu *et al.* (2024); Lee *et al.* (2023); Javed *et al.* (2025). Ces avancées justifient l'étude de méthodes hybrides capables de modéliser la complexité croissante des cyberattaques IoT.

Dans ce mémoire, nous proposons une analyse comparative complète de plusieurs méthodes d'apprentissage automatique et profond appliquées à la détection de cyberattaques dans des environnements IoT. Les expériences reposent sur le dataset CIC-IoT-2023, un benchmark récent intégrant diverses catégories d'attaques représentatives des menaces contemporaines. Les modèles étudiés incluent Random Forest et Logistic Regression, ainsi que des approches profondes telles que CNN, CNN–LSTM et Stacked CNN–LSTM.

Les résultats obtenus montrent que les méthodes hybrides exploitant conjointement des représentations spatiales et temporelles surpassent les approches traditionnelles dans la détection de cyberattaques complexes. L'analyse des matrices de confusion et des courbes ROC met en évidence les forces et limites de chaque modèle, notamment concernant la détection des attaques de reconnaissance, des attaques par force brute et des comportements se rapprochant du trafic légitime. De plus, ce travail intègre des considérations contemporaines telles que la confidentialité différentielle, l'apprentissage distribué et l'explicabilité, essen-

tielles pour une détection fiable dans des environnements IoT réels et distribués Rahman *et al.* (2024); Khan *et al.* (2024); Zhou *et al.* (2023).

1.1 PROBLÉMATIQUE SCIENTIFIQUE

La problématique centrale de ce mémoire peut être formulée comme suit :

> Comment concevoir et optimiser des modèles hybrides d'apprentissage profond capables de détecter efficacement des cyberattaques dans des environnements IoT massifs, hétérogènes et déséquilibrés, tout en maintenant une robustesse face à des menaces variées et évolutives ?

Cette question soulève plusieurs défis majeurs, en particulier la présence d'un déséquilibre important entre classes, la variabilité des flux réseau, la nécessité de capturer simultanément des dépendances spatiales et temporelles et les contraintes computationnelles inhérentes aux dispositifs IoT.

1.2 QUESTIONS DE RECHERCHE

Les questions de recherche qui guident ce travail sont les suivantes :

- Les méthodes hybrides CNN–LSTM surpassent-elles les modèles CNN ou LSTM simples dans la détection de cyberattaques IoT ?
- Comment les modèles classiques tels que Random Forest et Logistic Regression se comparent-ils aux approches profondes sur le dataset CIC-IoT-2023 ?
- Quelles catégories d'attaques sont correctement détectées et lesquelles demeurent difficiles à identifier ?

- Quels facteurs influencent le plus les performances : prétraitement, équilibrage, profondeur du modèle ou choix méthodologique ?
- Comment expliquer les décisions des modèles dans un contexte IoT distribué ?

1.3 OBJECTIFS DU MÉMOIRE

OBJECTIF GÉNÉRAL

Le présent travail vise à développer, entraîner et évaluer des méthodes hybrides d'apprentissage profond pour la détection de cyberattaques dans les environnements IoT en utilisant le dataset CIC-IoT-2023.

OBJECTIFS SPÉCIFIQUES

- mettre en place un pipeline complet de prétraitement, normalisation et équilibrage ;
- transformer les données en séquences exploitables par les modèles profonds ;
- concevoir et entraîner les modèles CNN, CNN–LSTM et Stacked CNN–LSTM ;
- comparer les approches profondes aux modèles traditionnels ;
- analyser les performances à travers matrices de confusion et courbes ROC ;
- discuter les limites, implications et perspectives offertes par les approches étudiées.

1.4 CONTRIBUTIONS DU MÉMOIRE

Les principales contributions de ce mémoire sont les suivantes :

- développement d'un pipeline complet de traitement et d'équilibrage des données IoT ;

- conception de méthodes hybrides CNN–LSTM et Stacked CNN–LSTM adaptées au trafic séquentiel ;
- évaluation comparative entre modèles d'apprentissage automatique et d'apprentissage profond ;
- analyses détaillées mettant en évidence les attaques difficiles à classifier ;
- intégration des enjeux contemporains liés à la confidentialité, à l'apprentissage fédéré et à l'explicabilité.

1.5 ORGANISATION DU MÉMOIRE

Ce document est structuré comme suit :

- Le Chapitre 1 présente le contexte général, la problématique et les contributions.
- Le Chapitre 2 expose les concepts fondamentaux utiles à la compréhension du travail.
- Le Chapitre 3 propose une revue structurée des travaux connexes portant sur la détection de cyberattaques IoT.
- Le Chapitre 4 décrit le pipeline méthodologique, le dataset CIC-IoT-2023, les étapes de prétraitement et les modèles étudiés.
- Le Chapitre 5 présente l'ensemble des résultats expérimentaux.
- Le Chapitre 6 analyse les forces, limites et implications des modèles évalués.
- Le Chapitre 7 synthétise les résultats et propose des pistes futures.

CHAPITRE II

CONCEPTS FONDAMENTAUX

Ce chapitre présente les fondements théoriques nécessaires à la compréhension de ce mémoire. Il expose d'abord les principes de l'Internet des Objets et les menaces auxquelles ces systèmes sont exposés, puis introduit les approches utilisées pour analyser et détecter les cyberattaques dans les environnements IoT. Les méthodes d'apprentissage automatique et d'apprentissage profond les plus courantes sont ensuite décrites, ainsi que les modèles combinés exploitant conjointement les relations spatiales et temporelles des flux réseau. Enfin, des notions transversales telles que l'apprentissage fédéré, la confidentialité différentielle et l'explicabilité des modèles sont abordées.

2.1 SÉCURITÉ DE L'INTERNET DES OBJETS

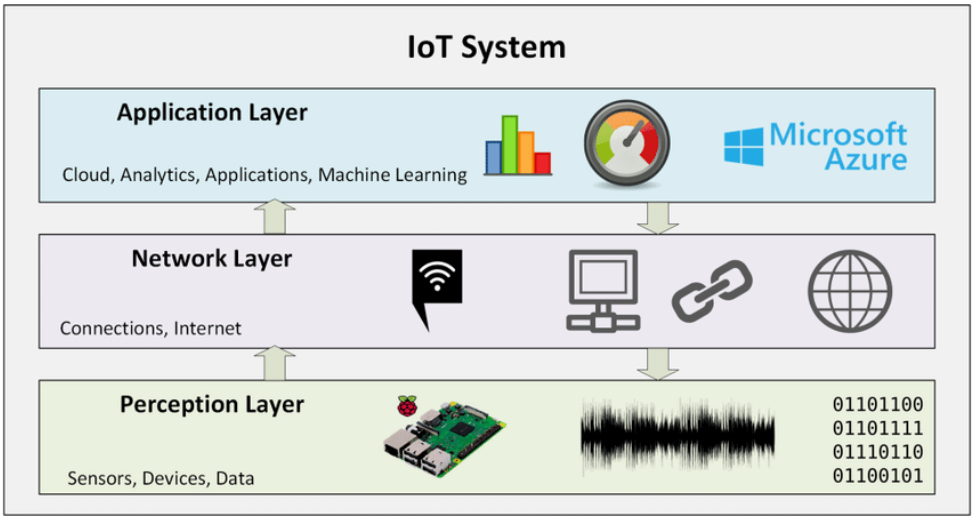

FIGURE 2.1 : Modèle en trois couches de l'Internet des Objets, adapté de Ayadi (2019).

L'Internet des Objets regroupe un ensemble d'équipements connectés capables de collecter, transmettre et parfois analyser des données. Il constitue aujourd'hui un envi-

ronnement hétérogène où coexistent capteurs industriels, dispositifs médicaux, objets domestiques intelligents et systèmes embarqués. Cette diversité rend l'écosystème IoT particulièrement difficile à sécuriser. En pratique, un même réseau IoT peut regrouper des objets utilisant des piles logicielles différentes, des mécanismes de communication variés, et des fréquences d'échantillonnage hétérogènes. Cette variété se reflète directement dans le trafic réseau, dont la distribution peut changer selon l'heure, l'usage, la charge, ou l'état des appareils.

L'augmentation massive du nombre de dispositifs connectés s'accompagne d'une multiplication des vecteurs de cyberattaques. Plusieurs travaux ont montré que les objets connectés souffrent d'une faible capacité de calcul, de mécanismes de mise à jour limités et d'une absence de normalisation Zolanvari *et al.* (2019). Ces faiblesses favorisent l'émergence d'incidents majeurs. Parmi les plus connus, les campagnes Mirai ont permis de détourner des milliers d'appareils pour mener des attaques distribuées à grande échelle Antonakakis *et al.* (2017). D'autres analyses ont mis en évidence les vulnérabilités ayant conduit à la propagation de WannaCry Mohurle & Patil (2017), ainsi que les mécanismes d'infiltration de Stuxnet ciblant des systèmes industriels Falliere *et al.* (2012). Ces événements illustrent que les environnements IoT ne sont pas seulement exposés à des attaques opportunistes, mais aussi à des attaques structurées, parfois longues, combinant reconnaissance, compromission progressive et perturbation des services.

Ces événements démontrent la fragilité structurelle des environnements IoT et soulignent la nécessité de développer des méthodes analytiques capables de repérer et comprendre des comportements malveillants variés malgré la diversité des équipements. Dans ce cadre, il devient essentiel de définir des classes d'attaque représentatives et adaptées au trafic IoT moderne, afin d'évaluer la capacité des modèles à distinguer des comportements parfois très proches sur le plan statistique.

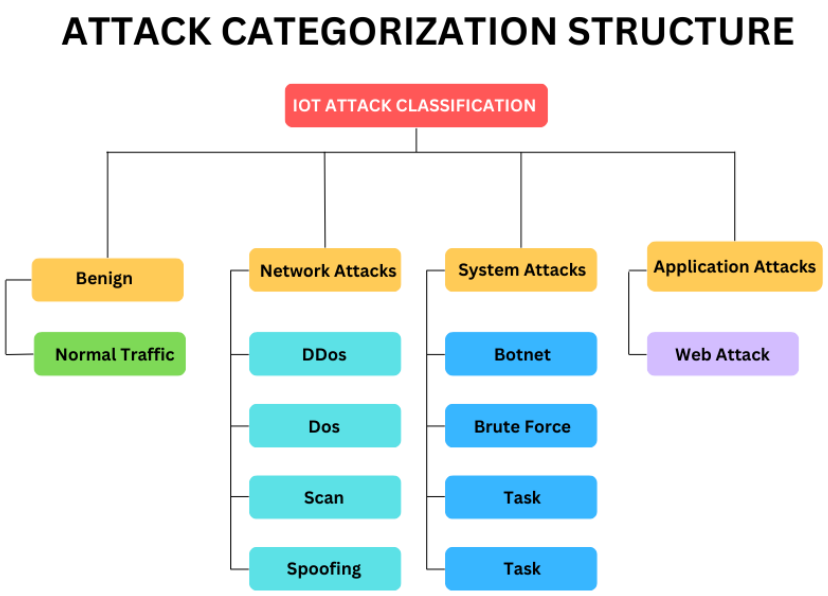

FIGURE 2.2 : Structure de catégorisation des cyberattaques IoT selon les huit classes étudiées et équilibrées dans ce mémoire.

La Figure 2.2 présente la structure de catégorisation retenue dans ce mémoire. Elle met en évidence les huit classes sur lesquelles repose l'analyse, et qui correspondent aux classes effectivement équilibrées lors du prétraitement. Le choix de ces classes répond à un objectif méthodologique précis : couvrir un spectre réaliste de cyberattaques observées dans des environnements IoT, tout en conservant une granularité suffisante pour analyser les confusions interclasses et la sensibilité des modèles aux attaques rares.

La classe *Benign* (ou *Normal Traffic*) est indispensable, car elle représente le comportement légitime de référence. Dans un réseau IoT, le trafic normal n'est pas uniforme : il varie selon les objets, les fréquences d'émission, et les services utilisés. Conserver une classe bénigne bien définie permet d'évaluer la capacité des modèles à limiter les fausses

alertes. Cette dimension est critique, car un système qui signale trop d'événements normaux comme des attaques devient inutilisable en contexte réel.

Les classes *DoS* et *DDoS* ont été retenues séparément car, bien qu'elles partagent un objectif commun de perturbation de service, elles présentent des dynamiques différentes. Une attaque DoS correspond souvent à une source unique ou limitée qui sature une ressource cible, tandis qu'une attaque DDoS implique un ensemble distribué d'émetteurs, généralement un botnet. Dans les réseaux IoT, cette distinction est importante car les schémas de trafic, la dispersion des adresses sources et la variabilité temporelle peuvent différer. Conserver les deux classes permet de vérifier si le modèle apprend des indices robustes ou s'il confond ces attaques à cause de signatures statistiquement proches.

La classe *Scan* correspond aux activités de reconnaissance. Elle est particulièrement pertinente en IoT car de nombreuses attaques débutent par une phase d'exploration, visant à identifier des ports ouverts, des services accessibles ou des dispositifs vulnérables. Les scans peuvent être rapides ou au contraire étalés dans le temps, ce qui les rend parfois difficiles à distinguer d'un trafic bénin, notamment quand des objets communiquent de manière intermittente. Inclure cette classe permet de tester la sensibilité des modèles à des attaques discrètes, souvent précurseurs d'attaques plus graves.

La classe *Spoofing* représente l'usurpation d'identité, un scénario fréquent dans les réseaux IoT où l'authentification peut être faible. Le spoofing peut prendre plusieurs formes, comme l'usurpation d'adresses, la falsification de paquets, ou l'imitation d'un comportement légitime. Cette classe est essentielle car elle met à l'épreuve la capacité des modèles à détecter des manipulations visant à ressembler au trafic normal, ce qui est un défi majeur en classification supervisée.

La classe *Botnet* a été retenue car elle correspond à une menace structurante du paysage IoT, illustrée par des cas comme Mirai Antonakakis *et al.* (2017). Un botnet ne se résume pas à un seul type de flux : il peut inclure des phases d'infection, de commande et contrôle, et de participation à des attaques coordonnées. Cette diversité interne rend la classe botnet complexe, mais aussi très représentative des menaces modernes. La présence de cette classe permet d'évaluer si le modèle généralise correctement face à des comportements composites.

La classe *Brute Force* a été retenue car l'IoT est fortement exposé aux tentatives d'authentification répétées, en raison de mots de passe faibles, de configurations par défaut et de services exposés. Les attaques par force brute produisent parfois un trafic régulier et répétitif, mais peuvent aussi être confondues avec des erreurs légitimes ou des tentatives d'accès normales. Cette classe teste donc la capacité des modèles à exploiter des patterns fins et à réduire les faux négatifs dans des scénarios d'intrusion progressive.

Enfin, la classe *Web Attack* représente les attaques au niveau applicatif. Même si certains objets IoT sont simples, de nombreux environnements IoT s'appuient sur des interfaces web, des tableaux de bord, des passerelles ou des API. Les attaques web permettent d'exploiter des failles applicatives, de voler des informations ou de prendre le contrôle de services. Inclure cette classe élargit l'évaluation au-delà des attaques purement volumétriques et introduit une dimension applicative essentielle pour refléter la réalité des plateformes IoT modernes.

Ainsi, la sélection de ces huit classes vise à garantir une couverture équilibrée entre trafic normal, attaques de disponibilité, reconnaissance, usurpation, compromission distribuée, intrusions par authentification et attaques applicatives. Cette catégorisation

permet également d'analyser des confusions pertinentes, notamment entre DoS et DDoS, entre scan et trafic bénin intermittent, ou encore entre spoofing et comportements légitimes.

2.2 ANALYSE ET DÉTECTION DES CYBERATTAQUES DANS L'IOT

La détection des cyberattaques vise à identifier des activités anormales ou malveillantes dans le trafic réseau. Dans un contexte IoT, l'enjeu est double : détecter des attaques fréquentes et volumétriques, tout en restant sensible à des attaques plus discrètes. Les attaques de reconnaissance, de spoofing ou de brute force peuvent générer des signaux faibles, difficiles à distinguer d'un trafic normal varié. Cette difficulté explique l'intérêt de méthodes capables d'apprendre des régularités complexes et de généraliser à des comportements changeants.

Historiquement, deux approches ont dominé les systèmes de détection :

Les méthodes *basées sur signatures* comparent le trafic observé à une base de motifs connus. Elles offrent une très bonne précision pour les attaques déjà documentées, mais échouent à identifier des comportements nouveaux, polymorphes ou modifiés. Cette limite est particulièrement critique dans les environnements IoT, où les menaces évoluent rapidement Zolanvari *et al.* (2019). De plus, ces méthodes nécessitent une mise à jour continue des signatures, ce qui est difficile à maintenir à grande échelle dans des infrastructures distribuées.

Les méthodes *fondées sur anomalies* apprennent un modèle du comportement normal du réseau, puis détectent les écarts. Elles sont plus adaptatives, mais souffrent souvent d'un taux élevé de faux positifs, en raison de la forte variabilité des flux IoT Boukhris *et al.* (2023). Leur déploiement réel est également complexe, car la définition d'un comportement « normal » est difficile dans un environnement hétérogène. En IoT, un comportement normal

peut changer avec une simple mise à jour d'objet, une modification de fréquence d'émission, ou une nouvelle application.

