SJER Scientific Journal of Engineering Research

futuristech PT. Teknologi Futuristik Indonesia

Date of publication March 18, 2026, date of current version March 18, 2026. e-ISSN: 3109-1725

Digital Object Identifier 10.64539/sjer.v2i2.2026.434

Article

Isolation Forest–Based Intrusion Detection for Cyber-Physical Systems

Godfrey Perfectson Oise[1,*], Susan Konyeha[2], Felix Oshiorenoya Uloko[3], Kevin Chinedu Pius[1], Enovwo Eferoba–Idio[1], Michael Uyiosa Edobor[4], Evans Mintah[5], Osahon Ukpebor[6], Oludare Sokoya[7], Tejiri Jessa[5]

[1] Department of Computing, Wellspring University, Edo State, Nigeria; e-mail: godfrey.oise@wellspringuniversity.edu.ng (G. P. Oise), kevin.pius@wellspringuiversity.edu.ng (K. C. Pius), eferoba-idio.enovwo@wellspringuniversity.edu.ng (E. Eferoba-Idio).

[2] Department of Data Science, University of Benin, Edo State, Nigeria; e-mail: susan.konyeha@uniben.edu (S. Konyeha).

[3] Department Computer Science, Veritas University, Abuja, Bwari 901101, Federal Capital Territory, Nigeria; e-mail: ulokof@veritas.edu.ng (F. O. Uloko).

[4] Department of Computer Science, University of Benin, Edo State, Nigeria; e-mail: u.edobor@yahoo.com (M. U. Edobor).

[5] College of Business, Westcliff University, Irvine, CA 92614, United States; e-mail: mintah1@gmail.com (E. Mintah), tejirijessa@gmail.com (T. Jessa).

[6] Department of Computer and Information Science, University of the Cumberlands, Williamsburg, KY 40769, United States; e-mail: iukpebor@gmail.com (O. Ukpebor).

[7] College of Business, Engineering & Technology, National University, San Diego, CA 92123, United States; e-mail: daresokoya@gmail.com (O. Sokoya).

[*] Correspondence Author

The authors received no financial support for the research, authorship, and/or publication of this article.

Abstract: Cyber-physical engineering systems (CPES) form the backbone of critical infrastructures such as power generation, industrial automation, and water treatment facilities. Because cyber intrusions in these environments can directly disrupt physical processes, reliable intrusion detection mechanisms are essential for maintaining operational safety and system resilience. However, many existing intrusion detection approaches rely on supervised learning techniques that require large volumes of labeled attack data, which are rarely available in real industrial environments. In addition, advanced detection methods often introduce significant computational overhead, limiting their practicality for deployment in resource-constrained cyber-physical systems. To address these challenges, this study proposes a one-class anomaly detection framework based on the Isolation Forest algorithm for monitoring cyber-physical engineering systems. The proposed approach learns the statistical distribution of normal operational behavior using multivariate sensor, actuator, and control signals, and identifies deviations from this learned pattern as potential cyber intrusions. The framework is evaluated using the Hardware-in-the-Loop–based Augmented Industrial Control System (HAI) Security Dataset, which provides realistic industrial process measurements under both normal and attack scenarios. Experimental results show that the model achieves overall accuracy (0.89) and strong performance in identifying normal operational states (F1-score = 0.94). However, attack detection shows moderate recall (0.48) but low precision (0.04) due to class imbalance and overlapping anomaly score distributions. These findings indicate that Isolation Forest serves as a computationally efficient baseline anomaly detection mechanism for real-time CPS monitoring, while highlighting the need for hybrid and temporally aware detection strategies to improve attack discrimination in industrial cyber-physical environments.

Keywords: Cyber-Physical Systems; Intrusion Detection; Isolation Forest; Anomaly Detection; Industrial Control Systems.

1. Introduction

Cyber-physical engineering systems represent a transformative paradigm in modern engineering, where physical processes are intricately integrated with computational intelligence and networked communication infrastructures. These systems form the backbone of critical applications such as industrial automation, intelligent energy generation and distribution, advanced manufacturing, and transportation networks [1]. By enabling continuous sensing, autonomous control, and real-time decision-making, cyber-physical systems (CPS) enhance operational efficiency, reliability, and safety. The fusion of physical and cyber components, however, significantly broadens the attack surface, exposing these systems to sophisticated cyber threats that can produce tangible physical consequences [2]. Unlike conventional information technology systems, cyberattacks on CPS can directly manipulate sensors, actuators, and control logic, disrupting system dynamics, compromising operational stability, and potentially causing irreversible physical damage. The dual cyber-physical nature of these systems therefore requires security solutions that go beyond traditional IT-centric approaches, addressing not only data integrity but also the real-time operational consequences of attacks. [3] Cybersecurity is crucial for protecting hosts, networks, and cloud infrastructure.

Developing effective intrusion detection systems (IDS) is challenging due to the complexity and volume of data, making it hard to identify abnormal behavior. This study evaluates anomaly-based IDS using machine learning and proposes an effective model to improve anomaly detection and secure cloud operations. [4] Modern cybersecurity depends on intrusion detection systems (IDS), but traditional methods face challenges like high false positives, scalability issues, and difficulty detecting zero-day attacks. Machine learning has improved IDS performance, though problems such as imbalanced datasets and processing overhead remain [5]. This study proposes an ML-based IDS using Extra Trees Classifier, Gaussian Naive Bayes, and AdaBoost, demonstrating improved detection accuracy, real-time performance, and versatility against diverse cyber threats compared to existing approaches [6]. Historically, intrusion detection in CPS relied on signature-based and rule-based methods, which, while effective against previously observed attack patterns, lack adaptability and are unable to detect novel or evolving attacks [7]. These approaches also struggle in highly dynamic systems characterized by non-linear interactions among multiple subsystems and continuously varying operational conditions.

