Journal of Engineering Research and Reports

Volume 26, Issue 3, Page 161-173, 2024; Article no.JERR.114125
ISSN: 2582-2926

Enhancing Industrial Control System Security: An Isolation Forest-based Anomaly Detection Model for Mitigating Cyber Threats

Md. Saif Mahmud [a], Md. Ashikul Islam [b], Md. Maruf Rahman [a], Debashon Chakraborty [c], Shaharier Kabir [d], Abu Shufian [d*] and Protik Parvez Sheikh [d]

[a] Department of Business, Engineering & Technology, Texas A&M University-Texarkana, Texas, USA.
[b] Department of Electrical Engineering, Lamar University, Texas, USA.
[c] Department of College of Business, Lamar University, Texas, USA.
[d] Department of Electrical & Electronic Engineering, American International University-Bangladesh, Dhaka, Bangladesh.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JERR/2024/v26i31102

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/114125

Original Research Article

Received: 29/12/2023
Accepted: 03/03/2024
Published: 06/03/2024

ABSTRACT

In the evolving landscape of industrial control systems (ICS), the sophistication of cyber threats has necessitated the development of advanced anomaly detection mechanisms to safeguard critical infrastructure. This study introduces a novel anomaly detection model based on the Isolation Forest algorithm, tailored for the complex environment of ICS. Unlike traditional detection methods that

**Corresponding author: Email: shufian.eee@aiub.edu, shufian.eee@gmail.com;*

often rely on predefined thresholds or patterns, our model capitalizes on the Isolation Forest's ability to efficiently isolate anomalies in high-dimensional datasets, making it particularly suited for the dynamic and intricate data generated by ICS. Leveraging the HAI dataset, which encompasses operational data from a realistic ICS testbed augmented with a Hardware-In-the-Loop (HIL) simulator, this research demonstrates the model's effectiveness in identifying both known and novel cyber threats across various ICS components. Our findings reveal that the Isolation Forest-based model outperforms traditional anomaly detection techniques in terms of detection accuracy, false positive rate, and computational efficiency. Furthermore, the model exhibits a remarkable ability to adapt to the evolving nature of cyber threats, underscoring its potential as a robust tool for enhancing the security posture of ICS. Through a detailed analysis of its application in detecting sophisticated attacks represented in the HAI dataset, this study contributes to the ongoing discourse on improving ICS security and presents a compelling case for the adoption of machine learning-based anomaly detection solutions in industrial settings.

Keywords: Anomaly detection; industrial control systems (ICS, isolation forest algorithm, cyber-physical systems (CPS); hardware-in-the-loop (HIL) simulation; adaptive threat detection.

1. INTRODUCTION

In the modern era, Industrial Control Systems (ICS) have emerged as the backbone of critical infrastructure, enabling the automation and efficient management of industrial processes across a diverse range of sectors. These systems integrate devices, networks, and controllers into cohesive frameworks that control complex operations from power generation to water treatment and transportation systems [1]. The evolution of ICS has been pivotal in advancing operational efficiency, reliability, and safety in industrial operations, making it an indispensable element of contemporary society.

At the core of ICS architecture are various control systems, including but not limited to Supervisory Control and Data Acquisition (SCADA) systems and Distributed Control Systems (DCS). SCADA systems are designed to collect and analyze data in real-time, facilitating remote monitoring and control over large geographical areas [2]. This capability is crucial for utilities and critical infrastructures, such as power grids and water distribution networks, where operational integrity and reliability are paramount. On the other hand, DCS are typically employed in manufacturing plants and process industries, like chemical processing and oil refining, to regulate production processes and ensure consistency and quality. These systems' distributed nature allows for centralized control room management while supporting local process control, enhancing both operational flexibility and system redundancy [3].

The intricate networks and systems that comprise Industrial Control Systems are not just critical; they are the lifelines of modern infrastructure, supporting everything from electricity distribution to water purification and transportation. The inherent complexity and interconnectedness of these systems mean that a single point of failure can trigger a cascade of failures across the network, leading to widespread operational disruption, economic losses, and potential harm to public safety and the environment. This domino effect underscores the paramount importance of fault detection within ICS. Fault detection in ICS is not merely about identifying malfunctions or breakdowns in hardware; it's about recognizing any deviation from normal operation that could indicate a potential security threat or system vulnerability [4]. The capability to detect these faults promptly ensures that corrective measures can be taken before minor issues escalate into major system failures or, worse, full-blown disasters. However, the challenge of fault detection in such complex and dynamic environments is significant. Traditional security mechanisms, while foundational to system security, offer limited protection against sophisticated cyber threats.

The integration of ICS into critical infrastructures signifies their importance but also highlights the potential risks and vulnerabilities associated with their operation. The reliance on digital networks and computer-based control logic exposes these systems to cyber threats, ranging from data breaches to targeted attacks aimed at disrupting industrial operations [5]. The consequences of such incidents are far-reaching, potentially leading to operational downtime, economic losses, environmental damage, and even endangering human lives. Therefore, the security of ICS is not just a technical issue but a national

security concern, necessitating robust and resilient protective measures. In light of the increasing complexity and sophistication of cyber threats, traditional security mechanisms such as firewalls, intrusion detection systems, and regular patching practices, though necessary, are no longer sufficient to guarantee the security of ICS. The dynamic and evolving nature of cyber threats requires a proactive and adaptive approach to ICS security, emphasizing the importance of advanced anomaly detection techniques capable of identifying and mitigating previously unknown threats [6].

