2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)

An Integrated Approach to Enhancing Equipment Anomaly Detection Efficiency in Large Language Models Using Multiple Machine Learning Algorithms

Ziwen Jin
Key Laboratory of Networked Control Systems, Chinese Academy of Sciences
Shenyang, China
University of Chinese Academy of Sciences
Beijing, China
jinziwen@sia.cn

Jianming Zhao *
State Key Laboratory of Robotics
Shenyang Institute of Automation
Chinese Academy of Sciences
Shenyang, China
Key Laboratory of Networked Control Systems, Chinese Academy of Sciences
Shenyang, China
University of Chinese Academy of Sciences
Beijing, China
zhaojianming@sia.cn

Wenyu Li
Key Laboratory of Networked Control System, Chinese Academy of Sciences
Shenyang, China
liwenyu@sia.cn

Chuan Sheng
State Key Laboratory of Robotics
Shenyang Institute of Automation, Chinese Academy of Sciences
Shenyang, China
Key Laboratory of Networked Control Systems, Chinese Academy of Sciences
Shenyang, China
shengchuan@sia.cn

Tieliang Sun
National Petroleum Pipeline Network Group Co., Ltd
Beijing, China
stl@pipechina.com.cn

Feng Lv
National Petroleum Pipeline Network Group Co., Ltd
Beijing, China
lf@pipechina.com.cn

***Abstract*— As the application of Industrial Control Systems (ICS) in cyberspace continues to expand, cyberattacks targeting ICS have become increasingly sophisticated and frequent. Such attacks not only threaten the normal operation of systems but also pose significant risks, including substantial economic losses and social impacts. This paper focuses on enhancing the cybersecurity protection capabilities of ICS, particularly improving the detection efficiency in identifying typical threats such as Denial-of-Service (DoS), Man-in-the-Middle (MITM) attacks, and malware infiltration. We introduce large language models as a novel analytical tool to conduct deep analysis of network traffic, aiming to achieve more accurate identification of anomalous behaviors and establish faster response mechanisms. Through a series of experiments using datasets, the effectiveness of this method has been validated. However, our research also uncovers limitations in the existing technological framework, such as the extensive computational resources required for model training and the need to improve real-time processing performance. To address these issues, the paper further explores optimization measures, including refining algorithm structures and reducing feature dimensions, as well as other improvements. Moreover, suggestions for future research directions are proposed, aiming to advance the security and reliability of ICS operations and support continuous socio-economic development.**

Keywords—Industrial Control Systems (ICS) Security, Anomaly Detection, Machine Learning Algorithm, Large Language Model

I. INTRODUCTION

Industrial Control Systems (ICS) are systems used for the automation operations, process control, and monitoring of industrial infrastructure, comprising automated control components and real-time data acquisition and monitoring elements [1]. ICS include Supervisory Control and Data Acquisition (SCADA) systems, Distributed Control Systems (DCS), Programmable Logic Controllers (PLC), and Process Control Systems (PCS). These systems are widely applied in critical industries such as power, water, oil and gas, nuclear energy, chemicals, transportation, pharmaceuticals, food, and discrete manufacturing.

Historically, ICS were deployed in physically isolated areas with a focus on functional implementation rather than cybersecurity design. This closed environment limited the capabilities for remote operation and maintenance. As industrialization and informatization deepens, an increasing number of ICS have been connected to enterprise networks to support remote control and supervision, thereby exposing them to public networks and increasing the risk of cyberattacks. Such exposure not only threatens personal safety and the security of production environments but can also result in significant economic losses. Therefore, ensuring the security and stability of ICS is essential for maintaining normal production and safeguarding facility security.

Common attack methods include, but are not limited to, Distributed Denial of Service (DDoS), Man-in-the-Middle (MITM) attacks, and malicious code injection. DDoS attacks overwhelm the target system's processing capability or bandwidth with multi-source traffic to prevent normal access; MITM attacks involve interception and alteration of communication information, threatening data integrity and

This work was supported by National Key R&D Program of China (2023YFB3107700)
979-8-3315-0976-7/25/$31.00 ©2025 IEEE

confidentiality; and malicious code injection compromises system functionality or steals data, such as SQL injection, cross-site scripting and code injection attacks [2-5].

