IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 April 2024, accepted 19 July 2024, date of publication 22 July 2024, date of current version 1 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3432454

RESEARCH ARTICLE

A Comparative Study of Lightweight Machine Learning Techniques for Cyber-Attacks Detection in Blockchain-Enabled Industrial Supply Chain

SHEREEN ISMAIL[1], (Member, IEEE), SALAH DANDAN[2], DIANA W. DAWOUD[3], (Senior Member, IEEE), AND HASSAN REZA[2]

[1]Merit Network Inc., University of Michigan, Ann Arbor, MI 48108, USA
[2]School of Electrical Engineering and Computer Science, University of North Dakota, Grand Forks, ND 58202, USA
[3]College of Engineering and Information Technology, University of Dubai, Dubai, United Arab Emirates

Corresponding author: Shereen Ismail (drismail@umich.edu)

ABSTRACT The security of Industrial Supply Chain (ISC) has emerged through the integration of Industrial Internet of Things (IIoT) and Blockchain (BC) technology. This new era involves effectively protecting IIoT systems from various threats and ensuring their smooth operation and resilience against potential cyber-attacks. Within the ISC ecosystem, combining machine learning (ML)-based security models for cyber-attack detection can play a crucial role in enhancing the ISC security and proactively identifying potential threats. This paper presents a BC-enabled ISC that embed ML security model integrated within a multi-layered approach. We conducted a comparative study and performance analysis of several ML classification techniques, with a focus on supervised methods to identify the lightweight model for cyber-attack detection suitable for deployment in resource-constrained IIoT environment. We investigate the performance of Gaussian Naive Bayes (NB), K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), and three ensemble techniques, namely Bagging, Stacking, and Boosting. The study employs the WUSTL-IIOT-2021 imbalanced dataset, which contains samples representing four types of attacks, including denial of service (DoS), SQL injection, reconnaissance, and backdoor. The paper addresses the imbalance in class representation by customizing the dataset for training and testing the ML models. Both Mutual Information (MI) and Extra-trees (ET) are applied as a one-stage ensemble feature selection. The performance of the ML models are investigated using classification accuracy (Acc), precision, recall, F1 score, Matthews correlation coefficient (MCC), model size (Mem), training time (TT) and prediction time (PT).

INDEX TERMS Industrial supply chain, industrial Internet of Things, security, cyber-attacks, detection, machine learning, blockchain, lightweight.

I. INTRODUCTION

The Industrial Supply Chain (ISC) is a complex network of organizations, resources, and activities involved in the production and distribution of goods and services in the industrial sector. It encompasses all steps from the procurement of raw materials, manufacturing, distribution, and up to the delivery of finished products to customers (fig. 1). The adoption of the Industrial Internet of Things (IIoT) in the Supply Chain (SC) has the potential to drive significant growth and transformation in various industries. In 2020, the global IIoT market in the SC was valued at over $13 billion, and it was projected to grow at a Compound Annual Growth Rate (CAGR) of around 11% from 2021 to 2027 [1]. However, the

The associate editor coordinating the review of this manuscript and approving it for publication was Mueen Uddin.

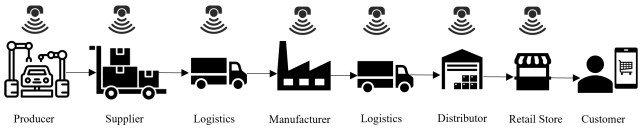


FIGURE 1. Generic ISC components.

meteoric rise of IIoT devices in the SC industry has caused concern. These systems face security and privacy issues, along with various kinds of cyber-attacks that can have severe consequences. The growing sophistication of cyber threats becomes a significant concern for ISC systems that affect security, privacy, control, availability, and reliability [2], [3]. Examples of attacks incidents, documented in the literature, include Stuxnet in 2010, NotPetya in 2017, SolarWinds in 2020, and the Colonial Pipeline Ransomware Attack in 2021 [2], [4], [5]. These attacks have the potential to disrupt critical infrastructure, compromise sensitive data, and have severe economic and security implications.