L'évolution des menaces et la complexité croissante des flux IoT ont conduit la communauté scientifique à se tourner vers l'apprentissage automatique et l'apprentissage profond pour renforcer la capacité de détection. Les méthodes modernes utilisent des représentations dérivées directement des données, permettant une meilleure adaptation aux comportements changeants. Des modèles plus avancés exploitent également les propriétés temporelles des flux, ce qui est essentiel pour identifier des séquences d'actions malveillantes organisées.

Enfin, plusieurs travaux explorent des approches combinées permettant de réduire les faux positifs, d'améliorer la détection des attaques rares et de mieux gérer les environnements IoT distribués Vu Dinh *et al.* (2023). Dans ce mémoire, le fait d'équilibrer explicitement les huit classes de la Figure 2.2 permet d'éviter qu'un modèle obtienne de bonnes performances globales uniquement en privilégiant les classes majoritaires, et favorise une évaluation plus juste, centrée sur la capacité de généralisation et la robustesse interclasse.

2.3 APPRENTISSAGE AUTOMATIQUE APPLIQUÉ À L'ANALYSE DES CYBERATTAQUES

L'apprentissage automatique constitue une première étape efficace pour analyser les flux réseau IoT. Des modèles tels que Random Forest, Support Vector Machines ou XGBoost sont fréquemment utilisés Owusu *et al.* (2024). Ils offrent de bons résultats lorsque les caractéristiques sont soigneusement extraites. Dans un cadre IoT, ces modèles

sont particulièrement appréciés car ils s'adaptent bien aux données tabulaires issues de statistiques de flux et peuvent être entraînés relativement rapidement.

Cependant, plusieurs limites demeurent :

- ils ne prennent pas en compte les dépendances temporelles entre paquets ;
- leurs performances dépendent fortement de la qualité de l'ingénierie des caractéristiques ;
- ils sont vulnérables au déséquilibre des jeux de données IoT, où certaines attaques sont très rares Sun *et al.* (2023).

Ces limites motivent l'usage croissant de méthodes profondes capables d'apprendre automatiquement des représentations plus robustes, notamment lorsque les frontières entre classes sont subtiles, comme pour la reconnaissance ou le spoofing.

2.4 MÉTHODES PROFONDES POUR LA DÉTECTION DES CYBERATTAQUES

L'apprentissage profond s'est imposé comme une approche essentielle pour analyser les cyberattaques IoT. Contrairement aux méthodes classiques, les modèles profonds apprennent eux-mêmes des représentations discriminantes. Ils sont donc moins dépendants des étapes manuelles de sélection des caractéristiques, et peuvent capturer des interactions complexes entre variables qui seraient difficiles à modéliser explicitement.

CNN : EXTRACTION LOCALE DE MOTIFS

Les modèles CNN sont particulièrement performants pour extraire des régularités locales dans les flux réseau. Lee et al. Lee *et al.* (2023) montrent qu'ils surpassent souvent les modèles classiques dans la détection de comportements subtils. Leur principal atout

réside dans leur capacité à capturer des motifs discriminants même lorsque le trafic comporte du bruit. Leur temps d'apprentissage reste raisonnable, ce qui facilite leur utilisation dans des environnements contraints. Leur limite principale concerne l'absence de prise en compte explicite des dépendances temporelles, ce qui peut réduire la sensibilité aux attaques progressives.

LSTM : MODÉLISATION DES SÉQUENCES

Les modèles LSTM ont été introduits pour capturer la dynamique temporelle des flux. Ils sont particulièrement efficaces pour identifier des attaques progressives, programmées en séquence ou réparties dans le temps Xu *et al.* (2024). Cependant, leur coût computationnel élevé et leur besoin en données séquentielles propres limitent parfois leur utilisation dans les environnements IoT, surtout lorsque les séquences sont courtes ou irrégulières.

MODÈLES GÉNÉRATIFS

Les autoencodeurs variationnels et les GAN ont émergé comme outils puissants pour détecter des anomalies rares. Ils apprennent à modéliser le comportement normal du trafic, ce qui permet d'identifier les écarts subtils. CTVAE Vu Dinh *et al.* (2023) reconstruit efficacement les séquences normales. S2CGAN Wang *et al.* (2023) améliore la détection d'attaques minoritaires en générant des données synthétiques. Ces modèles offrent de bonnes performances mais nécessitent des ressources importantes et un réglage délicat.

2.5 COMBINAISONS CNN–LSTM POUR L'ANALYSE DES CYBERATTAQUES

Les approches combinant convolution et traitement séquentiel exploitent simultanément les motifs locaux et la dynamique temporelle. Elles obtiennent généralement de

meilleures performances dans les environnements IoT où les cyberattaques présentent des schémas complexes et évolutifs Javed *et al.* (2025). Les versions empilées renforcent encore cette capacité en superposant plusieurs niveaux d'extraction.

2.6 CONFIDENTIALITÉ, APPRENTISSAGE DISTRIBUÉ ET EXPLICABILITÉ

Les environnements IoT sont distribués, hétérogènes et souvent sensibles en matière de données. La détection des cyberattaques doit donc être compatible avec des exigences fortes de confidentialité, d'interprétabilité et de distribution du calcul.

APPRENTISSAGE FÉDÉRÉ

L'apprentissage fédéré permet d'entraîner un modèle global sans centraliser les données. Il réduit considérablement les risques de fuite d'information Khan *et al.* (2024) tout en permettant d'apprendre à partir de comportements locaux différents Mukherjee *et al.* (2024).

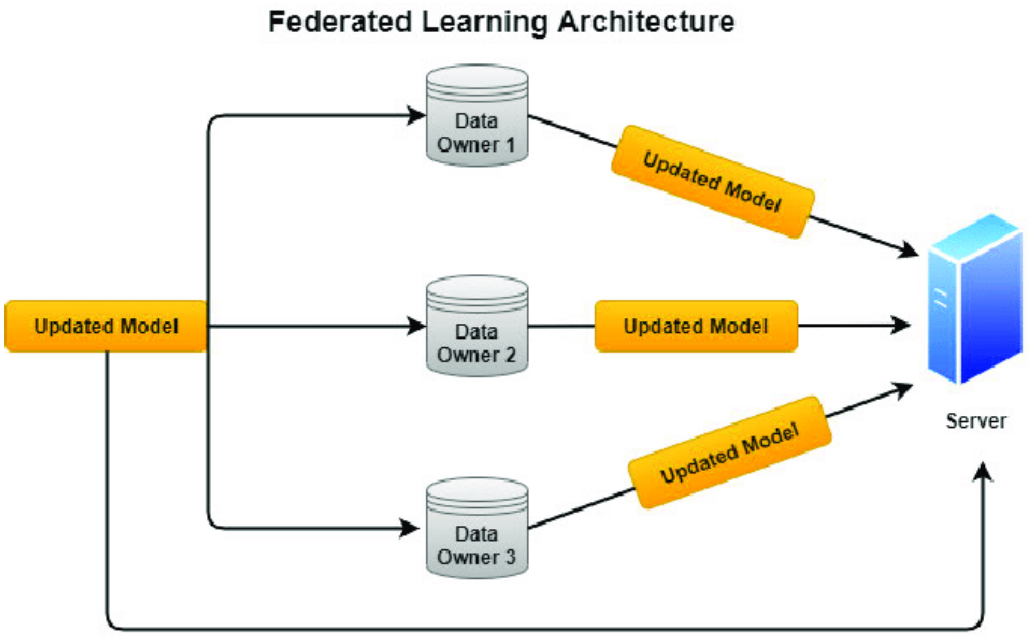

FIGURE 2.3 : Fonctionnement général de l'apprentissage fédéré. Khan *et al.* (2024)

CONFIDENTIALITÉ DIFFÉRENTIELLE

La confidentialité différentielle protège les données personnelles en ajoutant un bruit contrôlé aux paramètres. Elle limite les attaques d'inférence sans dégrader excessivement la performance Zhou *et al.* (2023). Dans un contexte IoT, cette dimension est importante car les données peuvent provenir de capteurs domestiques, industriels ou médicaux, et peuvent révéler des informations sensibles sur les comportements ou l'environnement.

EXPLICABILITÉ

L'explicabilité est indispensable pour comprendre pourquoi un modèle décide qu'un trafic est malveillant ou non. Rahimi et al. Rahimi *et al.* (2024) montrent que des méthodes comme SHAP ou LIME facilitent l'analyse et renforcent la confiance dans les modèles. Dans un cadre d'évaluation multi-classes, l'explicabilité aide également à analyser les confusions interclasses, par exemple lorsque des attaques de reconnaissance sont confon-

dues avec du trafic normal, ou lorsque des comportements de spoofing imitent des activités légitimes.

CHAPITRE III

TRAVAUX CONNEXES

Ce chapitre rassemble les principaux travaux qui ont marqué l'évolution récente de la détection des cyberattaques dans les environnements IoT. Il met en relief les approches retenues par la communauté scientifique, leurs apports respectifs ainsi que les limites observées dans différents contextes d'évaluation. Cette revue permet de situer l'étude présentée dans ce mémoire au regard des avancées actuelles et d'identifier les orientations méthodologiques qui guident le choix des modèles examinés dans les chapitres suivants.

3.1 ÉVOLUTION DES APPROCHES DE DÉTECTION DES CYBERATTAQUES DANS LES ENVIRONNEMENTS IOT

La croissance spectaculaire des dispositifs connectés a stimulé un intérêt majeur pour la détection des cyberattaques dans les environnements IoT. Les premiers travaux ont principalement adapté des mécanismes issus des systèmes de sécurité réseau traditionnels utilisés dans les réseaux d'entreprise, reposant sur l'analyse statistique du trafic et sur des modèles supervisés relativement simples. Zolanvari et al. Zolanvari *et al.* (2019) montrent toutefois que ces approches initiales présentent d'importantes limites : faible capacité à gérer des volumes massifs de données, sensibilité au bruit, performances dégradées en présence de classes déséquilibrées et difficulté à détecter des cyberattaques émergentes dont les signatures ne sont pas encore documentées.

L'évolution rapide des menaces IoT a profondément orienté les recherches vers des solutions plus intelligentes et adaptatives. Les cyberattaques de grande ampleur ou furtives mettent en évidence la nécessité de dépasser les approches basées sur des signatures

statiques. Plusieurs auteurs soulignent que le trafic IoT peut présenter des motifs très subtils, distribués ou évolutifs, rendant inefficaces les mécanismes traditionnels. Hou et al. Hou *et al.* (2024) montrent par exemple que la détection moderne des cyberattaques doit intégrer des modèles capables d'adapter dynamiquement leurs paramètres en fonction des variations comportementales du réseau. De même, la thèse de Gharbi Gharbi (2024) illustre la difficulté pour les approches conventionnelles de suivre des comportements d'attaque polymorphes ou faiblement corrélés à des indicateurs statistiques simples.

Cette dynamique a favorisé une transition vers des modèles d'apprentissage profond, mieux adaptés pour capturer des motifs complexes dans des environnements hétérogènes. Lee et al. Lee *et al.* (2023) démontrent que les architectures neuronales surpassent systématiquement les modèles traditionnels grâce à leur capacité d'extraction automatique de caractéristiques discriminantes. Srinivas et al. Srinivas *et al.* (2024) mettent en avant, à travers une analyse approfondie, que les approches hybrides combinant des mécanismes convolutionnels et séquentiels représentent aujourd'hui l'une des pistes les plus prometteuses pour capter simultanément les motifs locaux et la dynamique temporelle du trafic IoT.

Cette transition marque une étape essentielle dans l'évolution des approches modernes de détection des cyberattaques, qui cherchent désormais à allier précision, adaptabilité et robustesse dans des environnements massivement distribués et hautement variables.

3.2 APPROCHES FONDÉES SUR L'APPRENTISSAGE AUTOMATIQUE TRADITIONNEL

Les méthodes classiques d'apprentissage automatique ont constitué une étape importante dans l'évolution des approches de détection des cyberattaques IoT. Ces approches

reposent généralement sur des modèles tels que Random Forest, SVM, Decision Trees, AdaBoost ou encore XGBoost, et ont montré une efficacité notable dans l'analyse de flux réseau structurés. Leur succès initial s'explique par leur capacité à traiter des données tabulaires à haute dimension et à modéliser des relations non linéaires entre caractéristiques.

Owusu et al. Owusu *et al.* (2024) démontrent que l'algorithme XGBoost se distingue particulièrement dans la détection de cyberattaques de type brute force, reconnaissance ou spoofing. Leur étude met en évidence l'importance cruciale de l'ingénierie des caractéristiques : les performances des modèles classiques dépendent en grande partie de la qualité des attributs extraits manuellement, ce qui limite leur capacité à s'adapter automatiquement à des motifs complexes.

Sun et al. Sun *et al.* (2023) confirment également que les méthodes d'ensemble surpassent généralement les modèles linéaires dès lors que les flux IoT présentent des variations locales marquées. Cependant, ces approches ne capturent pas les dépendances temporelles essentielles pour identifier des cyberattaques évolutives ou coordonnées, ce qui constitue une limitation majeure dans les environnements IoT modernes.

Un autre défi mis en avant dans la littérature concerne le déséquilibre important des jeux de données IoT. Zolanvari et al. Zolanvari *et al.* (2019) montrent que les modèles classiques ont tendance à privilégier les classes majoritaires, entraînant une sous-détection des cyberattaques rares telles que les reconnaissances silencieuses ou certaines formes de spoofing. Cette sensibilité au déséquilibre, combinée à l'absence de mécanismes pour modéliser la dynamique temporelle, explique pourquoi les approches traditionnelles atteignent rapidement leurs limites face à la complexité croissante des réseaux IoT.

3.3 MODÈLES PROFONDS POUR LA DÉTECTION DES CYBERATTAQUES IOT : TENDANCES ET AVANCÉES

L'arrivée des architectures profondes a marqué un tournant dans la détection des cyberattaques. Contrairement aux modèles traditionnels, les réseaux neuronaux apprennent automatiquement des représentations complexes à partir des données brutes. Cette capacité est particulièrement pertinente pour les environnements IoT, où les relations entre les caractéristiques peuvent être non linéaires, hiérarchiques et fortement dépendantes du contexte.

3.3.1 RÉSEAUX CNN ET EXTRACTION HIÉRARCHIQUE DES CARACTÉRISTIQUES

Les CNN (Convolutional Neural Networks) ont été parmi les premiers modèles profonds appliqués à la détection des cyberattaques IoT. Leur force principale réside dans la capacité à extraire automatiquement des motifs locaux robustes. Lee et al. Lee *et al.* (2023) montrent que les CNN surpassent systématiquement les modèles traditionnels dans plusieurs benchmarks IoT, notamment grâce à l'utilisation de filtres multi-échelles capables de capturer les variations fines du trafic. Cette capacité d'extraction hiérarchique des caractéristiques permet d'obtenir des représentations compactes et discriminantes, particulièrement adaptées aux environnements IoT où les flux sont souvent bruités, irréguliers et fortement hétérogènes.

Plusieurs travaux confirment également que les CNN offrent une excellente scalabilité lors du traitement de grands volumes de données, ce qui en fait des candidats efficaces pour les systèmes de détection des cyberattaques déployés en périphérie réseau. Gharbi Gharbi (2024) souligne par exemple que les CNN réduisent considérablement la dépendance à

l'ingénierie manuelle des caractéristiques, une étape souvent lourde dans les approches traditionnelles. Toutefois, malgré leur efficacité structurelle, une limite persistante concerne leur incapacité à modéliser explicitement la dimension temporelle. Lorsque les motifs d'attaque évoluent sur plusieurs paquets successifs, les CNN tendent à perdre l'information séquentielle, ce qui réduit leur performance face à des cyberattaques progressives, furtives ou polymorphes. C'est cette limitation majeure qui a motivé l'émergence d'approches combinées intégrant des mécanismes séquentiels ou attentionnels.

3.3.2 RÉSEAUX LSTM ET MODÉLISATION DES DÉPENDANCES TEMPORELLES

Les réseaux LSTM ont été introduits pour pallier les limites des architectures strictement convolutionnelles, en particulier leur incapacité à capturer la dynamique temporelle des flux réseau. Xu et al. Xu *et al.* (2024) démontrent que les LSTM, utilisés seuls ou combinés à des CNN, permettent de modéliser la structure séquentielle du trafic IoT, améliorant la détection de cyberattaques telles que les floods progressifs, les reconnaissances lentes ou les scans étalés sur plusieurs intervalles temporels. Leur capacité à mémoriser des dépendances de moyen et long terme en fait des modèles particulièrement adaptés aux environnements où les variations d'un paquet à l'autre constituent un indice clé d'activité malveillante.

