applied sciences

MDPI

Article

Cyberattack Detection Systems in Industrial Internet of Things (IIoT) Networks in Big Data Environments

Abdullah Orman

Department of Computer Technologies, Ankara Yıldırım Beyazıt University, Ankara 06010, Turkey; aorman@aybu.edu.tr

Abstract: The rapid expansion of the Industrial Internet of Things (IIoT) has revolutionized industrial automation and introduced significant cybersecurity challenges, particularly for supervisory control and data acquisition (SCADA) systems. Traditional intrusion detection systems (IDSs) often struggle to effectively identify and mitigate complex cyberthreats, such as denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks. This study proposes an advanced IDS framework integrating machine learning, deep learning, and hybrid models to enhance cybersecurity in IIoT environments. Using the WUSTL-IIoT-2021 dataset, multiple classification models—including decision tree, random forest, multilayer perceptron (MLP), convolutional neural networks (CNNs), and hybrid deep learning architectures—were systematically evaluated based on key performance metrics, including accuracy, precision, recall, and F1 score. This research introduces several key innovations. First, it presents a comparative analysis of machine learning, deep learning, and hybrid models within a unified experimental framework, offering a comprehensive evaluation of various approaches. Second, while existing studies frequently favor hybrid models, findings from this study reveal that the standalone MLP model outperforms other architectures, achieving the highest detection accuracy of 99.99%. This outcome highlights the critical role of dataset-specific feature distributions in determining model effectiveness and calls for a more nuanced approach when selecting detection models for IIoT cybersecurity applications. Additionally, the study explores a broad range of hyperparameter configurations, optimizing model effectiveness for IIoT-specific intrusion detection. These contributions provide valuable insights for developing more efficient and adaptable IDS solutions in IIoT networks.

Keywords: Industrial Internet of Things (IIoT); SCADA security; DoS attacks; intrusion detection; cybersecurity; big data; deep learning; machine learning; hybrid model

Academic Editor: Stefan Fischer

Received: 1 February 2025
Revised: 9 March 2025
Accepted: 10 March 2025
Published: 13 March 2025

Citation: Orman, A. Cyberattack Detection Systems in Industrial Internet of Things (IIoT) Networks in Big Data Environments. *Appl. Sci.* **2025**, *15*, 3121. https://doi.org/10.3390/app15063121

1. Introduction

Over the last decade, the proliferation of the Internet of Things (IoT) across various domains, including healthcare, agriculture, smart cities, environmental monitoring, and industry, has significantly influenced the global economy. By facilitating communication between physical devices via the internet, IoT has enabled the emergence of innovative services and applications [1]. Projections indicate that by 2025, IoT-related developments will generate an economic impact ranging from USD 3.9 trillion to 11.1 trillion [2].

IIoT extends the IoT paradigm to the manufacturing sector to enhance operational efficiency. IIoT involves the integration of intelligent devices with management platforms, while supervisory control and data acquisition (SCADA) systems serve as fundamental

components. The transition of traditional SCADA architectures into IoT-enabled frameworks has fundamentally altered the landscape of cybersecurity threats [3].

Although integrating SCADA systems with IoT and cloud infrastructures provides cost-effectiveness and enhanced performance, it simultaneously introduces heightened cybersecurity risks, mainly due to the potential for remote access vulnerabilities [4]. IIoT devices are increasingly susceptible to various cyberthreats, including distributed denial-of-service (DDoS) attacks and malware infections. Consequently, research efforts concerning IIoT security have intensified recently [5].

The IIoT ecosystem is increasingly vulnerable to cybersecurity threats, including unauthorized access, data integrity breaches, denial-of-service (DoS) attacks, and protocol-specific vulnerabilities. To mitigate these risks, intrusion detection systems (IDSs) are critical in identifying anomalous network traffic, enabling proactive security measures [6]. Additionally, digital forensic incident response (DFIR) strategies are essential for protecting supervisory control and data acquisition (SCADA) systems from cyberthreats [7]. However, conventional security approaches have limitations, necessitating the development of next-generation detection methods. Recent advancements in machine learning and deep learning have significantly improved the accuracy and efficiency of IDS models, making them increasingly relevant in IIoT cybersecurity research [8–12].

A promising approach in this domain is the integration of hybrid deep learning models that enhance IDS performance against evolving cyberthreats, including distributed denial-of-service (DDoS) and DoS attacks. Hybrid deep learning models combine multiple neural network architectures or integrate traditional machine learning approaches to enhance predictive performance across various applications. Such methodologies leverage the strengths of distinct algorithms to address the shortcomings found in their isolated use [13]. By leveraging the strengths of multiple deep learning architectures, hybrid models improve detection accuracy while reducing false-positive rates—critical factors in complex and dynamic IIoT environments. Notably, Konatham et al. [14] introduced a hybrid model combining convolutional neural networks (CNNs) and gated recurrent units (GRUs), achieving 94.94% accuracy in detecting anomalies in IIoT edge computing environments. This highlights the effectiveness of integrating spatial and temporal feature extraction. Similarly, Marzouk et al. [15] developed a hybrid model incorporating metaheuristic algorithms for intrusion detection in clustered IIoT environments, demonstrating adaptability to different network architectures.

DDoS attacks remain a significant challenge in network security, requiring advanced detection mechanisms. Shaikh et al. [16] proposed a CNN–LSTM hybrid model specifically designed to detect DDoS attacks, utilizing CNNs for spatial feature extraction and LSTMs for temporal sequence analysis. This dual-method approach enhances detection accuracy while addressing the limitations of traditional IDS models, which often struggle with the complexity and volume of IIoT network data. Further refinement has been achieved through optimization techniques such as the satin bowerbird optimization algorithm, which improves data preprocessing for enhanced model compatibility [17].

Recent developments also incorporate federated learning to enhance privacy while maintaining detection performance. Huang et al. [18] introduced a federated learning-based IDS model that integrates CNNs with attention mechanisms, effectively addressing privacy concerns and accuracy challenges in IIoT environments. This approach is particularly relevant given the increasing emphasis on secure industrial networks.

Additionally, ensemble methods have demonstrated significant potential in improving predictive performance. Begum et al. [19] proposed a CNN–LSTM hybrid model that achieved 99% accuracy on the KDD-Cup dataset, reinforcing the effectiveness of ensemble

techniques in intrusion detection. Javeed et al. [20] further highlighted the importance of hybrid classifiers in detecting sophisticated cyberthreats in secure industrial environments.

This study employed the WUSTL-IIoT-2021 dataset to detect cyberthreats in IIoT systems, providing a robust platform for evaluating security mechanisms by simulating real-world industrial conditions [21]. As machine learning continues to demonstrate significant promise in cybersecurity, research in this area has expanded rapidly [22,23]

In this study, SCADA network traffic was analyzed using five machine learning models (CART, decision tree, logistic regression, naïve Bayes, random forest), five deep learning models (CNN, GRU, LSTM, RNN, MLP), and two hybrid models (CNN–LSTM, LSTM–CNN). A key contribution of this research is the comparative evaluation of machine learning, deep learning, and hybrid models within a unified experimental framework, providing a comprehensive analysis of their relative effectiveness. By systematically applying different hyperparameter configurations, the study aimed to refine model performance and establish a foundation for future research. The model demonstrating the highest accuracy in cyberattack detection was identified and evaluated for potential integration into network-based IDS solutions. When deployed in IDS environments, the proposed model offers superior detection accuracy and reduced error rates compared to conventional security mechanisms.

The results indicated that the multilayer perceptron (MLP) model outperformed other approaches, achieving an accuracy of 99.99%, surpassing similar studies in the literature. Contrary to widely held assumptions that hybrid models yield the highest performance, this study demonstrates that standalone models can achieve superior accuracy when applied to the WUSTL-IIoT-2021 dataset. These findings provide valuable insights for advancing cyberattack detection methodologies in IIoT environments.