Conventional security measures, such as authentication protocols and encryption, are designed to secure networks and systems against unauthorized access. While these measures are crucial for the foundational security of ICS, they are not infallible. Cyber attackers continually evolve their strategies and methods, developing malware and other malicious activities that can bypass these traditional defenses. The static nature of such conventional security mechanisms means they are often ill-equipped to identify or mitigate novel or sophisticated attacks that do not match known threat patterns. Moreover, the reliance on authentication and encryption does little to address the insider threat, where individuals with legitimate access intentionally or unintentionally cause harm to the system. This vulnerability highlights the need for security measures that go beyond perimeter defense and access control, advocating for a more dynamic and adaptive approach to ICS security. In response to the limitations of traditional security measures, anomaly detection emerges as a critical component of modern ICS security strategies. Unlike conventional methods that focus on preventing unauthorized access, anomaly detection aims to identify unusual patterns or behaviors within the system that could indicate a security threat or system malfunction [7].

The isolation forest algorithm represents a significant advancement in the field of anomaly detection [8]. This algorithm is particularly well-suited for identifying outliers in data, operating on the principle that anomalies are data points that are few and different. By isolating these points, the algorithm effectively identifies potential threats with a high degree of accuracy and efficiency. The isolation forest algorithm's ability to detect anomalies without the need for a detailed profile of normal operation makes it an invaluable tool for enhancing ICS security. Its implementation can serve as a dynamic and adaptive layer of defense, capable of detecting a wide range of threats, from sophisticated cyber-attacks to subtle system malfunctions that conventional measures might overlook. As industries continue to integrate advanced technologies and digital solutions into their operational frameworks, the role of ICS in managing and controlling industrial processes becomes increasingly critical. The need to ensure the security and reliability of these systems is paramount, driving the development of innovative security solutions designed to protect critical infrastructures from the ever-present threat of cyber-attacks [9]. The adoption of anomaly detection models, such as the Isolation Forest-Based Anomaly Detection Model, offers a promising path forward, enhancing the resilience of ICS against a wide range of cyber threats and ensuring the continued safe and efficient operation of critical infrastructures worldwide [10].

The integration of isolation forest algorithm, into ICS security frameworks represents a paradigm shift in how threats are identified and mitigated. By focusing on the detection of anomalies as indicators of potential threats, this approach offers a more flexible and responsive strategy for securing complex and dynamic industrial control systems. The implementation of such advanced detection methods complements traditional security measures, providing a comprehensive defense mechanism that enhances the resilience of ICS against both known and emerging cyber threats.

Incorporating anomaly detection into ICS security not only addresses the limitations of conventional mechanisms but also introduces a proactive stance in system defense. This proactive approach is crucial for anticipating and mitigating threats before they can cause significant damage, ensuring the continued safe and efficient operation of critical infrastructures. As such, the exploration and adoption of isolation forest-based anomaly detection models stand at the forefront of efforts to fortify ICS against the multifaceted landscape of cyber threats.

Implementing anomaly detection in ICS faces significant challenges, particularly in the development and training of machine learning models. A critical obstacle is the difficulty in generating labeled datasets that are essential for training these models. In real-world ICS

environments, simulating cyber-attacks or system failures to create these datasets poses a considerable risk of causing actual system failures, thereby compromising the integrity and safety of the systems involved.

A novel solution to this challenge is the use of Hardware-in-the-Loop (HIL) simulation. HIL simulation integrates real system components with simulated environments, allowing for the safe generation of labeled datasets that accurately reflect various operational scenarios, including attack patterns. This approach mitigates the risks associated with direct testing on operational systems, providing a robust platform for developing and refining anomaly detection models without compromising system integrity.

The evolving domain of Industrial Control Systems (ICS) security has garnered significant research interest, particularly in the development of robust anomaly detection mechanisms to mitigate the sophisticated cyber threats these systems face. The literature presents various approaches to this challenge, each contributing unique insights and methodologies pertinent to the field.

Zou et al. provided a practical perspective through a real case study in an industrial environment, highlighting the process of virus spread and worm propagation [11]. Their work emphasized the effectiveness of anomaly detection techniques in identifying malicious activities and aiding security administrators in enhancing ICS security. This case underscores the necessity of practical, real-world validations for theoretical models and approaches. Li et al. addressed the scarcity of attack data in power ICS by proposing a cross-domain anomaly detection method [3]. Utilizing the TrAdaBoost algorithm, they successfully transferred knowledge from related domains to the power ICS context, achieving lower error rates compared to using Long Short-Term Memory (LSTM) networks alone. Their approach is particularly relevant in scenarios where historical attack data is insufficient or non-existent. Wang et al. focused on the in-depth detection of abnormal behavior in power ICS, capturing and analyzing protocol-specific data packets to detect anomalies [12]. Their methodological framework for syntactic and semantic analysis, along with business command analysis, provides a comprehensive approach to identifying irregular behaviors and traditional network attacks such as malware and Trojan horses.

Zhao et al. introduced an anomaly detection model for ICS that combines the Gaining-sharing knowledge (GSK) algorithm with LSTM networks [13]. They utilized the GSK algorithm for feature selection, enhancing the accuracy and reducing the computational burden of the LSTM classifier. Moreover, they refined the GSK algorithm with the Taguchi method to optimize feature selection, further improving the model's efficiency and robustness as demonstrated on a real gas pipeline dataset. Zhang et al. proposed a control flow anomaly detection algorithm that operates by examining the business programs' control flow within ICS [14]. By creating a standard path set and matching current flows against this benchmark, their Control Flow Checking Path Matching (CFCPM) algorithm effectively detects deviations indicative of system anomalies, highlighting the algorithm's potential in recognizing concealed intrusion attacks.