Blockchain (BC) technology has emerged to effectively protect IIoT-ISC integrated systems from various threats and ensure their smooth operation and resilience against potential cyber-attacks. BC has a great applicability in SC to provide end-to-end traceability by automatically keep updating the data transaction records when a change is made along the overall SC network. In this way, stakeholders are able to track and trace the entire lifecycle of a product. This transparency helps identify any anomalies or suspicious activities in the ISC network.

Data science-enabled technologies associated with Industry 4.0 in IIoT devices and smart sensors facilitates real-time monitoring, data collection, and process optimization. IIoT involves the use of interconnected sensors, actuators, devices, and systems in industrial settings that can be used to collect network traffic logs, system event logs, and other relevant data streams. IIoT provides visibility and transparency, which can be leveraged for early threat detection and mitigation in SC. Data science is an integral part of modern ISC systems, applied to sourcing, storing, cleaning, and analyzing vast amounts of raw data mainly real-time and decision-making. Cyber-security data science focuses on the analysis of collected cyber-security data and used by Machine learning (ML) to build data-driven security models able to identify cyber-attacks which help organizations respond effectively to these threats [6], ultimately safeguarding the integrity and reliability of the SC [7].

ML techniques have been investigated in numerous studies to comprehend the typical behaviour of sensors and IIoT devices, labeling any deviation from this norm as suspicious [8]. Despite its wide use, ML continues to be an active area of research, necessitating ongoing investigation and performance analysis of ML as one of the efficient solutions to secure SC systems against cyber-attacks [3], particularly when integrated with resource-constrained devices in IIoT taking into consideration that the datasets used to develop the ML models for attack detection in IIoT should be collected in real world scenarios. We focus on the issues associated with developing ML classifiers using such datasets which are expected to have small amount of abnormal data samples compared to normal ones.

Several comprehensive datasets are available in the literature to develop and validate ML models for IoT/IIoT security, including Edge-IIoTset [9], DS2OS, UNSW-NB15 [10], Bot-IoT, X-IIoTID [11], and LATAM-DDOS-IOT [12]. However, we selected WUSTL-IIoT-2021 and its predecessor WUSTL-IIOT-2018 which have been collected through emulating real-world industrial systems.

In one aspect, this paper presents architectural security solution for a BC-enabled ISC system that embed a lightweight ML model in a ML security engine. While a second aspect of this paper is conducting a comparative study and performance analysis of various ML supervised techniques: Gaussian Naive Bayes (NB), K-Nearest Neighbors (KNN), Random Forest (RF), and Decision Trees (DT) vs. three ensemble algorithms: Bagging, Stacking, and Boosting to detect cyber-attacks targeting IIoT in ISC ecosystem. The aim is to pinpoint lightweight ML model that strikes a balance between efficiency and resource usage, making it well-suited for deployment in resource-constrained IIoT. We apply lightweight methodology in terms of pre-processing and feature selection. We use the WUSTL-IIOT-2021 dataset, which contains samples corresponding to four kinds of attacks, including Denial of Service (DoS), Reconnaissance, Command injection, and Backdoors. This dataset suffers from a lack of data samples for some attack classes, in addition to a large imbalance towards normal samples which have more prevalence in the dataset. Therefore, we customize the dataset for training and testing the ML models for better class representation. The evaluation is performed in terms of classification accuracy (Acc), Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), as well as Model Size (Mem), Training Time (TT), and Prediction Time (PT) metrics.