Given that many industrial control devices have low power and limited resources, directly deploying security features may degrade performance or lead to downtime. Moreover, the original designs did not adequately account for the requirements for new security services, which might affect high availability and reliability. Thus, anomaly detection, a non-intrusive method, is widely recognized as it has minimal impact on system availability and real-time performance. To enhance the security protection capabilities of ICS, especially to improve the detection efficiency of the aforementioned typical threats, large language models are introduced as a new analytical tool. They perform deep analysis of network traffic to achieve more precise identification of anomalous behaviors and establish faster response mechanisms.

II. Literature Review

In the study by Zhao et al. [6], a security probing method for industrial control devices based on fingerprint space construction was proposed to address the challenges of acquiring fingerprints and their labels from industrial control devices, as well as the low fingerprint resolution inherent in existing technologies. This method establishes an initial fingerprint space for industrial control devices and employs anomaly detection algorithms to filter out non-industrial fingerprints, while simultaneously enriching label information. Ultimately, a cluster analysis is conducted on industrial control devices within the target network to determine their label information. The research findings indicate that this method enhances the accuracy of identifying device manufacturers and types, contributing to improved automated asset management in industrial control systems. It also facilitates the development of a cybersecurity collaborative defense system.

In the study by Zhang et al. [7], the exploration of leveraging Large Language Models (LLMs) for automated network intrusion detection in wireless communication networks was conducted, with an enhancement of LLMs' performance achieved through in-context learning. Experimental results show that for the GPT-4 model, only 10 in-context learning examples are sufficient to significantly improve detection accuracy and F1-Score to over 95%. This indicates that even with limited task-specific data, pre-trained LLMs can achieve excellent performance, thereby providing an effective solution for intrusion detection in wireless communication networks.

In the study by Rung-Ching Chen et al. [8], the investigation into feature selection via machine learning methods was conducted, with a specific emphasis on the Random Forest (RF) algorithm. The article presents experiments on three high-dimensional datasets—Bank Marketing, Car Evaluation, and Smartphone-Based Human Activity Recognition—to evaluate the performance of classifiers such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Linear Discriminant Analysis (LDA). The results demonstrate that Random Forest outperforms all other classifiers across all experimental groups, particularly in feature selection, where RF's varImp method proves more effective than Boruta and Recursive Feature Elimination (RFE). The article highlights the importance of feature selection for simplifying models, reducing training time, preventing overfitting, and alleviating the curse of dimensionality.

In the study by Alduailij et al. [9], a machine learning-based approach for detecting Distributed Denial of Service (DDoS) attacks was proposed, aiming to reduce misclassification errors and improve detection accuracy. The article applies Mutual Information (MI) and Random Forest Feature Importance (RFFI) for feature selection, and evaluates various models—including Random Forest (RF), Gradient Boosting, Weighted Voting Ensemble, K-Nearest Neighbors (KNN), and Logistic Regression—for classification. The experimental results ** reveal that RF achieves the highest accuracy (0.99) and the lowest misclassification rate**, with only one attack misclassified as normal. Compared to existing methods, this approach demonstrates superior performance in accuracy and misclassification rate, highlighting its effectiveness and advantages in DDoS detection.

III. Methodology

Organizing traffic extracted from Industrial Control Systems (ICS) into fingerprint vectors for further analysis and monitoring is an effective method for anomaly detection and identification in industrial control networks. With the rapid development of large language models (LLMs), they have been widely adopted across various domains. However, the computational resources required for training LLMs are substantial, their training time is typically long, and their real-time processing capabilities remain limited. On one hand, reducing computational resources and training time can lower operational costs; on the other hand, real-time processing capability is a critical parameter that determines whether ICS can promptly respond to anomalies to prevent severe losses.