The collective insights from these studies inform the current research, which seeks to enhance ICS security through an Isolation Forest-Based Anomaly Detection Model. The literature underscores the importance of addressing the unique challenges of ICS environments, such as high-dimensional datasets and the need for real-time detection capabilities. The proposed model builds on these foundations, aiming to deliver a solution that is not only accurate and efficient but also capable of adapting to the dynamic threat landscape of ICS. This model serves as a second layer of defense, complementing traditional security measures with a dynamic and adaptive approach to threat detection. This model leverages the isolation forest algorithm's efficiency in identifying data anomalies, offering a promising solution to detecting sophisticated cyber threats in ICS environments. The adoption of HIL simulation for generating labeled datasets enables the training of supervised machine learning models under realistic yet controlled conditions, ensuring the reliability and effectiveness of the anomaly detection model [15]. By incorporating data from various ICS components and levels, the proposed model achieves a comprehensive understanding of normal and anomalous system behaviors, enhancing its accuracy and sensitivity in threat detection.

The motivation behind this research is rooted in the growing vulnerability of ICS to cyber-attacks, including insider threats and stealthy, sophisticated attacks that conventional security measures fail to address. The critical nature of

ICS and their role in supporting essential services and infrastructure makes them attractive targets for cybercriminals, posing significant risks to national security, public safety, and economic stability. The development of advanced anomaly detection models, such as the proposed Isolation Forest-Based Anomaly Detection Model, is driven by the urgent need to enhance the resilience of ICS against these evolving cyber threats, ensuring the continuity and reliability of critical infrastructures. The remainder of this paper is organized as follows: Section 2 describes the model architecture, including the HAI dataset, the Isolation Forest algorithm, and the evaluation metrics. Section 3 discusses the implementation of ICS test bed and data forming. Section 4 presents the results of our experiments, performance analysis. Finally, Section 5 concludes the paper with a summary of our contributions and the broader significance of our work.

2. MODEL ARCHITECTURE

2.1 Dataset Analysis

In Fig. 1. presented illustrates a comprehensive method for securing Industrial Control Systems (ICS) against cyber threats through advanced data analysis and machine learning. It begins with data collection from SCADA systems [16] and Hardware-in-the-Loop (HIL) simulations, creating a rich dataset that includes both normal operational data and simulated anomaly events. This dataset undergoes data engineering to refine features and normalize the data scale, ensuring optimal input for model training.

Isolation algorithm then trained on this curated dataset. The trained models are tasked with classifying system behavior into normal or malicious activities, enhancing the ICS's ability to detect and respond to anomalies and potential cyber threats effectively. This structured approach leverages the strengths of both empirical data and simulated scenarios to bolster the ICS's defensive capabilities without compromising the system's operational integrity. This open access dataset is available in [17].

2.2 Feature Engineering

The critical task of feature selection for Machine Learning Intrusion Detection System within an Industrial Control System environment [18], we adopted a methodical approach to isolate the most significant predictors for our model. Our methodology was anchored in the utilization of a filter-based feature selection technique, leveraging the Pearson correlation coefficient as a metric to discern the linear relationship between potential features and the target variable. This step was instrumental in identifying features with a strong correlation to the target, thereby enhancing the predictive power of the model while concurrently streamlining computational efficiency to a vital consideration for real-time application. To obviate the issue of multicollinearity and the inclusion of redundant data, features exhibiting high inter-correlations were either amalgamated or the least correlated ones with respect to the target were excluded, contingent on their correlation coefficients. Further, we implemented the MinMaxScaler for data normalization, ensuring that the feature values were proportionately scaled within a bounded range, thus facilitating a consistent and expedient learning process. This meticulous selection and scaling of features poised our model to accurately discern between normal operations and potential security breaches, ensuring robustness and agility in our anomaly detection mechanism.

2.3 Isolation Forest Algorithm

The Isolation Forest, an ensemble method, distinguishes itself by isolating anomalies instead of profiling normal data points. Its primary advantage lies in the minimal requirement of preprocessing and its inherent speed, which is crucial for real-time anomaly detection in ICS [19].

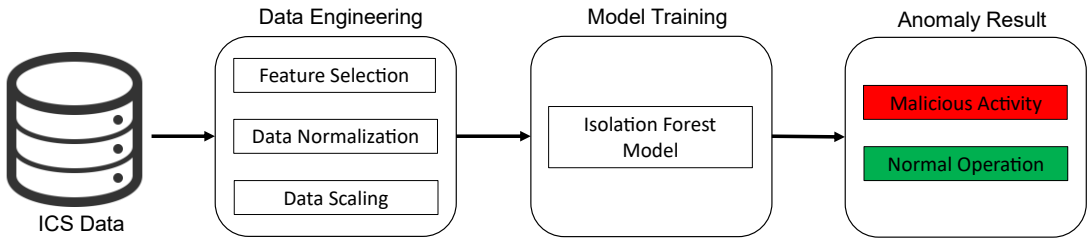

Fig. 1. Data analysis flow diagram

The algorithm operates on the principle of recursive partitioning. It constructs numerous random decision trees, termed 'isolation trees' or 'i-trees', to isolate observations. The core idea is that anomalies are few and different and thus easier to isolate from the rest of the sample. An isolation tree is grown by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. This partitioning process continues recursively until each observation is isolated, or until the tree reaches a predefined limit [8].