Anomaly-based detection methods emerged as a response, leveraging statistical models to characterize normal system behavior and flag deviations [8]. Although promising, these methods often exhibit high false-positive rates, limited scalability to high-dimensional multivariate data, and insufficient modeling of temporal dependencies that are inherent to cyber-physical processes. [9] Traditional IDSs are important for digital security, but often suffer from high false positives and slow responses to new threats. This study proposes an IDS combining Particle Swarm Optimization (PSO) with machine learning and neural networks to improve detection accuracy and efficiency. Using NSL-KDD and CICIDS datasets, feature selection and classifier training with PSO reduced dimensionality without harming performance. The Random Forest with PSO achieved 99.99% accuracy and F1-score, demonstrating the effectiveness of evolutionary optimization for real-time intrusion detection in dynamic cyber environments. [10] Proposes a GAN-LSTM hybrid model for anomaly detection in Cyber-Physical Systems, addressing class imbalance and temporal dependencies. Tested on SWaT and WADI datasets, it outperforms state-of-the-art methods with high accuracy, precision, recall, and F1-scores, effectively detecting rare and complex attacks using LSTM temporal learning and GAN-based oversampling. [11] Proposes an Optimized Isolation Forest-based IDS (OIFIDS) for the Industrial IoT, capable of efficiently handling heterogeneous and streaming data. Using a modified Harris Hawks Optimization, it reduces dimensionality and learning time while improving detection. Evaluated on multiple datasets, OIFIDS effectively manages irrelevant features, concept drift, and scenarios with no anomalies, outperforming existing IDS approaches.

Machine learning (ML) approaches have advanced intrusion detection research by enabling data-driven modeling of complex CPS behavior [12]. Supervised algorithms can learn discriminative patterns from historical operational data, but they require extensive labeled attack datasets, which may be scarce or incomplete in realistic scenarios [13]. Unsupervised and semi-supervised techniques, including clustering, principal component analysis, one-class classifiers, and tree-based anomaly detection methods, address these limitations by learning representations of normal behavior and identifying deviations indicative of cyberattacks [5]. These approaches are particularly suitable for CPS environments, where unknown or rare attacks are common, and labeled datasets are often limited. Recent research has also explored deep learning architectures, such as autoencoders and recurrent networks, to capture the temporal and spatial correlations in multivariate sensor and actuator streams [14]. While effective, these methods introduce substantial computational overhead, require large datasets for training, and often lack interpretability, which is critical for safety-critical industrial systems. A critical limitation of prior work is its narrow focus on detection accuracy, without sufficient consideration of computational efficiency, real-time feasibility, scalability, and system-level operational impact. Many existing studies treat intrusion detection as an isolated cyber problem, neglecting how attacks propagate through the physical components and affect system dynamics [15]. This fragmentation constrains the practical applicability of detection frameworks in real-world engineering systems, where both cyber and physical consequences must be considered [16], [17].

To address these gaps, there is a pressing need for semi-supervised, anomaly-based frameworks that balance detection performance with operational constraints [10], [18], effectively capture high-dimensional multivariate and temporal CPS data, and provide interpretable results for safety-critical environments. The present study proposes a data-driven, semi-supervised intrusion detection framework using the Isolation Forest algorithm. This approach leverages the HIL-based Augmented ICS (HAI) Security Dataset, which provides a rich set of time-continuous measurements from multiple industrial control processes, including boiler, turbine, and water-treatment systems [19]. The dataset captures both normal and attack scenarios, with multiple processes synchronized via a Hardware-in-the-Loop simulator, yielding highly coupled, correlated, and realistic data streams [20]. The Isolation Forest algorithm is particularly suited for this context, as it efficiently identifies anomalies in high-dimensional data without requiring labeled attack samples, making it ideal for industrial CPS applications where attack types are diverse and evolving [21], [22]. By learning the distribution of normal operational behavior, the model can detect deviations indicative of cyberattacks, while maintaining computational efficiency suitable for real-time deployment. In response to these challenges, this study examines a semi-supervised, anomaly-based intrusion detection framework for cyber-physical engineering systems using the Isolation Forest algorithm [23]. The proposed approach models normal operational behavior from high-dimensional sensor and actuator data, targeting realistic industrial scenarios characterized by limited labeled attack data and strict real-time requirements. The framework is evaluated on the Hardware-in-the-Loop–based Augmented Industrial Control System (HAI) Security Dataset, which captures interconnected industrial processes under normal and attack conditions. The contributions of this work include the development of a computationally efficient intrusion detection framework, an objective evaluation that highlights the effects of class imbalance and false alarms, and practical insights into the trade-offs between attack detection performance and operational reliability in safety-critical systems.

This study makes three primary contributions to intrusion detection research in cyber-physical engineering systems. First, it proposes a one-class anomaly detection framework based on the Isolation Forest algorithm capable of modeling high-dimensional multivariate sensor and actuator data in complex CPS environments. Second, the framework is evaluated not only in terms of detection performance but also with respect to computational efficiency and suitability for real-time monitoring, addressing practical deployment constraints in industrial systems. Third, the study provides insights into the practical applicability and limitations of machine learning–based anomaly detection for CPS, highlighting its role as a lightweight baseline monitoring mechanism.

The remainder of this paper is organized as follows. Section 2 presents the methodology, including dataset description, preprocessing, the Isolation Forest model, and evaluation metrics. Section 3 reports the experimental results. Section 4 discusses the implications for cyber-physical systems. Section 5 concludes the study and outlines future research directions.

2. Methodology

This study adopts a data-driven, semi-supervised machine learning framework to detect cyber intrusions in cyber-physical engineering systems (CPES). The methodology integrates dataset acquisition, preprocessing, feature standardization, anomaly detection modeling, and performance evaluation. The objective is to develop an intrusion detection mechanism capable of identifying abnormal system behavior while maintaining computational efficiency suitable for real-time industrial environments. The research workflow of the proposed framework is illustrated in Figure 1, which summarizes the sequential stages from dataset acquisition to performance evaluation.

Figure 1 illustrates the overall workflow of the proposed intrusion detection framework. The process begins with the acquisition of the HAI Security Dataset, followed by data preprocessing and feature standardization to prepare the data for machine learning analysis. The Isolation Forest model is then trained using normal operational data to learn the baseline behavior of the cyber-physical system. After training, anomaly scores are computed for unseen observations, and samples exceeding the anomaly threshold are classified as potential cyberattacks. Finally, the model's performance is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score.