A. CASE STUDY: A PROPOSED BC-ENABLED ISC SECURITY ARCHITECTURE

Recent years have witnessed the rapid development of the IIoT systems integrated with ISC. Meanwhile, the incorporation of BC technology has emerged as a promising solution for secure identification and authentication of IIoT devices [13]. BC technology provides a secure framework for storing, processing, and sharing data obtained from these devices [14]. By using a distributed, decentralized, and shared ledger, BC offers a viable approach to address security risks inherent in IIoT [15]. In this sense, IIoT and BC can facilitate the realization of secure ISC, particularly at the intersection of data analysis and ML for enhancing the attack detection capabilities [16].

This section presents a multi-layered security architecture for the BC-enabled ISC which consists of the following

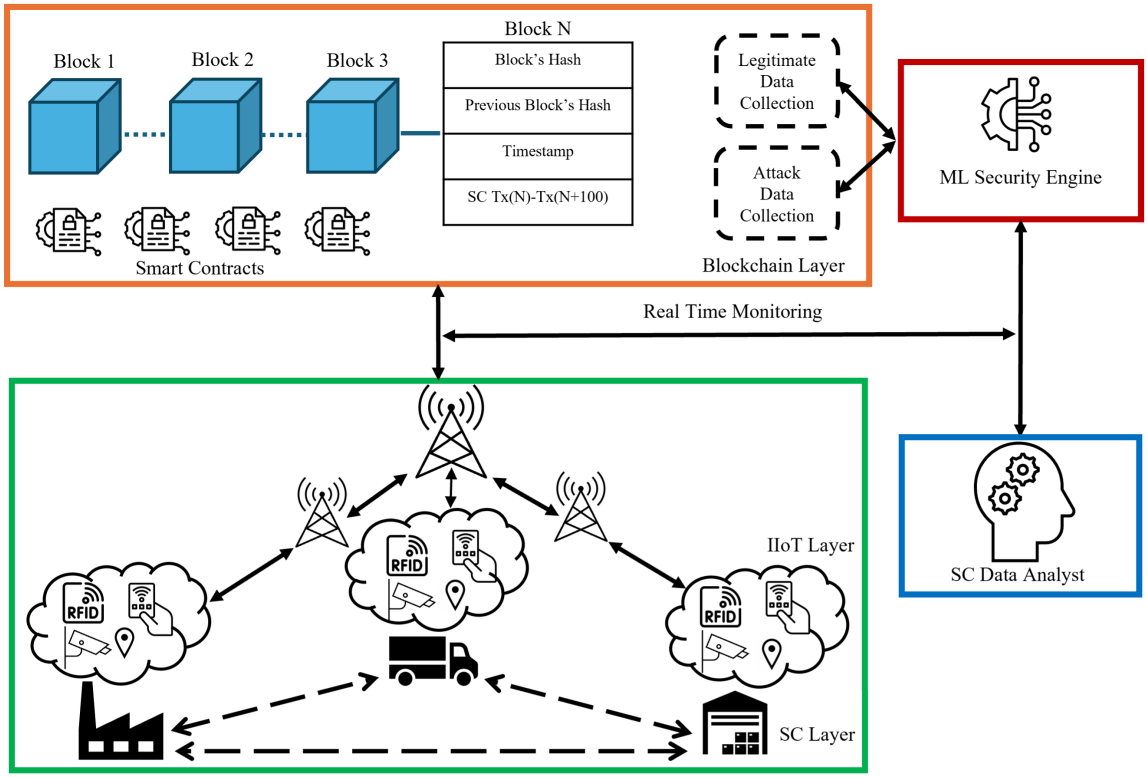


FIGURE 2. BC-enabled ISC integrated with ML security engine.

components: SC layer, IoT layer, BC layer, SC data analyst, and ML security engine. We describe the architecture and discuss the interaction among its different components as depicted in fig. 2.

At the SC layer, the process of manufacturing, transportation, and distribution is monitored, forming the backbone of the entire system. The IIoT layer incorporates sensors and smart devices embedded throughout the SC different stages, facilitating real-time data collection and communication. These devices relay critical information on the status and conditions of goods, equipment, and facilities holds significant potential for deriving useful behavioral patterns of nodes to enhance the detection of attacks.