The path length from the root node to the terminating node serves as a measure of normality; shorter paths indicate anomalies. For a dataset D, with n samples, an isolation tree iT is built on a random subset of data of size ψ, and the process is repeated to create an ensemble of t trees. The anomaly score is computed as (1):

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (1)$$

where x is the instance to be scored, $E(h(x))$ is the average path length of x over the forest of isolation trees, and $c(n)$ is the average path length of unsuccessful search in a Binary Search Tree (BST) given by (2):

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \quad (2)$$

Here, $H(i)$ is the harmonic number and can be estimated by $ln(i) + 0.5772156649$ (Euler's constant). Anomalies are then determined based on a threshold set on the anomaly score.

In our model, the Isolation Forest algorithm was trained and tuned to optimize for both recalls, to minimize the number of missed detections (false negatives), and precision, to minimize the number of false alerts (false positives), which are particularly disruptive in an ICS context. The mathematical robustness of the algorithm combined with its computational efficiency makes it an excellent candidate for real-time anomaly detection in complex and data-intensive environments such as ICS.

2.4 Performance Evaluation Metrics

For the performance evaluation of the Isolation Forest algorithm within the ICS anomaly detection framework, a suite of metrics to provide a comprehensive assessment of the model's effectiveness. These metrics were chosen to capture various aspects of model performance, including its accuracy in predicting anomalies, the rate of false positives, the model's sensitivity, and its overall error rate [20].

2.4.1 Accuracy

This metric assesses the overall correctness of the model and is calculated as the ratio of correctly predicted instances (both normal and anomalous) to the total number of instances. The formula is given by (3):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

2.4.2 Precision

Often referred to as the positive predictive value, this metric evaluates the proportion of true positive predictions in all positive predictions. It is crucial for determining the reliability of the anomaly detection in ICS, where false positives can be costly. Precision is defined as (4):

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

2.4.3 Recall (sensitivity or true positive rate)

This measures the model's ability to correctly identify all the actual anomalies. High recall is necessary for ICS to ensure that no actual threat goes unnoticed. It is calculated by (5):

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

2.4.4 F1 score

The F1 Score is the harmonic mean of precision and recall, providing a balance between the two in cases where an even trade-off is desired. It is particularly useful when the class distribution is uneven. The F1 score is computed as (6):

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

3. RESULTS AND DISCUSSION

The Fig. 2 presents a Correlation Matrix, a quantitative tool that displays the correlation coefficients between variables in a dataset, indicating the degree to which they are linearly related. Each cell in the matrix shows the correlation coefficient between two variables, ranging from -1 to 1. A value of 1 implies perfect positive correlation, meaning as one variable increases, the other does likewise. A value of -1 indicates a perfect negative correlation, where an increase in one variable corresponds to a decrease in the other.

A value of 0 suggests no linear correlation. In this matrix, shades of blue represent the strength of correlation, with darker shades indicating stronger relationships. For example, variables P1_B2016 and P1_B4005 show a very high positive correlation (0.94), hinting that they may change together, while P1_B3004 and P1_B3005 exhibit almost no correlation (-0.07), suggesting no linear relationship in their changes.

The variable labeled 'Attack' appears to have little to no linear correlation with the other variables, as indicated by its predominantly light shading. This matrix is crucial for identifying relationships within data, which can inform feature selection for machine learning models, particularly in contexts such as ICS security where understanding variable relationships is key to detecting anomalies.

A The confusion matrix generated from the model evaluation present in Fig.3., a compelling narrative of the model's efficacy. A total of 85,515 normal instances were correctly classified (true positives), indicating a robust capability of the model to recognize the standard operation patterns of the ICS. Notably, there were no instances of normal behavior misclassified as anomalies (false positives), reinforcing the model's precision.

However, the model did not perform flawlessly in identifying all anomalous instances. The model misclassified 885 anomalies as normal behavior, suggesting areas for improvement in the model's sensitivity to subtle irregularities. The absence of true negatives in the confusion matrix indicates that the model did not correctly identify any of the anomalous instances. This result could point to a potential overfitting to the normal instances or a need for further refinement of the model's parameters to enhance its detection sensitivity.

Fig. 4, represents the anomaly score of the dataset. There are two sets of data points represented by different colors: green and red, which are labeled "Normal" and "Anomaly" respectively. The green line at the very bottom indicates a normal state that is constant over time. The red points are scattered above, presumably indicating moments where anomalies were detected over the given time period.