2.1. Research Stages

The research methodology follows a structured pipeline consisting of several stages designed to ensure reliable anomaly detection in cyber-physical systems. The process begins with dataset acquisition, where the Hardware-in-the-Loop–based Augmented Industrial Control System (HAI) Security Dataset is obtained. This dataset provides realistic operational measurements from interconnected industrial processes under both normal and attack conditions. Next, data preprocessing is conducted to prepare the dataset for machine learning analysis. This stage involves handling missing values, cleaning inconsistent records, and preserving the temporal structure of the data. Following preprocessing, feature standardization is applied to

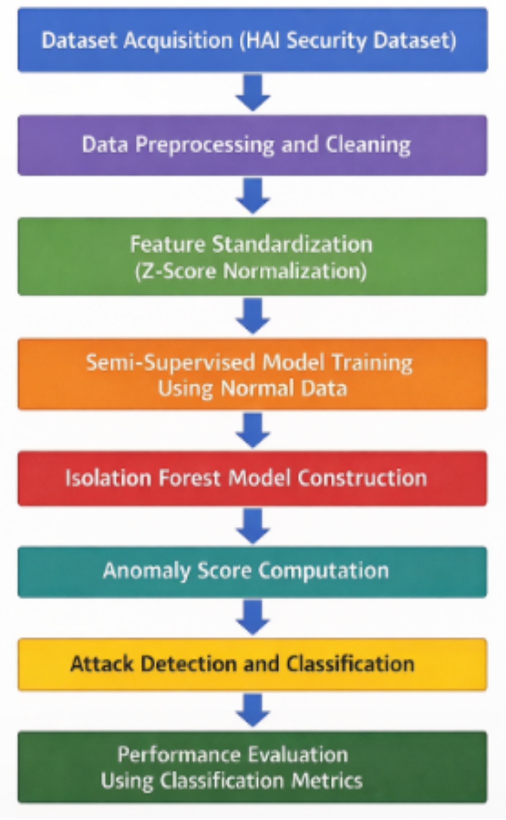
Dataset Acquisition (HAI Security Dataset)
Data Preprocessing and Cleaning
Feature Standardization
(Z-Score Normalization)
Semi-Supervised Model Training
Using Normal Data
Isolation Forest Model Construction
Anomaly Score Computation
Attack Detection and Classification
Performance Evaluation
Using Classification Metrics

Figure 1. Research stages of the proposed intrusion detection framework.

normalize continuous sensor measurements using z-score normalization. This ensures that variables with different measurement scales contribute equally during anomaly detection.

Subsequently, model training is performed using only normal operational data to learn the baseline behavior of the cyber-physical system. This approach reflects realistic industrial environments where labeled attack samples are limited. After training, the Isolation Forest model is used to compute anomaly scores for unseen observations. Samples with high anomaly scores are identified as potential cyber intrusions. Finally, the system performs attack classification and performance evaluation, where the predicted outcomes are compared with actual labels using standard evaluation metrics such as accuracy, precision, recall, and F1-score.

2.2. Dataset Description

The experimental evaluation in this study is conducted using the HAI Security Dataset [24], which serves as a benchmark dataset for cybersecurity research in industrial control systems. The dataset was collected from a Hardware-in-the-Loop (HIL) augmented industrial control system testbed, designed to emulate real industrial processes involved in power generation and water-treatment operations. The experimental environment integrates multiple subsystems, including boilers, turbines, and water-treatment units, which interact through industrial communication protocols and control architectures.

Each subsystem is operated using industrial-grade control hardware, including Emerson Ovation Distributed Control Systems (DCS), GE Mark VIe controllers, and Siemens S7 programmable logic controllers (PLCs). The dataset records time-continuous measurements from sensors, actuators, and control signals, capturing the operational dynamics of the system under both normal and attack scenarios. Multiple versions of the dataset include diverse cyberattack scenarios such as manipulation of sensor signals, actuator control interference, and command injection attacks. These characteristics make the dataset suitable for evaluating machine learning models designed for anomaly detection in cyber-physical systems.

2.3. Data Preprocessing and Feature Standardization

Data preprocessing is performed to ensure compatibility with machine learning models and to enhance the reliability of anomaly detection. First, missing or inconsistent values in the dataset are handled using forward-fill and interpolation techniques, which preserve the continuity of time-series observations while minimizing bias introduced by missing data. Next, continuous sensor measurements are normalized using z-score standardization, defined as:

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

where x represents the original feature value, μ denotes the mean of the feature, and σ represents the standard deviation. This transformation ensures that all features share a common scale, preventing variables with large magnitudes from dominating the anomaly detection process. Discrete actuator states are retained in their original form to preserve their operational meaning. For the semi-supervised learning setup adopted in this study, only normal operational data is used during model training, while the testing dataset contains both normal and attack samples. The labels are consolidated into a binary indicator representing the presence or absence of cyberattacks across the system.

2.4. Isolation Forest–Based Intrusion Detection Model

The intrusion detection problem in this study is formulated as a one-class anomaly detection task, where the model learns the distribution of normal system behavior and identifies deviations that may correspond to cyber intrusions. This formulation reflects realistic industrial environments in which labeled attack data are scarce or incomplete. Consequently, the model is trained exclusively on normal operational data, while both normal and attack observations are used during testing to evaluate detection performance. To implement this approach, the Isolation Forest algorithm is employed due to its efficiency in handling high-dimensional data and its effectiveness for anomaly detection in large-scale datasets. Unlike distancebased or density-based anomaly detection techniques, Isolation Forest identifies anomalies by isolating observations through random partitioning of the feature space using tree structures.

Given a dataset $X = \{x_1, x_2, \dots, x_n\}$, the algorithm constructs an ensemble of isolation trees. Each tree is built by recursively partitioning the data through random selection of a feature and a split value within the feature's range. Observations that differ significantly from the majority of the data tend to be isolated earlier in the partitioning process, resulting in shorter path lengths within the tree structure. For a given observation xxx, the anomaly score is computed based on the average path length across all trees in the ensemble. Data points with shorter average path lengths are assigned higher anomaly scores, indicating a greater likelihood of being anomalous. A threshold is then applied to the anomaly score to classify observations as either normal behavior or potential cyber intrusions. This approach provides a computationally efficient and scalable baseline mechanism for detecting anomalous behavior in cyber-physical engineering systems. The overall procedure of the proposed Isolation Forest–based intrusion detection is summarized in Algorithm 1.

2.5. Mathematical Formulation of Isolation Forest

Let the cyber-physical system dataset be represented as a multivariate feature space:

$$X = \{x_1, x_2, \dots, x_n\}, \quad x_i \in \mathbb{R}^d \tag{2}$$

where n denotes the number of observations and d represents the number of features describing the system state (e.g., sensor readings, actuator states, and control variables).