This paper proposes a methodology based on a real-world experimental dataset from a water treatment plant's industrial control environment. In practical applications, anomaly detection systems in water treatment plants typically integrate sensors, PLCs (Programmable Logic Controllers), and intelligent algorithms to rapidly identify water quality issues, equipment malfunctions, or other anomalies, triggering alarms or emergency responses within seconds. A faster response time for anomaly detection is crucial to enhance water treatment quality, mitigate equipment damage, and prevent accidents.

To address these challenges, this paper proposes a hybrid method combining multiple machine learning algorithms to assess fingerprint vector feature importance, retain significant features, and compress the vectors. The compressed vectors are then fine-tuned using large language models, aiming to reduce training resource requirements and shorten training time.

A. Acquisition of Fingerprint Vectors

Following the methodology proposed by Zhao et al. [6], the process begins by combining collected industrial control device attributes—such as IP and TCP protocol header fields, timestamp options, and other metadata—with device labels (including IP addresses, manufacturer names, and device types) to form complete fingerprint vectors. Subsequently, these vectors are organized into an industrial control device fingerprint space. Next, network traffic from the target ICS is exported, and

device-specific fingerprint vectors (DF) are extracted based on IP addresses using Formula 1. Clustering algorithms classify these vectors, and a voting mechanism determines cluster label information. This method enables security probing and device identification within the network, profiling devices for enhanced anomaly detection and network health monitoring.

$$DF = \{ITTL, IPDF, IDD, IWS, MSS, WSC, SAP, ILRT, TON, TSCON, TCF, RTD\} \quad (1)$$

The fingerprint vectors obtained through the method proposed in this paper consist of 12 elements. However, when combined with the massive data flow in industrial control systems, the volume of data used for identification and detection remains substantial. To address this issue, we propose employing a Random Forest (RF) algorithm to analyze the importance of these fingerprint vectors, as illustrated in Figure 1. The result of this analysis is a feature importance ranking chart for the generated vectors, shown in Figure 2.

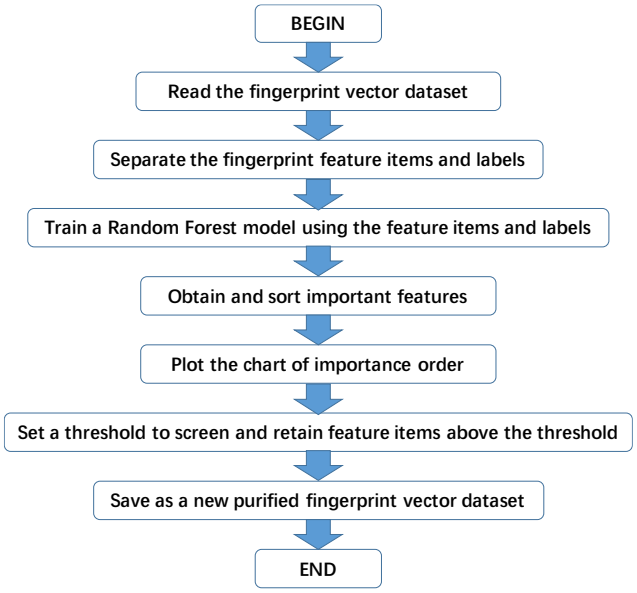

Fig. 1. Flowchart of the Fingerprint Vector Purification Process Using the Random Forest Algorithm

