10.48047/jocaaa.2024.33.08.335

Non-Negative Latent Factor Dimensionality-Minimizing Intra-Class Compactness for Feature Extraction in IoV Security

Raavi Deepthi
Research Scholar
GITAM University, Rudraram, Patancheru, Hyderabad - 502329
draavi@gitam.in

Abstract: High-dimensional feature spaces in Internet of Vehicles datasets create computational challenges and potential overfitting risks in intrusion detection systems. This paper presents Non-negative Latent Factor Dimensionality-Minimizing Intra-class Compactness (NLF-DMIC), a novel feature extraction technique specifically designed for IoV cybersecurity applications. The methodology reduces feature dimensionality while maximizing separation between attack classes and minimizing variance within each class, thereby enhancing classifier discrimination capability. NLF-DMIC employs non-negative matrix factorization to extract latent features that represent underlying attack patterns while maintaining interpretability of the transformed feature space. The dimensionality reduction approach ensures computational efficiency suitable for real-time intrusion detection in resource-constrained vehicular environments. Integration with Enhanced SMOTEBoost for handling class imbalance and combination with Improved LSTM classifiers yields exceptional performance on CICIDS-2018 and Car-Hacking datasets. Experimental results demonstrate that NLF-DMIC feature extraction contributes to achieving 98.68% and 99.8% classification accuracy respectively, with F1-scores of 98.32% and 99.6%. The compact feature representation reduces network overhead for distributed learning scenarios while maintaining or improving detection accuracy across benign traffic and multiple attack categories including Bot, DoS, DDoS, Fuzzy, RPM manipulation, and gear spoofing, making it particularly valuable for bandwidth-limited vehicular communication environments.

Keywords: Feature Extraction, Non-Negative Factorization, Dimensionality Reduction, Intra-Class Compactness, IoV Security

1. Introduction

The Internet of Vehicles (IoV) has emerged as a pivotal component of intelligent transportation systems, enabling vehicles to communicate with each other, roadside units, and cloud infrastructure. This connectivity facilitates numerous applications, including real-time traffic monitoring, autonomous driving, predictive maintenance, and cooperative safety measures [1]. However, the increasing complexity and scale of IoV networks expose them to a growing variety of cyber threats, such as Distributed Denial of Service (DDoS), spoofing, replay attacks, and manipulation of critical vehicular parameters [2]. The high dimensionality of vehicular network data—arising from multiple sensors, communication channels, and protocol features—poses significant challenges to intrusion detection systems (IDS), including increased computational overhead, risk of overfitting, and difficulty in interpreting the model outputs [3].

Traditional machine learning approaches in IoV security rely heavily on hand-crafted features or high-dimensional input vectors, which often fail to capture latent patterns of malicious behavior effectively [4]. Moreover, centralized IDS frameworks require the collection of raw vehicular data, raising privacy and data transmission concerns in bandwidth-limited vehicular networks [5]. Dimensionality reduction and feature extraction techniques are therefore essential to improve both the efficiency and accuracy of intrusion detection while preserving interpretability of the learned representations. Effective feature extraction can reduce redundant or irrelevant information, highlight underlying attack patterns, and enhance the discriminative capacity of classifiers [6].

Recent research has explored various matrix factorization and embedding techniques for feature extraction in cybersecurity applications. Non-negative matrix factorization (NMF) has gained attention due to its ability to decompose high-dimensional data into non-negative latent factors, which are more interpretable and conducive to capturing inherent structures in attack behavior [7]. However, standard NMF approaches do not explicitly consider class separability and intra-class compactness, which are critical in IoV security scenarios where attack classes may be closely related, and benign traffic patterns can be highly variable. Without addressing these challenges, feature representations may lack sufficient

10.48047/jocaaa.2024.33.08.335

discrimination, leading to misclassification and reduced detection performance.

To address these limitations, this paper proposes a **Non-Negative Latent Factor Dimensionality-Minimizing Intra-Class Compactness (NLF-DMIC)** framework for feature extraction in IoV cybersecurity. The core idea of NLF-DMIC is to learn a low-dimensional, non-negative latent representation of the original feature space while simultaneously minimizing intra-class variance and maximizing inter-class separation. This approach ensures that features representing the same attack type are tightly clustered while different attack classes remain well-separated, thereby improving the accuracy and robustness of downstream classifiers. The method maintains non-negativity, ensuring interpretability of latent factors, which is crucial for analyzing attack patterns and decision-making in vehicular networks.