The Isolation Forest algorithm constructs an ensemble of t isolation trees. Each tree is generated by recursively partitioning a randomly selected subset of the dataset. At each node of the tree:

A feature f is randomly selected from the feature set.

A split value p is randomly chosen within the range of that feature.

This process continues until either the observation is isolated or the maximum tree depth is reached.

Let $h(x)$ denote the path length of an observation x, defined as the number of edges traversed from the root node to the terminating node in a tree. The average path length across all trees in the forest is given by:

$$E[h(x)] \tag{3}$$

The anomaly score for an observation x is computed as:

$$s(x,n) = 2^{-\frac{E[h(x)]}{c(n)}} \tag{4}$$

where $c(n)$ represents the average path length of unsuccessful searches in a Binary Search Tree and is defined as:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \tag{5}$$

and $H(i)$ denotes the harmonic number approximated as:

$$H(i) \approx \ln(i) + 0.5772156649 \tag{6}$$

The anomaly score $s(x,n)$ lies in the range:

$$0 < s(x,n) < 1 \tag{7}$$

Observations with scores close to 1 are more likely to be anomalies, while values closer to 0 indicate normal behavior.

A classification threshold τ is then applied to determine whether a data point is anomalous:

$$y = \begin{cases} 1, & s(x,n) \geq \tau \quad \text{(anomaly / attack)} \\ 0, & s(x,n) < \tau \quad \text{(normal)} \end{cases} \tag{8}$$

This formulation allows the Isolation Forest to efficiently detect anomalous observations in high-dimensional cyber-physical datasets while maintaining low computational complexity.

2.6. Model Training and Evaluation Strategy

Model training is conducted using only normal operational data to establish a baseline representation of system behavior. This approach reflects realistic industrial environments where labeled attack data is scarce or unavailable. During testing, the trained model analyzes datasets containing both normal and attack samples to evaluate its capability to detect anomalous events. Detection results are assessed using both classification metrics and time-series-aware evaluation techniques. In addition to predictive performance, computational efficiency is also evaluated by measuring model training time and inference latency. These measurements are important for determining whether the proposed intrusion detection framework can be deployed in real-time cyber-physical monitoring systems.

2.7. Evaluation Metrics

The performance of the proposed intrusion detection framework is assessed using standard classification metrics derived from the confusion matrix. Let: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives. The performance of the intrusion detection model is evaluated using standard classification metrics.

Algorithm 1. Isolation Forest–Based Intrusion Detection.

```
Input: $D_{train}$ – Normal training data
$D_{test}$ – Testing dataset
T – Number of trees
$\psi$ – Subsample size
$\theta$ – Anomaly threshold
Output: $y_{pred}$ – Predicted labels
$s(x)$ – Anomaly scores
Step 1: Data Preprocessing
    Handle missing values using forward-fill or interpolation
    Standardize continuous features using z-score normalization
Step 2: Model Training
    For i = 1 to T
        Randomly sample $\psi$ instances from $D_{train}$
        Build isolation tree by recursively splitting features
Step 3: Compute Anomaly Score
    For each sample x in $D_{train}$
        Compute path length h(x)
        Compute anomaly score s(x)
Step 4: Classification
    If s(x) ≥ θ
        Label as Attack
    Else
        Label as Normal
```

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Accuracy measures the proportion of correctly classified observations.

Precision:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Precision evaluates how many predicted attack instances correspond to actual attacks.

Recall:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

Recall measures the ability of the model to correctly identify actual attack events.

F1-Score:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

The F1-score provides a balanced measure of detection performance by combining precision and recall.

2.8. Implementation Environment

The proposed intrusion detection framework is implemented using the Python programming language. Data preprocessing and analysis are conducted using the pandas and NumPy libraries, while the Isolation Forest algorithm is implemented using scikit-learn. The HAI Security Dataset is accessed via KaggleHub, and experiments are conducted in a reproducible computational environment. The modular design of the framework allows it to be extended to other cyber-physical datasets and anomaly detection algorithms.

3. Results

The experimental results position the Isolation Forest model as a computationally efficient baseline anomaly detection framework for cyber-physical engineering systems rather than a high-precision attack classifier. The strong performance observed for the normal class (F1 = 0.94) indicates that the model effectively captures the dominant operational behavior of the system. However, the low precision for the attack class reflects the difficulty of distinguishing malicious behavior from legitimate operational variability in complex industrial environments. Many cyber-physical attacks are designed to remain within physically plausible operating ranges, which reduces the statistical separability between normal and malicious states. Consequently, point-wise anomaly detectors such as Isolation Forest may identify deviations but still produce high false-positive rates. These results therefore highlight the need for hybrid detection frameworks that integrate anomaly detection with temporal modeling, imbalance-aware learning strategies, or system-level contextual analysis.

Table 1 depicts that the model achieves high overall accuracy (0.89) due to excellent performance on the dominant Normal class (F1 = 0.94). However, detection of the minority Attack class is poor, with very low precision (0.04) and F1-score (0.08), despite a moderate recall (0.48). The gap between macro and weighted averages reflects severe class imbalance, indicating the model is biased toward normal traffic and is unreliable for attack detection without imbalance-aware improvements.

Figure 2 shows that the model correctly classifies most *Normal* instances (76,119 true negatives), but misclassifies a substantial number of normal samples as attacks (9,396 false positives). For the *Attack* class, 427 attacks are correctly detected (true positives), while 458 attacks are missed and labeled as normal (false negatives). These results confirm a strong bias toward the majority normal class and explain the low attack precision observed in Table 1, where many predicted attacks correspond to normal traffic despite moderate attack recall.

Figure 3 shows moderate classification capability, with an AUC of 0.71, indicating performance better than random guessing but limited class separability. The trade-

Table 1. Classification Report.