D'autres travaux récents, comme ceux de Gharbi Gharbi (2024), confirment l'intérêt des approches récurrentes pour capturer des motifs temporels subtils, notamment lorsque les cyberattaques ne présentent pas de caractéristiques discriminantes immédiates mais émergent progressivement dans le temps. Cependant, plusieurs limites ont été observées. Les modèles LSTM nécessitent souvent davantage de ressources computationnelles, un temps d'entraînement plus long, et une régularisation soigneuse pour éviter les phénomènes

de surapprentissage. De plus, la nature souvent courte, bruitée et irrégulière des séquences IoT réduit la capacité du LSTM à exploiter pleinement son potentiel, car la profondeur temporelle disponible est parfois insuffisante pour révéler des dépendances significatives. Ces défis ont encouragé l'exploration de modèles combinés et de variantes plus légères, mieux adaptées aux contraintes des dispositifs IoT.

3.3.3 MODÈLES GÉNÉRATIFS ET AUTOENCODEURS

Les modèles génératifs ont récemment suscité un intérêt croissant dans la détection des cyberattaques IoT en raison de leur capacité à identifier des comportements anormaux même lorsqu'ils disposent de peu d'échantillons d'attaques. Vu Dinh et al. Vu Dinh *et al.* (2023) introduisent CTVAE, un autoencodeur variationnel compact conçu pour modéliser les dépendances temporelles des séquences réseau. En apprenant à reconstruire fidèlement les flux normaux, CTVAE permet de détecter de manière efficace les écarts subtils, ce qui le rend particulièrement pertinent pour les environnements IoT où les cyberattaques peuvent être rares mais critiques.

Wang et al. Wang *et al.* (2023) proposent S2CGAN, un GAN semi-supervisé capable de générer des échantillons synthétiques pour renforcer la représentation des classes minoritaires. Cette approche améliore la détection des cyberattaques rares et réduit la sensibilité au déséquilibre massif typique des jeux de données IoT. D'autres travaux connexes, tels que ceux de Ramirez et al. Ramirez *et al.* (2024), confirment l'intérêt des modèles génératifs pour augmenter artificiellement les données et stabiliser l'entraînement des classifieurs profonds.

Malgré leur potentiel, ces modèles présentent certaines limites. Leur coût computationnel reste élevé, ce qui complique leur déploiement sur des dispositifs IoT à faible

puissance. De plus, la qualité de la génération dépend fortement de l'équilibre entre reconstruction et diversité des données produites. Ces contraintes encouragent l'exploration de variantes plus légères et de stratégies combinées associant modèles génératifs et approches séquentielles.

3.4 APPROCHES HYBRIDES CNN–LSTM : CONTRIBUTIONS RÉCENTES

Les approches combinées CNN–LSTM représentent aujourd'hui l'une des pistes les plus étudiées dans la littérature dédiée à la détection des cyberattaques IoT. Elles visent à exploiter simultanément l'extraction de motifs locaux et la modélisation temporelle des flux réseau.

Srinivas et al. Srinivas *et al.* (2024) proposent un panorama complet des approches combinées et montrent qu'elles surpassent la plupart des modèles simples, notamment sur des cyberattaques complexes. Leur étude met également en évidence la sensibilité de ces méthodes aux paramètres d'entraînement, tels que la longueur des séquences d'entrée et le degré d'équilibrage des classes.

Javed et al. Javed *et al.* (2025) introduisent une approche CNN–LSTM adaptative, capable d'ajuster dynamiquement sa profondeur en fonction de la complexité du trafic. Cette stratégie améliore de manière notable le F1-score de plusieurs catégories de cyberattaques IoT difficiles.

Xu et al. Xu *et al.* (2024) montrent également que l'hybridation améliore la précision tout en réduisant les faux positifs, mais soulignent que la performance dépend fortement du choix des séquences et de la qualité du prétraitement.

3.5 SÉCURITÉ, CONFIDENTIALITÉ ET APPRENTISSAGE DISTRIBUÉ

Avec la croissance des systèmes IoT distribués, l'attention de la recherche s'est progressivement élargie aux enjeux de confidentialité et de sécurité des modèles de détection des cyberattaques eux-mêmes. La centralisation des données, longtemps considérée comme une étape incontournable de l'entraînement, expose désormais les systèmes à des risques élevés de fuite d'informations sensibles et d'attaques visant à extraire des caractéristiques du trafic. Pour répondre à ces défis, de nouvelles approches privilégient des mécanismes d'apprentissage réparti capables de préserver la confidentialité tout en maintenant une capacité de détection élevée. L'apprentissage fédéré, par exemple, permet à plusieurs dispositifs de collaborer sans partager leurs données brutes, réduisant ainsi la surface d'attaque. Parallèlement, des stratégies de protection fondées sur l'ajout contrôlé de bruit ou sur des filtrages locaux renforcent la résilience des modèles contre les attaques d'inférence. Ces évolutions témoignent d'un mouvement croissant vers des systèmes de détection des cyberattaques plus robustes, capables d'opérer dans des environnements distribués, hétérogènes et soumis à des contraintes strictes de confidentialité.

3.5.1 APPRENTISSAGE FÉDÉRÉ

Khan et al. Khan *et al.* (2024) démontrent que l'apprentissage fédéré permet de former un système global de détection des cyberattaques sans centraliser les données, réduisant ainsi le risque de fuite d'information. Mukherjee et al. Mukherjee *et al.* (2024) montrent également que la fédération améliore la scalabilité et la résilience des systèmes distribués de détection des cyberattaques. Au-delà de ces avantages, plusieurs travaux soulignent que l'apprentissage fédéré constitue une réponse directe aux contraintes inhérentes aux systèmes IoT, où les ressources computationnelles sont limitées et les communications coûteuses. L'entraînement local sur les dispositifs permet de diminuer la charge réseau,

tout en tenant compte des particularités du trafic propre à chaque nœud. La littérature met aussi en lumière des défis persistants, tels que l'hétérogénéité des données locales, la synchronisation des modèles et la vulnérabilité aux attaques de type empoisonnement. Malgré ces limites, l'apprentissage fédéré s'impose comme une voie prometteuse pour développer des systèmes distribués de détection des cyberattaques conciliant performance, confidentialité et déploiement en environnements réels.

3.5.2 CONFIDENTIALITÉ DIFFÉRENTIELLE

Zhou et al. Zhou *et al.* (2023) introduisent des mécanismes de confidentialité différentielle adaptés aux modèles de détection des cyberattaques, permettant de protéger les données sensibles tout en préservant des performances acceptables. Cette approche est particulièrement pertinente pour les environnements impliquant des données personnelles. En pratique, la confidentialité différentielle consiste à ajouter un bruit contrôlé aux gradients ou aux paramètres du modèle afin d'empêcher la réidentification ou l'inférence sur les données d'entraînement. Les travaux récents montrent que ces mécanismes renforcent significativement la robustesse face aux attaques d'extraction de modèles et aux analyses inverses. Cependant, l'ajout de bruit entraîne un compromis entre niveau de protection et précision du modèle, ce qui nécessite un réglage fin des budgets de confidentialité. Dans les systèmes IoT, où les données sont particulièrement hétérogènes et souvent sensibles, cette technique constitue une solution viable pour concilier apprentissage efficace et exigence de protection réglementaire.

3.5.3 EXPLICABILITÉ DES MODÈLES PROFONDS

Plusieurs travaux récents se concentrent également sur l'explicabilité des systèmes de détection des cyberattaques basés sur des modèles profonds. Rahimi et al. Rahimi *et al.* (2024) montrent que les outils d'explicabilité (SHAP, LIME, gradients intégrés) sont essentiels pour interpréter les décisions de modèles complexes et détecter d'éventuels comportements inattendus. Leur étude souligne que l'explicabilité est indispensable pour le déploiement opérationnel des modèles, notamment dans des contextes critiques tels que l'industrie ou la santé. En offrant une visibilité sur les attributs qui influencent les prédictions, ces techniques permettent d'améliorer la confiance des analystes, de diagnostiquer les erreurs et de repérer des biais potentiels. Elles jouent également un rôle important dans la validation réglementaire et l'audit de sécurité, en facilitant la démonstration de conformité. Malgré leurs avantages, les méthodes explicatives souffrent encore de limites, notamment en termes de stabilité, de granularité et de coût computationnel, ce qui encourage la poursuite des recherches dans ce domaine.

3.6 LIMITES OBSERVÉES DANS LA LITTÉRATURE

L'analyse de l'état de l'art met en évidence un ensemble de limites structurelles qui traversent la majorité des approches existantes pour la détection des cyberattaques dans les environnements IoT. Un premier constat concerne la forte sensibilité des modèles au déséquilibre des classes. Les attaques rares, souvent essentielles en contexte opérationnel comme les attaques de reconnaissance silencieuse ou les variantes discrètes de spoofing, demeurent difficiles à détecter en raison de leur faible représentation dans les jeux de données. Cela entraîne une chute marquée du rappel et peut conduire à une incapacité totale à identifier certaines catégories d'attaques critiques. Les modèles d'ensemble tels

que Random Forest ou XGBoost permettent d'atténuer légèrement ce phénomène, mais leur efficacité devient limitée lorsque l'hétérogénéité des données atteint un niveau élevé.

Une autre limite récurrente observée dans la littérature est la prise en compte insuffisante des contraintes computationnelles propres aux dispositifs IoT. De nombreux travaux se concentrent sur des modèles profonds très coûteux en mémoire et en puissance de calcul, comme les réseaux convolutionnels profonds, les approches hybrides ou les autoencodeurs variationnels. Cependant, les objets connectés disposent souvent de ressources extrêmement limitées, ce qui rend leur déploiement difficile, voire impossible. Cette absence de réalisme limite fortement la transférabilité des résultats de la recherche vers des scénarios concrets.

Les approches hybrides, incluant notamment les combinaisons CNN et LSTM, présentent également une dépendance marquée aux hyperparamètres. La profondeur du réseau, la taille des filtres, la fenêtre temporelle utilisée ou encore les techniques de régularisation influencent de manière importante la stabilité de l'entraînement et les performances obtenues. Plusieurs études rapportent des variations importantes de résultats selon les configurations, ce qui complique la reproductibilité et empêche une comparaison équitable entre modèles.

De plus, la littérature comporte encore peu de travaux analysant de manière systématique le dataset CIC IoT 2023, pourtant l'un des plus récents et des plus complets pour la détection des cyberattaques IoT. Les études antérieures portent principalement sur des jeux de données plus anciens tels que CIC IDS 2017 ou sur des jeux synthétiques. Cette rareté limite la compréhension des défis spécifiques posés par CIC IoT 2023, notamment la forte corrélation entre les caractéristiques et la structure non linéaire des distributions.

Enfin, l'analyse des erreurs interclasses reste souvent superficielle. Les confusions fines, telles que celles observées entre les attaques DoS et DDoS ou entre différentes formes

de reconnaissance, sont rarement examinées en profondeur. Pourtant, ces situations représentent des cas fréquents et problématiques dans les réseaux IoT réels. Plus généralement, les enjeux liés au déploiement opérationnel sont peu étudiés. La latence, la consommation énergétique, la capacité de mise à jour en ligne ou la robustesse face aux attaques visant directement le modèle sont des aspects encore largement sous-explorés.

- forte sensibilité des modèles au déséquilibre des classes
- faible prise en compte des contraintes computationnelles propres aux dispositifs IoT
- dépendance élevée aux hyperparamètres pour les approches hybrides
- rareté des travaux évaluant de manière systématique le dataset CIC IoT 2023
- manque d'analyse fine des erreurs interclasses telles que DoS contre DDoS ou reconnaissance subtile
- faible prise en compte des enjeux de déploiement réel

3.7 POSITIONNEMENT DU PRÉSENT MÉMOIRE

Au regard des limites identifiées dans la littérature, le présent mémoire se positionne comme une contribution complémentaire et analytique aux travaux existants sur la détection des cyberattaques dans les environnements IoT. Alors que de nombreuses études se concentrent sur une famille restreinte de modèles ou sur des configurations expérimentales spécifiques, ce travail adopte une approche comparative systématique, intégrant à la fois des modèles classiques d'apprentissage automatique et des architectures profondes hybrides combinant des mécanismes convolutifs et séquentiels.

L'originalité de l'approche proposée réside dans la mise en place d'un pipeline méthodologique rigoureux et homogène, conçu pour garantir une comparaison équitable entre des modèles de nature différente. Ce pipeline inclut un prétraitement approfondi des

données, une normalisation cohérente des caractéristiques, une réduction de corrélation afin de limiter les redondances informationnelles, ainsi qu'un équilibrage strict des classes. Ces choix méthodologiques visent à isoler l'impact réel des architectures de modélisation sur la détection des cyberattaques, indépendamment des biais induits par la distribution initiale des données.

Le travail s'appuie sur le dataset récent CIC IoT 2023, encore peu exploré dans les études contemporaines, ce qui constitue une contribution pertinente au regard de la complexité de ce jeu de données. Ce dataset se distingue par la diversité de ses scénarios de cyberattaques, la forte corrélation entre certaines caractéristiques et la structure non linéaire des distributions, représentant ainsi un cadre d'évaluation exigeant et plus proche de conditions réalistes de trafic IoT. L'exploitation de ce jeu de données permet d'analyser le comportement des modèles dans un contexte plus représentatif des environnements actuels.

Une analyse approfondie des performances est menée à travers l'étude des matrices de confusion, des courbes ROC et d'un examen systématique des erreurs spécifiques à chaque modèle. Cette analyse fine permet non seulement de comparer les métriques globales, mais également de mettre en évidence les confusions interclasses récurrentes et les limites structurelles propres à chaque approche de détection. Elle contribue ainsi à une meilleure compréhension des mécanismes internes des modèles et de leur capacité à distinguer des cyberattaques aux signatures proches.

Le mémoire se distingue également par l'exploration explicite des limites des architectures profondes dans un contexte de contraintes computationnelles et d'entraînement accéléré. Ce choix expérimental vise à refléter des conditions plus proches de celles rencontrées dans les environnements IoT réels, où les ressources sont limitées et où les modèles doivent souvent être entraînés ou mis à jour dans des délais restreints. Cette perspective

pragmatique permet d'évaluer non seulement la performance brute des modèles, mais aussi leur viabilité potentielle dans des scénarios opérationnels.

- il propose une évaluation comparative approfondie de modèles classiques et de modèles hybrides CNN–LSTM pour la détection des cyberattaques IoT ;
- il met en œuvre un pipeline expérimental rigoureux intégrant prétraitement, normalisation, réduction de corrélation et équilibrage strict des classes ;
- il s'appuie sur le dataset récent CIC IoT 2023, encore peu étudié dans la littérature ;
- il fournit une analyse détaillée fondée sur les matrices de confusion et les courbes ROC afin d'identifier les confusions interclasses ;
- il explore les limites structurelles des architectures profondes dans des conditions d'entraînement accélérées et contraintes.

Cette analyse positionne clairement la contribution du mémoire au sein des travaux existants sur la détection des cyberattaques dans l'Internet des Objets et prépare la transition vers le Chapitre 4, consacré à la description détaillée de la méthodologie expérimentale adoptée.

CHAPITRE IV
MÉTHODOLOGIE

Ce chapitre décrit de manière détaillée la méthodologie adoptée pour concevoir, entraîner et évaluer des modèles dédiés à la **détection des cyberattaques dans l'Internet des Objets (IoT)**. Contrairement aux approches classiques centrées exclusivement sur la détection d'intrusions, ce travail adopte une vision plus large et plus représentative des menaces actuelles en considérant explicitement les différentes catégories de cyberattaques ciblant les réseaux et dispositifs IoT.

Dans ce travail, les termes *détection d'intrusions* et *détection des cyberattaques* sont utilisés de manière complémentaire, la première étant considérée comme un cas particulier de la seconde dans le contexte spécifique des environnements IoT.

L'approche méthodologique s'articule autour d'un ensemble d'étapes structurées allant de l'exploitation du jeu de données CIC-IoT-2023 jusqu'à l'analyse comparative des performances de plusieurs familles de modèles. Ces familles incluent des modèles classiques d'apprentissage automatique (Random Forest, XGBoost), ainsi que des modèles d'apprentissage profond capables de capturer des relations complexes et temporelles dans les flux réseau, notamment les CNN, CNN–LSTM et Stacked CNN–LSTM. L'objectif est de mettre en place une démarche rigoureuse, reproductible et adaptée aux contraintes spécifiques des environnements IoT, telles que l'hétérogénéité des dispositifs, la variabilité du trafic et le déséquilibre entre les classes de cyberattaques.

Afin de structurer clairement cette démarche, un workflow méthodologique global a été défini. Celui-ci permet de visualiser l'enchaînement logique des différentes étapes,

depuis les données brutes jusqu'à l'analyse finale des résultats. La Figure 4.1 présente ce workflow général, qui constitue le fil conducteur de l'ensemble de ce chapitre.