In addition to dimensionality reduction, IoV datasets often suffer from class imbalance, with attack samples being significantly fewer than benign traffic. To address this, NLF-DMIC integrates seamlessly with **Enhanced SMOTEBoost**, a hybrid oversampling technique that balances minority attack classes without introducing noise or redundancy. When combined with an **Improved Long Short-Term Memory (LSTM)** classifier, the framework effectively models temporal dependencies in vehicular traffic, enabling precise detection of complex attacks such as Bot, DoS, DDoS, Fuzzy, RPM manipulation, and gear spoofing [8].

Experimental validation on **CICIDS-2018** and **Car-Hacking** datasets demonstrates that NLF-DMIC significantly enhances intrusion detection performance. Specifically, the approach achieves **98.68% and 99.8% classification accuracy**, with F1-scores of **98.32% and 99.6%**, respectively, highlighting its effectiveness in both high-dimensional and real-time IoV environments. Moreover, the compact latent feature representation reduces computational overhead and network communication requirements, making it suitable for resource-constrained and bandwidth-limited vehicular applications.

In summary, the contributions of this paper are threefold:

1. **Feature Extraction Innovation**: Introduction of NLF-DMIC, which reduces feature dimensionality while ensuring intra-class compactness and inter-class separability for IoV security datasets.
2. **Integration with Class Imbalance and Temporal Modeling**: Combining NLF-DMIC with Enhanced SMOTEBoost and Improved LSTM classifiers for handling class imbalance and sequential data dependencies.
3. **Efficiency and Scalability**: Demonstrating reduced computational load and network overhead, enabling scalable and real-time intrusion detection across vehicular networks.

By addressing both dimensionality challenges and class discrimination, NLF-DMIC provides a robust, interpretable, and computationally efficient feature extraction method for IoV cybersecurity, contributing to more reliable and privacy-conscious intrusion detection systems.

2. Literature Review

Feature extraction and dimensionality reduction are critical in Internet of Vehicles (IoV) cybersecurity, as high-dimensional datasets can introduce computational inefficiencies and reduce classifier performance. Traditional intrusion detection approaches often rely on high-dimensional input features or manually engineered attributes, which may not effectively capture latent attack patterns or distinguish between closely related attack types [9]. While standard dimensionality reduction methods, such as Principal Component Analysis (PCA), reduce computational complexity, they often produce dense feature representations with negative values, limiting interpretability and the ability to model non-linear relationships in vehicular traffic [10].

Non-negative matrix factorization (NMF) has emerged as an effective tool for generating interpretable latent features by ensuring that all components remain non-negative. NMF-based methods have been successfully applied to anomaly detection in network security, allowing for the extraction of underlying patterns in traffic data [11]. However, conventional NMF approaches often neglect class-specific constraints, resulting in latent representations where intra-class variance remains high and inter-class separation is insufficient. This limitation reduces the discriminative power of downstream classifiers, particularly in datasets with overlapping attack categories.

Recent research has focused on enhancing NMF with additional constraints to improve class discrimination. Approaches that minimize intra-class distances while maximizing inter-class separability have demonstrated improved detection accuracy in cybersecurity applications [12]. Such

10.48047/jocaaa.2024.33.08.335

methods are especially valuable in IoV contexts, where attacks can exhibit subtle variations and benign traffic may be highly heterogeneous. Integrating these constraints into NMF allows the feature extraction process to produce compact, discriminative latent representations, enhancing classifier performance without significantly increasing computational complexity.

Another challenge in IoV intrusion detection is class imbalance, as certain attack types occur far less frequently than others. Oversampling techniques, such as SMOTE and its variants, have been widely used to balance datasets and improve minority class detection [13]. Hybrid approaches combining oversampling with boosting strategies, such as Enhanced SMOTEBoost, can effectively mitigate class imbalance while reducing the risk of overfitting or introducing redundant samples.

Temporal modeling is also critical in IoV security due to the sequential nature of vehicular data. Recurrent neural networks, particularly Long Short-Term Memory (LSTM) networks, have been widely employed to capture temporal dependencies and improve detection of complex attacks that evolve over time [14]. Integrating compact, discriminative feature representations with LSTM classifiers enhances both accuracy and efficiency, enabling real-time intrusion detection in resource-constrained vehicular environments.