Class	Precision	Recall	F1-Score	Support
Normal	0.99	0.89	0.94	85515
Attack	0.04	0.48	0.08	885
Accuracy			0.89	86400
Macro Avg	0.52	0.69	0.51	86400
Weighted Avg	0.98	0.89	0.93	86400

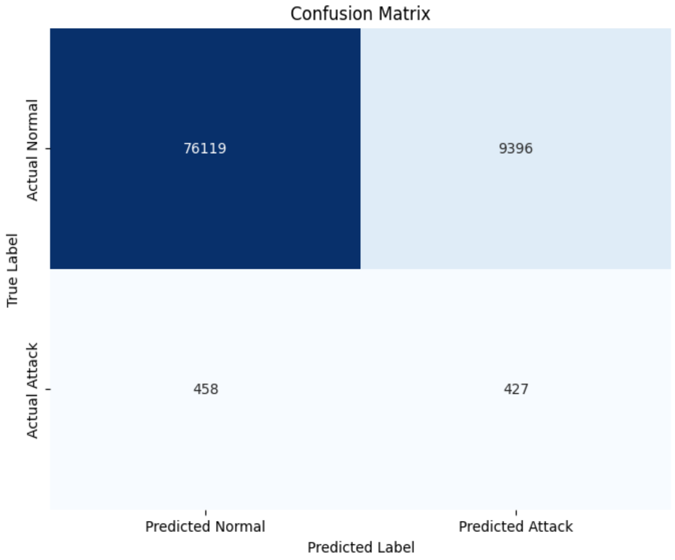

Figure 2. Confusion matrix summary.

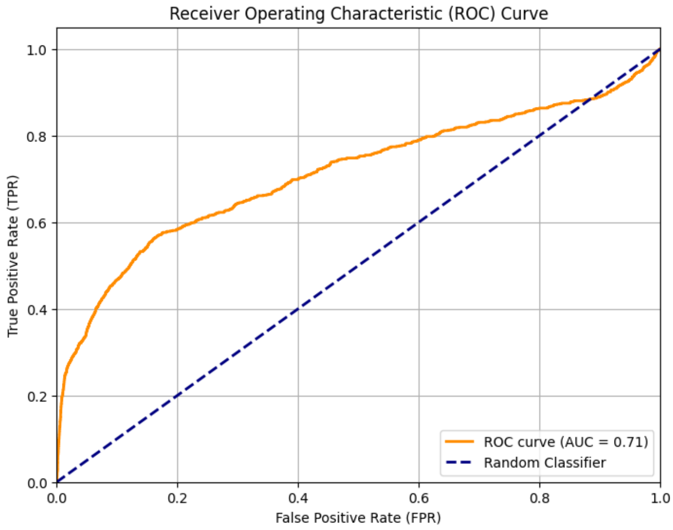

Figure 3. ROC curve.

off between true positives and false positives remains pronounced, suggesting that improving attack detection increases false alarms and that further optimization is needed for reliable intrusion detection.

Figure 4 shows substantial overlap, with both classes concentrated around similar score ranges. Although attack samples tend to have slightly lower anomaly scores (indicating higher anomaly), the lack of clear separation limits discriminative power. This overlap explains the high false-positive rate and low attack precision observed in earlier results, suggesting that the anomaly scoring function does not sufficiently distinguish attacks from normal behavior and requires refinement or additional features for improved detection.

Figure 5 illustrates the temporal variation of anomaly scores across the observed time period, with red markers indicating the timestamps corresponding to known attack events. Several anomaly score peaks occur near attack intervals, suggesting that the model can detect certain deviations associated with malicious activity. However, elevated anomaly scores are also observed during periods of normal system operation, indicating the presence of false positives. This behavior is consistent with the overlapping anomaly score distributions shown in Figure 4 and the low precision reported in Table 1. The figure therefore highlights that while the model is capable of identifying deviations in system behavior, the absence of clear separation between normal and attack patterns limits its discriminative reliability. These findings suggest that incorporating temporal aggregation or sequence-based validation mechanisms may improve detection stability in cyber-physical monitoring environments.

4. Discussion

This section provides a deeper interpretation of the experimental results by situating the performance of the proposed Isolation Forest–based intrusion detection framework within the broader landscape of anomaly detection methods commonly applied to cyber-physical engineering systems (CPES), including autoencoder-based models, one-class support vector machines (OC-SVM), and deep temporal architectures. The achieved overall accuracy of 0.89, driven by strong performance on the Normal class (F1 = 0.94), demonstrates that Isolation Forest effectively captures the dominant operational distribution of the CPES. Unlike reconstruction-based autoencoders, which learn normal behavior by minimizing reconstruction error, Isolation Forest relies on random feature partitioning and path-length statistics to isolate rare observations. This distinction is critical in CPES contexts: reconstruction-based models often struggle when normal data exhibit high variability or multimodal behavior, leading to inflated reconstruction errors and unstable thresholds. The strong normal-class performance observed here suggests that Isolation Forest is comparatively robust to such variability, making it suitable for industrial environments with frequent operational transients.

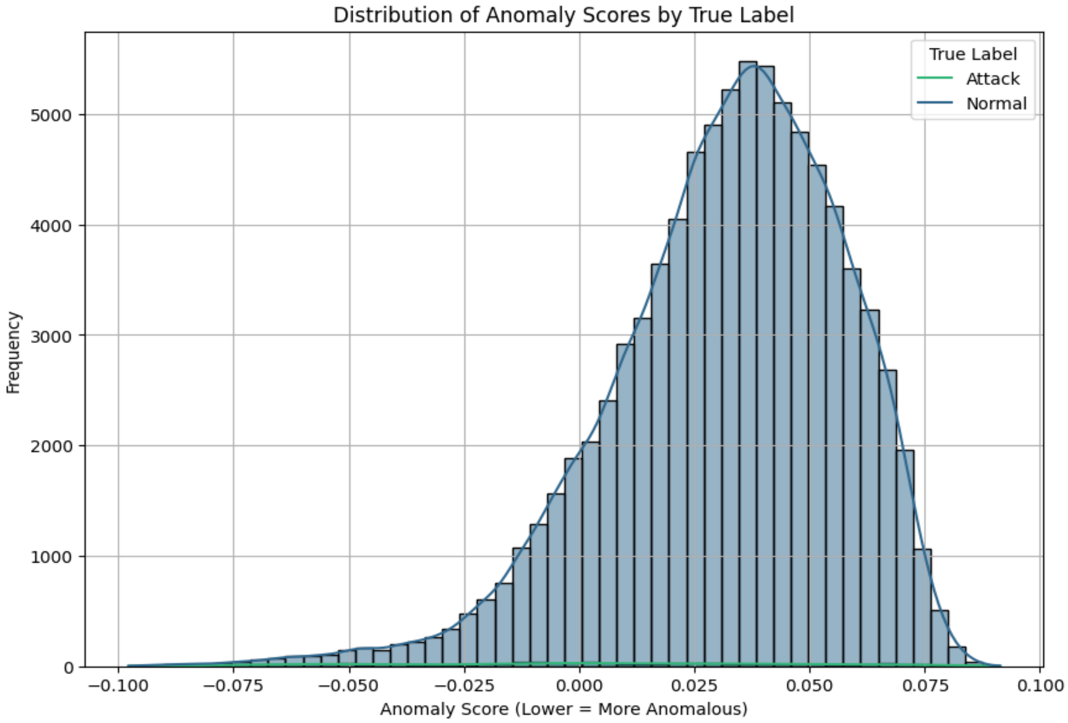

Figure 4. Anomaly score distribution.