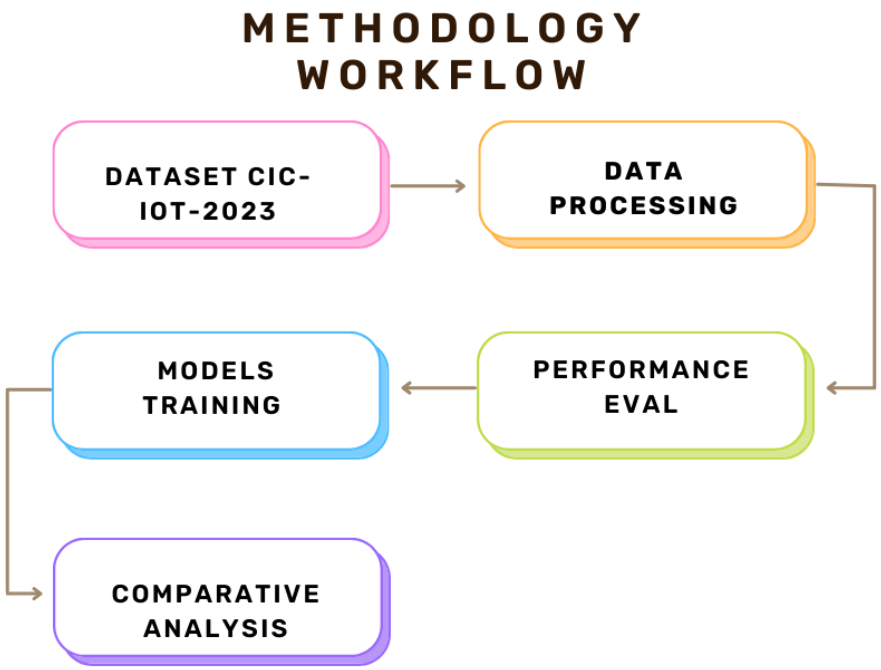

FIGURE 4.1 : Workflow méthodologique pour la détection des cyberattaques IoT.

Ce workflow met en évidence cinq phases principales et interdépendantes : l'exploitation du jeu de données CIC-IoT-2023, le traitement et la préparation des données, l'entraînement des modèles de détection, l'évaluation des performances et l'analyse comparative des résultats. Cette structuration méthodologique vise à refléter fidèlement le cycle complet de conception d'un système de détection des cyberattaques adapté aux environnements IoT, depuis l'acquisition des données brutes jusqu'à l'interprétation finale des performances obtenues.

La première phase repose sur l'exploitation du dataset CIC-IoT-2023, sélectionné pour sa richesse, sa diversité et son caractère représentatif des environnements IoT modernes. Ce jeu de données regroupe des flux réseau issus de dispositifs hétérogènes et inclut plusieurs catégories de cyberattaques réalistes, ce qui en fait une base pertinente pour l'étude de mécanismes de détection robustes. La seconde phase concerne le traitement et la préparation des données, étape essentielle pour corriger les imperfections inhérentes aux données réelles, telles que le bruit, les valeurs manquantes, les incohérences typologiques et le déséquilibre entre classes. Cette phase conditionne directement la stabilité et la capacité de généralisation des modèles de détection.

La troisième phase est dédiée à l'entraînement des modèles, incluant aussi bien des approches classiques d'apprentissage automatique que des architectures profondes capables de capturer des motifs complexes et des dynamiques temporelles propres aux cyberattaques IoT. La quatrième phase, centrée sur l'évaluation des performances, permet de mesurer de manière quantitative l'efficacité des modèles à détecter correctement les comportements malveillants tout en limitant les erreurs critiques. Enfin, la phase d'analyse comparative vise à confronter les résultats obtenus afin d'identifier les architectures les plus adaptées aux contraintes spécifiques des environnements IoT. L'ensemble de ce workflow garantit une démarche méthodologique cohérente, reproductible et orientée vers une détection fiable et opérationnelle des cyberattaques dans l'Internet des Objets.

4.1 VUE D'ENSEMBLE DU PIPELINE MÉTHODOLOGIQUE

La méthodologie adoptée repose sur un pipeline séquentiel structurant la détection et la classification des cyberattaques dans les environnements IoT. Ce pipeline vise à transformer des flux réseau bruts, fortement hétérogènes et bruités, en représentations exploitables permettant d'identifier des comportements malveillants variés.

Quatre grandes étapes structurent ce pipeline :

- le prétraitement et le nettoyage des données issues du trafic IoT ;
- la transformation, la normalisation et l'équilibrage statistique des classes de cyberattaques ;
- l'entraînement de modèles d'apprentissage automatique et d'apprentissage profond ;
- l'évaluation comparative des performances à l'aide de métriques adaptées aux données déséquilibrées.

Cette organisation garantit une transition cohérente entre les différentes phases du workflow présenté précédemment. Elle permet également d'assurer que chaque étape repose sur des données correctement préparées, limitant ainsi les biais expérimentaux et facilitant la reproductibilité des expériences. Le pipeline général suivi dans ce travail est illustré à la Figure 4.2.

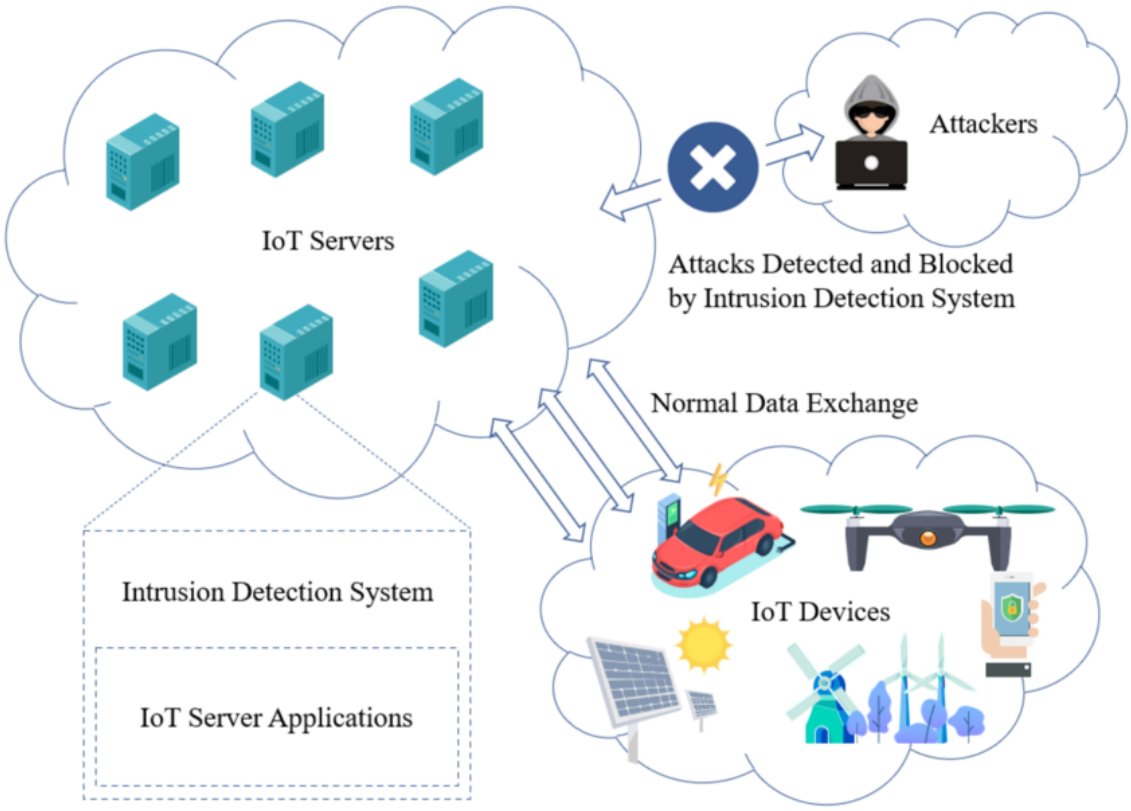

FIGURE 4.2 : Pipeline général pour la détection des cyberattaques dans un environnement IoT, adapté de Zolanvari *et al.* (2019).

4.2 PRÉPARATION ET PRÉTRAITEMENT DES DONNÉES

Le jeu de données CIC-IoT-2023 regroupe des flux réseau générés dans des environnements IoT réalistes, intégrant une grande diversité de dispositifs, de protocoles et de scénarios d'attaque. Comme c'est le cas pour la majorité des jeux de données issus de contextes réels, il présente plusieurs défis majeurs : présence de bruit, valeurs manquantes, incohérences de format et déséquilibre important entre les différentes catégories de cyberattaques.

Le prétraitement des données constitue une étape fondamentale du workflow méthodologique. Il vise à corriger ces irrégularités afin de fournir une base de données cohérente, stable et exploitable par les algorithmes d'apprentissage supervisé. Cette étape est parti-

culièrement critique pour les modèles d'apprentissage profond, dont la convergence et la capacité de généralisation dépendent fortement de la qualité des données d'entrée.

4.2.1 PIPELINE DE PRÉTRAITEMENT

La Figure 4.3 illustre le pipeline de prétraitement appliqué au jeu de données CIC-IoT-2023. Celui-ci regroupe plusieurs opérations successives permettant de transformer les données brutes en matrices d'apprentissage adaptées aux modèles étudiés.

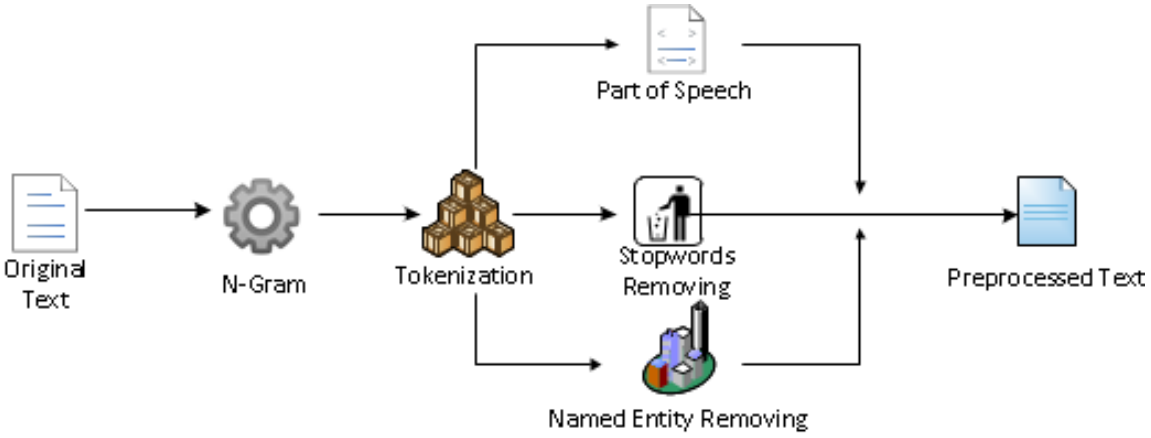

FIGURE 4.3 : Pipeline de prétraitement des données appliqué au jeu CIC-IoT-2023 Hou *et al.* (2024).

Ce pipeline inclut la conversion typologique des attributs, le nettoyage des valeurs aberrantes, la normalisation des caractéristiques numériques, l'encodage des étiquettes de classes et la réduction de la redondance entre variables.

4.2.2 NETTOYAGE ET TRANSFORMATION DES DONNÉES

Le nettoyage débute par la conversion de l'ensemble des attributs vers des formats numériques compatibles avec les algorithmes d'apprentissage. Les valeurs non convertibles sont supprimées afin de garantir une cohérence statistique lors de l'analyse. Les lignes

contenant des valeurs manquantes (*NaN*) ou infinies ($\pm\infty$) sont également éliminées afin d'éviter toute instabilité numérique lors de l'entraînement des modèles profonds.

Les caractéristiques restantes sont normalisées à l'aide de la méthode *StandardScaler*, permettant d'obtenir des distributions centrées et réduites. Cette normalisation est essentielle pour stabiliser l'apprentissage des couches convolutionnelles et récurrentes. Les étiquettes de classes sont ensuite encodées sous forme d'entiers afin d'être correctement interprétées par les algorithmes de classification.

Une analyse de corrélation basée sur le coefficient de Pearson est également réalisée afin d'identifier les attributs redondants. Les caractéristiques présentant une corrélation supérieure à 0.95 sont supprimées, conformément aux recommandations de travaux récents sur la détection des cyberattaques IoT Vu Dinh *et al.* (2023); Abdulaziz *et al.* (2023). Cette étape contribue à réduire la dimensionnalité et à limiter les risques de surapprentissage.

4.2.3 ÉQUILIBRAGE DES CLASSES DE CYBERATTAQUES

Le jeu de données CIC-IoT-2023 présente un déséquilibre massif entre les différentes catégories de cyberattaques, certaines classes telles que DoS ou DDoS étant largement surreprésentées par rapport à des attaques plus rares comme le spoofing, les attaques applicatives ou les tentatives de force brute.

Dans un contexte de détection des cyberattaques, ce déséquilibre peut conduire les modèles à privilégier les classes dominantes et à négliger des menaces pourtant critiques. Afin de pallier ce problème, une stratégie d'équilibrage contrôlée basée sur un *random undersampling* des classes majoritaires a été appliquée. Cette approche permet de réduire les biais statistiques tout en conservant des échantillons représentatifs des comportements malveillants.

À l'issue de cette étape, le jeu de données est équilibré selon huit classes représentant les principales familles de cyberattaques IoT :

- Benign ;
- DDoS ;
- DoS ;
- Scan ;
- Spoofing ;
- Web Attack ;
- Brute Force ;
- Botnet.

Cet équilibrage est indispensable pour garantir que les modèles apprennent effectivement les motifs discriminants associés à chaque type de cyberattaque, y compris celles qui sont faiblement représentées dans les données brutes.

4.3 MODÉLISATION ET CLASSIFICATION DES CYBERATTAQUES

Dans le cadre de la **détection des cyberattaques dans l'Internet des Objets**, la phase de modélisation consiste à définir une stratégie cohérente permettant d'associer un flux réseau observé à une classe de comportement (bénin ou malveillant) et, le cas échéant, à une **catégorie d'attaque** précise. Cette étape ne se limite pas à l'application d'un classificateur : elle implique de structurer la problématique de classification, de clarifier la nature des signaux disponibles dans les données, et de positionner les approches retenues par rapport aux grandes familles méthodologiques reconnues dans la littérature.

Les systèmes de détection sont généralement décrits selon trois grandes catégories : les approches basées sur signatures, celles basées sur anomalies et les approches hybrides.

Cette taxonomie est largement reprise dans les travaux de synthèse portant sur les environnements IoT, car elle permet de distinguer les mécanismes qui recherchent des motifs explicites d'attaque de ceux qui apprennent un comportement de référence et détectent les écarts Boukhris *et al.* (2023). La Figure 4.4 présente cette classification conceptuelle.

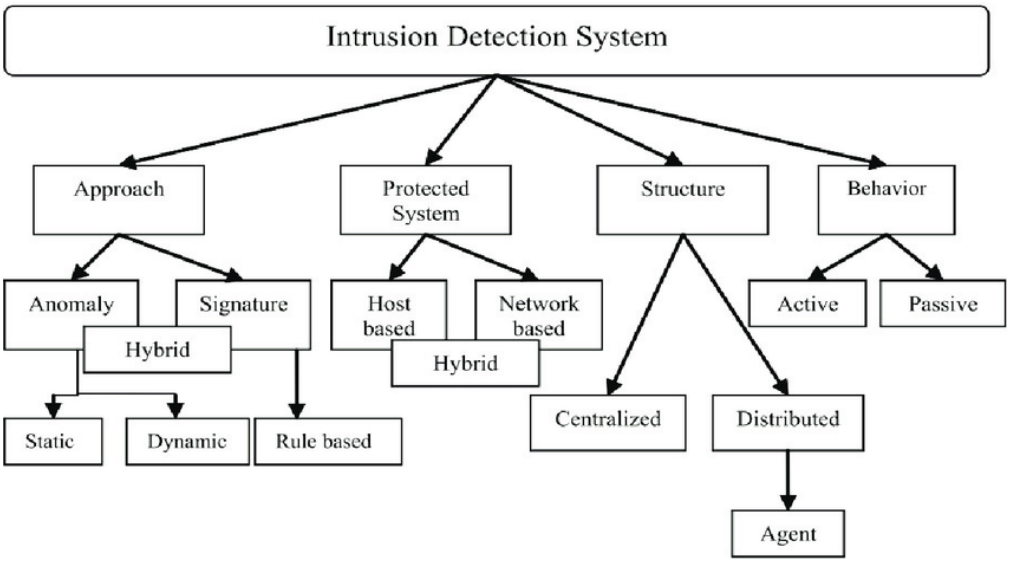

FIGURE 4.4 : Catégories principales des approches de détection appliquées aux cyberattaques (signatures, anomalies, approches hybrides) Boukhris *et al.* (2023).

Les méthodes basées sur signatures reposent sur des règles ou motifs connus et sont très performantes pour reconnaître des attaques déjà documentées, mais elles peinent à généraliser face à des variantes, des attaques polymorphes ou des comportements émergents. À l'inverse, les approches basées sur anomalies modélisent le comportement normal et identifient les déviations ; elles sont potentiellement plus adaptatives, mais peuvent générer un taux de faux positifs plus élevé, en particulier dans les environnements IoT où le trafic est hautement variable. Les approches hybrides cherchent à combiner les avantages des deux familles en associant robustesse face aux attaques inconnues et meilleure maîtrise des fausses alertes.