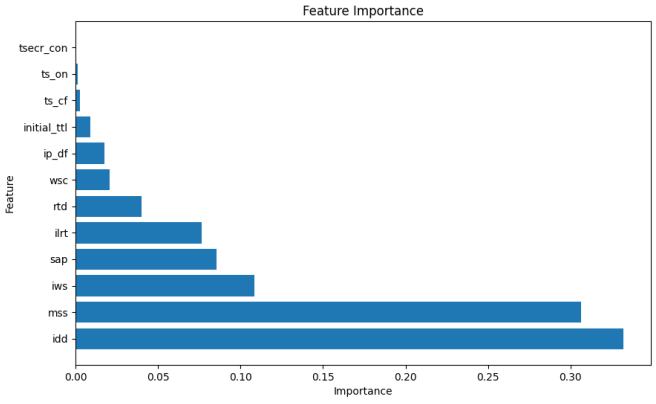

Fig. 2. Chart of Feature Importance Ranking for Fingerprint Vectors

Features with importance scores greater than 5% (as determined by RF) were selected as effective features for the industrial control equipment fingerprint vector, yielding five features that met this criterion. To mitigate the risk of missing critical information inherent in the data when relying on a single feature selection method, we applied Recursive Feature Elimination (RFE) and Linear Mutual Information (LMI) to score the original fingerprint vector features. The final selection of features was based on the average importance scores from RF, RFE, and LMI. As shown in Table 1, the five features with the highest composite scores were retained to form the final fingerprint vector, as described in Formula 2. After refinement, the fingerprint vector size was reduced to 41.67% of its original dimensions, significantly alleviating the computational burden during large model fine-tuning.

TABLE I. FEATURE IMPORTANCE SCORING TABLE

Features	*RF SCORES*	*REF SCORES*	*LMI SCORES*	*TOTAL SCORES*
idd	0.3321	0.1410	0.1817	0.6548
mss	0.3062	0.1538	0.1903	0.6503
ilrt	0.0764	0.1282	0.1554	0.3600
iws	0.1086	0.1154	0.1197	0.3437
rtd	0.0402	0.0897	0.1477	0.2777
sap	0.0854	0.1026	0.0409	0.2288
wsc	0.0205	0.0769	0.0557	0.1531
ip_df	0.0177	0.0641	0.0525	0.1342
initial_ttl	0.0088	0.0256	0.0414	0.0759
ts_cf	0.0027	0.0513	0.0052	0.0592
ts_on	0.0012	0.0385	0.0050	0.0446
tsecr_con	0.0003	0.0128	0.0044	0.0175

$$DF = \{IDD, MSS, ILRT, IWS, RTD\} \quad (2)$$

B. *Training Methods for Large Models*

From the original fingerprint vector dataset, 3,000 samples were randomly selected for fine-tuning large language models under both "adequate" and "inadequate" conditions, as illustrated in Figure 3. For testing, 1,148 samples were randomly selected from the test dataset and input into the fine-tuned models to evaluate anomaly detection performance.

The purified fingerprint vectors were also used to extract 3,000 training samples, which were fine-tuned using the same models. The test samples were aligned with training features and processed through the fine-tuned models to assess detection accuracy again.

This study employed public datasets and evaluation results from authoritative institutions to test three 4-bit quantized models—Gemma-7b-bnb-4bit, Llama-3-8b-bnb-4bit, and Mistral-7b-bnb-4bit—using Unsloth. After applying Unsloth acceleration, GPU resource consumption remained below 8GB, and fine-tuning speed was approximately doubled compared to non-accelerated training. Various large language models were

tested to evaluate the applicability of purifying industrial control device fingerprint vectors using multiple machine learning algorithms.

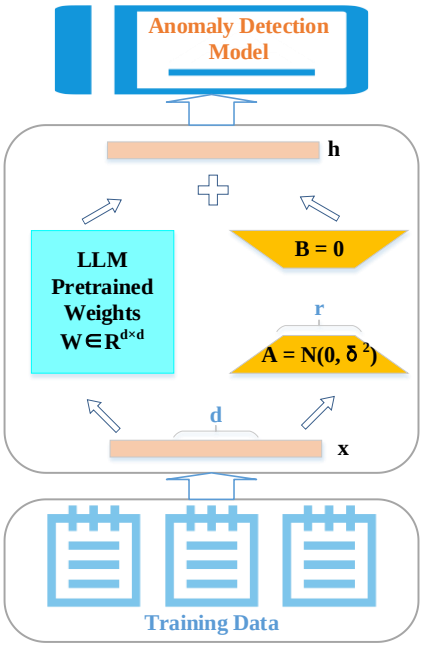

Fig. 3. Specific process of fine-tuning large language model

C. *Evaluation Criteria*

Precision measures the proportion of true positive predictions among all positive predictions, reflecting the model's accuracy for positive classes. Recall evaluates the proportion of actual positives correctly identified, highlighting the model's ability to detect all positive instances. The F1-Score is the harmonic mean of precision and recall, balancing both metrics for comprehensive evaluation. AUC-PR (Area Under the Precision-Recall Curve) assesses classifier performance across classification thresholds, particularly useful for imbalanced datasets where minority class detection is critical. AUC-PR values range from 0 to 1, with higher values indicating superior model performance.