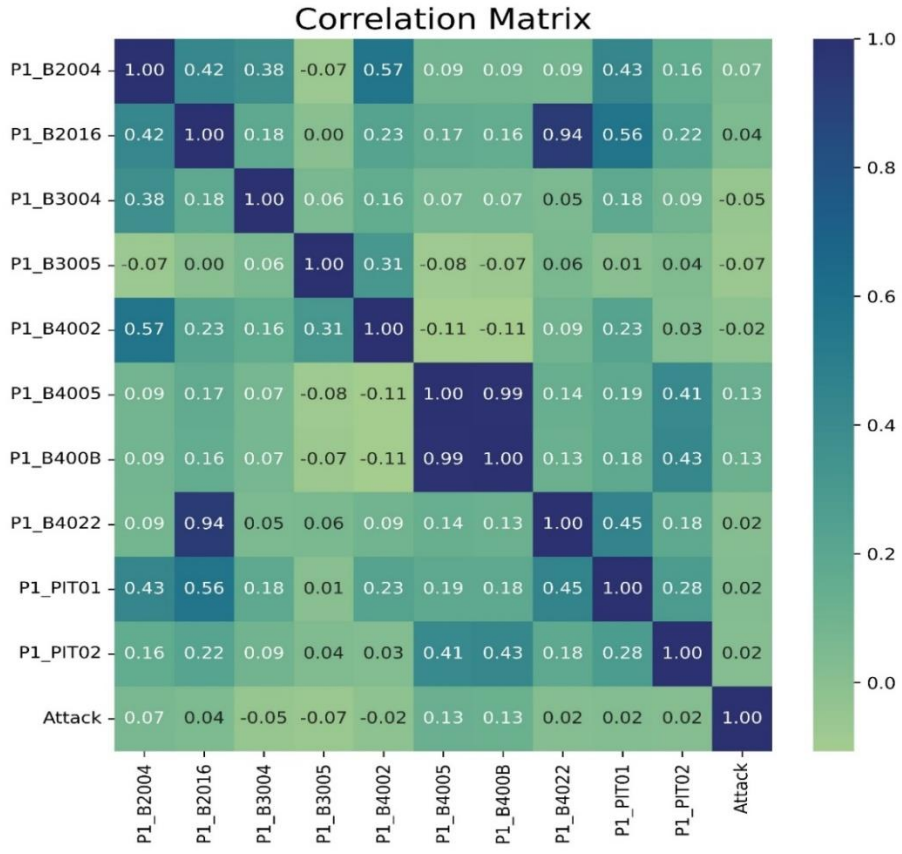

Fig. 2. Correlation matrix

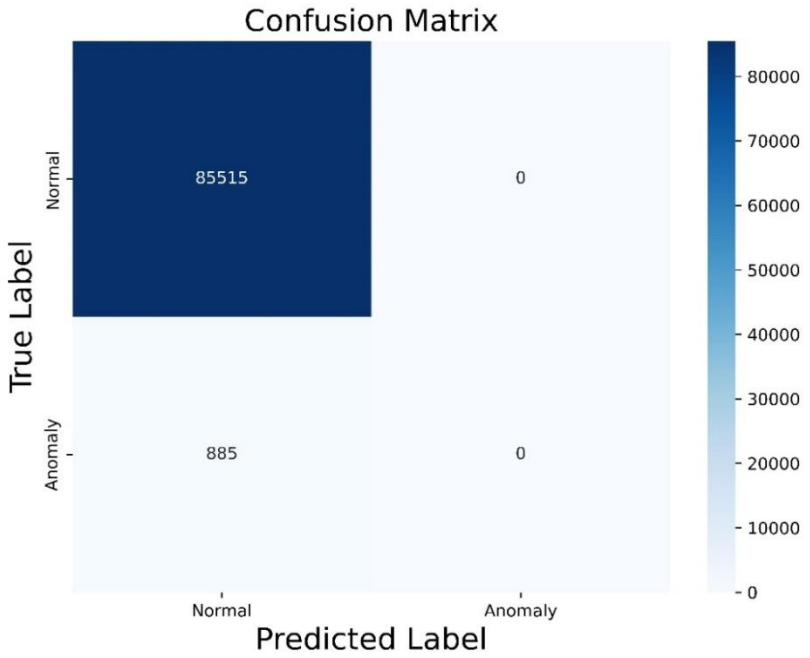

Fig. 3. Confusion matrix

These anomalies are all scored at a state of 1, which might indicate a binary state where 1 represents the presence of an anomaly. And major 7 point are representing the attack in the system. Meaning 7 malicious activities are formed.

In the Fig. 5, specifically examining the heat-exchanger outlet pressure, the Isolation Forest algorithm exhibited notable efficacy. The time-series data, represented graphically, illustrates the pressure readings over a continuous operational period marked against timestamps. The green line depicts the normal operational state of the pressure measurements, labeled as P1_PIT01. Superimposed upon this, in red, are the instances identified as anomalies by the model. The analysis detected a discernible pattern of sporadic spikes in pressure, which significantly deviated from the established norm. These deviations were systematically classified as anomalies, as indicated by the red markers. The frequency and magnitude of these outliers are critical, as they may signify potential malfunction or external interference within the system. It is observed that the pressure readings occasionally surged beyond the 1.5 bar threshold, a parameter we had previously determined as indicative of anomalous behavior based on the operational characteristics of the heat exchanger.

In Fig. 6, the Isolation Forest algorithm's performance is showcased through the analysis of the heat-exchanger outlet pressure over time. The graphical representation of the data features a blue line that tracks regular pressure levels, labeled P1_B2016, juxtaposed with red markers that the algorithm has identified as anomalies. These marked anomalies correspond to noticeable and intermittent pressure spikes that stray from the normal pattern. Such deviations are significant as they could signal possible system malfunctions or security breaches. Notably, instances where the pressure exceeded the pre-established threshold of 1.4 bar were automatically flagged by the model, aligning with our defined criteria for abnormal behavior linked to the system's thermal output operations.

The temporal distribution of the anomalies did not suggest a periodic or systematic occurrence, thereby eliminating the likelihood of these events being attributed to regular maintenance or predictable operational adjustments. Such irregularity in the distribution underscores the necessity for real-time monitoring and immediate response to maintain system integrity and safety. The result highlights the Isolation Forest algorithm's strength in real-time anomaly detection in ICS environments. By promptly identifying these pressure aberrations, the model serves as a critical component of a proactive ICS monitoring system, aiming to mitigate potential risks associated with pressure deviations in the heat-exchanger mechanism. This effective detection of anomalies underscores the potential

of employing such machine learning techniques for the enhancement of predictive maintenance and the prevention of unscheduled downtimes in industrial settings.

In evaluating the performance metrics of the Isolation Forest algorithm applied within our industrial control system context, the results affirm a high level of accuracy and precision. The model achieved an accuracy rate of 98.98%, illustrating its effectiveness in correctly classifying the vast majority of data points. Precision, a measure of the model's ability to return relevant instances, stood at an exceptional 99.98%, indicating that almost all instances predicted as anomalies were indeed true anomalies. This precision is critical in industrial settings, as false positives can lead to unnecessary and costly operational interruptions.