Dans ce mémoire, la stratégie retenue s'inscrit dans une logique **hybride** au sens méthodologique, dans la mesure où les modèles étudiés combinent des mécanismes complémentaires d'extraction de motifs et d'apprentissage de dépendances, afin d'améliorer la discrimination entre plusieurs catégories de cyberattaques. Plus précisément, une attention particulière est portée aux modèles capables d'exploiter à la fois (i) des **motifs locaux** (corrélations entre attributs d'un même segment de trafic) et (ii) des **structures temporelles** (évolution des flux sur une fenêtre de temps). Cette combinaison est particulièrement pertinente pour l'IoT, où certaines attaques sont caractérisées par des répétitions, des vagues de trafic ou des séquences d'actions distribuées dans le temps (p. ex. scans progressifs, botnets, force brute, attaques par déni de service).

4.4 ENTRAÎNEMENT DES MODÈLES

L'entraînement des modèles constitue une étape centrale du workflow méthodologique présenté en début de chapitre, car il conditionne directement la capacité des algorithmes à apprendre des représentations discriminantes du trafic IoT. Dans le contexte de la **détection des cyberattaques**, l'objectif n'est pas uniquement d'obtenir une bonne performance globale, mais aussi de garantir une capacité de généralisation sur des attaques aux signatures proches, sur des classes minoritaires et sur des comportements susceptibles d'évoluer.

Trois grandes familles de modèles ont été évaluées afin de couvrir un spectre représentatif de stratégies de détection : (i) des modèles d'apprentissage automatique classiques robustes sur données tabulaires, (ii) des modèles d'apprentissage profond orientés extraction automatique de représentations, et (iii) des variantes combinées permettant de mieux exploiter la structure temporelle des flux :

- **Modèles ML classiques** : Random Forest, XGBoost Owusu *et al.* (2024). Ces modèles sont choisis pour leur efficacité sur des caractéristiques tabulaires, leur capacité à gérer des frontières de décision non linéaires, et leur coût computationnel généralement plus faible. Ils constituent également une base de comparaison solide pour mesurer l'apport réel des modèles profonds.
- **Modèles DL** : CNN, CNN–LSTM et Stacked CNN–LSTM Xu *et al.* (2024). Les CNN ciblent l'extraction de motifs locaux et de corrélations entre caractéristiques, alors que les composantes récurrentes (LSTM) visent à modéliser la dynamique temporelle des flux, en particulier lorsque des attaques se manifestent sous forme de séquences ou de répétitions dans le temps.
- **Modèles hybrides optimisés** : architectures récentes inspirées de variantes légères (p. ex. LightCNN) Pham *et al.* (2024) et de méthodes adaptatives de fusion temporelle Javed *et al.* (2025). Ces approches sont considérées afin d'évaluer dans quelle mesure des mécanismes de fusion plus riches ou des modèles plus compacts peuvent améliorer la détection de cyberattaques complexes, notamment lorsqu'elles sont faiblement représentées.

L'objectif de cette diversité est de comparer de manière rigoureuse la performance des modèles lorsque ceux-ci sont confrontés à un dataset IoT hétérogène, comportant des interactions complexes entre caractéristiques, et présentant un déséquilibre important entre classes. Dans ce contexte, la comparaison doit être menée sur une base expérimentale homogène : mêmes partitions d'entraînement/validation/test, mêmes données prétraitées, et mêmes métriques d'évaluation, afin de s'assurer que les différences observées reflètent les capacités intrinsèques des modèles plutôt que des effets liés au traitement des données.

Pour les modèles d'apprentissage profond, l'entraînement suit des pratiques standardisées visant à assurer stabilité, reproductibilité et convergence. Tous les modèles DL ont été optimisés à l'aide de l'algorithme Adam, reconnu pour sa capacité à adapter dynamiquement le taux d'apprentissage en fonction des gradients, et à accélérer la convergence dans des espaces de grande dimension. Afin de limiter le surapprentissage et d'améliorer la généralisation, des mécanismes de régularisation ont été intégrés : le *dropout* réduit la co-adaptation des neurones en désactivant aléatoirement une fraction des unités pendant l'entraînement, tandis que la *batch normalization* stabilise la distribution interne des activations, ce qui améliore la robustesse et la vitesse de convergence.

Un critère d'*early stopping* a été appliqué afin d'éviter un entraînement excessif, en interrompant automatiquement l'apprentissage lorsque la performance sur le jeu de validation cesse de s'améliorer. Cette stratégie est particulièrement pertinente dans les données IoT où certaines classes peuvent induire un surapprentissage rapide, notamment lorsque des motifs dominants apparaissent fréquemment. En complément, un ajustement adaptatif du taux d'apprentissage via un scheduler permet de raffiner progressivement les paramètres du modèle lorsque la phase d'optimisation atteint un plateau.

Pour les modèles exploitant la dimension temporelle (CNN–LSTM et Stacked CNN–LSTM), une attention particulière est portée au format des données d'entrée. Les observations sont restructurées sous forme de séquences, ce qui permet aux unités LSTM de capturer des dépendances à court et moyen terme dans le trafic (p. ex. répétitions, changements progressifs, bursts), caractéristiques de plusieurs familles de cyberattaques. Dans les versions empilées, la superposition de plusieurs blocs convolutionnels renforce d'abord l'extraction hiérarchique de motifs, puis les couches récurrentes apprennent la dynamique temporelle associée à ces motifs.

Enfin, l'entraînement des modèles est effectué sur des données prétraitées, normalisées et équilibrées, afin que les performances observées reflètent la capacité réelle des algorithmes à distinguer les classes de cyberattaques et non des biais induits par la distribution initiale. Cette configuration garantit des comparaisons équitables entre les modèles, et permet d'analyser plus finement les forces et limites de chaque famille, en particulier sur les classes rares ou difficiles à distinguer.

4.5 MÉTRIQUES DE PERFORMANCE

L'évaluation des modèles de **détection des cyberattaques dans l'Internet des Objets** repose sur un ensemble de métriques quantitatives permettant d'analyser, sous différents angles, la capacité d'un classificateur à distinguer correctement les flux bénins des activités malveillantes. Dans les environnements IoT, cette étape est particulièrement critique en raison de la forte hétérogénéité du trafic, du déséquilibre marqué entre les classes et de la présence d'attaques discrètes ou progressives. Les métriques constituent ainsi un socle méthodologique essentiel pour juger non seulement de la performance globale des modèles, mais également de leur fiabilité opérationnelle face à des scénarios réalistes de cyberattaques Rahman *et al.* (2024).

Dans ce mémoire, quatre mesures fondamentales sont utilisées : l'*Accuracy*, la *Precision*, le *Recall* et le *F1-score*. Chacune de ces métriques capture un aspect complémentaire du comportement des modèles et leur interprétation conjointe est indispensable pour évaluer de manière équilibrée les performances. En particulier, dans le contexte IoT, certaines cyberattaques (telles que le spoofing, les scans discrets ou les attaques par force brute à faible intensité) sont rares ou difficilement distinguables des flux légitimes, ce qui impose une analyse fine au-delà d'une simple mesure globale de justesse.

DÉFINITIONS MATHÉMATIQUES

Accuracy : mesure globale de justesse

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

L'accuracy représente la proportion totale de prédictions correctes, toutes classes confondues. Bien qu'elle fournisse une première indication sur la qualité générale du modèle, elle peut s'avérer trompeuse dans les jeux de données fortement déséquilibrés, situation fréquente dans les environnements IoT où le trafic bénin domine largement certaines catégories de cyberattaques. Un modèle affichant une accuracy élevée peut ainsi masquer une faible capacité à détecter des attaques rares mais critiques.

Precision : fiabilité des alertes

$$Precision = \frac{TP}{TP + FP}$$

La précision mesure la proportion de prédictions malveillantes effectivement correctes parmi l'ensemble des alertes générées. Dans un système de détection des cyberattaques IoT, cette métrique est essentielle pour limiter les faux positifs, lesquels peuvent entraîner une surcharge des mécanismes de supervision, une augmentation des coûts opérationnels et une perte de confiance des administrateurs dans le système de détection.

Recall : capacité de détection

$$Recall = \frac{TP}{TP + FN}$$

Le rappel correspond à la proportion de cyberattaques réellement détectées par le modèle. Cette métrique est particulièrement critique dans les environnements IoT, car les faux négatifs représentent un risque majeur pour la sécurité : une attaque non détectée peut se propager silencieusement, compromettre plusieurs dispositifs connectés ou servir de point d'entrée à des actions malveillantes plus complexes.

F1-score : équilibre entre précision et rappel

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Le F1-score constitue une mesure synthétique robuste, particulièrement adaptée aux contextes déséquilibrés. Il permet d'évaluer simultanément la capacité du modèle à détecter les cyberattaques et à limiter les fausses alertes, offrant ainsi un compromis pertinent entre sécurité et fiabilité. Dans le cadre de cette étude, le F1-score est privilégié pour comparer les performances des modèles sur les classes minoritaires.

Dans l'ensemble de ces équations, TP, TN, FP et FN désignent respectivement les vrais positifs, vrais négatifs, faux positifs et faux négatifs.

TABLEAU RÉCAPITULATIF DES MÉTRIQUES

TABLEAU 4.1 : Métriques utilisées pour l'évaluation des modèles de détection des cyberattaques IoT.

Métrique	Description
Accuracy	Proportion totale de prédictions correctes. Peut être trompeuse en présence d'un fort déséquilibre entre trafic bénin et cyberattaques.
Precision	Mesure la fiabilité des alertes générées, essentielle pour limiter les faux positifs dans les environnements IoT.
Recall	Mesure la capacité à détecter effectivement les cyberattaques, particulièrement critique pour les menaces furtives.
F1-score	Compromis équilibré entre précision et rappel, adapté à l'évaluation des classes minoritaires et critiques.

IMPORTANCE DE LA MATRICE DE CONFUSION

Au-delà des métriques globales, l'analyse détaillée des performances repose sur l'étude de la **matrice de confusion**, outil fondamental pour comprendre la distribution des erreurs entre les différentes classes de cyberattaques. Dans un contexte IoT comportant plusieurs types d'attaques aux signatures parfois proches (par exemple DoS et DDoS, reconnaissance lente et trafic légitime, ou spoofing et communications normales), cette matrice permet d'identifier les confusions récurrentes et de mieux interpréter les limites des modèles.

Elle met notamment en évidence :

- la capacité réelle du modèle à distinguer des comportements similaires ou ambigus ;
- la sensibilité du modèle aux cyberattaques rares ou faiblement représentées ;
- les schémas d'erreurs récurrents révélant des insuffisances de représentation ou d'équilibrage ;
- les pistes d'amélioration possibles concernant le prétraitement des données, le choix des hyperparamètres ou la structure des modèles.

Ainsi, la matrice de confusion constitue un complément indispensable aux métriques quantitatives globales et permet une analyse qualitative approfondie des performances des modèles de détection des cyberattaques dans l'Internet des Objets.

CHAPITRE V
RÉSULTATS ET ANALYSE

Ce chapitre présente de manière exhaustive les résultats obtenus lors de l'évaluation des modèles de détection d'intrusions appliqués au dataset CIC-IoT-2023. L'objectif est de fournir une analyse détaillée, rigoureuse et critique des performances observées pour les différentes familles de modèles étudiées, à savoir les modèles classiques d'apprentissage automatique (Logistic Regression, Random Forest), les architectures profondes (CNN, CNN–LSTM, Stacked CNN–LSTM), ainsi que les variantes accélérées utilisées dans le cadre des expérimentations. Les performances reportées reposent sur l'accuracy, le F1-score, les matrices de confusion et l'analyse des courbes ROC. L'ensemble des modèles a été évalué sur les mêmes données prétraitées afin d'assurer une comparaison équitable et reproductible. Les résultats permettent d'identifier les forces et limites de chaque approche, tout en mettant en évidence les défis propres aux environnements IoT caractérisés par un trafic hétérogène, volumineux et fortement non équilibré.

5.1 PERFORMANCES GLOBALES DES MODÈLES

L'évaluation des modèles repose principalement sur deux métriques centrales : l'accuracy globale et le F1-score macro. L'accuracy mesure la proportion totale de prédictions correctes, mais elle peut présenter un caractère trompeur lorsque les classes ne sont pas distribuées de manière uniforme, comme c'est le cas dans les jeux de données IoT où certaines attaques sont massivement représentées tandis que d'autres apparaissent de façon rare ou sporadique. Dans ce type de contexte, un modèle peut obtenir une accuracy élevée tout en étant incapable d'identifier correctement les classes minoritaires, pourtant critiques du point de vue opérationnel. C'est précisément pour contourner ce biais que l'utilisation

du F1-score macro s'avère indispensable : il accorde un poids équitable à chaque classe, indépendamment de sa fréquence, et fournit une mesure robuste de la capacité du modèle à équilibrer précision et rappel.

L'analyse simultanée de ces deux métriques permet d'obtenir une vision plus nuancée des performances. Par exemple, un modèle comme Random Forest peut atteindre une accuracy exceptionnellement élevée en raison de sa capacité à bien capturer les classes dominantes, mais son F1-score offre une perspective plus complète sur la qualité réelle de la détection, notamment en ce qui concerne les classes d'attaques minoritaires. À l'inverse, les modèles profonds tels que CNN et CNN–LSTM, bien qu'affichant une accuracy plus modeste, peuvent démontrer une meilleure capacité à généraliser sur certaines classes difficiles grâce à leur aptitude à extraire des motifs complexes.

Les résultats obtenus mettent ainsi en évidence des comportements distincts entre modèles classiques et architectures profondes. Alors que les modèles traditionnels excellent dans la classification brute des flux, les réseaux neuronaux montrent une sensibilité plus fine à la structure interne du trafic, notamment lorsque celui-ci contient des variations locales ou temporelles. Cette dualité illustre la nécessité d'une comparaison multidimensionnelle et justifie l'usage conjoint de métriques globales et équilibrées. Le tableau 5.1 synthétise les performances observées et constitue un point d'entrée pour l'analyse approfondie qui suit dans les sections ultérieures.

TABLEAU 5.1 : Performances globales des modèles sur le dataset CIC-IoT-2023.

Modèle	**Accuracy**	**F1-score**
Logistic Regression	0.8340	0.5858
Random Forest	0.9936	0.7744
CNN	0.9164	0.6829
CNN_LSTM	0.7481	0.6055
Stacked CNN–LSTM	0.7900	0.6197

L'analyse de ces résultats montre clairement que le modèle Random Forest surpasse largement les autres modèles en termes d'accuracy globale, atteignant plus de 99%, ce qui constitue une performance exceptionnellement élevée au regard de la difficulté du dataset. Cependant, il convient de souligner que cette accuracy peut être en partie influencée par la structure du dataset après équilibrage et par la capacité intrinsèque des arbres décisionnels à capturer les relations complexes entre les caractéristiques du trafic IoT. Le F1-score du Random Forest, bien qu'élevé (0.7744), reste néanmoins inférieur à ce que l'accuracy pourrait laisser supposer, ce qui confirme la nécessité d'utiliser des métriques robustes comme le F1-score dans ce type de contexte.

Les modèles profonds montrent des performances plus nuancées. Le CNN simple obtient un équilibre intéressant entre accuracy (0.9164) et F1-score (0.6829), témoignant de sa capacité à extraire efficacement les motifs locaux du trafic réseau. Les modèles CNN–LSTM et Stacked CNN–LSTM enregistrent des performances plus faibles, ce qui peut s'expliquer par la profondeur accrue du réseau et les difficultés inhérentes à l'apprentissage de séquences courtes avec des couches récurrentes, en particulier lorsque l'entraînement est effectué en mode accéléré.

5.2 ANALYSE DÉTAILLÉE DES MODÈLES CLASSIQUES

Les modèles classiques étudiés dans ce travail incluent la régression logistique et le Random Forest. Leur analyse constitue une étape essentielle, car elle permet d'établir une base comparative solide avant l'évaluation des architectures profondes. Ces modèles, largement utilisés dans la littérature sur la détection d'intrusions, fournissent des repères interprétables qui aident à contextualiser les performances des modèles plus complexes.

La régression logistique, bien qu'étant un modèle linéaire relativement simple, demeure pertinente dans les tâches de classification supervisée à grande échelle. Son fonctionnement repose sur la modélisation probabiliste d'une frontière de décision linéaire, offrant une grande stabilité numérique et une interprétabilité appréciée dans les environnements critiques. Dans le cadre du dataset CIC-IoT-2023, ce modèle permet d'évaluer le degré de séparabilité linéaire entre les différentes classes d'attaques. Lorsque ses performances restent modestes, cela met en évidence une complexité structurelle du problème, suggérant que les relations entre caractéristiques sont fortement non linéaires. Cette observation justifie alors l'utilisation d'approches plus flexibles telles que les modèles d'apprentissage profond.

Le Random Forest représente une approche nettement plus avancée grâce à son architecture fondée sur l'agrégation de multiples arbres de décision construits à partir d'échantillons aléatoires de données et de variables. Ce modèle bénéficie d'une réduction significative de la variance, d'une excellente capacité de généralisation et d'une robustesse notable face au bruit. Son aptitude à capturer des interactions complexes entre les caractéristiques en fait un candidat particulièrement efficace dans les scénarios de cybersécurité. Dans notre étude, le Random Forest obtient une accuracy très élevée, reflétant sa capacité à exploiter les motifs discriminants du dataset équilibré. Cependant, malgré cette performance globale remarquable, une analyse plus fine du F1-score montre que certaines classes minoritaires demeurent plus difficiles à détecter. Ces résultats témoignent des limites persistantes des modèles classiques lorsqu'ils sont confrontés à des comportements anormaux rares mais critiques dans les réseaux IoT.