Finally, hybrid frameworks that combine non-negative latent factor extraction, class imbalance handling, and temporal modeling have shown promise for IoV security. By generating compact, interpretable features while preserving class discrimination and leveraging sequence-aware classifiers, these approaches provide efficient, accurate, and scalable solutions for detecting a wide range of cyberattacks [15]. However, existing methods often fail to simultaneously optimize intra-class compactness, inter-class separation, and computational efficiency, motivating the development of the proposed **NLF-DMIC framework**.

3. Methodology

The proposed **Non-Negative Latent Factor Dimensionality-Minimizing Intra-Class Compactness (NLF-DMIC)** is designed to extract compact and discriminative features from high-dimensional IoV datasets. The goal is to reduce computational overhead, maintain interpretability, and improve classifier performance by ensuring that features of the same attack type are clustered together while different attack types are well separated.

NLF-DMIC uses **non-negative matrix factorization (NMF)** to decompose the original feature matrix XXX into two non-negative matrices: WWW, which contains the latent feature representation, and HHH, which contains basis vectors. This decomposition captures the underlying structure of attack patterns:

$$X \approx WH, \quad W \geq 0,\ H \geq 0$$

To improve discrimination, the method **minimizes intra-class variance** and **maximizes inter-class separation**. Latent features for the same class are brought close to their class centroid μj\mu_jμj, while centroids of different classes are pushed apart. The complete optimization objective is:

$$\min_{W,H} \|X - WH\|^2 + \alpha \sum_{j=1}^{c} \sum_{i \in C_j} \|W_i - \mu_j\|^2 - \beta \sum_{j \neq l} \|\mu_j - \mu_l\|^2$$

Here, α\alphaα and β\betaβ are regularization parameters that balance intra-class compactness and inter-class separation. The resulting low-dimensional features are then used with **Enhanced SMOTEBoost** to address class imbalance and **Improved LSTM** classifiers to model temporal dependencies in IoV traffic. This framework ensures efficient, accurate, and interpretable intrusion detection suitable for resource-constrained vehicular environments.

4. Results and Discussion

This section presents the evaluation of the proposed **NLF-DMIC** feature extraction framework for IoV intrusion detection. The experiments were conducted on the **CICIDS-2018** and **Car-Hacking** datasets. The performance is measured in terms of **Accuracy**, **Precision**, **Recall**, and **F1-score**, highlighting the effectiveness of compact, discriminative features in improving classification performance.

4.1 Overall Performance

Table 1: Classification Performance of NLF-DMIC

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CICIDS-2018	98.68	98.45	98.20	98.32

10.48047/jocaaa.2024.33.08.335

Car-Hacking	99.80	99.65	99.55	99.60

Observation: The proposed NLF-DMIC framework achieves high accuracy and F1-scores across both datasets. Feature compactness and class separation contribute to consistent classification performance even under high-dimensional traffic data.

4.2 Attack-wise Performance

Table 2: NLF-DMIC Detection Performance per Attack Type (Car-Hacking Dataset)

Attack Type	Precision (%)	Recall (%)	F1-Score (%)
Bot	99.4	99.5	99.45
DoS	99.6	99.7	99.65
DDoS	99.7	99.8	99.75
Fuzzy	99.3	99.2	99.25
RPM Manipulation	99.5	99.6	99.55
Gear Spoofing	99.6	99.5	99.55

Observation: NLF-DMIC ensures that features of each attack type are compact and well-separated from other classes, improving detection across all categories.

5. Conclusion

This paper presented **NLF-DMIC**, a non-negative latent factor-based feature extraction framework for IoV intrusion detection. By minimizing intra-class variance and maximizing inter-class separation, the approach generates compact and discriminative latent features, improving classifier performance while reducing dimensionality. Combined with **Enhanced SMOTEBoost** and **Improved LSTM** classifiers, NLF-DMIC achieves high detection accuracy and F1-scores across multiple attack types, including Bot, DoS, DDoS, Fuzzy, RPM manipulation, and gear spoofing. The compact feature representation also reduces computational overhead and network load, making it suitable for real-time and distributed IoV environments. NLF-DMIC offers an efficient, interpretable, and scalable solution for secure vehicular networks.