The BC layer adds an additional dimension to security by providing an immutable and transparent ledger. It ensures the SC data integrity, creating a decentralized platform that resists data tampering and enhances traceability. Every transaction within the SC is securely recorded on the BC, reducing the risk of data manipulation or unauthorized access and smart contracts receive all BC transactions as function calls, generating activities, and facilitating access for transaction-involved parties to exchange control track and receive alerts in the event of a violation.

SC Data Analyst has a core role and acts as the central intelligence hub. Linked to IIoT layer for real-time data feeds, BC layer for secure and transparent transaction records, and ML security engine. SC Data Analyst is equipped to continuously monitor, actively aggregate, and analyze the incoming traffic from the different network parities, working with the ML security engine to identify any malicious activities and promptly initiates necessary actions to mitigate their effects.

B. RELATED WORK

Significant research efforts are apparent in securing Internet of Things (IoT), with a focus on industrial applications such as manufacturing, logistics and SC management. This emphasis is highlighted by multiple surveys, including [17], [18], [19], [20], [21], [22], [23], [24], exploring potential techniques and solutions to improve the security of IIoT applications in these specific industrial environments by considering directions include employing ML, BC, and integrated ML-BC techniques as shown in table 1.

Starting with the work of [25] that identifies and addresses the specific challenges of applying ML to secure IIoT. A particular emphasis is placed on the reliance on commercially available or publicly accessible datasets that may not specifically represent the challenges of IIoT, and the imbalanced dataset problem, a critical issue where the number of minority samples (indicative of attacks) is much lower than the majority class (normal behavior). To bridge this gap, the paper conducts real cyber-attacks, gathering a dataset with both normal and attack traffic.

This is achieved through a testbed that emulates an IIoT control system used in industrial reservoirs, featuring a water level and turbidity monitoring system. However, the choice of attack tools, as the study employs Kali Linux for command injection attacks, may not fully represent the variety of methods that attackers might employ in real-world scenarios. Subsequently, the study evaluates the efficacy of Artificial Neural Network (ANN) on this imbalanced IIoT dataset. The problem of an imbalanced dataset has also been addressed in [26], in addition to the lack of explainability in Intrusion Detection Systems (IDSs) within the context of IIoT. The authors emphasize the unique challenges faced by IIoT, such as low vendor commitment to security, human factors, human safety concerns, compliance and regulations, data storage complexities, industrial sabotage possibilities, and real-time process control vulnerabilities. The proposed system in [26] aims to fill these gaps by introducing an efficient and interpretable IDS for IIoT. It utilizes Recursive Feature Elimination (RFE) to condense the original 48-feature IIoT traffic dataset (WUSTL-IIOT-2021) into an 11-feature version. While specifically training on IIoT data the classifiers, including RF, Logistic Regression (LR), DT, and NB, using Shapley additive explanation (SHAP) to enhance IDS interpretability. However, a potential concern lies in the experimental environment's mismatch with the actual memory and processing power available in IIoT devices, which may impact the generalizability of efficiency results to real-world scenarios. In [27], an IDS is presented that employs the Particle Swarm Optimization (PSO) and Bat algorithm (BA) for feature selection, in conjunction with the RF classifier to identify and classify malicious behaviors in IIoT-based network traffic. It is crucial to consider that the efficacy of the proposed model, particularly its reliance on BA for feature selection, may exhibit variability across diverse datasets and scenarios of cyber-attacks. Emphasizing on the need for transparency and explainability in decisions made by Artificial Intelligence (AI) models, the authors of [28] proposed a model named Transparency Relying Upon Statistical Theory (TRUST), designed to be a universal Explainable AI (XAI) solution. The TRUST employs factor analysis to transform input features into latent variables, utilizing Mutual Information (MI) to rank and select the most influential ones, referred to as representatives of the classes. The model employs multi-modal Gaussian distributions to determine the likelihood of a new sample belonging to each class. The results of the IIoT case study showcase TRUST's capability to provide explanations for new random samples with an average success rate of 98%. In a comparative analysis with Local Interpretable Model-agnostic Explanations (LIME), a popular XAI model, TRUST exhibits superiority in terms of performance, speed, and the methodology of explainability. It is important to note, however, that while the assumption of mutual independence among representatives simplifies the model, it may not always accurately capture the complexities of real-world scenarios. Another strategy to mitigate the challenge of unbalanced datasets is investigated in [29]. This approach involves employing standardized and normalized numerical features, in conjunction with one-hot encoded categorical features, to improve the overall efficacy of the considered ML models, namely RF, DT, KNN, Support Vector Machine (SVM), LR, and NB, thus improving their accuracy in predictions. However, it should be noted that the article lacks a comprehensive discussion of evaluation metrics such as Precision, Recall, and the F1-score, focusing primarily on MCC. In [9], the authors generated a dataset, named Edge-IIoTset, which is intended for use in ML-based IDSs and supports two modes of operation: centralized and federated learning. Then, the authors proceeded to conduct a thorough primary exploratory data analysis, that involved analyzing the performance of DT, RF, SVM, KNN, and Deep Neural Network (DNN), with a focus on their efficacy when applied to the Edge-IIoTset dataset. The evaluation is conducted in both centralized and federated learning modes, providing insights into the effectiveness of different models in handling cyber-security challenges in IoT and IIoT environments. While the paper highlights the superiority of federated learning in terms of performance enhancement, it acknowledges that achieving higher accuracies may necessitate multiple rounds for convergence. a more comprehensive comparative study on the effectiveness of traditional ML algorithms, such as DT, RF, KNN, NB, LR, and SVM, sheds light on their individual performances in predicting cyber-attacks within the context of Industry 4.0.