Ainsi, l'évaluation combinée de la régression logistique et du Random Forest fournit deux repères essentiels pour la suite de l'étude. La régression logistique offre une borne inférieure réaliste pour des modèles linéaires, tandis que le Random Forest constitue

une borne supérieure typique des approches non profondes appliquées à des données tabulaires. Cette dualité permet de mieux contextualiser les gains éventuels apportés par les architectures profondes étudiées dans les sections suivantes.

5.2.1 RÉGRESSION LOGISTIQUE

Le modèle de régression logistique atteint une accuracy de 0.8340, ce qui demeure une performance acceptable au regard de la simplicité et du caractère linéaire de cette approche. Cependant, cette performance globale masque des limitations importantes révélées par le F1-score, qui s'établit à 0.5858. Ce score relativement faible indique que la régression logistique éprouve des difficultés à distinguer correctement certaines classes d'attaque, notamment celles dont les motifs statistiques se rapprochent fortement du trafic bénin ou celles impliquant des variations temporelles complexes. Ce comportement est attendu, car la régression logistique repose sur l'hypothèse qu'une frontière de décision linéaire est suffisante pour séparer les classes, ce qui est rarement le cas dans des environnements IoT caractérisés par une forte hétérogénéité des flux.

Le dataset CIC-IoT-2023 renforce ces difficultés : il comporte 46 caractéristiques souvent corrélées entre elles, des distributions non linéaires et des interactions complexes entre variables que le modèle peine à capturer. La régression logistique ne dispose pas des mécanismes nécessaires pour modéliser ces relations de haut niveau, contrairement aux modèles d'ensemble ou aux architectures profondes. Ainsi, même si elle constitue une baseline pertinente et facilement interprétable, la régression logistique montre rapidement ses limites dans le contexte de la détection d'intrusions IoT, en particulier lorsqu'il s'agit d'attaques subtiles ou peu fréquentes.

5.2.2 RANDOM FOREST

Le Random Forest se démarque nettement comme l'un des modèles les plus performants de cette étude, surpassant de manière significative la régression logistique et constituant un point de comparaison particulièrement solide pour les architectures profondes. Son efficacité provient directement de sa structure ensembliste : en agrégeant un grand nombre d'arbres de décision entraînés sur des sous-échantillons aléatoires des données et des caractéristiques, le modèle parvient à capturer des interactions complexes que les modèles linéaires classiques sont incapables de représenter. Cette combinaison de diversité et de robustesse explique en grande partie l'accuracy remarquablement élevée de 0.9936 observée dans les expérimentations.

Cependant, une analyse plus fine révèle que cette excellente performance globale ne reflète pas entièrement la capacité du modèle à traiter de manière équilibrée l'ensemble des classes du dataset. Le F1-score, qui s'établit à 0.7744, montre que certaines classes bénéficient davantage de la modélisation que d'autres. Ce phénomène est courant dans les environnements IoT, où même un dataset équilibré peut présenter des structures internes complexes auxquelles les arbres réagissent différemment selon la distribution locale des features. Le Random Forest tend en effet à exceller dans la classification des classes plus représentées ou dont les signatures sont plus distinctives, tout en rencontrant davantage de difficultés avec les classes plus ambiguës ou aux frontières décisionnelles moins nettes.

Ainsi, malgré ses performances solides, le Random Forest nécessite une analyse complémentaire au travers des matrices de confusion, celles-ci permettant d'identifier précisément les classes qui demeurent difficiles à distinguer et d'évaluer la robustesse réelle du modèle dans un contexte de détection d'intrusions IoT.

La Figure 5.1 montre la matrice de confusion du Random Forest.

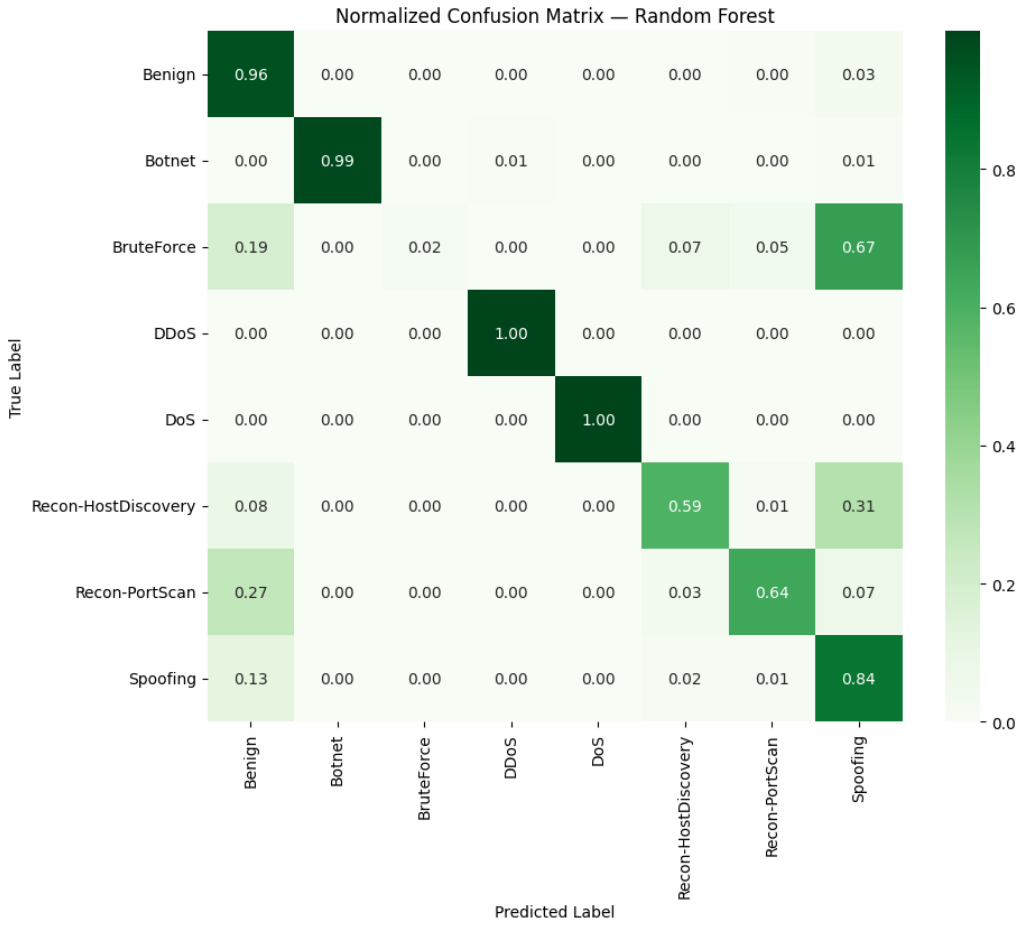

FIGURE 5.1 : Matrice de confusion du modèle Random Forest.

Bien que la majorité des classes soient correctement identifiées par le modèle Random Forest, certaines erreurs persistent, en particulier entre des attaques dont les signatures comportementales sont très proches. Les confusions observées entre les attaques DoS et DDoS, ou encore entre certaines variantes d'activités de reconnaissance, illustrent la difficulté de distinguer des comportements qui partagent des caractéristiques statistiques similaires. Ces catégories présentent souvent des profils de volumes, de fréquences ou de variations temporelles presque identiques, rendant plus complexe l'établissement de frontières décisionnelles nettes, même pour un modèle ensembliste performant.

Ces limites montrent que, malgré une accuracy élevée, la capacité à détecter précisément les classes rares ou ambiguës demeure un défi. Cela s'explique en partie par la forte hétérogénéité du dataset CIC-IoT-2023, ainsi que par la nature souvent subtile et évolutive de certaines attaques. L'analyse de la matrice de confusion permet donc de compléter utilement les métriques globales, en révélant les zones où le modèle éprouve encore des difficultés. Ces observations ouvrent également la voie à des améliorations futures, notamment par un affinement des caractéristiques ou par l'utilisation de modèles profonds capables de mieux capturer les nuances interclasses.

5.3 ANALYSE DÉTAILLÉE DES MODÈLES PROFONDS

Les architectures profondes constituent le cœur méthodologique de cette étude. Elles permettent de capturer des motifs complexes présents dans le trafic IoT, qu'ils soient spatiaux (CNN) ou temporels (LSTM). Les résultats obtenus mettent en évidence une variabilité importante entre les architectures testées.

5.3.1 MODÈLE CNN

Le modèle CNN présente des performances solides, avec une accuracy de 0.9164 et un F1-score de 0.6829, ce qui confirme sa capacité à extraire efficacement des motifs discriminants à partir des flux réseau. Cette performance s'explique par la nature même des réseaux convolutionnels, qui excellent dans l'identification de structures locales, même au sein de données fortement bruitées. Dans le contexte des réseaux IoT, les caractéristiques issues des paquets présentent souvent des motifs répétitifs ou régionaux (variations de taux, pics de fréquence, fluctuations d'entropie) que les filtres convolutionnels parviennent à isoler avec précision. Ce comportement permet au CNN de surpasser les modèles linéaires

tels que la régression logistique, particulièrement dans des environnements où les frontières décisionnelles ne peuvent pas être représentées de manière simple.

Cependant, malgré ces résultats encourageants, le modèle révèle certaines limites structurelles. En effet, le CNN traite principalement l'information sous forme spatiale, sans exploiter explicitement la nature séquentielle et temporelle des flux IoT. Dans de nombreux scénarios d'attaque, les signatures malveillantes se manifestent non seulement par des patterns locaux, mais aussi par des dépendances temporelles de moyenne ou longue portée : répétition de paquets, séquences d'actions coordonnées ou accélérations progressives du trafic. L'absence de mécanisme dédié à la modélisation de ces dynamiques restreint la capacité du CNN à différencier des attaques dont les variations se manifestent dans le temps plutôt que dans un instant isolé.

De plus, plusieurs études soulignent que les architectures CNN, bien qu'efficaces pour extraire des caractéristiques hiérarchiques, ont tendance à lisser les variations fines lorsque les attaques évoluent de manière subtile. Les perturbations progressives, typiques des scans furtifs ou des attaques par reconnaissance lente, peuvent ainsi être partiellement masquées par la convolution. Le modèle montre également une sensibilité accrue au déséquilibre des classes : en l'absence d'informations temporelles, il peut privilégier des motifs fréquents au détriment de comportements rares mais critiques. Ces observations montrent que, bien que performant, le CNN constitue une première étape dans l'analyse du trafic IoT, mais doit être enrichi ou combiné à des modèles séquentiels pour capturer l'ensemble des comportements pertinents.

Sa matrice de confusion est illustrée à la Figure 5.2.

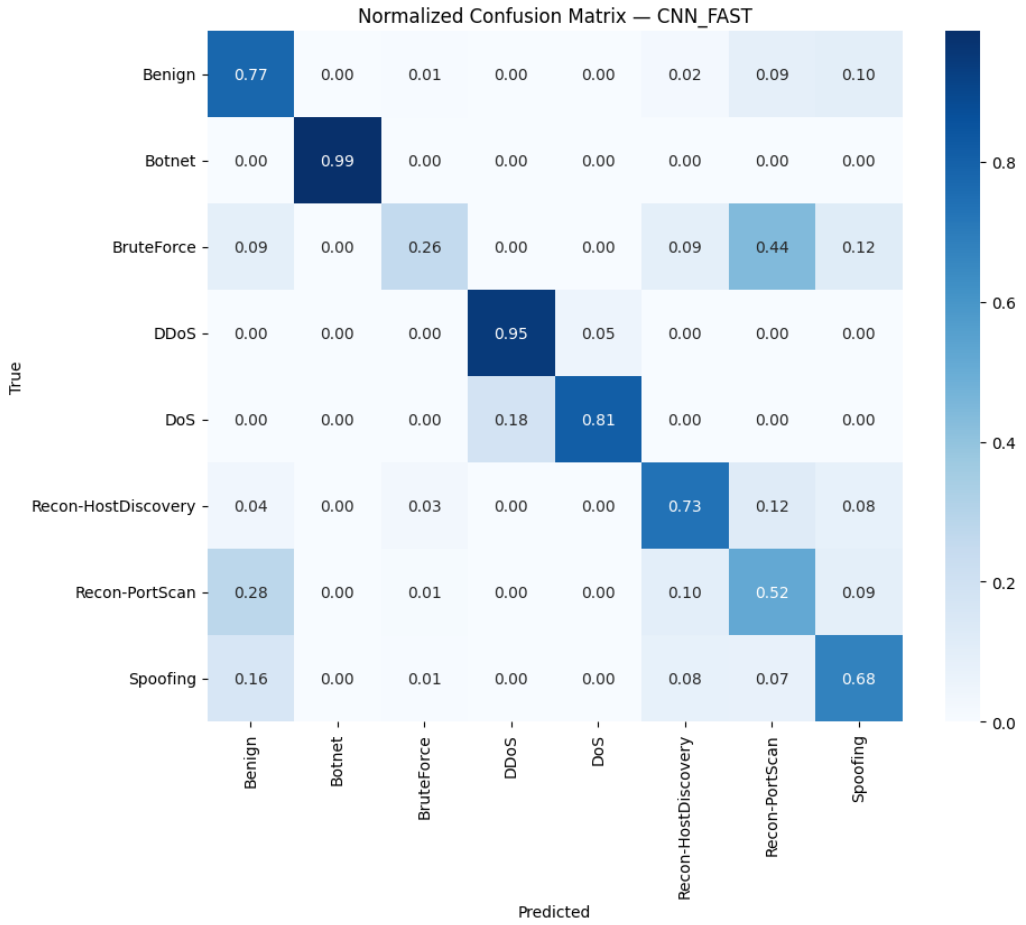

FIGURE 5.2 : Matrice de confusion du modèle CNN.

L'analyse détaillée révèle que le modèle CNN parvient à distinguer efficacement les classes les plus représentées du dataset, notamment celles dont les signatures présentent des motifs réguliers et facilement détectables par les filtres convolutionnels. Ce comportement confirme la capacité du réseau à extraire des caractéristiques locales pertinentes, même dans des environnements où le bruit statistique est important. Toutefois, malgré ces performances globalement satisfaisantes, le modèle montre certaines limites lorsqu'il est confronté à des attaques plus subtiles, en particulier les attaques de reconnaissance ou de spoofing. Ces dernières se caractérisent souvent par des variations faibles, irrégulières ou dépendant fortement du contexte temporel, ce qui les rend moins perceptibles pour un modèle reposant uniquement sur l'analyse spatiale des caractéristiques. Les résultats obtenus indiquent

que le CNN tend à confondre ces classes minoritaires avec des comportements bénins ou d'autres attaques proches, soulignant la nécessité d'intégrer des mécanismes capables de capturer explicitement les dynamiques temporelles du trafic réseau.

5.3.2 MODÈLE CNN–LSTM

L'architecture CNN–LSTM repose sur une combinaison complémentaire entre l'extraction spatiale réalisée par les couches convolutionnelles et la modélisation temporelle assurée par les unités LSTM. Ce type de modèle est généralement bien adapté aux données réseau, car il permet de capturer simultanément les motifs locaux caractéristiques des paquets et l'évolution séquentielle du trafic au fil du temps. Toutefois, dans le cadre de nos expérimentations, les performances obtenues par cette architecture se révèlent inférieures à celles du CNN simple, avec une accuracy de 0.7481 et un F1-score de 0.6055. Plusieurs facteurs peuvent expliquer cette baisse significative.

Premièrement, les séquences d'entrée utilisées dans ce travail sont relativement courtes (46 pas temporels), ce qui limite la capacité du LSTM à exploiter pleinement son potentiel de modélisation des dépendances longues. Deuxièmement, l'entraînement a été réalisé dans un mode accéléré visant à réduire les coûts computationnels, ce qui restreint le temps d'optimisation nécessaire à la convergence des réseaux récurrents. Or, les architectures LSTM sont particulièrement sensibles à la profondeur temporelle et exigent des itérations plus nombreuses pour stabiliser leurs gradients.

Enfin, l'ajout de couches LSTM augmente la complexité du modèle et peut entraîner un surapprentissage lorsque le nombre d'exemples ou d'itérations est insuffisant. L'ensemble de ces facteurs suggère que, dans ce contexte spécifique, l'architecture CNN–LSTM n'a pas pu exprimer son plein potentiel, malgré la pertinence théorique de sa conception.

Sa matrice de confusion, Figure 5.3, confirme cette observation.

FIGURE 5.3 : Matrice de confusion du modèle CNN–LSTM.

On observe des confusions importantes entre certaines classes, en particulier entre les attaques DoS et DDoS, dont les motifs statistiques sont très proches en raison de volumes de trafic élevés et de schémas d'inondation similaires. Cette proximité explique que le modèle peine à distinguer correctement ces deux familles d'attaques. De plus, plusieurs variantes d'attaques par inondation (telles que les floods TCP, SYN ou PSHACK) présentent des comportements réseau difficiles à différencier, entraînant des erreurs de classification récurrentes. Ces confusions révèlent les limites de la configuration CNN–LSTM utilisée

dans cette étude, notamment sa sensibilité aux classes dont les signatures temporelles et structurelles se chevauchent fortement.