Furthermore, this study makes significant contributions to the cybersecurity literature by providing a detailed analysis of common attack types in SCADA systems, offering a valuable resource for researchers and developers working in this domain. Using a dataset obtained directly from real IIoT devices ensures the practical applicability of the proposed model across both SCADA and IIoT environments. Unlike traditional attack detection techniques, integrating artificial intelligence-based methodologies introduces an innovative perspective, with the effectiveness of the proposed approach validated through comparative analysis with existing studies.

In addition to evaluating machine learning and deep learning models, the study also explores hybrid architectures, presenting new insights for cybersecurity experts. The study lays a solid foundation for future research by testing and optimizing various hyperparameters. A comprehensive assessment of SCADA security vulnerabilities and cyberthreats is conducted, enhancing awareness of potential risks. Moreover, advanced big data analytics in this research contribute significantly to data processing methodologies and cyberthreat detection.

This study sought to answer the following research questions:

- To strengthen cybersecurity in IIoT environments, which machine learning, deep learning, and hybrid models exhibit the highest performance in cyberattack detection?
- How are these models compared and evaluated regarding key performance metrics such as accuracy, F1 score, precision, and recall?

The structure of this paper is as follows. Section 2 provides a review of related literature, Section 3 details the methodology and dataset, Section 4 describes the implementation and evaluation processes, Section 5 presents and analyzes the results, and Section 6 concludes the study with discussions on future research directions.

2. Related Work

IIoT systems form a complex structure that integrates various devices and technologies, making them vulnerable to cyberattacks. In particular, denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks can seriously affect the operation of IIoT networks, and such attacks have the potential to turn off production systems [24,25]. Many IIoT devices have outdated technology and are equipped with insufficient security measures. This can make it easier for attackers to infiltrate the network [26]. In addition, since IIoT systems usually operate over cloud-based infrastructures, cloud computing systems also become a target for attackers [27,28].

The WUSTL-IIoT-2021 dataset has become a widely recognized benchmark for evaluating intrusion detection systems (IDSs) in IIoT environments. Its consistent adoption across numerous studies enables direct comparisons between different models, algorithms, and methodologies within a standardized experimental framework [29].

For instance, Alani et al. (2022) [30] introduced a deep learning-based IDS for industrial IoT, employing a multi-stage process. Initially, the dataset was randomly partitioned into training and testing subsets, with 75% allocated for training a deep learning classifier over 75 epochs using a batch size of 1000. The remaining 25% was used in the testing phase to evaluate the classifier's generalization capability. To further enhance generalization, a third stage incorporated 10-fold cross-validation. The resulting model, DeepIIoT, demonstrated exceptional performance on the WUSTL-IIoT-2021 dataset, achieving accuracy exceeding 99% with false-positive and false-negative rates of 0.069% and 0.032%, respectively. Notably, DeepIIoT outperformed alternative approaches tackling similar challenges.

In 2023, Mohy-Eddine et al. [31] introduced an intrusion detection model emphasizing advanced preprocessing techniques. Their approach integrated feature selection using the Pearson correlation coefficient (PCC) and outlier detection via the isolation forest (IF) algorithm. The PCC was employed as a feature reduction method to improve model convergence, lower computational costs, and enhance training efficiency. The IF algorithm effectively identified outliers within the Bot-IoT and WUSTL-IIoT-2021 datasets, significantly improving model performance, particularly in handling imbalanced datasets. Similarly, Babayigit and Abubaker (2024) [32] proposed a hybrid framework for detecting malicious activities in IIoT environments, incorporating minimum description length (MDL) and transfer learning (TL). This framework standardized dimensions and distributions across diverse IIoT datasets, enabling a unified feature representation. A CNN-GRU model was trained on the integrated datasets, with Bayesian optimization (BO) applied for hyperparameter tuning. Experimental results demonstrated the framework's effectiveness in developing a robust deep learning model capable of generalizing across multiple datasets, emphasizing the crucial role of dataset quality in ensuring reliable training and testing outcomes.

Eid et al. (2024) [33] developed a machine learning-based intrusion detection system (IDS) for IIoT networks, evaluating six machine learning algorithms: decision tree (DT), random forest (RF), k-nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), and naïve Bayes (NB). All models performed well, with Matthews correlation coefficient (MCC) scores exceeding 99%. Alani (2023) [9] investigated a flow-based IDS for IIoT utilizing classifiers such as RF, LR, DT, and Gaussian naïve Bayes (GNB). The study employed a preprocessed dataset and applied recursive feature elimination to refine the feature set to 11 attributes. A 10-fold cross-validation approach ensured robust generalization with minimal accuracy variance.

Popoola et al. (2023) [34] introduced a federated deep learning (FDL) model for intrusion detection in consumer-centric IoT, achieving accuracy of 0.9997 for DT with an inference time of 0.1517 μs. Their comparative analysis of multiple datasets, including

X-IIoTID, Edge-IIoTset, and WUSTL-IIoT-2021, demonstrated the superiority of FDL models over centralized deep learning (CDL) models, providing timely and privacy-preserving intrusion detection. Alzahrani and Aldhyani (2023) [35] examined AI-driven cybersecurity enhancements for industrial control systems, comparing machine learning techniques (KNN, RF) and deep learning architectures (CNN-GRU). Their study achieved exceptionally high accuracy rates—99.99% for KNN and RF and 99.98% for CNN-GRU—on the WUSTL-IIoT-2018 dataset, further confirming the effectiveness of AI-based intrusion detection strategies in IIoT security.

Similarly, Dina et al. (2023) [36] proposed a deep learning-based intrusion detection model utilizing feedforward neural networks (FNNs) and convolutional neural networks (CNNs). Their evaluation using the WUSTL-IIoT-2021 dataset demonstrated superior performance, with the FNN-focal model achieving accuracy of 98.95%. Diaba et al. (2023) [37] analyzed the impact of manipulated datasets on machine learning models to assess cybersecurity risks in power systems. Based on the WUSTL-IIoT-2021 and WUSTL-IIoT-2018 datasets, their findings revealed that manipulated datasets led to reduced accuracy, increased prediction errors, and longer training times for all algorithms except the boosted tree algorithm.

Xu et al. (2023) [38] proposed an IoT intrusion detection system based on machine learning, employing the XGBoost classifier in conjunction with two-stage feature selection methods: binary gray wolf optimization (BGWO) and recursive feature elimination with XGBoost (RFE-XGBoost). Experiments conducted on five publicly available datasets demonstrated the approach's superior accuracy, recall, precision, and F1 score performance. However, the study also highlighted challenges in scaling the method to large datasets due to computational complexity, memory constraints, efficiency, generalization ability, and robustness.

Ahakonye et al. (2024) [39] introduced the trees bootstrap aggregation (TBA) algorithm for detecting and classifying IoT-SCADA network traffic, focusing on the IEC-104 network communication protocol. Their study demonstrated TBA's high precision in identifying different network traffic types while reducing false-acceptance rates in heterogeneous IIoT sensor data. Bekbulatova et al. (2023) [40] addressed IIoT security concerns by proposing an anomaly-based IDS for detecting zero-day attacks. Their approach utilized semi-supervised learning on large-scale, unlabeled IIoT network traffic, implementing the DeepSAD model in a federated learning framework. While the centralized model outperformed the federated approach in detecting DoS attacks, variations in client performance within the federated setting indicated potential areas for future optimization.

Gaber et al. (2023) [41] made a notable contribution to IIoT cybersecurity by integrating particle swarm optimization (PSO) and the bat algorithm (BA) for feature selection, significantly improving the efficiency of IIoT-based traffic classification. Their study, which utilized the WUSTL-IIoT dataset, achieved accuracy of 99.99% and precision of 99.96%, substantially enhancing computational efficiency. Eid et al. (2024) [33] systematically evaluated machine learning models, investigating preprocessing techniques and dataset imbalances for IDSs in IIoT environments. Their study demonstrated that applying the synthetic minority oversampling technique (SMOTE) improved binary classification accuracy, with random forest and decision trees achieving 99.98%. The study also introduced a novel multi-class classification approach using SMOTE, enhancing detection performance for various attack types, with RF, DT, and LR achieving near-perfect accuracy.