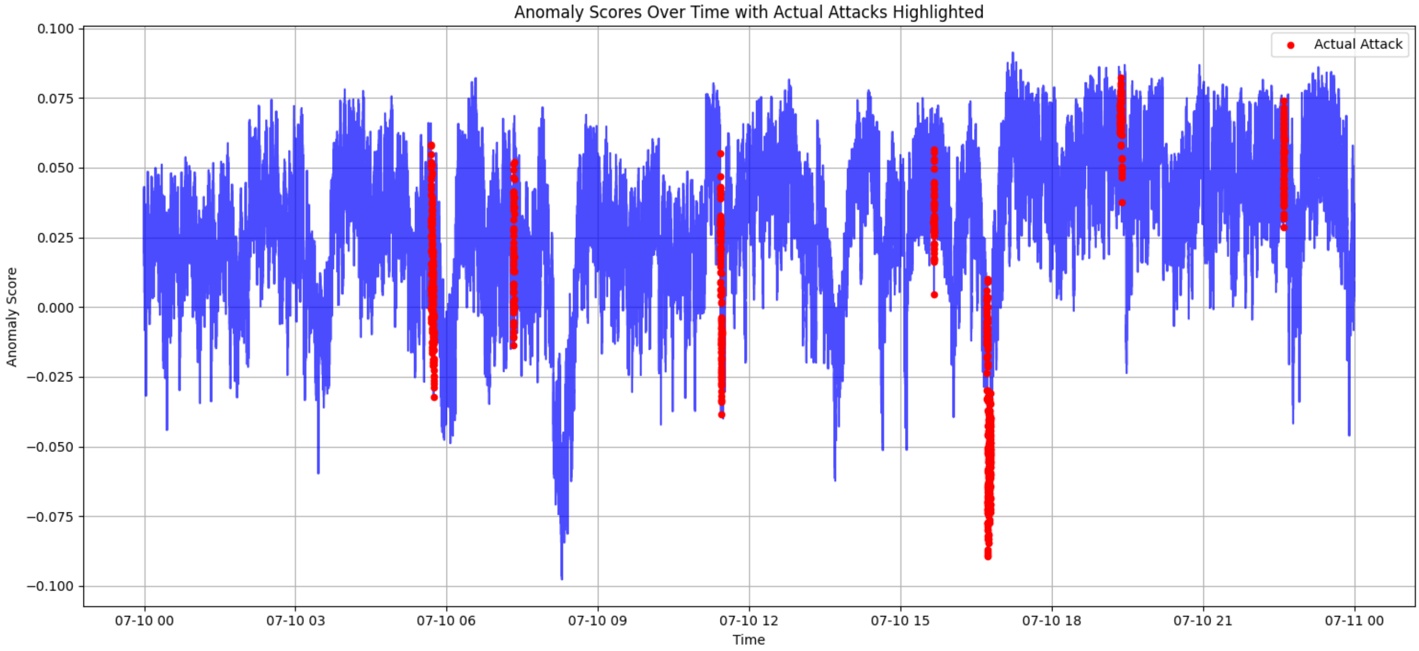

Figure 5. Anomaly Scores Over Time with Actual Attacks Highlighted.

However, the low precision (0.04) and F1-score (0.08) for the Attack class highlight intrinsic limitations common to many unsupervised and semi-supervised methods. Unlike OC-SVM, which attempts to learn a tight boundary around normal data in a high-dimensional kernel space, Isolation Forest does not explicitly define a decision boundary. This allows it to scale efficiently and avoid kernel sensitivity, but it also limits its ability to separate attack samples that remain close to the normal data manifold. OC-SVMs, although potentially offering sharper boundaries, are highly sensitive to kernel choice and hyperparameter tuning and scale poorly with the dimensionality and size of industrial CPS datasets, which constrains their practicality for real-time deployment. Deep autoencoder-based and recurrent models, such as LSTM autoencoders, are theoretically better suited to capturing complex temporal and cross-variable dependencies inherent in CPES [9]. By explicitly modeling sequential dynamics, these approaches can detect attacks that manifest as subtle but persistent deviations over time [25].

However, these advantages come at the cost of substantial computational overhead, long training times, and reliance on large volumes of representative normal data. Moreover, deep models often lack interpretability and are sensitive to noise and concept drift, which can complicate validation and operational trust in safety-critical systems. In contrast, the Isolation Forest results presented here demonstrate competitive baseline performance with minimal training complexity and stable inference latency, reinforcing its suitability as a first-line monitoring mechanism [26], [27]. The confusion matrix and ROC analysis further emphasize that attack detection in CPES is constrained less by classifier choice than by the fundamental overlap between normal and attack-induced system behavior. Many industrial attacks are engineered to remain within physically plausible operating bounds, reducing the effectiveness of point-wise anomaly detectors, regardless of their underlying learning paradigm.

The overlapping anomaly score distributions observed in Figure 4 illustrate this limitation and explain why threshold-based detectors, whether Isolation Forest, OC-SVM, or reconstruction-error–based autoencoders, exhibit similar trade-offs between missed detections and false alarms. The temporal alignment between anomaly score peaks and attack events in Figure 5 suggests that incorporating temporal context is essential for improved discrimination. Deep temporal models achieve this implicitly through sequence learning, but similar benefits may be obtained more efficiently through hybrid approaches that combine Isolation Forest with temporal aggregation, change-point detection, or post-hoc sequence validation. Such strategies preserve the computational efficiency and scalability of tree-based methods while addressing their inability to model persistence and causal structure directly [9], [28].