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+FP+TN+FN} \\ F1\text{-}Score &= 2*\frac{Precision*Recall}{Precision+Recall} \\ Recall &= \frac{TP}{TP+FN} \end{aligned} \tag{3}$$

IV. Result Analysis

The Llama model was fine-tuned for 100 epochs using both the original training dataset and the refined dataset. Subsequently, anomaly detection was performed on the original test dataset and the refined test dataset, with the results depicted in Figure 4. The experiments demonstrated that training time decreased from 356 seconds to 253 seconds (28.93% reduction) when using the refined dataset. Additionally, the single-response time during testing decreased from 0.78 seconds to 0.7 seconds (10.26% reduction), showing significant improvements in both training and testing speeds.

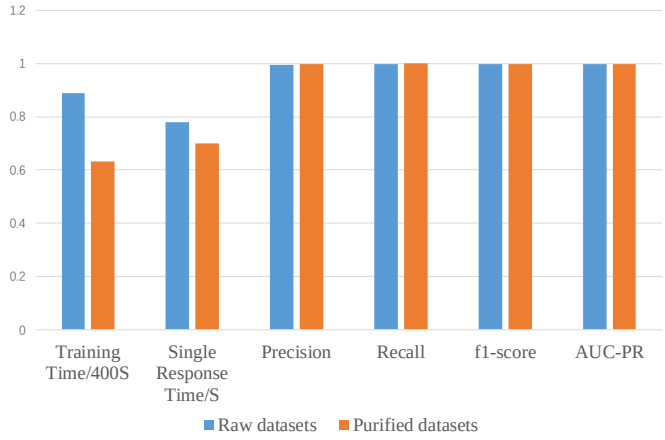

Fig. 4. Comparison of experimental results of fully trained Llama model before and after data purification

In terms of performance metrics, the accuracy, recall, F1-score, and AUC-PR values for anomaly detection using the original dataset were 99.65%, 99.77%, 99.77%, and 99.86%, respectively. In contrast, those using the refined dataset achieved 99.83%, 100.00%, 99.88%, and 99.88%. All metrics showed improvements, indicating that dataset refinement not only enhanced training and testing efficiency but also further optimized model performance.

The Llama model was fine-tuned for 30 epochs using both the original training dataset and the refined dataset. Subsequently, anomaly detection was performed on the original test dataset and the refined test dataset, with the results depicted in Figure 5. The experiments revealed that training time decreased from 104 seconds to 74 seconds (28.85% reduction) when using the refined dataset. Additionally, the single-response time during testing decreased from 0.76 seconds to 0.72 seconds (5.26% reduction), demonstrating significant improvements in both training and testing speeds.

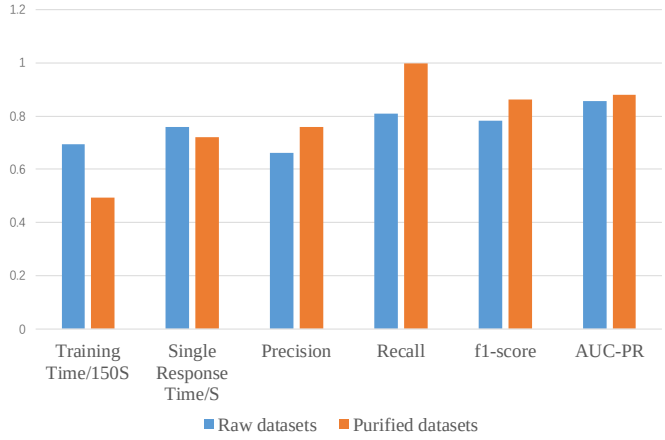

Fig. 5. Comparison of experimental results of insufficient training Llama model before and after data purification