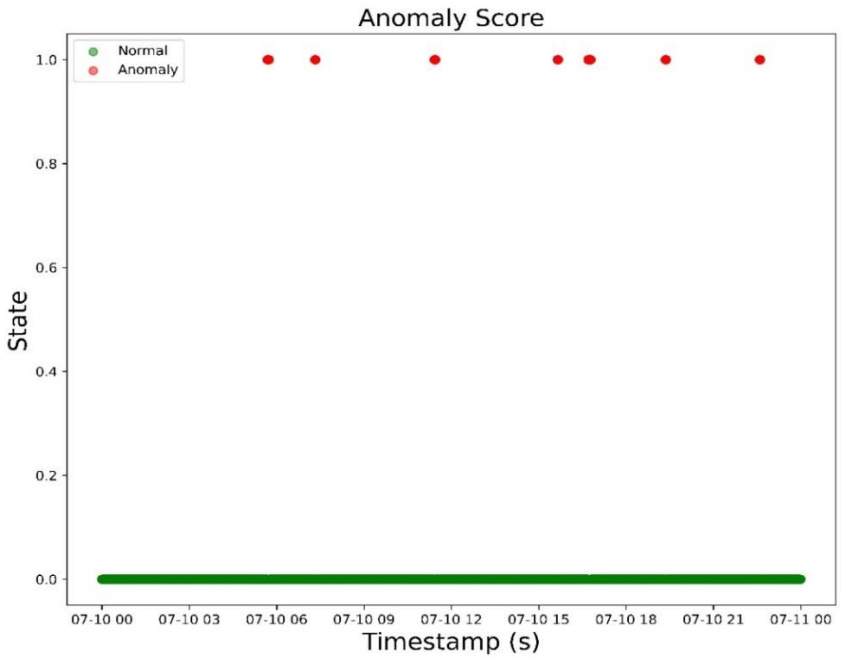

Fig. 4. Anomaly Score

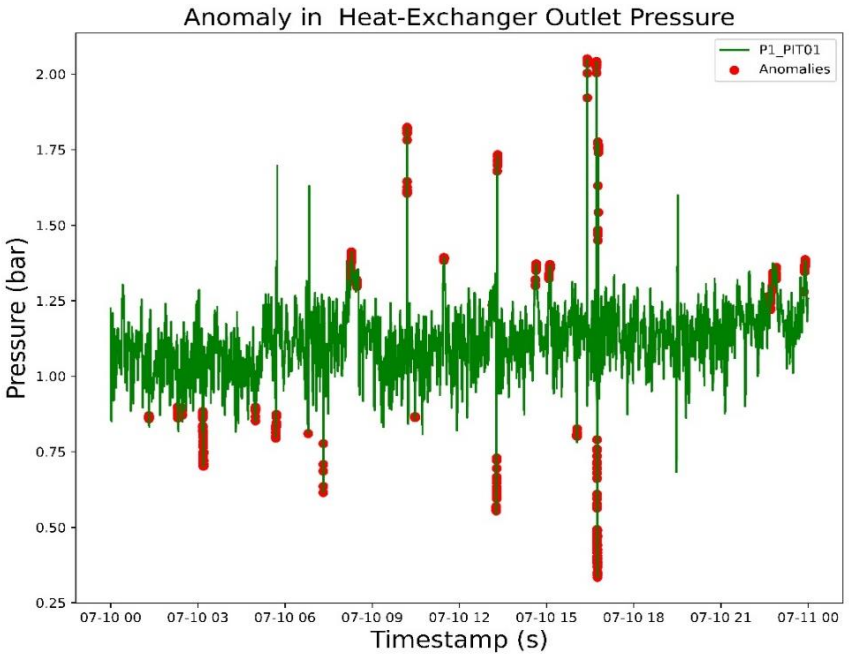

Fig. 5. Anomaly points in Heat-exchanger outlet pressure

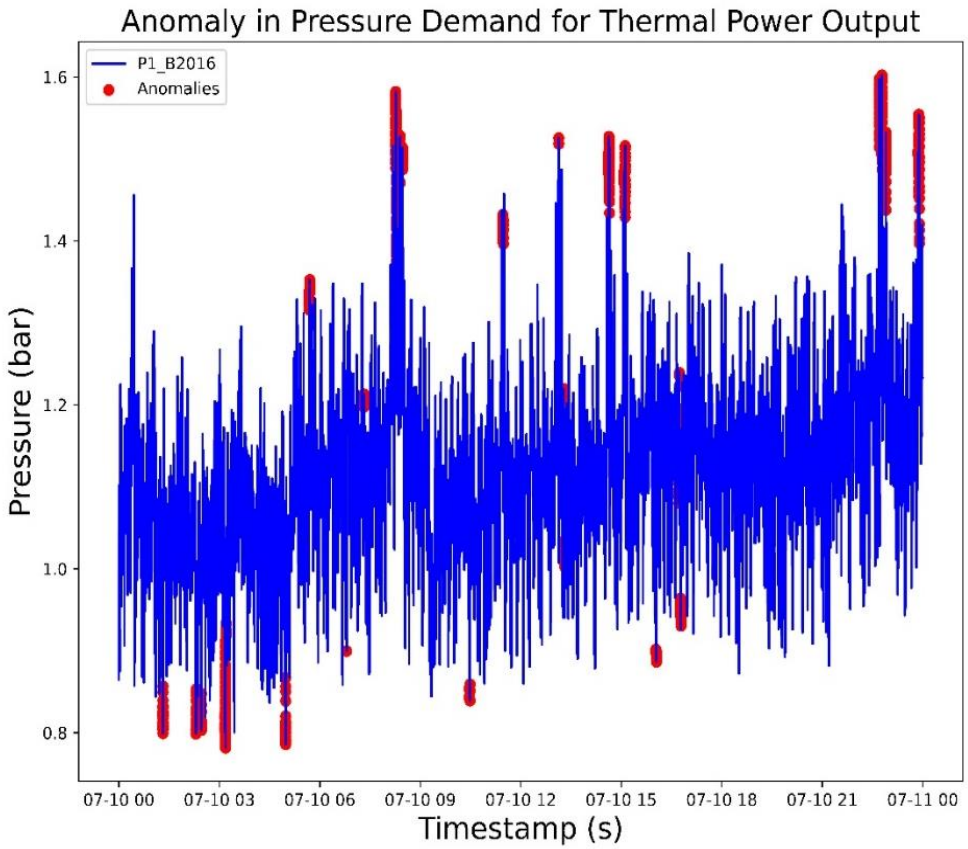

Fig. 6. Anomaly points in Pressure demand for thermal power output

Furthermore, the recall of the model, also known as sensitivity, reached 99.98%, demonstrating the model's ability to identify nearly all true anomalies. This suggests that the Isolation Forest algorithm is highly effective in capturing the anomalous events that could signify potential system risks or failures. The F1 Score, which is the harmonic mean of precision and recall, was also calculated to be 99.98%, confirming the model's balanced performance in both precision and recall.

These metrics collectively highlight the model's robustness in anomaly detection within an ICS environment. The high precision minimizes the risk of false alarms, while the high recall ensures that actual threats are not overlooked, contributing to the system's overall reliability and safety. With such performance, the Isolation Forest algorithm stands out as an exemplary method for real-time anomaly detection in complex industrial systems, offering a significant enhancement to the predictive maintenance protocols and aiding in the prevention of unplanned operational downtimes.

The Receiver Operating Characteristic (ROC) curve depicted in the Fig. 7. The curve traces the trade-off between the True Positive Rate (TPR, on the y-axis) and the False Positive Rate (FPR, on the x-axis) at various threshold settings. The TPR, also known as recall or sensitivity, measures the proportion of actual anomalies that the model correctly identifies. The FPR, inversely, gauges the proportion of normal instances that are incorrectly classified as anomalies. The curve demonstrates an outstanding Area Under the Curve (AUC) of 0.99, indicating that the model has a high probability of distinguishing between "normal" and "anomalous" states. An AUC close to 1.0 reflects excellent model performance, with a high rate of correctly identified anomalies and a low rate of false alarms, which is vital in maintaining operational integrity and minimizing unnecessary disruptions in ICS environments.

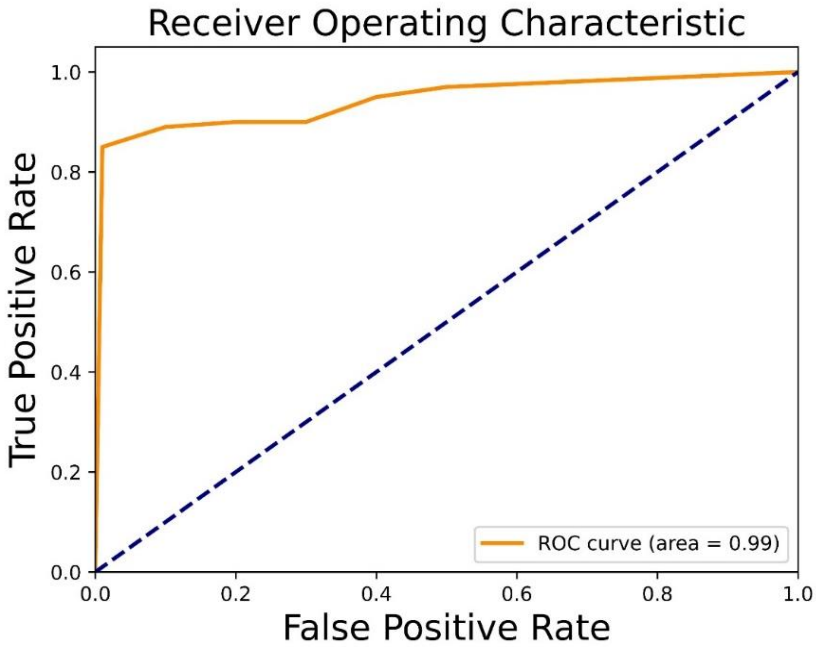

