Machine Learning Algorithms for Detecting Intra-Vehicular Data Falsification

Hope Leticia Nakayiza, Love Allen Chijioke Ahakonye †, Dong-Seong Kim, Jae Min Lee,
IT Convergence Engineering, † ICT Convergence Research Center,
Kumoh National Institute of Technology Gumi, South Korea
hopeleticia, loveahakonye, dskim, ljmpaul@kumoh.ac.kr

***Abstract*—The Internet of Vehicles (IoV) emphasizes the crucial role of Intrusion Detection Systems (IDS) in strengthening security with Machine learning (ML) algorithms, promising enhanced IDS performance by offering real-time anomaly detection capabilities. This study evaluates ML algorithms for accurately detecting intra-vehicular data falsification. Combining effective data preprocessing, the simulation results demonstrate enhanced ML model performance in notably detecting intrusions leveraging the recently published CICIoV2024 dataset.**

***Index Terms*—Internet of Vehicles, machine learning, intrusion detection**

I. Introduction

The Internet of Vehicles (IoV) transforms transportation systems by integrating vehicles, road infrastructure, and communication networks, enhancing safety, efficiency, and convenience. IoV optimizes traffic management, detects congestion, and enables traffic pattern recognition and incident detection in urban areas [1]. It also supports vehicle interactions with external entities for monitoring road conditions and estimating speeds, thus improving traffic surveillance services [2]. However, IoV's interconnected nature poses security challenges like distributed denial of service and other cyber-attacks targeted at vehicle communication channels [3], [4]. These attacks can lead to vehicle downtime and traffic collisions, highlighting the critical need for robust security measures in IoV environments [2].

Intrusion detection systems (IDS) are critical for securing Internet of Vehicles (IoVs), monitoring vehicle operations, detecting attacks, and aiding mitigation [2]. Leveraging machine learning (ML) algorithms, IDS can achieve precise threat identification through real-time data analysis and anomaly detection [5], [6]. This study evaluates ML algorithms for intrusion detection in IoV environments to enhance cybersecurity and ensure system resilience against evolving threats.

Existing IoV security datasets have limitations and must include essential features for developing effective security solutions [1]. The newly developed CICIoV2024 dataset addresses these gaps by introducing new attack types [1]. Building upon their work, this preliminary study aims to enhance the performance of ML algorithms for IoV intrusion detection leveraging considerable data preprocessing to detect the attack types of the recent CICIoV2024 dataset effectively. This work advances ML applications in IoV cybersecurity, evaluating a new baseline, and paves the way for optimizing ML models and enhancing feature analysis.

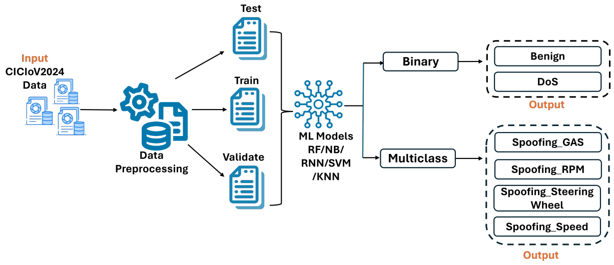


Fig. 1: Process workflow of the IoV attack detection using ML algorithms

II. System Methodology

Representative ML algorithms for intrusion detection were evaluated using the CICIoV2024 dataset [1] [1]. Derived from trials on a 2019 Ford vehicle's unaltered internal framework, it consists of intra-vehicular communication from Electronic Control Units. It covers five attack types: gas and RPM spoofing, steering and speed manipulation, and denial of service via the CAN-BUS protocol.

Fig. 1 illustrates the workflow and system model of the proposed scheme. The evaluation of the ML models involved three primary phases: training, testing, and validation, simultaneously for binary and multiclass classification with the same parameter settings. The evaluation metrics of computation time, accuracy, recall, precision, and F1 score determined the most efficient model. Enhancing the works of [1] with substantial data preprocessing by replacing all missing values with mean, standardizing data features to a mean of 0 and standard deviation of 1, and one-hot encoding all categorical values to numerical labels significantly improved the model performance.

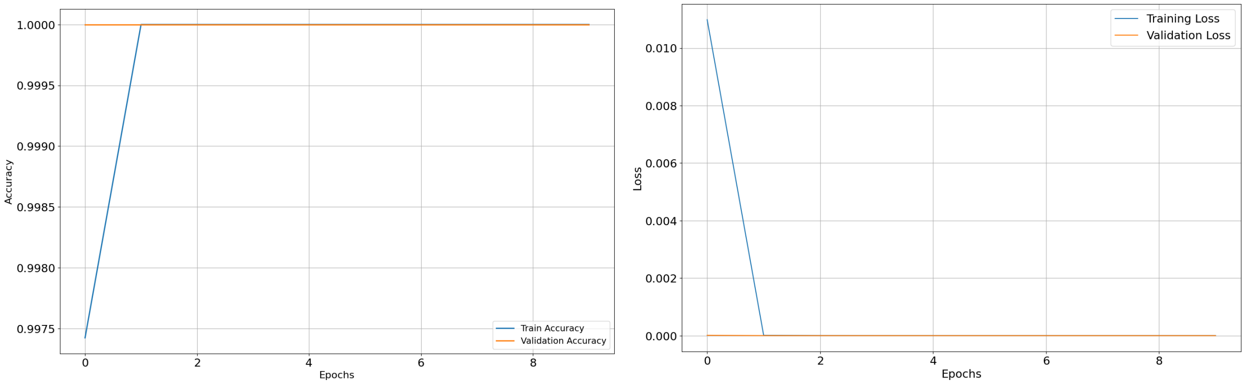
III. Performance Evaluation

The experimentation results demonstrate that the evaluated ML algorithms significantly detect IoV attacks, as shown in Table I highlighting the performance of the various algorithms in the binary and multiclass IoV attack scenarios. The models