In terms of performance metrics, the accuracy, recall, F1-score, and AUC-PR values for anomaly detection using the original dataset were 66.20%, 80.92%, 78.30%, and 85.57%, respectively. In contrast, those using the refined dataset achieved 75.87%, 99.88%, 86.18%, and 87.87%. All metrics showed

improvements, indicating that dataset refinement can enhance training and testing efficiency even with insufficient training epochs. Moreover, the model's performance demonstrated more significant improvements compared to scenarios with sufficient training epochs.

The Mistral model was fine-tuned for 100 epochs using both the original training dataset and the refined dataset. Subsequently, anomaly detection was performed on the original test dataset and the refined test dataset, with the results depicted in Figure 6. The experiments revealed that training time decreased from 428 seconds to 288 seconds (32.71% reduction) when using the refined dataset. Additionally, the single-response time during testing decreased from 0.88 seconds to 0.77 seconds (12.50% reduction), demonstrating significant improvements in both training and testing speeds.

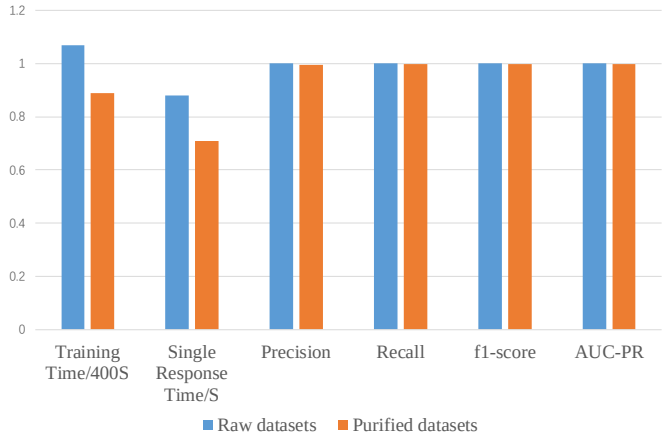

Fig. 6. Comparison of experimental results of fully trained Mistral model before and after data purification

In terms of performance metrics, the accuracy, recall, F1-score, and AUC-PR values for anomaly detection using the original dataset were 100.00%, 100.00%, 100.00%, and 100.00%, respectively. In contrast, those using the refined dataset achieved 99.65%, 99.77%, 99.77%, and 99.86%. While all metrics remained at a high level, the results indicate that dataset refinement can enhance training and testing efficiency under sufficient training epochs without compromising model performance.

The Mistral model was fine-tuned for 30 epochs using both the original training dataset and the refined dataset. Subsequently, anomaly detection was performed on the original test dataset and the refined test dataset, with the results depicted in Figure 7. The experiments revealed that training time decreased from 125 seconds to 84 seconds (32.80% reduction) when using the refined dataset. Additionally, the single-response time during testing decreased from 0.76 seconds to 0.69 seconds (9.21% reduction), demonstrating significant improvements in both training and testing speeds.

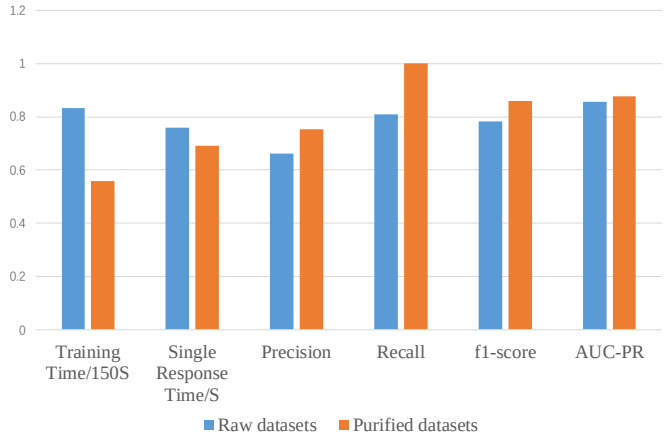

Fig. 7. Comparison of experimental results of insufficient training Mistral model before and after data purification