Integrating BC and ML techniques to enhance security capabilities in IIoT has been addressed in [30], [31], and [32]. The study presented in [30] offered a comprehensive cyber-security approach for IIoT networks. Emphasis is placed on the imperative requirement for robust controls to safeguard the integrity of vital information disseminated within these networks. The paper substantiates its claims by presenting empirical evidence that highlights the effectiveness of these proposed solutions compared to traditional methods. Nevertheless, the performance of the proposed model may encounter limitations in scenarios that involve new or previously unseen attack patterns. A comparable methodology has been explored in [31], introducing a layered architecture that integrates BC and ML for IIoT applications in smart manufacturing. The architecture consists of five layers: sensing, network, transport with BC components, application, and advanced services including BC data, ML model, and cloud. Using BC for the acquisition of sensor access control data, while the system employs ML for proficient attack detection, distinguishing a variety of malicious activities. It's noteworthy, however, that the study relies on the categorization of dataset attributes into six groups, potentially oversimplifying the representation of the IIoT environment. While in [32], the integration of BC with IIoT, a lightweight and decentralized BC architecture, a specialized BC service layer, and the application of Proof of Authentication (PoAh) in the main BC network is proposed.

The evaluation metrics include consensus algorithm performance, resource utilization, energy efficiency, and service execution time. The practical implementation of the proposed framework in a real-world industrial setting, specifically a fruit processing plant. Though the paper emphasizes the advantage of a fee-free transaction processing model in the proposed architecture, a more comprehensive examination of potential challenges or trade-offs concerning network sustainability and participant motivation would enhance the overall clarity of the proposed architecture.