Fig. 7. Receiver Operating Characteristic (ROC) curve

4. CONCLUSION

In conclusion, this research has successfully demonstrated the viability of employing the Isolation Forest algorithm as an advanced anomaly detection model within the realm of Industrial Control System (ICS) security. Through meticulous adaptation and application within the ICS domain, our model has shown exceptional aptitude in the identification of anomalous behavior, thereby offering a robust enhancement to the security posture of critical infrastructure systems. By utilizing a comprehensive and diverse dataset, augmented by Hardware-In-the-Loop (HIL) simulation, the study has underscored the model's capability to not only detect established cyber threats but also adapt to emerging ones, ensuring its relevance and efficacy in the face of an ever-evolving cyber threat landscape. The model's performance is quantitatively underscored by its impressive metrics: an accuracy of 98.98%, precision and recall both at an outstanding 99.98%, and an F1 Score of 99.98%. Which stands as a testament to the potential of machine learning techniques in fortifying the resilience of ICS against sophisticated cyber threats. The findings of this study contribute a significant discourse in the ongoing efforts to safeguard our vital infrastructures, presenting the Isolation Forest-based anomaly detection as an indispensable tool in the arsenal against cyber adversaries.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. F Kargl, Van Der Heijden RW, König H, Valdes A, Dacier MC. Insights on the security and dependability of industrial control systems. IEEE Secur Priv. 2014;12(6):75–78.
   DOI:10.1109/MSP.2014.120.
2. Fan X, Fan K, Wang Y, Zhou R. Overview of cyber-security of industrial control system. International Conference on Cyber Security of Smart Cities, Industrial Control System and Communications, SSIC 2015 – Proceedings; 2015.
   DOI:10.1109/SSIC.2015.7245324.
3. Y Li et al. Cross-domain Anomaly Detection for Power Industrial Control System. ICEIEC 2020 - Proceedings of 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication. 2020;383–386.
   DOI:10.1109/ICEIEC49280.2020.9152334
4. Paridari K, O'Mahony N, El-Din Mady A, Chabukswar R, Boubekeur M, Sandberg H, A framework for attack-resilient industrial control systems: Attack detection