In terms of performance metrics, the accuracy, recall, F1-score, and AUC-PR values for anomaly detection using the original dataset were 66.20%, 80.92%, 78.30%, and 85.57%, respectively. In contrast, those using the refined dataset achieved 75.35%, 100.00%, 85.94%, and 87.67%. While all metrics improved, the results indicate that dataset refinement can enhance efficiency under limited training epochs while significantly boosting model performance.

The Gemma model was fine-tuned for 100 epochs using both the original training dataset and the refined dataset. Subsequently, anomaly detection was performed on the original test dataset and the refined test dataset, with the results depicted in Figure 8. The experiments revealed that training time decreased from 452 seconds to 320 seconds (29.20% decrease) when using the refined dataset. Additionally, the single-response time during testing decreased from 0.81 seconds to 0.72 seconds (11.11% decrease), demonstrating significant improvements in both training and testing speeds.

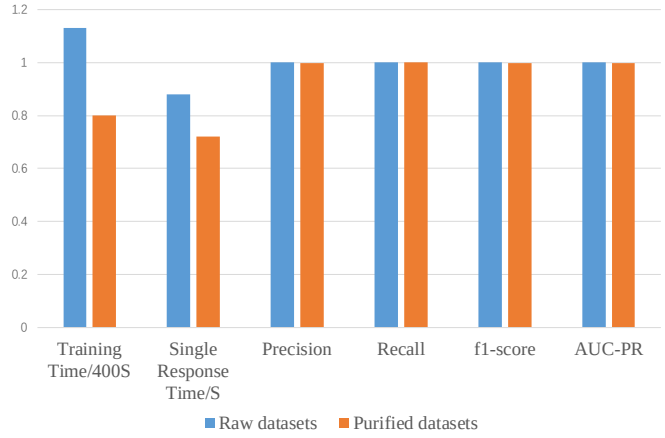

Fig. 8. Comparison of experimental results of fully trained Gemma model before and after data purification

In terms of performance metrics, the accuracy, recall, F1-score, and AUC-PR values for anomaly detection using the original dataset were 99.83%, 100.00%, 99.88%, and 99.88%, respectively. In contrast, those using the refined dataset achieved 99.91%, 100.00%, 99.94%, and 99.94%. While all metrics

remained at a high level, the results indicate that dataset refinement can enhance training and testing efficiency under sufficient training epochs while further improving model performance.

The Gemma model was fine-tuned for 30 epochs using both the original training dataset and the refined dataset. Subsequently, anomaly detection was performed on the original test dataset and the refined test dataset, with the results depicted in Figure 9. The experiments revealed that training time decreased from 132 seconds to 94 seconds (28.79% decrease) when using the refined dataset. Additionally, the response time per test instance was reduced from 0.81 seconds to 0.72 seconds (11.11% decrease), demonstrating significant improvements in both training and testing speeds.

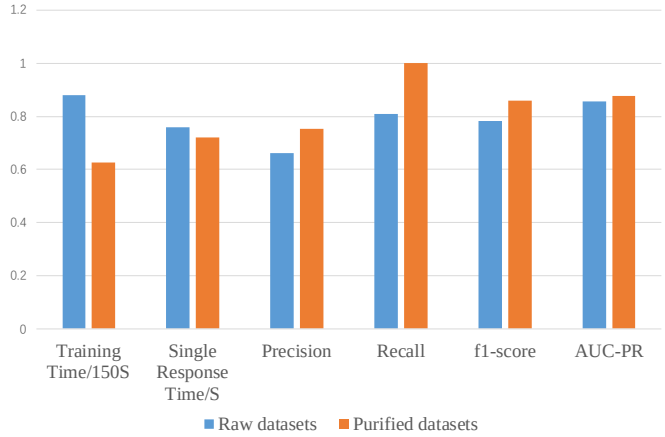

Fig. 9. Comparison of experimental results of insufficient training Gemma model before and after data purification