Casajús-Setién et al. (2023) [42] proposed an anomaly detection-based IDS framework using a transformer model, significantly advancing IIoT cybersecurity. Utilizing the WUSTL-IIoT-2021 dataset, their research demonstrated the model's effectiveness in analyzing sequential network flows using a streamlined multi-head attention mechanism.

Ye et al. (2024) [43] introduced an ensemble framework incorporating an enhanced harmony search algorithm (HBO) for feature selection in IDSs. Their approach, tested on datasets such as NSL-KDD, WUSTL-IIoT-2021, and HAI, improved intrusion detection accuracy by nearly 15% on large-scale datasets while significantly reducing the number of original features.

Lastly, Saxena and Mittal (2023) [44] conducted a comprehensive review of existing IIoT network datasets and advanced persistent threat (APT) attack characteristics, proposing a standardized evaluation framework for benchmarking IDS models. Their assessment, applied to datasets including WUSTL-IIoT-2021, underscored the dataset's significance in advancing IIoT security research and developing effective countermeasures against sophisticated cyberthreats.

In summary, studies utilizing the WUSTL-IIoT-2021 dataset frequently adapt their methodologies and consistently report high accuracy in intrusion detection. Many of the proposed models evaluated on this dataset achieve accuracy nearing 99%, reflecting a prevailing trend in developing robust and effective intrusion detection systems for IIoT environments. Several studies have demonstrated superior performance compared to the existing literature, highlighting the reliability and effectiveness of these approaches. The strong emphasis on high accuracy underscores the critical need for advanced, trustworthy security solutions to protect IIoT networks. As a widely adopted benchmark, the WUSTL-IIoT-2021 dataset plays a key role in these evaluations.

As shown in Table 1, the WUSTL-IIoT-2021 dataset has been analyzed in various studies. However, as it is relatively new compared to other datasets in the field, research directly focused on it remains limited. This study is expected to contribute to its growing adoption. Furthermore, this study stands out by examining three different modeling approaches—machine learning, deep learning, and hybrid models—making it one of the few in this domain to do so. The analysis also incorporates 12 different models, a rare approach in existing research. While hybrid methods have generally outperformed standalone deep learning techniques in previous studies, the specific models employed in this research demonstrated superior performance on the WUSTL-IIoT-2021 dataset.

Table 1. Comparison of other work on intrusion detection.

Reference	Authors	Year	Type	Dataset	Accuracy (%)	F1 Score (%)
[30]	Alani et al.	2022	MPL	WUSTL-IIOT-2021	99.97	99.97
[31]	Mohy-Eddine et al.	2023	IDS	WUSTL-IIOT-2021	94.29	93.96
[32]	Babayigit and Abubaker	2024	DL model	Combined dataset (Edge-IIoTSet, WUSTL-IIoT-2021, and XIIoTID)	94.83	94.65
[33]	Eid et al.	2023	ML-based IDS model	WUSTL-IIOT-2021	99.97	
[9]	Alani	2023	System	WUSTL-IIOT-2021	99.98	99.98
[34]	Popoola et al.	2023	FDL model	X-IIoTID, Edge-IIoTset, and WUSTL-IIoT-2021	99.39	99.90
[35]	Alzahrani and Aldhyani	2023	CNN-GRU model	WUSTL-IIOT-2021	99.98	
[36]	Dina et al.	2023	DL model	WUSTL-IIOT-2021	98.95 (FNN) 98.21 (CNN)	68.48 (FNN) 70.50 (CNN)
[37]	Diaba et al.	2023	ML algorithms	WUSTL-IIoT-2018 & WUSTL-IIOT-2021	99	
[40]	Bekbulatova et al.	2023	Semi-supervised approach: DeepSAD	WUSTL-IIOT-2021	99.93	100
[41]	Gaber et al.	2023	Machine learning (ML) algorithm	WUSTL-IIOT-2021	99.99	99.96
[33]	Eid et al.	2024	SMOTE-based multi-class balancing technique	WUSTL-IIOT-2021	99.99	95.4–99.98
[42]	Casajús-Setién et al.	2023	NIDS framework	WUSTL-IIOT-2021	96.37	92.51
[43]	Ye et al.	2023	Harmony search algorithm (HBO)-based feature selection (FS)	NSL-KDD, WUSTL-IIOT-2021, and HAI	98.72 (precision) 96.36 (recall)	97.27

3. Methods

This study aimed to detect cyberattacks in IIoT networks, which are increasingly adopted, interconnected, and expected to play a growing role in cybersecurity research as network sizes expand. The proposed framework consists of four key stages—data preprocessing, data splitting, classification, and evaluation—as illustrated in Figure 1. The raw dataset was prepared for classification algorithms following the data preprocessing phase.

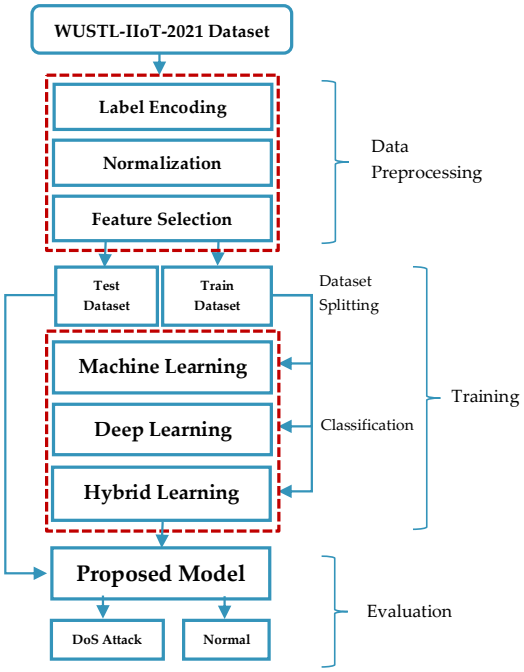

Figure 1. Schematic diagram of research methodology.

Subsequently, machine learning, deep learning, and hybrid models were applied to the preprocessed dataset. All modeling tasks were executed within the Google Colab environment to overcome hardware limitations, leveraging Google's computational infrastructure. The analysis used widely recognized libraries, including Pandas, MLlib, Scikit-learn, and PyCharm. Apache Spark (OpenJDK 8 headless ver 3.0.0) was selected as the computing platform and Python (ver 3.10) was used as the primary programming language. Hyperparameter tuning was performed for the deep learning models, with optimal values determined through iterative testing and alignment with commonly adopted parameters in the literature. The models categorized network traffic into normal and attack classes, and their performance was assessed on an independent test set that was not utilized during training. The model achieving the highest accuracy in attack detection was identified and considered for integration into network-based intrusion detection systems (IDSs). When implemented in IDSs, the developed model demonstrates superior detection rates and reduced error margins, surpassing traditional intrusion detection techniques.

A binary classification approach was employed to distinguish between normal and attack traffic. Since 90% of the attacks in the dataset comprised denial-of-service (DoS) traffic, the characteristics of other attack types, which exhibited high similarity to DoS patterns, were categorized accordingly. The dataset was divided into two subsets to ensure objective evaluation—70% for training and 30% for testing—with the test set exclusively reserved for performance assessment. The models were evaluated based on F1 score, accuracy, recall, and precision metrics to assess their effectiveness in detecting attack traffic.

During the final evaluation phase, the performance of the proposed method was rigorously analyzed, and the results demonstrated high accuracy in detecting DoS attacks, underscoring the model's effectiveness in IIoT cybersecurity applications.