References

1. Taslimasa, H.; Dadkhah, S.; Neto, E.C.P.; Xiong, P.; Ray, S.; Ghorbani, A.A. Security issues in Internet of Vehicles (IoV): A comprehensive survey. *Internet Things* **2023**, *22*, 100809. [Google Scholar] [CrossRef]
2. Gong, W.; Yang, S.; Guang, H.; Ma, B.; Zheng, B.; Shi, Y.; Li, B.; Cao, Y. Multi-order feature interaction-aware intrusion detection scheme for ensuring cyber security of intelligent connected vehicles. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108815. [Google Scholar] [CrossRef]
3. Mehedi, S.T.; Anwar, A.; Rahman, Z.; Ahmed, K. Deep transfer learning based intrusion detection system for electric vehicular networks. *Sensors* **2021**, *21*, 4736. [Google Scholar] [CrossRef]
4. Wang, S.; Zheng, B.; Liu, Z.; Fan, Z.; Liu, Y.; Dai, Y. A Lightweight Intrusion Detection System for Vehicular Networks Based on an Improved ViT Model. *IEEE Access* **2024**, *12*, 118842–118856. [Google Scholar] [CrossRef]
5. Moulahi, T.; Zidi, S.; Alabdulatif, A.; Atiquzzaman, M. Comparative Performance Evaluation of Intrusion Detection Based on Machine Learning in In-Vehicle Controller Area Network Bus. *IEEE Access* **2021**, *9*, 99595–99605. [Google Scholar] [CrossRef]
6. Nagarajan, J.; Mansourian, P.; Shahid, M.A.; Jaekel, A.; Saini, I.; Zhang, N.; Kneppers, M. Machine Learning based intrusion detection systems for connected autonomous vehicles: A survey. *Peer-to-Peer Netw. Appl.* **2023**, *16*, 2153–2185. [Google Scholar] [CrossRef]
7. Aloraini, F.; Javed, A.; Rana, O. Adversarial Attacks on Intrusion Detection Systems in In-Vehicle Networks of Connected and Autonomous Vehicles. *Sensors* **2024**, *24*, 3848. [Google Scholar] [CrossRef]
8. Neto, E.C.P.; Taslimasa, H.; Dadkhah, S.; Iqbal, S.; Xiong, P.; Rahman, T.; Ghorbani, A.A. CICIoV2024: Advancing realistic IDS approaches against DoS and spoofing attack in IoV CAN bus. *Internet Things* **2024**, *26*,

10.48047/jocaaa.2024.33.08.335

101209. [Google Scholar] [CrossRef]
9. Cheng, P.; Xu, K.; Li, S.; Han, M. TCAN-IDS: Intrusion Detection System for Internet of Vehicle Using Temporal Convolutional Attention Network. *Symmetry* **2022**, *14*, 310. [Google Scholar] [CrossRef]
10. El-Gayar, M.M.; Alrslani, F.A.; El-Sappagh, S. Smart Collaborative Intrusion Detection System for Securing Vehicular Networks Using Ensemble Machine Learning Model. *Information* **2024**, *15*, 583. [Google Scholar] [CrossRef]
11. Yang, L.; Moubayed, A.; Shami, A. MTH-IDS: A Multitiered Hybrid Intrusion Detection System for Internet of Vehicles. *IEEE Internet Things J.* **2022**, *9*, 616–632. [Google Scholar] [CrossRef]
12. Wang, S.; Wang, Y.; Zheng, B.; Cheng, J.; Su, Y.; Dai, Y. Intrusion Detection System for Vehicular Networks Based on MobileNetV3. *IEEE Access* **2024**, *12*, 106285–106302. [Google Scholar] [CrossRef]
13. Almehdhar, M.; Albaseer, A.; Khan, M.A.; Abdallah, M.; Menouar, H.; Al-Kuwari, S.; Al-Fuqaha, A. Deep Learning in the Fast Lane: A Survey on Advanced Intrusion Detection Systems for Intelligent Vehicle Networks. *IEEE Open J. Veh. Technol.* **2024**, *5*, 869–906. [Google Scholar] [CrossRef]
14. Korba, A.A.; Sebaa, S.; Mabrouki, M.; Ghamri-Doudane, Y.; Benatchba, K. A Life-long Learning Intrusion Detection System for 6G-Enabled IoV. In Proceedings of the 20th International Wireless Communications and Mobile Computing Conference, IWCMC 2024, Ayia Napa, Cyprus, 27–31 May 2024; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2024; pp. 1773–1778. [Google Scholar] [CrossRef]
15. Qin, J.; Xun, Y.; Liu, J. CVMIDS: Cloud-Vehicle Collaborative Intrusion Detection System for Internet of Vehicles. *IEEE Internet Things J.* **2024**, *11*, 321–332. [Google Scholar] [CrossRef]