In terms of performance metrics, the accuracy, recall, F1-score, and AUC-PR values for anomaly detection using the original dataset were 75.35%, 100.00%, 85.94%, and 87.67%, respectively. However, those using the refined dataset remained identical at 75.35%, 100.00%, 85.94%, and 87.67%. All metrics remained stable, indicating that dataset refinement can enhance training and testing efficiency even with limited training epochs, while maintaining consistent model performance.

V. Conclusion

The method of purifying industrial control equipment fingerprint features using multiple machine learning algorithms—specifically retaining significant features for anomaly detection—proves effective across various large language models and under both adequate and inadequate fine-tuning conditions. This approach not only accelerates model fine-tuning speed but also reduces the per-item anomaly detection latency of the detection model. Furthermore, it maintains detection accuracy and may even slightly enhance model performance in cases of inadequate fine-tuning.

The training and testing datasets employed in this study are relatively small, meaning that even without the proposed method, model fine-tuning time remains short. However, in real-world industrial scenarios, continuous collection and training of data from all network devices are essential to ensure system reliability. Such scenarios may involve datasets orders of magnitude larger than our experimental data (e.g., 3 million data points versus 3,000 in this study). Our method reduces fine-tuning time for the three models by 15%-30%, significantly improving efficiency. In high-risk environments like water conservation facilities, thermal power plants, oil refineries, and chemical plants, rapid anomaly detection and response are critical for personnel to mitigate risks promptly.

Finally, the proposed anomaly detection method, which leverages inherent equipment fingerprint characteristics, demonstrates a unique advantage: fingerprint traits are intrinsic properties of industrial control equipment and remain stable provided the equipment functions normally, even if network conditions change or new threats emerge. During practical deployment, fingerprints can be captured during system commissioning or stable operation phases to build a secure database, which trains a large-scale fingerprint model. After system updates, maintenance, or equipment replacements, re-capturing fingerprints and updating the database ensures sustained accuracy and detection effectiveness.

References

[1] Alladi T, Chamola V, Zeadally S. Industrial control systems: Cyberattack trends and countermeasures[J]. Computer Communications, 2020, 155: 1-8.

[2] Hosseini S, Azizi M. The hybrid technique for DDoS detection with supervised learning algorithms[J]. Computer Networks, 2019, 158: 35-45.

[3] Abomhara M, Køien G M. Security and privacy in the Internet of Things: Current status and open issues[C]//2014 international conference on privacy and security in mobile systems (PRISMS). IEEE, 2014: 1-8.

[4] D'Orsaneo J, Tummala M, McEachen J, Martin B. Analysis of Traffic Signals on an SDN for Detection and Classification of a Man-in-the-Middle Attack[C]//2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS). IEEE, 2018: 1-9.

[5] Guan Y, Ge X. Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks[J]. IEEE Transactions on Signal and Information Processing over Networks, 2017, 4(1): 48-59.

[6] Zhao J, Jin Z, Zeng P, Sheng C, Wang T. An Anomaly Detection Method for Oilfield Industrial Control Systems Fine-Tuned Using the Llama3 Model[J]. Applied Sciences, 2024, 14(20): 9169.

[7] Zhang H, Sediq A B, Afana A, Erol-Kantarci M. Large Language Models in Wireless Application Design: In-Context Learning-enhanced Automatic Network Intrusion Detection[J]. arXiv preprint arXiv:2405.11002, 2024.

[8] Chen R C, Dewi C, Huang S W, Caraka R E Selecting critical features for data classification based on machine learning methods[J]. Journal of Big Data, 2020, 7(1): 52.

[9] Alduailij M, Khan Q W, Tahir M, Sardaraz M, Alduailij M, Malik F Machine-learning-based DDoS attack detection using mutual information and random forest feature importance method[J]. Symmetry, 2022, 14(6): 1095.