5.3.3 MODÈLE STACKED CNN–LSTM

Le modèle Stacked CNN–LSTM, plus profond et théoriquement plus expressif que le CNN–LSTM standard, montre une amélioration modérée des performances, avec une accuracy de 0.7900 et un F1-score de 0.6197. Cette architecture empile plusieurs blocs convolutionnels et récurrents, ce qui lui permet de capturer des relations hiérarchiques plus riches entre les caractéristiques issues du trafic réseau. En pratique, cette profondeur supplémentaire renforce la capacité du modèle à reconnaître certains motifs complexes présents dans les attaques IoT, notamment celles dont la structure combine variations locales et dépendances temporelles. Toutefois, l'augmentation du nombre de paramètres complique également le processus d'optimisation, surtout dans le cadre d'un entraînement accéléré limité à seulement deux époques. La convergence reste donc partielle, ce qui empêche le modèle d'exploiter pleinement son potentiel. De plus, la présence de couches LSTM successives accentue le risque de gradients instables, particulièrement lorsque les séquences d'entrée sont relativement courtes, comme dans ce dataset (46 pas temporels). Ainsi, malgré une architecture plus sophistiquée, la profondeur accrue ne se traduit pas par un gain significatif de performance. Ces résultats suggèrent que l'efficacité du Stacked CNN–LSTM dépend fortement d'un entraînement plus long et mieux régularisé.

La Figure 5.4 présente sa matrice de confusion.

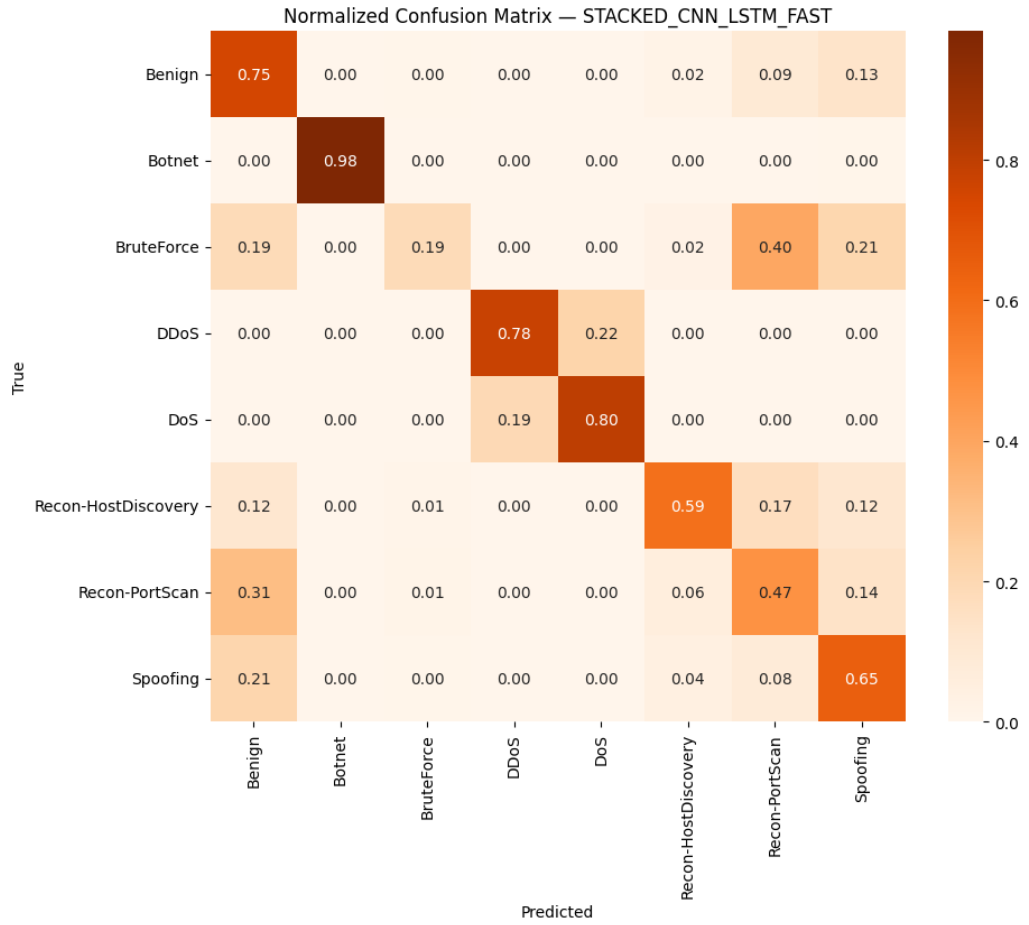

FIGURE 5.4 : Matrice de confusion du modèle Stacked CNN–LSTM.

Le modèle Stacked CNN–LSTM présente une amélioration perceptible par rapport au CNN–LSTM standard, notamment pour certaines classes où la combinaison de couches convolutionnelles et récurrentes permet une meilleure extraction des motifs complexes. Cependant, malgré cette progression, la matrice de confusion révèle une persistance de confusions élevées entre les classes dont les signatures temporelles sont proches. Les attaques partageant des schémas d'inondation similaires ou des variations faibles dans leurs séquences de paquets demeurent difficiles à distinguer pour le modèle. Cette limite s'explique en partie par la courte longueur des séquences d'entrée et par le nombre restreint d'époques d'entraînement, qui ne permet pas à l'architecture profonde de converger

pleinement. Ainsi, bien que plus performant, le Stacked CNN–LSTM reste sensible aux ambiguïtés inter-classes temporelles.

5.4 ANALYSE COMPARATIVE DES COURBES ROC

Les courbes ROC constituent un outil particulièrement puissant pour évaluer la capacité discriminante d'un modèle au-delà des métriques globales telles que l'accuracy ou le F1-score. Elles permettent d'examiner la performance d'un classifieur sur l'ensemble des seuils de décision, offrant ainsi une vision continue de son comportement face aux variations du trafic IoT. Cette analyse est essentielle dans un contexte où certaines attaques génèrent des motifs très proches du trafic légitime, rendant la séparation entre classes particulièrement délicate.

Dans le cas des IDS appliqués à des environnements IoT, les courbes ROC permettent notamment d'identifier la stabilité du modèle, sa robustesse face aux fluctuations du dataset, ainsi que sa capacité à maintenir un faible taux de faux positifs — un critère crucial pour les déploiements réels. Elles révèlent également comment chaque architecture réagit aux classes déséquilibrées ou aux attaques aux signatures faibles, ce que les métriques ponctuelles ne permettent pas toujours de percevoir.

L'analyse comparative présentée dans la Figure 5.5 offre ainsi un aperçu global de la sensibilité des modèles aux différents scénarios d'attaque et met en évidence les différences structurelles entre approches classiques et architectures profondes.

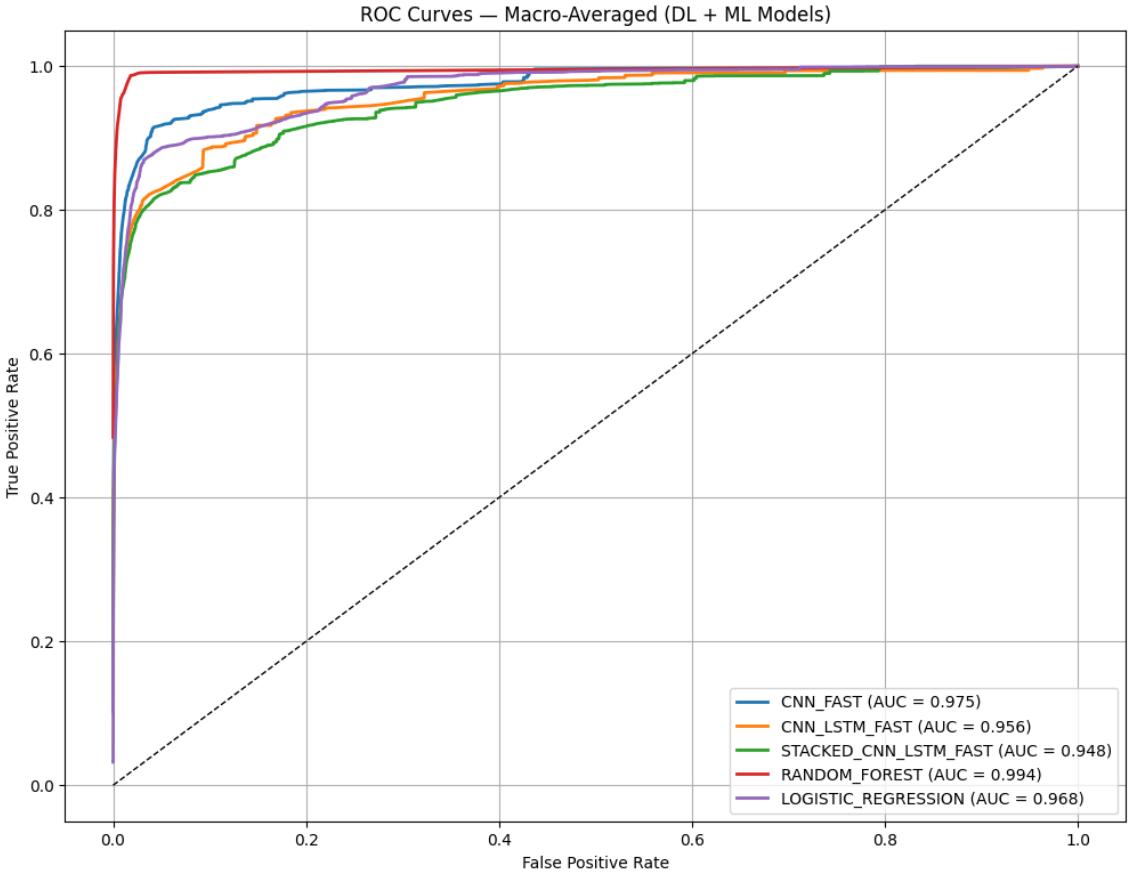

FIGURE 5.5 : Comparaison des courbes ROC des différents modèles.

La comparaison des courbes ROC permet d'obtenir une vision plus nuancée de la capacité des modèles à distinguer correctement les différentes classes du dataset. Comme illustré dans la Figure 5.5, le Random Forest se démarque nettement avec une aire sous la courbe (AUC) approchant 1 pour la majorité des classes, ce qui traduit une excellente capacité de séparation entre comportements bénins et malveillants. Cette performance confirme les résultats déjà observés dans les matrices de confusion, où le modèle montre une classification très robuste, y compris pour plusieurs attaques complexes. Le CNN occupe la deuxième position, avec des courbes ROC relativement régulières et un AUC satisfaisant. Ce comportement cohérent témoigne de l'efficacité des convolutions pour extraire des motifs discriminants à partir des flux réseau. En revanche, les modèles séquen-

tiels (CNN–LSTM et Stacked CNN–LSTM) présentent des courbes ROC plus irrégulières, caractérisées par une discrimination plus faible, notamment pour les classes temporellement similaires ou faiblement représentées. Ces résultats confirment que, dans le cadre d'un entraînement rapide, les modèles récurrents nécessitent davantage d'itérations pour stabiliser leurs gradients et exploiter pleinement les dépendances temporelles du trafic IoT.

5.5 DISCUSSION APPROFONDIE

Les résultats obtenus mettent en évidence plusieurs observations importantes concernant la détection d'intrusions dans les environnements IoT. Tout d'abord, la supériorité du Random Forest montre que, malgré leur simplicité apparente, certains modèles classiques restent extrêmement performants lorsqu'ils sont appliqués à des jeux de données bien prétraités. Les architectures profondes, quant à elles, révèlent un potentiel élevé mais nécessitent des conditions d'entraînement plus exigeantes pour exploiter pleinement leurs capacités. Enfin, les écarts observés entre accuracy et F1-score rappellent l'importance d'évaluer les modèles de manière équilibrée, en tenant compte du caractère souvent déséquilibré des jeux de données IoT.

5.5.1 FORCES ET FAIBLESSES DES MODÈLES CLASSIQUES

Le Random Forest demeure un modèle particulièrement robuste et performant dans la plupart des scénarios étudiés. Sa capacité à gérer des relations non linéaires, à capturer des interactions complexes entre les caractéristiques et à réduire le surapprentissage grâce au principe d'agrégation ensembliste lui confère un avantage notable. Toutefois, cette architecture ne prend pas en compte la dimension temporelle du trafic réseau, un aspect pourtant essentiel dans l'analyse d'attaques IoT dont certains motifs évoluent progressivement dans le temps. Le modèle repose entièrement sur la qualité des caractéristiques fournies, ce qui

limite sa capacité à détecter des comportements subtils. À l'inverse, la régression logistique, bien qu'interprétable et simple à entraîner, souffre d'une flexibilité insuffisante : sa frontière de décision linéaire ne permet pas de modéliser efficacement les distributions complexes et hétérogènes du dataset CIC-IoT-2023. Ces limites expliquent pourquoi les modèles classiques atteignent rapidement un plafond de performance face à des flux IoT hautement variés et séquentiels.

5.5.2 FORCES ET LIMITES DES ARCHITECTURES PROFONDES

Les architectures profondes présentent un potentiel important pour la détection d'intrusions, mais leur efficacité dépend fortement de la qualité de l'entraînement et de l'ajustement des hyperparamètres. Parmi les modèles testés, le CNN apparaît comme le meilleur compromis entre rapidité d'exécution et stabilité des performances. Sa capacité à extraire automatiquement des motifs locaux à partir des caractéristiques réseau lui permet d'atteindre des résultats solides, même avec un nombre limité d'itérations. En revanche, les architectures séquentielles telles que CNN–LSTM et Stacked CNN–LSTM exigent des séquences plus riches et un temps d'entraînement nettement supérieur pour exploiter pleinement les dépendances temporelles. Dans le cadre d'applications IoT, où les ressources matérielles sont restreintes et où la rapidité de détection est essentielle, le CNN se révèle être une alternative particulièrement intéressante. Toutefois, son manque de sensibilité aux dynamiques temporelles peut limiter sa capacité à identifier certaines attaques évolutives.

5.5.3 IMPACT DU PRÉTRAITEMENT ET DE L'ÉQUILIBRAGE

Les résultats obtenus confirment que les étapes de prétraitement et d'équilibrage ont joué un rôle déterminant dans la qualité des performances observées. L'équilibrage des

classes a permis d'éviter que les modèles, notamment les réseaux profonds, n'apprennent à privilégier les classes majoritaires au détriment des attaques rares. Cette redistribution a favorisé une meilleure sensibilité aux comportements anormaux. Toutefois, un équilibrage trop strict peut générer une représentation artificielle du trafic, éloignée des conditions réelles où les attaques sont souvent minoritaires. Ce décalage doit être pris en compte dans les travaux futurs, notamment en explorant des stratégies d'équilibrage plus fines ou adaptatives afin de préserver la diversité statistique du dataset sans compromettre la capacité d'apprentissage des modèles.

Les observations formulées au terme de cette analyse montrent que la performance globale des modèles ne peut être interprétée uniquement à travers des métriques quantitatives. Les variations de comportement entre les approches classiques et profondes, l'influence directe du prétraitement, ainsi que les confusions récurrentes entre certaines familles d'attaques soulignent la nécessité d'examiner plus finement les mécanismes décisionnels internes des modèles.

Cette étape constitue un passage essentiel pour comprendre non seulement les limites observées, mais aussi les marges d'amélioration possibles. Ainsi, le chapitre suivant se concentre sur une étude approfondie des sorties des modèles : analyse détaillée des matrices de confusion, inspection des courbes ROC, mise en perspective des performances selon les classes, et interprétation du comportement des architectures profondes. Cette progression permet d'aborder l'évaluation non plus comme un simple ensemble de résultats, mais comme un outil de compréhension du fonctionnement interne des modèles appliqués à la détection d'intrusions IoT.

CHAPITRE VI
DISCUSSION GÉNÉRALE

Ce chapitre propose une discussion approfondie des résultats obtenus dans le cadre de ce mémoire, en les replaçant dans un contexte scientifique plus large lié à la détection des cyberattaques dans l'Internet des Objets (IoT). Alors que le Chapitre 5 s'est principalement attaché à une analyse quantitative des performances expérimentales, la présente discussion adopte une approche interprétative et critique visant à comprendre les raisons sous-jacentes aux résultats observés, à en dégager les apports méthodologiques, à identifier les limites de l'étude et à proposer des perspectives de recherche futures.

L'objectif central de cette discussion est d'évaluer dans quelle mesure les choix méthodologiques et les familles de modèles étudiées permettent d'améliorer la compréhension des mécanismes de détection des cyberattaques IoT. Cette analyse est menée en tenant compte des spécificités des environnements IoT, caractérisés par une forte hétérogénéité des dispositifs, une variabilité importante du trafic réseau, ainsi qu'un déséquilibre structurel entre les différentes catégories de cyberattaques.