3.1. Dataset

The WUSTL-IIoT-2021 dataset was developed to facilitate the detection of cyberattacks targeting IIoT networks, supporting cybersecurity research. Created by Zolanvari et al. [29], this dataset was explicitly designed to enhance the detection and classification of denial-of-service (DoS) attacks. It was generated using an IIoT testbed that closely replicates real-world industrial systems, enabling the execution of actual cyberattacks. The dataset encompasses 2.7 GB of network traffic data collected over approximately 53 h. Before analysis, the dataset underwent preprocessing, including removing missing values, corrupted entries, and extreme outliers. Although it originally contained six distinct attack types, a binary classification approach was adopted, distinguishing between "attack traffic" and "normal traffic." Given the high feature similarity among different attack types, all attack instances were labeled class 1, while normal traffic was designated class 0. The characteristics of the dataset are presented in Table 2.

Table 2. Specifics of the developed dataset.

Dataset	WUSTL-IIoT
Number of observations	1,194,464
Number of features	41
Number of attack samples	87,016
Number of normal samples	1,107,448

The dataset was intentionally designed to be imbalanced to ensure a realistic representation of industrial network environments. Attack scenarios, including command injection, reconnaissance, and DoS attacks, were executed against the testbed to capture a diverse range of malicious activities. Notably, attack traffic constitutes less than 8% of the dataset, aligning with real-world industrial control system conditions. Table 3 provides statistical insights into the dataset, with an average data rate of 419 kbit/s and an average packet size of 76.75 bytes. Since DoS attacks typically generate high volumes of network traffic, approximately 90% of recorded attack instances were allocated to this category. In contrast, other attack types were comparatively infrequent, generating only limited traffic.

Table 3. Statistical information about traffic types in the dataset.

Traffic's Type	Percentage (%)
Normal Traffic	92.72
Total Attack Traffic	7.28
Command Injection Traffic	0.31
DoS Traffic	89.98
Reconnaissance Traffic	9.46
Backdoor Traffic	0.25

The dataset includes various host-specific attributes, such as source and destination IP addresses. However, incorporating these features during model training may lead to overfitting, limiting the model's ability to generalize to unseen data. Additionally, specific attributes, such as flow start and end times, do not directly contribute to attack detection. Consequently, features such as StartTime, Last-Time, SrcAddr, DstAddr, slpld, and

dlpld were removed to prevent model over-learning and improve detection performance. Following this refinement process, the total number of features was reduced to 42.

Feature selection plays a critical role in constructing an adequate dataset for intrusion detection. The selected features exhibited significant changes during attack phases compared to normal network behavior. If a feature remains static across both attack and normal states, even the most advanced detection algorithms will fail to identify anomalies. The final dataset contains 42 features, with their descriptions detailed in Table 4.

Table 4. Detailed description of the attributes of the dataset.

Features	Type	Descriptions
Mean Flow (mean)	Float	Average duration of the active flows
Source Port (Sport)	Integer	Source port number
Destination Port (Dport)	Integer	Destination port number
Source Packets (Spkts)	Integer	Source/Destination packet count
Destination Packets (Dpkts)	Integer	Destination/Source packet count
Total Packets (Tpkts)	Integer	Total transaction packet count
Source Bytes (Sbytes)	Integer	Source/Destination bytes count
Destination Bytes (Dbytes)	Integer	Destination/Source bytes count
Total Bytes (TBytes)	Integer	Total transaction bytes count
Source Load (Sload)	Float	Source bits per second
Destination Load (Dload)	Float	Destination bits per second
Total Load (Tload)	Float	Total bits per second
Source Rate (Srate)	Float	Source packets per second
Destination Rate (Drate)	Float	Destination packets per second
Total Rate (Trate)	Float	Total packets per second
Source Loss (Sloss)	Float	Source packets retransmitted/dropped
Destination Loss (Dloss)	Float	Destination packets retransmitted/dropped
Total Loss (Tloss)	Float	Total packets retransmitted/dropped
Total Percent Loss (Ploss)	Float	Percentage packets retransmitted/dropped
Source Jitter (ScrJitter)	Float	Source jitter in milliseconds
Destination Jitter (DrcJitter)	Float	Destination jitter in milliseconds
Source Interpacket (SIntPkt)	Float	Source interpacket arrival time in milliseconds
Destination Interpacket (DIntPkt)	Float	Destination interpacket arrival time in milliseconds
Protocol (Proto)	Char	Transaction protocol
Duration (Dur)	Integer	Record total duration
TCP RTT (TcpRtt)	Float	TCP connection setup round-trip time, sum of “synack” and “ackdat.”
Idle Time (Idle)	Float	Time since the last packet activity. This value is useful in real-time processing and is the current time–last time.
Sum (sum)	Integer	Total accumulated duration of aggregated records
Min (min)	Integer	Minimum duration of aggregated records
Max (max)	Integer	Maximum duration of aggregated records
Source Diff Serve Byte (sDSb)	Integer	Source different serve byte value
Source TTL (sTtl)	Float	Source → Destination TTL value
Destination TTL (dTtl)	Float	Destination → Source TTL value
Source App Byte (SAppBytes)	Integer	Source → Destination application bytes
Destination App Byte (DAppBytes)	Integer	Destination → Source application bytes
Total App Byte (TotAppByte)	Integer	Total application bytes
SYN_Ack (SynAck)	Float	TCP connection setup time, the time between the SYN and the SYN_ACK packets
Run Time (RunTime)	Float	Total active flow run time. This value is generated through aggregation and is the sum of the record’s duration.
Source TOC (sTos)	Integer	Source TOS byte value
Source Jitter (SrcJitAct)	Float	Source idle jitter (ms)
Destination Jitter (DstJitAct)	Float	Destination active jitter (ms)

The WUSTL-IIoT-2021 dataset contains 1,194,464 samples. It contains four types of attacks and is labeled into six classes: normal traffic, total attack traffic, command injection

traffic, DoS traffic, reconnaissance traffic, and backdoor traffic. Four of these are DoS attack types. Table 5 shows the data distribution by class. Additionally, Table 6 shows the count, mean, std, min, and max values of the dataset.

Table 5. Number of samples in each class of WUSTL-IIoT-2021 dataset.

Class	Number of Samples	Percentage (%)
Normal	1,107,448	92.72
DoS	87,016	7.28

Table 6. Statistical analysis of the WUSTL-IIoT-2021 dataset.

Attribute Name	Count	Mean	Std	Min	Max
Mean	1,017,794	0.13	0.69	0	5.00
Sport	1,017,794	54,446.18	12,063.55	0	2,765,721.00
Dport	1,017,794	790.94	3301.91	0	65,522.00
SrcPkts	1,017,794	161.63	50,431.25	0	27,739,670.00
DstPkts	1,017,794	16.90	1130.98	0	309,216.00
TotPkts	1,017,794	170.71	50,431.55	0	27,739,670.00
DstBytes	1,017,794	7286.89	733,509.15	0	82,515,592.00
SrcBytes	1,017,794	18,336.94	4,506,185.84	0	2,108,646,246.00
TotBytes	1,017,794	287,451.57	19,674,969.34	0	2,143,724,556.00
SrcLoad	1,017,794	15,710,849.65	83,370,738.15	0	1,156,000,000.00
DstLoad	1,017,794	222,130.87	7,976,189.49	0	688,000,000.00
Load	1,017,794	15,932,980.52	83,717,629.02	0	1,184,000,000.00
SrcRate	1,017,794	31,092.80	166,016.73	0	1,000,000.00
DstRate	1,017,794	415.26	15,458.57	0	1,000,000.00
Rate	1,017,794	31,539.27	166,697.32	0	3,000,000.00
SrcLoss	1,017,794	2.29	26.48	0	2105.00
DstLoss	1,017,794	2.55	51.79	0	5008.00
Loss	1,017,794	7.55	2742.31	0	2,765,721.00
pLoss	1,017,794	20.14	8.08	0	60.00
SrcJitter	1,017,794	469.68	467.45	0	47,216.66
DstJitter	1,017,794	13.88	139.15	0	36,793.16
SIntPkt	1,017,794	82.53	558.58	0	9999.34
DIntPkt	1,017,794	8.29	70.96	0	9023.43
Proto	1,017,794	95.11	1620.82	0	35,020.00
Dur	1,017,794	0.55	326.62	0	329,509.91
TcpRtt	1,017,794	0.00	0.05	0	3.04
IdleTime	1,017,794	1,548,269,699.62	29,485,577.59	0	1,548,888,448.00
Sum	1,017,794	0.20	0.80	0	5.10
Min	1,017,794	0.20	0.80	0	5.10
Max	1,017,794	0.20	0.80	0	5.10
sDSb	1,017,794	0.00	0.18	0	51.00
sTtl	1,017,794	128.96	24.64	0	255.00
dTtl	1,017,794	58.28	18.65	0	128.00
SAppBytes	1,017,794	219.87	2853.92	0	99,793.00
DAppBytes	1,017,794	6738.53	727,371.25	0	81,823,144.00
TotAppByte	1,017,794	674,045.81	42,102,963.44	0	4,293,700,396.00
SynAck	1,017,794	0.00	0.05	0	3.04
RunTime	1,017,794	0.20	0.80	0	5.10
sTos	1,017,794	0.01	0.73	0	207.00
SrcJitAct	1,017,794	61.89	414.07	0	4999.44
DstJitAct	1,017,794	0.27	4.98	0	769.52
Target	1,017,794	0.07	0.26	0	1.00