It is evident from the discussion presented in this section that the growth of attack detection solutions with a specific focus on SC systems, as discussed in [33] and [34], is limited. The contribution of this paper lies in proposing an approach that integrates cyber threat intelligence (CTI) processes with ML techniques to enhance cyber threat predictive analytics. Known attacks are used through CTI techniques to gather threat intelligence, while ML techniques are used to learn from the dataset and predict unknown cyber threats in SC systems. Clearly, the effectiveness of the proposed model may be constrained when dealing with evolving cyber threats that differ significantly from known attack patterns. On the other hand, the work in [34] proposed automated ML framework streamlines various processes, including data processing, model construction, hyperparameter optimization, and inference deployment. While demonstrating commendable results, it's essential to note a potential limitation in the use of categorical data. The need for encoding these data types into numerical values adds a pre-processing step, which may introduce complexities and influence the subsequent stages of the modeling process. The presented results highlight the impact of factors such as sampling method, encoding categorical values, feature selection, and hyperparameter optimization on the performance of ML methods.

This study selects a lightweight ML model that achieves a balance between efficiency and resource usage to be embedded in the ML security engine integrated with the proposed BC-enabled ISC system depicted in fig. 4. The selection of the lightweight ML model follows a comprehensive comparative study and performance analysis of various ML supervised techniques. These models are assessed for their effectiveness in detecting cyber-attacks targeting IIoT in the ISC ecosystem, especially uncommon ones such as Command Injection and Backdoor. We employ a lightweight methodology in terms of pre-processing and feature selection. We consider the WUSTL-IIOT-2021 dataset, which is highly imbalanced and reflects the characteristics of real IIoT systems. Therefore, we customize the dataset for training and testing the ML models for better class representation.

II. ML METHODOLOGY

Figure 3 represents the methodology followed in this study to investigate the ML models performance which includes dataset selection, data pre-processing, feature selection, ML algorithms, and performance evaluation metrics.

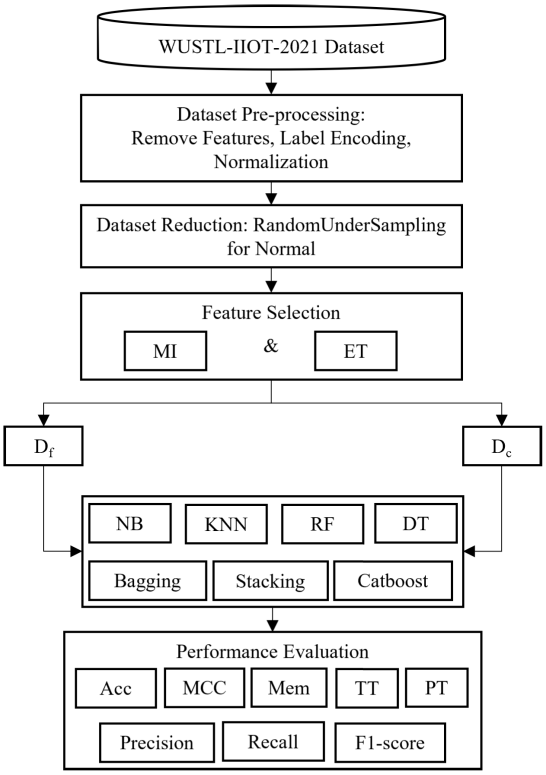


FIGURE 3. ML methodology.

A. DATASET DESCRIPTION

The WUSTL-IIoT-2021 dataset contains data samples designed for cyber-security research and constructed using IIoT network test-bed rather than general IoT devices to simulate several cyber-attack scenarios, as detailed in [36]. This work focuses on four kinds of cyber-attacks: DoS, Reconnaissance, Command Injection, and Backdoor, which can be defined as follows:

- DoS attack which can disrupt operations without much knowledge or effort by overwhelming critical systems or networks, causing downtime [37]. This can lead to production delays, order fulfillment issues, and financial losses.
- Reconnaissance attack involves gathering information about a target system or network to identify vulnerabilities, including function code scan, address scan, points scan, and device identification attack [38]. In the ISC, attackers might use reconnaissance to scout for weaknesses in manufacturing processes or logistics systems.
- Command Injection attack can be particularly dangerous in ISCs, where malicious actors inject unauthorized commands into software systems. This can lead to unauthorized control over critical processes or equipment.
- Backdoors attack is a significant security threat comes in the form of hidden or unauthorized access points which can allow attackers to manipulate with the

TABLE 1. Related work of ML and BC for the security of IIoT.