The results position Isolation Forest as a computationally efficient and operationally robust baseline for CPES intrusion detection, particularly when interpretability, scalability, and real-time constraints are prioritized. However, the findings also confirm that no single anomaly detection paradigm is sufficient for reliable attack detection in complex cyber-physical systems. Future intrusion detection frameworks should therefore adopt hybrid and hierarchical designs that integrate lightweight anomaly detectors with temporal and imbalance-aware mechanisms, bridging the gap between practical deplorability and high-fidelity attack discrimination.

5. Conclusion

This study examined a one-class anomaly detection framework for intrusion detection in cyber-physical engineering systems using the Isolation Forest algorithm. The proposed approach models normal system behavior from high-dimensional sensor, actuator, and control signals and is designed for realistic industrial environments where labeled attack data are limited and real-time monitoring is required. Experimental evaluation using the Hardware-in-the-Loop–based Augmented Industrial Control System (HAI) Security Dataset demonstrates that the model effectively captures normal operational patterns and achieves strong performance in identifying legitimate system states, resulting in high overall accuracy. These results confirm that Isolation Forest provides a computationally efficient and scalable solution suitable for continuous monitoring in industrial cyber-physical environments. However, the results also reveal important limitations in attack detection performance. Severe class imbalance and significant overlap between normal and attack data distributions result in low attack precision and a high false-positive rate, despite moderate recall.

These findings suggest that many cyberattacks in cyber-physical engineering systems remain statistically similar to legitimate operational behavior, making them difficult to distinguish using point-wise anomaly detection alone. Consequently, a trade-off exists between improving attack detection sensitivity and maintaining acceptable false alarm levels, which limits the effectiveness of standalone anomaly detection models in safety-critical environments. This work provides an objective assessment of Isolation Forest as a lightweight baseline anomaly detection mechanism for cyber-physical intrusion detection. While the proposed framework offers strong computational efficiency and practical deployability, achieving reliable attack discrimination will likely require hybrid detection architectures that incorporate temporal modeling, imbalance-aware learning strategies, or domain-informed detection mechanisms. Future research should therefore focus on integrating these complementary techniques to improve detection reliability and support the development of more resilient cybersecurity solutions for cyber-physical engineering systems.

5. Declarations

5.1. Author Contributions

Godfrey Perfectson Oise: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing – Original Draft, Visualization; **Susan Konyeha:** Investigation, Data Curation, Writing – Review & Editing; **Felix Oshiorenoya Uloko:** Methodology, Formal Analysis, Validation, Writing – Review & Editing; **Kevin Chinedu Pius:** Investigation, Resources, Data Curation, Writing – Review & Editing; **Enovwo Eferoba-Idio:** Investigation, Data Curation, Resources, Writing – Original Draft; **Michael Uyiosa Edobor:** Resources, Data Curation, Investigation, Writing – Review & Editing;

Evans Mintah: Supervision, Formal Analysis, Writing – Review & Editing; **Osahon Ukpebor:** Supervision, Project Administration, Writing – Review & Editing; **Oludare Sokoya:** Validation, Visualization, Writing – Review & Editing; **Tejiri Jessa:** Resources, Data Curation, Writing – Review & Editing; All authors have read and approved the final version of the manuscript.

5.2. Institutional Review Board Statement

Not applicable.

5.3. Informed Consent Statement

Not applicable.

5.4. Data Availability Statement

The datasets used and/or analyzed during the current study are available at: https://www.kaggle.com/datasets/icsdataset/hai-security-dataset.

5.5. Acknowledgment

Not applicable.

5.6. Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

6. References

[1] I. A. Khan, M. Keshk, D. Pi, N. Khan, Y. Hussain, and H. Soliman, "Enhancing IIoT networks protection: A robust security model for attack detection in Internet Industrial Control Systems," *Ad Hoc Networks*, vol. 134, p. 102930, Sep. 2022. https://doi.org/10.1016/j.adhoc.2022.102930.

[2] P. Verma, D. O'Shea, T. Newe, N. Mehta, N. Bharot, and J. G. Breslin, "ABIDS-VEM: leveraging an equilibrium optimizer and data ramification in association with ensemble learning for anomaly-based intrusion detection system," *J. Supercomput.*, vol. 81, no. 7, p. 856, May 2025. https://doi.org/10.1007/s11227-025-07292-w.

[3] K. Shanthi and R. Maruthi, "Machine Learning Approach for Anomaly-Based Intrusion Detection Systems Using Isolation Forest Model and Support Vector Machine," in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, Aug. 2023, pp. 136–139. https://doi.org/10.1109/ICIRCA57980.2023.10220620.

[4] C. K. Reddy, G. Keerthi, G. Pranay, and A. J. A, "Machine Learning based Enhanced Intrusion Detection for Cybersecurity," in *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, IEEE, Feb. 2025, pp. 1400–1407. https://doi.org/10.1109/ICSADL65848.2025.10933487.

[5] G. P. Oise, B. S. Olanrewaju, O. A. Orukpe, K. C. Pius, and A. O. Airhiavbere, "A Convolutional Neural Network Framework for Intelligent Intrusion Detection," *Scientific Journal of Computer Science*, vol. 2, no. 1, pp. 50–59, Feb. 2026. https://doi.org/10.64539/sjcs.v2i1.2026.404.

[6] S. A. Oyedotun, G. P. Oise, and C. E. Ozobialu, "Towards Intelligent Cybersecurity in SCADA and DCS Environments: Anomaly Detection Using Multimodal Deep Learning and Explainable AI," *Journal of Science Research and Reviews*, vol. 2, no. 3, pp. 20–31, Jul. 2025. https://doi.org/10.70882/josrar.2025.v2i3.76.

[7] M. Rani and Gagandeep, "An Efficient Network Intrusion Detection System Based on Feature Selection Using Evolutionary Algorithm Over Balanced Dataset," *Mobile Radio Communications and 5G Networks*, 2022, pp. 179–193. https://doi.org/10.1007/978-981-16-7018-3_15.

[8] A. Y. Hussein, P. Falcarin, and A. T. Sadiq, "IoT Intrusion Detection Using Modified Random Forest Based on Double Feature Selection Methods," in *International Conference on Emerging Technology Trends in Internet of Things and Computing*, 2022, pp. 61–78. https://doi.org/10.1007/978-3-030-97255-4_5.