In this study, evaluations were made using the WUSTL-IIoT-2021 dataset. This dataset was selected because it includes DoS attacks specific to IIoT. It has been preferred because it has recently been widely used in machine learning studies.

3.2. Data Preprocessing

The data preprocessing phase consists of three key stages: encoding, normalization, and feature selection.

The first stage, encoding, involves converting textual attributes within the dataset into numerical values to facilitate processing by artificial intelligence algorithms. Additionally,

class labels indicating the category of each data entry are numerically represented. In cases where class labels are initially in textual format, they are transformed into numerical equivalents. For instance, the Traffic_Type attribute in the dataset comprises textual categories such as normal, command injection, DoS, reconnaissance, and backdoor. Given that 92% of the dataset consists of normal traffic and 90% of the remaining attack traffic corresponds to DoS attacks, all other attack types were also categorized as DoS attacks. Since textual representations hinder computational efficiency, these labels were converted into numerical values. Through one-hot encoding, the Traffic_Type feature was transformed into the target feature, as shown in Table 7. In this conversion, rows labeled 0 represent normal traffic, while those labeled 1 correspond to attack traffic.

Table 7. Digital transformation table.

Name	Label Number
Normal Traffic	0
Attack Traffic	1

The second stage, normalization, standardizes the numerical values of features within a 0–1 range to prevent attributes with large numerical values from disproportionately influencing the model's outcomes. Normalization ensures that all features contribute equitably to the learning process, enhancing model stability and performance. After normalization, the dataset is prepared for use in machine learning models. This step is particularly critical in mitigating the dominance of high-value features, which could otherwise introduce biases in algorithmic calculations. Normalization was applied uniformly to all features, ensuring that their values remained within the 0–1 range.

The third stage, feature selection, involves eliminating non-contributory features that increase computational complexity and strain local hardware resources. An excessive number of features can lead to higher processing costs, increased energy consumption, and extended computation times. Therefore, feature reduction is often implemented to enhance model efficiency by streamlining the dataset.

In this study, random forest, logistic regression, and ExtraTreesClassifier algorithms were employed to evaluate the significance of each feature. The outcomes of these three feature selection techniques were compared, and the results indicated that feature reduction was unnecessary for this dataset. All 42 features were deemed relevant and retained for model training. By analyzing the impact of these features on classification accuracy, this study provides valuable insights for future research and contributes to the broader literature on intrusion detection in IIoT environments.

4. Experiments and Evaluations

In this section, the WUSTL-IIoT-2021 dataset was analyzed using five machine learning models and five deep learning models, and two hybrid models were used in addition to the models. Google infrastructure was used to overcome hardware limitations and perform all modeling operations in the Google Colab environment. Popular libraries such as Pandas, MLib, Scikit-learn, and PyCharm were used in the analysis processes. The Apache Spark platform was preferred as the working environment, and Python was used as the programming language. Hyperparameters were used in the analysis of deep learning models. The most appropriate values of these parameters were determined by testing different values and considering the values commonly used in the literature. The classification process used the binary (normal and attack traffic) approach. To increase the objectivity of the results, the data set was divided into training and testing: 70% was reserved for model training, and the remaining 30% was reserved for testing. The part

reserved for testing was never used with models other than testing (e.g., training). While detecting the attack traffic, the results were evaluated and compared according to the F1 score, accuracy, recall, and precision values. The proposed method was compared with five different machine and deep learning algorithms. In addition, two hybrid models were used for evaluation, and the results were interpreted.

4.1. Model Parameters and Training Configurations

This section presents the parameters used for the machine learning, deep learning, and hybrid models in this study. Each table provides a structured overview of the models, their configurations, and key hyperparameters.

Table 8 below summarizes the study's machine learning models and their key parameter settings.

Table 8. Machine learning models and parameters.

Model	Parameters
Decision Tree	Maximum depth: 5
Logistic Regression	Run with default parameters
Naïve Bayes	Smoothing: 1.0 (multinomial model)
Random Forest	Maximum depth: 5
Classification and Regression Tree (CART)	Maximum depth: 5

Table 9 presents the deep learning models, their input, hidden, and output layer configurations, and additional settings.

Table 9. Deep learning models and parameters.

Model	Input Layer	Hidden Layer(s)	Output Layer	Additional Settings
CNN	Conv1D, pool size: 2, 32 filters	Dense (128 neurons)	5 neurons, softmax	0.1 dropout, kernel size: 2, ReLU
GRU	64 neurons	150 neurons	5 neurons, softmax	0.1 dropout
LSTM	128 neurons	-	5 neurons, softmax	0.1 dropout
RNN	128 neurons	128 neurons	5 neurons, softmax	0.1 dropout
MLP	-	3 layers (64, 128, 256 neurons)	1 neuron, sigmoid	Adam optimizer, binary cross-entropy loss function

Table 10 outlines the hybrid models used in this study, detailing the combination of deep learning architectures and their respective configurations.

Table 10. Hybrid models and parameters.

Model	Input Layer Details	Output Layer Details
CNN–LSTM	CNN input: 64, 128, 5 neurons, kernel size: 3, 32 filters	LSTM output: 128 neurons, softmax
LSTM–CNN	LSTM input: 128 neurons, 0.1 dropout	CNN output: kernel size: 3, 32 filters, 128 neurons, pool size: 2

Table 11 provides an overview of the training configurations applied to all deep learning models.

Adam dynamically adjusts the learning rate, improving optimization efficiency and model convergence. To prevent overfitting, the epoch count should be increased carefully.

Table 11. Optimization and training settings.

Parameter	Value
Number of Epochs	10
Mini-Batch Size	100
Optimization Algorithm	Adam

4.2. Evaluation Parameters

The most commonly used metrics in the literature—F1 score, accuracy, recall, and precision—were employed to evaluate the classification models' performance. These metrics serve as standard benchmarks for assessing and comparing the effectiveness of different classification algorithms and are widely utilized in various machine learning and deep learning applications [45,46].

These evaluation parameters are derived from the confusion matrix, which consists of four key components:

- True Positive (TP): The number of correctly classified attack instances.
- True Negative (TN): The number of correctly classified normal instances.
- False Positive (FP): Normal instances incorrectly classified as attacks.
- False Negative (FN): Attack instances mistakenly classified as normal.

The evaluation metrics are computed as follows.