Ref.	BC	ML Model	Dataset	Contribution	Limitations
[25]		ANN	Simulated	Examined the efficiency of ANN in detecting anomalies under different dataset imbalance ratios.	Using Kali Linux for command injection attacks. This limitation in the choice of attack tools may not fully represent the variety of methods that attackers might employ in real-world scenarios.
[26]		RF LR DT NB	WUSTL-IIOT-2021	Utilized RFE to create a reduced set of features for improved efficiency and interpretability.	The potential mismatch between the experimental implementation environment and the actual memory and processing power available in IIoT devices. This could impact the generalizability of the efficiency results to real-world scenarios.
[27]		RF	WUSTL-IIoT-2021	Utilized PSO and BA for selecting relevant features from the IoT-based network traffic data.	The proposed model relies on BA for feature selection, and its effectiveness may vary across different datasets and attack scenarios.
[28]		ANN	WUSTL-IIoT	Proposed TRUST XAIas a model-agnostic, high-performing, and suitable XAI model for numerical applications. It provides transparency without compromising the selection or performance of the primary AI.	The assumption of mutual independence among representatives simplifies the model but may not always reflect real-world scenarios.
[9]		DT RF SVM KNN DNN	Edge-IIoTset	Provided a primary exploratory data analysis of the Edge-IIoTset dataset, where the performance of traditional ML and DL techniques is evaluated in centralized and federated learning modes.	While federated learning can lead to performance enhancement, it may require several rounds to converge to higher accuracies.
[35]		DT RF KNN NB LR SVM		Utilized ML algorithms to detect cyber attacks in Industry 4.0, specifically focusing on the performance of different algorithms in predicting attacks.	The paper does not provide details about the dataset used for training the ML algorithms, which could impact the generalizability of the results.
[30]	✓	KNN	UNSWNB15	Proposed a system that relies on three layers: Cloud, Edge, and Network. The collector node, positioned at the Edge layer, serves as the central point for communications from the edge, and designated as the execution site for both ML and BC algorithms.	The model's performance may be limited in scenarios with novel or previously unseen attack patterns.
[31]	✓	ANN DT RF NB AdaBoost SVM	TON_IoT	Introduced a five-layered architecture for IIoT applications, encompassing sensing, network/protocol, transport enforced with BC, application, and advanced services layers.	The categorization of dataset attributes into six groups may oversimplify the representation of the IIoT network.
[32]	✓			Introduced a BC-based architecture that performs five industrial operations: user registration, device registration, sensor data recording, data acquisition, and actuation.	While the paper acknowledges the absence of transaction processing fees, a more thorough discussion of potential trade-offs related to network sustainability and participant incentivization would enhance the clarity of the proposed architecture.
[33]		XGBoost SGB GBoost	Kaggle	Proposed an integrated approach that combines CTI processes with boosting ensemble learning, specifically using XGBoosting and Stochastic Gradient Boosting algorithms, to predict and analyze cyber threats.	Limited when addressing dynamic cyber threats that substantially deviate from established attack patterns.
[34]		Multiple	Multiple	Presented empirical validation on the practical understanding of how ML can be effectively utilized in addressing real-world supply chain challenges.	The proposed automated machine learning framework is sensitive to hyperparameter settings. Finding the optimal hyperparameter configuration requires extensive experimentation and computational resources, which can be a limiting factor in practical, resource-constrained scenarios.

IIoT infrastructure.; bypassing security controls and authentication procedures [36].