[9] S. Sharma, S. Mohan, P. Aryan, and R. V. S. Devi, "Cybersecurity Optimization Using Particle Swarm Optimization and Machine Learning in Intrusion Detection Systems," in *2025 IEEE 4th*

International Conference for Advancement in Technology (ICONAT), IEEE, Sep. 2025, pp. 1–5. https://doi.org/10.1109/ICONAT66879.2025.11362892.

[10] M. S. Siddique, Md. A. R. Khan, I. Ahammad, N. Nath, J. R. Das, and F. Rahman, "An intelligent intrusion detection system for cyber-physical systems using GAN-LSTM networks," *Franklin Open*, vol. 11, p. 100281, Jun. 2025. https://doi.org/10.1016/j.fraope.2025.100281.

[11] S. A. Elsaid and A. Binbusayyis, "An optimized isolation forest based intrusion detection system for heterogeneous and streaming data in the industrial Internet of Things (IIoT) networks," *Discover Applied Sciences*, vol. 6, no. 9, p. 483, Sep. 2024. https://doi.org/10.1007/s42452-024-06165-w.

[12] G. P. Oise, O. C. Nwabuokei, O. J. Akpowehbve, B. A. Eyitemi, and N. B. Unuigbokhai, "Towards Smarter Cyber Defense: Leveraging Deep Learning for Threat Identification and Prevention," *FUDMA Journal of Sciences*, vol. 9, no. 3, pp. 122–128, Mar. 2025. https://doi.org/10.33003/fjs-2025-0903-3264.

[13] S. Agrawal *et al.*, "Federated Learning for intrusion detection system: Concepts, challenges and future directions," *Comput. Commun.*, vol. 195, pp. 346–361, Nov. 2022. https://doi.org/10.1016/j.comcom.2022.09.012.

[14] F. Folino, G. Folino, M. Guarascio, F. S. Pisani, and L. Pontieri, "On learning effective ensembles of deep neural networks for intrusion detection," *Information Fusion*, vol. 72, pp. 48–69, Aug. 2021. https://doi.org/10.1016/j.inffus.2021.02.007.

[15] S. A. Oyedotun *et al.*, "The Role of Internal Audit in Fraud Detection and Prevention: A Multi-Contextual Review and Research Agenda," *Journal of Science Research and Reviews*, vol. 2, no. 2, pp. 76–85, May 2025. https://doi.org/10.70882/josrar.2025.v2i2.51.

[16] N. B. Unuigbokhai *et al.*, "Advancements in Federated Learning for Secure Data Sharing in Financial Services," *FUDMA Journal of Sciences*, vol. 9, no. 5, pp. 80–86, May 2025. https://doi.org/10.33003/fjs-2025-0905-3207.

[17] G. P. Oise *et al.*, "Decentralized Deep Learning in Healthcare: Addressing Data Privacy with Federated Learning," *FUDMA Journal of Sciences*, vol. 9, no. 6, pp. 19–26, Jun. 2025. https://doi.org/10.33003/fjs-2025-0906-3714.

[18] Q. Lin, R. Ming, K. Zhang, and H. Luo, "Privacy-Enhanced Intrusion Detection and Defense for Cyber-Physical Systems: A Deep Reinforcement Learning Approach," *Security and Communication Networks*, vol. 2022, 2022. https://doi.org/10.1155/2022/4996427.

[19] S. Vladov *et al.*, "Neural Network DDoS Mitigation System With Forensic Audit Support for Cyber Police," *IEEE Access*, vol. 13, pp. 204628–204655, 2025. https://doi.org/10.1109/ACCESS.2025.3634478.

[20] Z. Yu, H. Gao, X. Cong, N. Wu, and H. H. Song, "A Survey on Cyber-Physical Systems Security," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21670–21686, Dec. 2023. https://doi.org/10.1109/JIOT.2023.3289625.

[21] M. Arafah, I. Phillips, A. Adnane, W. Hadi, M. Alauthman, and A. K. Al-Banna, "Anomaly-based network intrusion detection using denoising autoencoder and Wasserstein GAN synthetic attacks," *Appl. Soft Comput.*, vol. 168, no. 8, pp. 6247–6256, Jan. 2025. https://doi.org/10.1016/j.asoc.2024.112455.

[22] M. Wang, K. Zheng, Y. Yang, and X. Wang, "An Explainable Machine Learning Framework for Intrusion Detection Systems," *IEEE Access*, vol. 8, pp. 73127–73141, 2020. https://doi.org/10.1109/ACCESS.2020.2988359.

[23] A. Alabdulatif, N. N. Thilakarathne, and Z. K. Lawal, "A Review on Security and Privacy Issues Pertaining to Cyber-Physical Systems in the Industry 5.0 Era," *Computers, Materials and Continua*, vol. 80, no. 3, pp. 3917–3943, 2024. https://doi.org/10.32604/cmc.2024.054150.

[24] ICS Security Dataset, "HAI Security Dataset," 2023. https://www.kaggle.com/datasets/icsdataset/hai-security-dataset.

[25] V. Selvakkumaran and R. Anandan, "Performance Analysis on Intrusion Detection System using Fuzzy Neural Network Approach," *International Journal of Advanced Science and Engineering*, vol. 12, no. 2, pp. 5753–5763, Dec. 2025. https://doi.org/10.29294/ijase.12.2.2025.5753-5763.

[26] G. P. Oise, "E-ViTNet: A lightweight vision transformer with oppositional cat swarm optimization for automated E-Waste sorting," *Next Research*, vol. 6, p. 101373, Apr. 2026. https://doi.org/10.1016/j.nexres.2026.101373.

[27] S. M. Rajagopal, M. Supriya, and R. Buyya, "FedSDM: Federated learning based smart decision making module for ECG data in IoT integrated Edge–Fog–Cloud computing

environments," *Internet of Things (Netherlands)*, vol. 22, Jul. 2023. https://doi.org/10.1016/j.iot.2023.100784.

[28] A. Mohamed, J. Heilala, and N. S. Madonsela, "Machine Learning-Based Intrusion Detection Systems for Enhancing Cybersecurity," in *2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, IEEE, Aug. 2023, pp. 366–370. https://doi.org/10.1109/SmartTechCon57526.2023.10391626.