Accuracy measures the proportion of correctly classified instances ($TP + TN$) to the total number of instances ($TP + TN + FP + FN$).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision quantifies the proportion of correctly classified attack instances (TP) to the total predicted attack instances ($TP + FP$).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall (sensitivity) measures the proportion of correctly classified attack instances (TP) relative to all actual attack instances ($TP + FN$).

$$\text{Recall} = \frac{TP}{FN + FP} \tag{3}$$

F1 score represents the harmonic mean of precision and recall, providing a balanced measure that accounts for both false positives and false negatives.

$$\text{F1 Score} = \frac{2TP}{2TP + FP + FN} \tag{4}$$

While precision emphasizes the accuracy of positive classifications, recall provides insight into the model's ability to identify actual attack instances correctly. The F1 score serves as a comprehensive measure that balances both precision and recall, making it one of the most widely adopted performance indicators in classification tasks.

5. Results and Comparison

The results of this study were compared across traditional machine learning, deep learning, and hybrid learning algorithms. The dataset was split into 70% training and 30% testing. Commonly accepted hyperparameter values were initially assigned to deep learning algorithms, and fine-tuning was performed to achieve the best results.

Although the dataset contains six different attack types, a binary classification approach was adopted: attack traffic vs. normal traffic. This method was successful due to the high similarity between feature distributions of different attack types. The Adam optimization algorithm accelerated and enhanced the model's convergence to the global minimum. Adam combines momentum and RMSprop methods, ensuring a balanced and efficient optimization process.

Using 42 features from the WUSTL-2021-IIoT dataset, machine learning, deep learning, and hybrid learning models were analyzed. The results are presented in Figure 2, which compares various machine learning and deep learning algorithms regarding accuracy on the WUSTL-IIoT-2021 dataset.

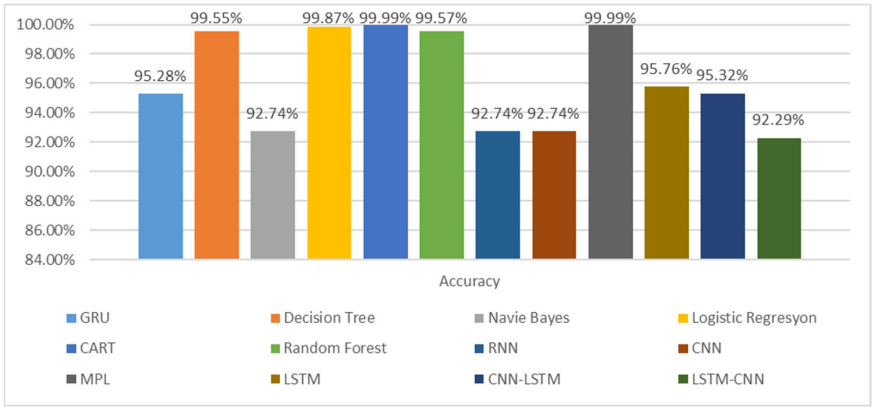

Figure 2. Comparison of accuracy.

MLP achieved the highest accuracy scores among deep learning models and CART and logistic regression among machine learning models, all exceeding 99.87% accuracy. However, the CNN–LSTM hybrid model performed worse than the standalone CNN and LSTM models on the WUSTL-IIoT-2021 dataset. While hybrid methods generally outperform individual deep learning techniques in the literature, standalone models performed better on this dataset, likely due to their unique characteristics.

The lowest-performing models were RNN, CNN, and naïve Bayes, with accuracy of 92.74%. These models struggled to classify DoS traffic and normal traffic instances accurately. A detailed comparison of classification metrics—accuracy, precision, recall, and F1 score—is presented in Figure 3 and Table 12.

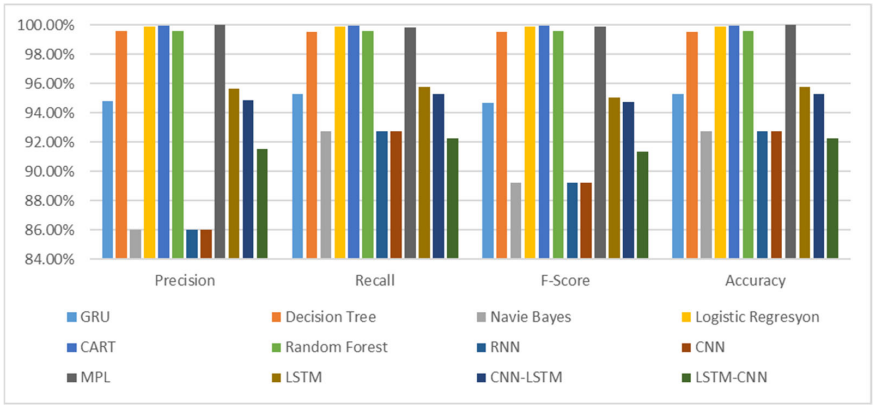

Figure 3. Comparison results according to precision, recall, and F1 score parameters.

The results showed that the LSTM–CNN model had the lowest accuracy, at 92.22%, while the MLP model achieved the highest accuracy, at 99.99%. The highest accuracy values are highlighted in bold. RNN and CNN exhibited the lowest success rates among deep

learning models, whereas CART was the best-performing machine learning model. A comparative analysis of all models is visually presented in Figure 3.

Table 12. Results of accuracy, precision, and F1 score parameters.

Model	Precision	Recall	F1 Score	Accuracy
GRU	0.94826	0.95276	0.94655	0.95276
Decision Tree	0.99576	0.995514	0.995573	0.995514
Naïve Bayes	0.860146	0.927441	0.892527	0.927441
Logistic Regression	0.998691	0.998694	0.998691	0.998694
CART	0.999863	0.999863	0.999863	0.999863
Random Forest	0.995927	0.995717	0.995769	0.995717
RNN	0.86015	0.92744	0.89253	0.92744
CNN	0.86015	0.92744	0.89253	0.92744
MLP	**0.999961**	**0.998081**	**0.999020**	**0.999985**
LSTM	0.95643	0.95763	0.95075	0.95763
CNN–LSTM	0.94848	0.95319	0.94746	0.95319
LSTM–CNN	0.91513	0.92289	0.91328	0.92289

The dataset is better suited for binary classification models. In the dataset, normal traffic is labeled 0, and attack traffic is labeled 1. The most commonly used models for binary classification in the literature include logistic regression, decision tree, random forest, and support vector machine (SVM). Our study examined these models along with additional approaches. While high classification performance was achieved with models suited for binary classification, MLP yielded the highest accuracy.

This is likely due to the heterogeneous nature of attack traffic within the dataset. Although the dataset consists of a simple 0–1 classification structure, the feature variations among different attack types increase complexity. As a result, simpler models such as logistic regression and decision tree struggle with classification. For complex datasets like this, more advanced models should be preferred. Therefore, this study explored binary classification models and various machine learning and deep learning models.

MLP model performance and hyperparameters: The MLP model achieved the highest accuracy (99.99%) using 10 epochs and a batch size of 100. It consists of three hidden layers with 64, 128, and 256 neurons. The output layer has a single neuron and employs the sigmoid activation function. The model was trained using binary cross-entropy as the loss function and Adam as the optimizer. The selected hyperparameters are listed in Table 13.

Table 13. Hyperparameters of MLP.

Hyperparameters	Values
Activation function	ReLU, sigmoid
Number of epochs	20
Units	64-128-256-1
Optimizer	Adam
Loss	Binary cross-entropy
Hidden layer	3
Accuracy	99.999%
Recall	99.998%
Precision	99.999%
F1 score	99.999%

ROC curve and performance evaluation: The receiver operating characteristic (ROC) curve in Figure 4 provides insights into the model's classification performance. The x-axis represents the false-positive rate (FPR), while the y-axis represents the true.positive rate (TPR). A high TPR and low FPR indicate strong model performance. The closer the ROC curve is to the top-left corner, the better the model's accuracy.