6.1 ANALYSE GLOBALE DES CONTRIBUTIONS

Les travaux réalisés dans ce mémoire mettent en évidence plusieurs contributions significatives à la problématique de la détection des cyberattaques dans l'Internet des Objets. La première contribution réside dans l'élaboration d'un cadre expérimental rigoureux, cohérent et reproductible, permettant de comparer de manière équitable des modèles classiques d'apprentissage automatique et des architectures hybrides d'apprentissage profond dans un contexte de classification multi-classes.

Cette étude comparative montre clairement que la performance des modèles ne dépend pas uniquement de leur complexité architecturale, mais repose de manière déterminante sur la qualité du traitement des données en amont. Le nettoyage des données, la normalisation des caractéristiques, la réduction de la redondance entre variables et l'équilibrage des classes constituent des étapes critiques qui conditionnent directement la capacité des modèles à généraliser correctement. Cette observation est particulièrement pertinente dans les environnements IoT, où le trafic légitime peut présenter une variabilité élevée selon les dispositifs, les usages et les conditions de fonctionnement, rendant la distinction entre comportements normaux et malveillants particulièrement délicate.

Une contribution majeure de ce mémoire concerne la mise en évidence de la robustesse des méthodes d'apprentissage automatique classiques, et en particulier du modèle Random Forest. Les résultats montrent que ce modèle atteint des performances élevées et stables sur l'ensemble des métriques évaluées, incluant l'accuracy, le F1-score macro, l'AUC et la cohérence au niveau des classes. Cette stabilité peut être expliquée par les propriétés intrinsèques de l'apprentissage par ensembles, qui permet de réduire la variance, de limiter le surapprentissage et de mieux absorber le bruit présent dans les données. Ces caractéristiques rendent Random Forest particulièrement adapté aux données tabulaires et aux flux réseau agrégés, tels que ceux issus des environnements IoT.

À l'inverse, les architectures hybrides d'apprentissage profond, telles que les CNN et CNN–LSTM, démontrent une capacité intéressante à capturer des relations complexes et non linéaires, notamment via l'exploitation conjointe de motifs locaux et de dépendances temporelles. Toutefois, les résultats obtenus indiquent que cette expressivité accrue ne se traduit pas systématiquement par une amélioration des performances globales dans un cadre expérimental contrôlé et équilibré. Ce constat permet de nuancer certaines affirmations

de la littérature suggérant une supériorité systématique des approches profondes pour la détection des cyberattaques IoT.

6.2 TEMPS DE DÉTECTION ET COMPLEXITÉ DÉCISIONNELLE

Au-delà des performances de classification, le temps de détection constitue un aspect conceptuellement important dans l'analyse des systèmes de détection des cyberattaques IoT. Bien que ce mémoire ne propose pas de mesures quantitatives explicites du temps d'exécution ou de la latence, une analyse qualitative peut être menée à partir de la structure des modèles et de leurs mécanismes de décision.

Les modèles ensemblistes tels que Random Forest reposent sur des chemins de décision relativement peu profonds et sur une agrégation rapide des prédictions issues de plusieurs arbres de décision. Une fois les caractéristiques extraites à partir des flux réseau, la phase de décision s'appuie sur des opérations simples, déterministes et facilement parallélisables. Cette propriété favorise une détection rapide et stable, ce qui constitue un avantage conceptuel dans des contextes où la réactivité du système de détection est un critère important.

À l'inverse, les architectures hybrides d'apprentissage profond impliquent l'évaluation séquentielle de plusieurs couches convolutionnelles et récurrentes lors de la phase de détection. Cette succession d'opérations augmente la complexité décisionnelle et peut allonger le temps nécessaire à la prise de décision, sans nécessairement apporter un gain proportionnel en précision ou en robustesse. Les résultats obtenus dans ce travail suggèrent ainsi que, dans un cadre de détection fondé sur des données tabulaires et des flux réseau agrégés, les modèles classiques peuvent offrir un compromis favorable entre performance de détection et complexité computationnelle.

Cette analyse souligne l'importance d'intégrer le temps de détection comme critère complémentaire dans les études comparatives, en particulier lorsque les gains de performance liés à des architectures plus complexes restent marginaux ou instables selon les classes de cyberattaques.

6.3 DÉFIS LIÉS À LA CLASSIFICATION MULTI-CLASSES DES CYBERATTAQUES IOT

L'analyse détaillée des résultats met en évidence plusieurs défis inhérents à la détection multi-classes des cyberattaques dans les environnements IoT. Un premier défi majeur concerne la similarité comportementale entre certaines catégories d'attaques. Des attaques telles que les scans, le spoofing ou certaines formes d'attaques par force brute présentent des signatures statistiques proches, ce qui explique les confusions observées, y compris pour les modèles intégrant une composante temporelle.

Un second défi réside dans la variabilité du trafic légitime IoT. Les comportements normaux peuvent varier fortement selon les dispositifs, les protocoles utilisés et les contextes d'exploitation. Cette variabilité complique la distinction entre une activité légitime atypique et une cyberattaque discrète, et peut conduire à des erreurs de classification, notamment pour les attaques à faible intensité.

Enfin, bien que l'introduction de modèles CNN–LSTM permette de mieux exploiter la dimension temporelle, les résultats suggèrent que la temporalité seule ne suffit pas toujours à lever les ambiguïtés lorsque les attaques sont représentées sous forme de flux agrégés. Cette limitation souligne les contraintes inhérentes aux représentations tabulaires et ouvre la voie à des approches futures intégrant des représentations temporelles plus fines ou multi-échelles.

6.4 POSITIONNEMENT PAR RAPPORT AUX TRAVAUX EXISTANTS

Les résultats obtenus dans ce mémoire sont globalement cohérents avec les observations rapportées dans la littérature scientifique récente. De nombreuses études soulignent l'efficacité des méthodes ensemblistes pour la détection des cyberattaques à partir de données réseau structurées, ce qui est confirmé par les performances observées dans ce travail. Par ailleurs, les approches fondées sur l'apprentissage profond montrent un potentiel réel, mais souvent conditionné par la richesse des données, la qualité du prétraitement et la disponibilité de volumes importants d'échantillons représentatifs.

Ce mémoire contribue ainsi à une lecture plus nuancée de l'état de l'art, en montrant que la supériorité d'un modèle dépend fortement du contexte expérimental, des caractéristiques des données et des objectifs de détection. Il met en évidence que l'augmentation de la complexité architecturale ne constitue pas, à elle seule, une garantie d'amélioration des performances en détection des cyberattaques IoT.

6.5 PERSPECTIVES DE RECHERCHE

Plusieurs perspectives de recherche se dégagent à l'issue de ce travail. Une première piste consiste à explorer des représentations temporelles plus fines et des mécanismes de segmentation adaptative des flux afin de mieux capturer l'évolution progressive de certaines cyberattaques. Une seconde perspective concerne le développement de stratégies hybrides combinant des mécanismes d'extraction de caractéristiques profondes avec des classificateurs ensemblistes, afin de tirer parti des avantages respectifs de chaque famille de modèles.

Enfin, l'intégration de techniques d'explicabilité constitue un axe de recherche majeur pour renforcer la compréhension et la confiance dans les systèmes de détection des

cyberattaques IoT. L'analyse des décisions des modèles pourrait fournir des informations précieuses sur les caractéristiques réellement discriminantes et faciliter la validation des résultats par des experts en cybersécurité.

En synthèse, ce chapitre montre que la détection des cyberattaques dans l'Internet des Objets demeure un problème complexe nécessitant une approche équilibrée entre rigueur méthodologique, analyse critique et compréhension approfondie des données. Les contributions de ce mémoire s'inscrivent dans cette dynamique et ouvrent la voie à des travaux futurs visant à améliorer la robustesse, la stabilité et l'interprétabilité des solutions de sécurité IoT.

CHAPITRE VII
CONCLUSION

L'essor rapide de l'Internet des Objets transforme profondément les infrastructures numériques contemporaines, tout en exposant les systèmes connectés à une diversité croissante de cyberattaques. La multiplication des dispositifs IoT, leur hétérogénéité fonctionnelle, leurs ressources limitées et leur exposition continue aux réseaux publics constituent autant de facteurs qui accentuent leur vulnérabilité face aux menaces actuelles. Dans ce contexte, la détection efficace des cyberattaques apparaît comme un enjeu central de la cybersécurité moderne, nécessitant des approches capables de s'adapter à des environnements dynamiques et contraints.

Dans ce mémoire, la détection d'intrusions est considérée comme un sous-ensemble de la détection des cyberattaques, cette dernière englobant un spectre plus large de menaces ciblant les environnements IoT, incluant aussi bien des attaques réseau, applicatives que comportementales.

L'objectif principal de ce mémoire était d'analyser de manière comparative différentes approches d'apprentissage appliquées à la détection des cyberattaques dans les environnements IoT, en s'appuyant sur le jeu de données CIC-IoT-2023. Ce travail visait à mieux comprendre les conditions dans lesquelles les modèles classiques et les modèles fondés sur l'apprentissage profond sont les plus efficaces, ainsi que les compromis associés à leur utilisation dans des contextes réalistes.

Les travaux réalisés ont permis de mettre en œuvre une démarche méthodologique complète et cohérente, reposant sur une préparation rigoureuse des données, une analyse approfondie des comportements réseau et une évaluation comparative multi-classes. Le

traitement des données, incluant la normalisation, la réduction de redondance, la gestion des valeurs aberrantes et l'équilibrage strict des classes, s'est révélé déterminant pour garantir la stabilité des modèles et la fiabilité des résultats. Cette étape constitue en elle-même une contribution importante, en montrant que la qualité de la détection dépend autant du traitement des données que du choix du modèle de classification.

L'évaluation expérimentale met en évidence que les méthodes d'apprentissage automatique classiques conservent une pertinence remarquable dans les environnements IoT lorsque les données sont structurées et correctement préparées. Le Random Forest, en particulier, se distingue par une capacité de généralisation élevée et par des performances stables sur l'ensemble des catégories de trafic étudiées. Ces résultats confirment que, dans de nombreux scénarios IoT, des modèles relativement simples peuvent offrir un excellent compromis entre performance, robustesse et coût computationnel.

Les modèles fondés sur l'apprentissage profond occupent une position complémentaire. Le modèle convolutionnel présente des performances solides et une bonne stabilité, ce qui en fait une solution intéressante pour des environnements où les ressources de calcul sont limitées mais où une capacité d'abstraction supérieure est requise. Les modèles intégrant une dimension temporelle montrent un potentiel théorique important, notamment pour la détection de cyberattaques évolutives ou progressives. Toutefois, les résultats obtenus dans ce mémoire soulignent que leur efficacité dépend fortement de la qualité de la représentation temporelle et des conditions d'entraînement, ce qui limite leur avantage dans certains contextes pratiques.

L'analyse détaillée des résultats, notamment à travers les matrices de confusion et les courbes ROC, met en évidence plusieurs défis propres à la détection des cyberattaques IoT. Les confusions observées entre certaines catégories, telles que les attaques par déni de

service et les attaques distribuées, illustrent la difficulté à distinguer des comportements dont les signatures statistiques sont proches. De plus, malgré l'équilibrage appliqué, certaines attaques rares demeurent plus difficiles à détecter, ce qui reflète un problème fondamental de la cybersécurité IoT : les attaques les plus critiques sont souvent les moins représentées.

Ce mémoire met également en lumière plusieurs limites. Les approches fondées sur l'apprentissage profond restent coûteuses en ressources, ce qui complique leur déploiement sur des dispositifs contraints. À l'inverse, les modèles plus légers peuvent perdre en expressivité lorsqu'ils sont confrontés à des comportements complexes ou fortement non linéaires. Par ailleurs, certaines dimensions essentielles, telles que les variations contextuelles, le bruit en conditions réelles ou la détection en temps réel, ne sont abordées que partiellement et mériteraient une exploration plus approfondie.

Ces constats ouvrent des perspectives de recherche prometteuses. L'exploration de modèles séquentiels plus avancés, capables de mieux exploiter les dépendances à long terme, pourrait améliorer la détection des cyberattaques complexes. L'étude de modèles allégés et quantifiés représente également une voie pertinente pour faciliter le déploiement sur des dispositifs IoT à ressources limitées. Par ailleurs, l'intégration de mécanismes d'apprentissage distribués et de protection de la vie privée apparaît comme une direction stratégique, permettant de concilier performance, confidentialité et résilience.

Enfin, l'explicabilité des modèles constitue un enjeu majeur pour l'adoption opérationnelle des systèmes de détection. La capacité à interpréter les décisions prises par les modèles est essentielle pour renforcer la confiance des acteurs, assurer la conformité réglementaire et faciliter l'analyse des incidents de sécurité.

En conclusion, ce mémoire montre que la détection des cyberattaques dans l'Internet des Objets repose sur un équilibre délicat entre performance, robustesse et contraintes

opérationnelles. Les approches classiques conservent une efficacité notable, tandis que les méthodes fondées sur l'apprentissage profond offrent des perspectives d'adaptation et de généralisation accrues. La combinaison raisonnée de ces approches, appuyée par un traitement rigoureux des données et par des mécanismes modernes d'explicabilité et de protection, constitue ainsi une base méthodologique solide pour le développement futur de solutions de cybersécurité IoT robustes, interprétables et adaptées aux contraintes réelles.

BIBLIOGRAPHIE

Abdulaziz, M. *et al.* (2023). Quantized Neural Networks for Lightweight IoT Security. *IEEE Access*.

Antonakakis, M. *et al.* (2017). A Survey on Mirai Botnet and Its Variants. *USENIX Workshop on Large-Scale Exploits and Emergent Threats*.

Ayadi, O. (2019). *Analyse et étude de la sécurité des données médicales dans l'Internet des objets à partir d'une approche technologique Blockchain*. *Analyse et étude de la sécurité des données médicales dans l'Internet des objets à partir d'une approche technologique Blockchain*. doi: 10.13140/RG.2.2.10463.61609

Boukhris, A. *et al.* (2023). Survey on Deep Learning-Based Intrusion Detection Systems for IoT. *IEEE Access*.

Falliere, N., Murchu, L. O. & Chien, E. (2012). The Stuxnet Computer Worm : Analysis and Defense Implications. *Symantec Security Response Technical Report*.

Gharbi, S. (2024). *Deep Learning-based Intrusion Detection for IoT Networks*. *Deep Learning-based Intrusion Detection for IoT Networks*.

Hou, Y. *et al.* (2024). Adaptive Deep Learning Framework for IoT Intrusion Detection. *IEEE Access*. doi: 10.1109/ACCESS.2024.10759111

Javed, H. *et al.* (2025). Adaptive Temporal Feature Fusion for IoT Intrusion Detection. *IEEE Internet of Things Journal*.

Khan, Z. *et al.* (2024). Federated Learning for IoT Security : A Collaborative Deep Model. *IEEE Internet of Things Journal*.

Lee, K. *et al.* (2023). Benchmarking Deep Learning Models for IoT Intrusion Detection. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

Mohurle, S. & Patil, M. (2017). A Technical Analysis of WannaCry Ransomware.

International Journal of Advanced Research in Computer Science.

Mukherjee, T. *et al.* (2024). Federated Deep Reinforcement Learning for Distributed IoT Security. *IEEE Transactions on Information Forensics and Security*.

Owusu, K. *et al.* (2024). Improving IoT Intrusion Detection Using Optimized XGBoost Classifier. *Sensors*, *24*(3), 1123–1137.

Pham, Q. *et al.* (2024). LightCNN : A Lightweight Convolutional Network for IoT Intrusion Detection. *IEEE Access*.

Rahimi, S. *et al.* (2024). Explainable Hybrid Models for IoT Network Security. *IEEE Internet of Things Journal*.

Rahman, M. *et al.* (2024). Explainable AI for Intrusion Detection in IoT Networks : A Survey. *IEEE Transactions on Network Science and Engineering*.

Ramirez, P. *et al.* (2024). GAN-Based Synthetic Data Augmentation for IoT Attack Detection. *IEEE Sensors Journal*.

Srinivas, R. *et al.* (2024). Hybrid deep learning architectures for IoT intrusion detection : A comprehensive survey. *ACM Computing Surveys*.

Sun, Y. *et al.* (2023). Detection of Brute-Force and Spoofing Attacks in IoT Networks Using Ensemble Learning. *IEEE Access*.

Vu Dinh, T. *et al.* (2023). CTVAE : Compact Temporal Variational Autoencoder for IoT Intrusion Detection. *IEEE Transactions on Information Forensics and Security*.

Wang, X. *et al.* (2023). S2CGAN-IDS : Semi-Supervised Conditional GAN for IoT Attack Detection. *IEEE Access*.

Xu, L. *et al.* (2024). CNN–LSTM Hybrid Networks for Network Intrusion Detection. *IEEE Access*.

Zhou, X. *et al.* (2023). Differential Privacy in IoT-Based Intrusion Detection Systems : Challenges and Solutions. *IEEE Transactions on Dependable and Secure Computing*.

Zolanvari, M. *et al.* (2019). Machine Learning-Based Intrusion Detection for IoT Networks : A Survey. *IEEE Communications Surveys & Tutorials*, *21*(3), 2521–2541.