[1] https://www.unb.ca/cic/datasets/iov-dataset-2024.html

TABLE I: Performance of the evaluated ML models in Binary and Multiclass of the new CICIoV2024 dataset

	Binary								Multiclass							
Metrics	**RF**	**NB**	**SVM**	**LR**	**AB**	**DNN**	**RNN**	**KNN**	**RF**	**NB**	**SVM**	**LR**	**AB**	**DNN**	**RNN**	**KNN**
Accuracy (%)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**99**	**100**	**95**	**100**	**100**	**95**	**100**	**100**	**100**
Precision (%)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**99**	**100**	**96**	**100**	**100**	**95**	**100**	**100**	**100**
Recall (%)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**99**	**100**	**95**	**100**	**100**	**95**	**100**	**100**	**100**
F1-score (%)	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**99**	**100**	**95**	**100**	**100**	**95**	**100**	**100**	**100**
Computation Time (s)	**63.942**	**7.233**	**18.893**	**23.9645**	**213.0526**	**804.54065**	**988.201**	**23436.15**	**8.109**	**8.781**	**1.071**	**11.4648**	**30.5439**	**142.9388**	**146.755**	**4.543**



(a) Multi-class Training Accuracy and Validation Accuracy

(b) Multi-class Training Loss and Validation Loss

Fig. 2: Graph showing the Training Accuracy and Validation Accuracy and the Training Loss and Validation Loss for Multi-class Classification using DNN

TABLE II: Comparative analysis of the evaluated ML models with the initial study on the new CICIoV2024 dataset

	Existing Study				*This Study*			
Algorithms	**LR**	**AB**	**DNN**	**RF**	**LR**	**AB**	**DNN**	**RF**
Accuracy (%)	95	87	95	95	100	95	100	100
Precision (%)	74	14	74	60	100	95	100	100
Recall (%)	68	17	68	68	100	95	100	100
F1-score (%)	63	15	63	62	100	95	100	100
Computaion Time (s)	nil	nil	nil	nil	11.46	30.54	142.93	63.94

performed differently in both classification instances, achieving significantly less training time during multiclass classification than the binary classification, showing the model's aptness in detecting any form of intra-vehicular falsification. Moreover, time efficiency is crucial when choosing a suitable ML algorithm in the IoV scenario. Figure 2 is a learning process graph that validates the absence of model overfitting and demonstrates the significant effect of the performed data preprocessing in enhancing the model performance compared to the existing studies.

IV. Conclusion

This study highlights the critical role of IDS in ensuring security within the IoV networks. This exploratory study highlights possible ML candidates for detecting intra-vehicular data falsification, leveraging the new CICIoV2024 dataset. The experimentation results indicate the significance of data preprocessing for improved model performance. Time efficiency is imperative when choosing an ML algorithm for intrusion detection. Our future direction is to optimize these models for real-time detection regarding the critical nature of connected vehicles.

Acknowledgment

This work was partly supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of IITP grant funded by the Korea government(MSIT) (IITP-2024-2020-0-01612, 25%) and by Priority Research Centers Program through the NRF funded by the MEST(2018R1A6A1A03024003, 25%) also This work was supported by (NRF-2022R1I1A3071844, 25%) and by the MSIT, Korea, under the ICAN support program(IITP-2024-00156394, 25%) supervised by the IITP

References

[1] E. Carlos Pinto Neto, H. Taslimasa, S. Dadkhah, S. Iqbal, P. Xiong, T. Rahman, and A. Ghorbani, "CICIOV2024: Advancing Realistic IDS Approaches Against DOS and Spoofing Attack in IoV Can Bus," *Internet of Things*, 2024.

[2] J. Nagarajan, P. Mansourian, M. A. Shahid, A. Jaekel, I. Saini, N. Zhang, and M. Kneppers, "Machine Learning Based Intrusion Detection Systems for Connected Autonomous Vehicles: A Survey," *Peer-to-Peer Networking and Applications*, vol. 16, no. 5, pp. 2153–2185, 2023.

[3] Y. Wang, G. Qin, M. Zou, Y. Liang, G. Wang, K. Wang, Y. Feng, and Z. Zhang, "A Lightweight Intrusion Detection System for Internet of Vehicles Based on Transfer Learning and MobileNetV2 with Hyper-Parameter Optimization," *Multimedia Tools and Applications*, vol. 83, pp. 1–23, 06 2023.

[4] G. Oluchi Anyanwu, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Optimization of RBF-SVM Kernel Using Grid Search Algorithm for DDoS Attack Detection in SDN-Based VANET," *IEEE Internet of Things Journal*, vol. 10, no. 10, pp. 8477–8490, 2023.

[5] L. A. C. Ahakonye, C. I. Nwakanma, J.-M. Lee, and D.-S. Kim, "Agnostic CH-DT Technique for SCADA Network High-Dimensional Data-Aware Intrusion Detection System," *IEEE Internet of Things Journal*, vol. 10, no. 12, pp. 10 344–10 356, 2023.

[6] G. O. Anyanwu, C. I. Nwakanma, J. M. Lee, and D.-S. Kim, "Novel Hyper-Tuned Ensemble Random Forest Algorithm for the Detection of False Basic Safety Messages in Internet of Vehicles," *ICT Express*, vol. 9, no. 1, pp. 122–129, 2023.