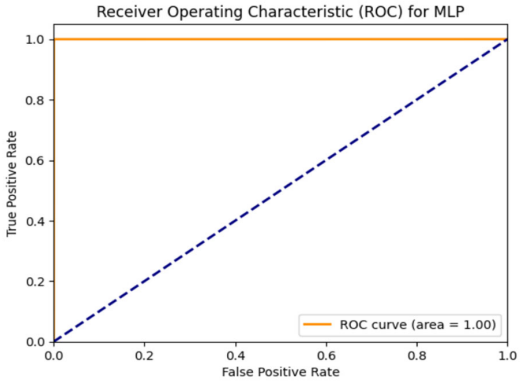

Figure 4. MLP ROC curve.

Recent studies show that the use of traditional methods for attack detection is decreasing and AI-based approaches are becoming more common. Hybrid models are increasingly being developed to minimize false positives. However, in this study, such hybrid approaches showed lower performance. This is thought to be due to the unique feature structure of the dataset.

As seen in Table 1, a comprehensive literature review revealed that this study achieved the highest accuracy on the WUSTL-2021-IIoT dataset using a wide range of models. These results show the importance of choosing the right classification model according to dataset complexity.

6. Conclusions and Future Works

Integrating SCADA systems with IIoT has significantly increased operational efficiency in various industrial sectors. However, this connection has also introduced numerous vulnerabilities related to DoS attacks and IDS. Understanding these threats and their mitigation mechanisms is crucial to maintaining the integrity and reliability of industrial operations. SCADA systems are critical for monitoring and controlling industrial processes, but their increased exposure to external networks has made them susceptible to various cyberthreats.

This study examined the network traffic of a SCADA system built with IIoT devices, and both normal and attack traffic were analyzed. This analysis, which was conducted using machine learning, deep learning, and hybrid learning models, aims to make a significant contribution considering the limited studies in the literature. Experimental studies were conducted in the Google Colab environment using the Apache Spark big data platform, and the performances of the models were evaluated with metrics such as accuracy, precision, recall, and F1 score. The MLP model achieved 99.99% accuracy, the CART model 99.98% accuracy, and the logistic regression model 99.86% accuracy, outperforming other methods in the literature. In this process, hyperparameter adjustments were performed for deep learning algorithms and parameter optimizations were performed for machine learning algorithms. The MLP model had the highest accuracy rate and was trained for 10 epochs using ReLU and softmax activation functions. It has a dropout rate of 0.5 and consists of three hidden layers containing 64, 128, and 256 units. The Adam algorithm was used for model optimization, and categorical cross-entropy was adopted as the loss function.

The proposed model successfully detects security threats in the data streams generated by SCADA systems through AI-driven anomaly detection. Additionally, in-depth analyses of information security, authentication mechanisms, and security technologies are presented, offering a comprehensive framework for protecting industrial control systems. The study reinforces security measures in this critical field by addressing SCADA security challenges within the broader IoT ecosystem.

Future studies can improve model performance by examining hybrid models and applying data-balancing techniques. Class imbalance problems can be addressed by in-

vestigating the effects of over- and undersampling methods. In addition, validating the developed models on different IIoT datasets is important to assess their generalizability. Integrating techniques such as hyperparameter optimization and automatic feature selection can increase the reliability and effectiveness of intrusion detection systems.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The WUSTL-IIOT-2021 dataset was used. Dataset link: https://www.cse.wustl.edu/~jain/iiot2/index.html (accessed 15 September 2024).

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BA	bat algorithm
CNN	convolutional neural network
CART	classification and regression tree
DFIR	digital forensic incident response
DDoS	distributed denial of service
DL	deep learning
DoS	denial of service
FDL	federated deep learning
FN	false negative
FP	false positive
GRU	gated recurrent unit
HBO	harmony search algorithm
ICS	industrial control system
IDS	intrusion detection system
IIoT	industrial internet of things
IoT	internet of things
KNN	k-nearest neighbor
LSTM	long short-term memory
LR	logistic regression
MCC	Matthews correlation coefficient
ML	machine learning
MLP	multilayer perceptron
NB	naïve Bayes
NIDS	network intrusion detection system
PSO	particle swarm optimization
RF	random forest
RNN	recurrent neural network
ROC	receiver operating characteristic
SCADA	supervisory control and data acquisition
SIEM	security information and event management
SMOTE	synthetic minority oversampling technique
SVM	support vector machine
TCP	transmission control protocol
TN	true negative
TP	true positive
TTL	time to live
XAI	explainable artificial intelligence

References

1. Farhan, L.; Shukur, S.T.; Alissa, A.E.; Alrweg, M.; Raza, U.; Kharel, R. A survey on the challenges and opportunities of the Internet of Things (IoT). In Proceedings of the 2017 Eleventh International Conference on Sensing Technology (ICST), Sydney, Australia, 4–6 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5.
2. Manyika, J.; Chui, M.; Bisson, P.; Woetzel, J.; Dobbs, R.; Bughin, J.; Aharon, D. *The Internet of Things: Mapping the Value Beyond the Hype*; McKinsey Global Institute: New York, NY, USA, 2015.
3. Panchal, A.C.; Khadse, V.M.; Mahalle, P.N. Security issues in IIoT: A comprehensive survey of attacks on IIoT and its countermeasures. In Proceedings of the 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), Lonavala, India, 23–24 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 124–130.
4. Sajid, A.; Abbas, H.; Saleem, K. Cloud-Assisted IoT-Based SCADA Systems Security: A Review of the State of the Art and Future Challenges. *IEEE Access* **2016**, *4*, 1375–1384. [CrossRef]
5. Alanazi, M.; Mahmood, A.; Chowdhury, M.J.M. SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues. *Comput. Secur.* **2023**, *125*, 103028. [CrossRef]
6. Tange, K.; De Donno, M.; Fafoutis, X.; Dragoni, N. A Systematic Survey of Industrial Internet of Things Security: Requirements and Fog Computing Opportunities. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 2489–2520. [CrossRef]
7. Zolanvari, M.; Teixeira, M.A.; Gupta, L.; Khan, K.M.; Jain, R. Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things. *IEEE Internet Things J.* **2019**, *6*, 6822–6834. [CrossRef]
8. Zolanvari, M.; Teixeira, M.A.; Jain, R. Effect of imbalanced datasets on security of industrial IoT using machine learning. In Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), Miami, FL, USA, 9–11 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 112–117.
9. Alani, M.M. An explainable efficient flow-based Industrial IoT intrusion detection system. *Comput. Electr. Eng.* **2023**, *108*, 108732. [CrossRef]
10. Dener, M.; Al, S.; Orman, A. STLGBM-DDS: An Efficient Data Balanced DoS Detection System for Wireless Sensor Networks on Big Data Environment. *IEEE Access* **2022**, *10*, 92931–92945. [CrossRef]
11. Okur, C.; Orman, A.; Dener, M. DDOS intrusion detection with machine learning models: N-BaIoT data set. In Proceedings of the International Conference on Artificial Intelligence and Applied Mathematics in Engineering, Baku, Azerbaijan, 20–22 May 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 607–619.
12. Dener, M.; Ok, G.; Orman, A. Malware Detection Using Memory Analysis Data in Big Data Environment. *Appl. Sci.* **2022**, *12*, 8604. [CrossRef]
13. Yan, K.; Wang, X.; Du, Y.; Jin, N.; Huang, H.; Zhou, H. Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy. *Energies* **2018**, *11*, 3089. [CrossRef]
14. Konatham, B.; Simra, T.; Amsaad, F.; Ibrahem, M.I.; Jhanjhi, N.Z. A Secure Hybrid Deep Learning Technique for Anomaly Detection in IIoT Edge Computing. *TechRxiv* **2024**. [CrossRef]
15. Marzouk, R.; Alrowais, F.; Negm, N.; Alkhonaini, M.A.; Hamza, M.A.; Rizwanullah, M.; Yaseen, I.; Motwakel, A. Hybrid Deep Learning Enabled Intrusion Detection in Clustered IIoT Environment. *CMC-Comput. Mater. Contin.* **2022**, *72*, 3763–3775. [CrossRef]
16. Shaikh, J.; Butt, Y.A.; Naqvi, H.F. Effective Intrusion Detection System Using Deep Learning for DDoS Attacks. *Asian Bull. Big Data Manag.* **2024**, *4*, 10-62019. [CrossRef]
17. Anbalagan, E.; Rao, D.P.S.V.S.; Alluri, D.A.; Nageswari, D.D.; Kalaivani, D.R. Improving Intrusion Detection using Satin Bowerbird Optimization with Deep Learning Model for IIoT Environment. *Int. J. Electr. Electron. Res.* **2024**, *12*, 219–227. [CrossRef]
18. Huang, J.; Chen, Z.; Liu, S.Z.; Zhang, H.; Long, H.X. Improved Intrusion Detection Based on Hybrid Deep Learning Models and Federated Learning. *Sensors* **2024**, *24*, 4002. [CrossRef] [PubMed]
19. Begum, S.K.S.; Pulikonda Venkata Yoga, A.; Kadavakollu Lakshmi, C.; Mandalapu, G. Ejection of Real-Time Malicious Intrusion and Attacks in IoT Empowered Infrastructure. *Int. Res. J. Adv. Eng. Hub (IRJAEH)* **2024**, *2*, 2456–2462. [CrossRef]
20. Javeed, D.; Gao, T.; Khan, M.T.; Shoukat, D. A Hybrid Intelligent Framework to Combat Sophisticated Threats in Secure Industries. *Sensors* **2022**, *22*, 1582. [CrossRef]
21. Sengupta, J.; Ruj, S.; Das Bit, S. A Comprehensive Survey on Attacks, Security Issues and Blockchain Solutions for IoT and IIoT. *J. Netw. Comput. Appl.* **2020**, *149*, 102481. [CrossRef]
22. Mohammadi, M.; Rashid, T.A.; Karim, S.H.T.; Aldalwie, A.H.M.; Tho, Q.T.; Bidaki, M.; Rahmani, A.M.; Hosseinzadeh, M. A comprehensive survey and taxonomy of the SVM-based intrusion detection systems. *J. Netw. Comput. Appl.* **2021**, *178*, 102983. [CrossRef]
23. Liu, Y.S.; Li, W.X.; Dong, X.W.; Ren, Z. Resilient Formation Tracking for Networked Swarm Systems Under Malicious Data Deception Attacks. *Int. J. Robust Nonlinear Control.* **2024**, *35*, 2043–2052. [CrossRef]
24. Sun, J.R.; Huang, C.T.; Hwang, M.S. A SYN flooding attack detection approach with hierarchical policies based on self-information. *ETRI J.* **2022**, *44*, 346–354. [CrossRef]