and controller reconfiguration. Proceedings of the IEEE. 2018;106(1):113–128.
DOI:10.1109/JPROC.2017.2725482.

5. Koay AMY, Ko RKL, Hettema H, Radke K. Machine learning in industrial control system (ICS) security: Current landscape, opportunities and challenges. J Intell Inf Syst. 2023;60(2):377–405.
DOI:10.1007/S10844-022-00753-1.
6. Xu J, Shi W, Zhang S, An Ensemble Learning Method with Feature Fusion for Industrial Control System Anomaly Detection. Proceedings of the 33rd Chinese Control and Decision Conference, CCDC. 2021;2563–2567.
DOI:10.1109/CCDC52312.2021.9602724.
7. Bae S, Hwang C, Lee T. Research on Improvement of Anomaly Detection Performance in Industrial Control Systems. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2021;13009 LNCS:76–87. DOI:10.1007/978-3-030-89432-0_7.
8. Kabir S, Shufian A, Zishan MSR, Isolation Forest Based Anomaly Detection and Fault Localization for Solar PV System. International Conference on Robotics, Electrical and Signal Processing Techniques. 2023;341–345.
DOI:10.1109/ICREST57604.2023.10070033
9. Kabir S, Md Oyon SS, Md Shahria N, Islam R, Md Hoque JAM, Shufian A. Integrating AE-CNN with Smart Relaying and SSCB for Enhanced Three-Phase Fault Detection and Mitigation. 2023 10th IEEE International Conference on Power Systems (ICPS). 2023;1–5.
DOI:10.1109/ICPS60393.2023.10428989
10. Peng Y et al. Cyber-Physical Attack-Oriented Industrial Control Systems (ICS) Modeling, Analysis and Experiment Environment. Proceedings - 2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2015. 2016;322–326.
DOI:10.1109/IIH-MSP.2015.110
11. Zou J, Jin X, Zhang L, Wang Y, Li B. A case study of anomaly detection in industrial environments. Proceedings - 22nd IEEE International Conference on Computational Science and Engineering and 17th IEEE International Conference on Embedded and Ubiquitous Computing, CSE/EUC. 2019;294–298.
DOI:10.1109/CSE/EUC.2019.00063
12. Wang B, Zhang J, Luo C, Yang L, Chen J, Ma H. Research on Deep Detection Technology of Abnormal Behavior of Power Industrial Control System. IEEE 6th Information Technology and Mechatronics Engineering Conference, ITOEC. 2022;256–1261.
DOI:10.1109/ITOEC53115.2022.9734439
13. Zhao H, Lei R, Fan F, Guo Y, Li Y. Abnormal Detection of Industrial Control System Based on LSTM and GSK Algorithm Customized by Taguchi Method. 2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence, CCAI. 2023;306–311.
DOI:10.1109/CCAI57533.2023.10201287
14. Zhang Z, Chang C, Lv Z, Han P, Wang Y. A control flow anomaly detection algorithm for industrial control systems, Proceedings - 2018 1st International Conference on Data Intelligence and Security, ICDIS. 2018;286–293.
DOI:10.1109/ICDIS.2018.00054
15. Zhao W, Peng Y, Xie F. Testbed techniques of industrial control system," Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, ICCSNT 2013. 2014;61–65.
DOI:10.1109/ICCSNT.2013.6967064
16. Oyon MSS, Shufian A, Kabir S, Islam MA, Mahin MSR, Mahmud MS. Three Phase Fault Analysis Using Thermal-Magnetic Circuit Breaker and Overcurrent Relay. IEEE Int. Conf. on Information and Communication Technology for Sustainable Development (ICICT4SD). 2023;269-273.
DOI:10.1109/ICICT4SD59951.2023.10303432.
17. icsdataset/hai: HIL-based Augmented ICS (HAI) Security Dataset. Accessed; 2024.
Available:https://github.com/icsdataset/hai
18. Mokhtari S, Yen KK. Measurement data intrusion detection in industrial control systems based on unsupervised learning. Applied Computing and Intelligence. 2021;1(1):61–74.
DOI:10.3934/ACI.2021004
19. Shufian A, Kabir S, Islam MA, Hoque MJAM, Adnan MA, Mohammad N. Grid-

tied Smart Microgrid with Heuristic Optimized Energy Management System (EMS), IEEE International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM). 2023;1-6.
DOI:10.1109/NCIM59001.2023.10212528

20. Xue F, Yan W. Multivariate Time Series Anomaly Detection with Few Positive Samples, Proceedings of the International Joint Conference on Neural Networks. 2022;2022.
DOI:10.1109/IJCNN55064.2022.9892091

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/114125