25. Mudassir, M.; Unal, D.; Hammoudeh, M.; Azzedin, F. Detection of Botnet Attacks against Industrial IoT Systems by Multilayer Deep Learning Approaches. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 2845446. [CrossRef]
26. Su, J.; Jiang, M.N. A Hybrid Entropy and Blockchain Approach for Network Security Defense in SDN-Based IIoT. *Chin. J. Electron.* **2023**, *32*, 531–541. [CrossRef]
27. Alsaedi, A.; Moustafa, N.; Tari, Z.; Mahmood, A.; Anwar, A. TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems. *IEEE Access* **2020**, *8*, 165130–165150. [CrossRef]
28. Yi, J.K.; Guo, L. AHP-Based Network Security Situation Assessment for Industrial Internet of Things. *Electronics* **2023**, *12*, 3458. [CrossRef]
29. Zolanvari, M.; Teixeira, M.A.; Gupta, L.; Khan, K.M.; Jain, R. WUSTL-IIOT-2021 Dataset for IIoT Cybersecurity Research. Available online: https://www.cse.wustl.edu/~jain/iiot2/index.html (accessed on 16 January 2025).
30. Alani, M.M.; Damiani, E.; Ghosh, U. DeepIIoT: An explainable deep learning based intrusion detection system for industrial IOT. In Proceedings of the 2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW), Bologna, Italy, 10 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 169–174.
31. Mohy-Eddine, M.; Guezzaz, A.; Benkirane, S.; Azrour, M. An effective intrusion detection approach based on ensemble learning for IIoT edge computing. *J. Comput. Virol. Hacking Tech.* **2023**, *19*, 469–481. [CrossRef]
32. Babayigit, B.; Abubaker, M. Towards a generalized hybrid deep learning model with optimized hyperparameters for malicious traffic detection in the Industrial Internet of Things. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107515. [CrossRef]
33. Eid, A.M.; Soudan, B.; Nasif, A.B.; Injadat, M. Comparative study of ML models for IIoT intrusion detection: Impact of data preprocessing and balancing. *Neural Comput. Appl.* **2024**, *36*, 6955–6972. [CrossRef]
34. Popoola, S.I.; Imoize, A.L.; Hammoudeh, M.; Adebisi, B.; Jogunola, O.; Aibinu, A.M. Federated deep learning for intrusion detection in consumer-centric Internet of Things. *IEEE Trans. Consum. Electron.* **2023**, *70*, 1610–1622. [CrossRef]
35. Alzahrani, A.; Aldhyani, T.H.H. Design of Efficient Based Artificial Intelligence Approaches for Sustainable of Cyber Security in Smart Industrial Control System. *Sustainability* **2023**, *15*, 8076. [CrossRef]
36. Dina, A.; Siddique, A.; Manivannan, D. A Deep Learning Approach for Intrusion Detection in Internet of Things Using Focal Loss Function. *Internet Things* **2023**, *22*, 100699. [CrossRef]
37. Diaba, S.Y.; Shafie-Khah, M.; Mekkanen, M.; Vartiainen, T.; Elmusrati, M. Risk Accessment of Machine Learning Algorithms on Manipulated Dataset in Power Systems. In Proceedings of the 2023 International Conference on Future Energy Solutions (FES), Vaasa, Finland, 12–14 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
38. Xu, B.Y.; Sun, L.; Mao, X.Q.; Ding, R.Y.; Liu, C.W. IoT Intrusion Detection System Based on Machine Learning. *Electronics* **2023**, *12*, 4289. [CrossRef]
39. Ahakonye, L.A.C.; Nwakanma, C.I.; Lee, J.-M.; Kim, D.-S. Trees Bootstrap Aggregation for Detection and Characterization of IoT-SCADA Network Traffic. *IEEE Trans. Ind. Inform.* **2024**, *20*, 5217–5228. [CrossRef]
40. Bekbulatova, V.; Morichetta, A.; Dustdar, S. FL-SERENADE: Federated Learning for SEmi-supeRvisEd Network Anomaly DEtection. A Case Study. In Proceedings of the 2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Abu Dhabi, United Arab Emirates, 14–17 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1072–1079.
41. Gaber, T.; Awotunde, J.B.; Folorunso, S.O.; Ajagbe, S.A.; Eldesouky, E. Industrial internet of things intrusion detection method using machine learning and optimization techniques. *Wirel. Commun. Mob. Comput.* **2023**, *2023*, 3939895. [CrossRef]
42. Casajús-Setién, J.; Bielza, C.; Larrañaga, P. Anomaly-based intrusion detection in iiot networks using transformer models. In Proceedings of the 2023 IEEE International Conference on Cyber Security and Resilience (CSR), Venice, Italy, 31 July–2 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 72–77.
43. Ye, Z.; Luo, J.; Zhou, W.; Wang, M.; He, Q. An ensemble framework with improved hybrid breeding optimization-based feature selection for intrusion detection. *Future Gener. Comput. Syst.* **2024**, *151*, 124–136. [CrossRef]
44. Saxena, A.; Mittal, S. Advanced Persistent Threat Datasets for Industrial IoT: A Survey. In Proceedings of the 2023 Second International Conference on Informatics (ICI), Noida, India, 23–25 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
45. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
46. He, H.; Ma, Y. *Imbalanced Learning: Foundations, Algorithms, and Applications*; Wiley: Hoboken, NJ, USA, 2013.