B. DATA PRE-PROCESSING

We used the original imbalanced WUSTL-IIOT-2021 dataset [39] that consists of 1,194,464 sample observations and 48 feature as shown in table 2 to extract a representative customized dataset for our simulations. For WUSTL-IIOT-2021 dataset, we deal with less than 8% attacks samples, which is closer to real world scenarios for ISC system where data samples are typically collected over an extended period of time from numerous sensors and IoT devices with varying sampling frequencies, producing high dimensional datasets.

Data pre-processing improves dataset usability. It involves eliminating redundant and unnecessary features, applying label encoding, and normalizing the data through scaling [40]. We start with the removal of 'StartTime', 'LastTime', 'SrcAddr', 'DstAddr', 'sIpId', 'dIpId' features. We apply label encoding to convert the string values into numerical format.

C. FEATURE SELECTION

The processing time of ML algorithms typically increases as the number of features, number of samples, dependencies among features, types of features, and nested feature categories increase. Concisely, feature selection is a method for eliminating irrelevant and redundant information to the greatest extent possible while retaining the most valuable discriminating features that significantly contribute to the detection process. This process reduces the data's dimensionality, enabling ML algorithms to converge fastly.

We apply MI and Extra-trees (ET) as a one-stage ensemble feature selection. MI is used to measure the dependencies between the output class and the input variables while ET which utilizes multiple randomized DTs, making it less sensitive to noise and irrelevant feature and assign an importance score to each feature. Both are used to rank and select the most relevant features for data classification.

The considered dataset is highly imbalanced with significantly low number of minority samples. This problem facing ML algorithms and represent a real barrier in which sometimes the trained models may not be able to detect the attack and how different the evaluation metrics would react to this problem [25], [41].

To have a represented dataset, We start by applying RandomUnderSampler to reduce the size of normal samples from 1,107,488 to 150,000. We calculate the *ClassWeight* metric before and after applying RandomUnderSampler, which can be represented as follows:

$$ClassWeight = \frac{N_{m_a}}{N_{m_i}} \tag{1}$$

where N_{m_a} represents the number of samples in majority class and N_{m_i} represents the number of samples in minority classes. for the original datatset, *ClassWeight* is 12.72 while after applying RandomUnderSampler, *ClassWeight* becomes 1.7 which is more than 7 times less. Assigning class weights where N_{m_i} considers only unrepresented classes results in generation of unique testing subsets for performance evaluation using imbalanced dataset.

Generally, a test set is used to evaluate the performance of the model fitted to the training set. Furthermore, to have a more accurate prediction specifically in terms of low represented classes, we prepare two distinct testing sets to evaluate the ML models: D_f which contains samples of all classes of DoS, Reconnaissance, Command injection, and Backdoors attacks as well as the reduced number of normal samples, while D_c exclusively comprises the samples of the severely unrepresented classes: Command injection and Backdoors, while excluding the samples of DoS and Reconnaissance attacks. Train-test split is necessary to segment data into subsets for model training and evaluation [42]. We adopt a 70:30 train-test splitting ratio to assess the the ML models performance, a ratio recommended by numerous studies for expediting model fitting and preserving optimal predictions [43].

D. MACHINE LEARNING ALGORITHMS

We investigate the performance of traditional ML supervised techniques: NB, KNN, RF, and DT vs. three ensemble algorithms: Bagging with a DT base algorithm, Stacking with level 0 estimators: LR, Nearest Centroid (NC) and LightGBM, and level 1 estimator LR, and Catboost selected from the boosting family aiming to pinout the lightweight model that achieves a balance between efficiency and resource utilization for effectively detecting attacks in ISC. NB is known for its simplicity and fast training, KNN excels in locally adapting to data patterns, RF and DT offer versatility and interpretability. Meanwhile, ensemble algorithms, including Bagging, Stacking, and CatBoost harness the strengths of multiple models to improve predictive performance [44].

III. SIMULATION RESULTS AND DISCUSSION

We used Google Collaboratory and Python programming on the customized dataset (table 2) to assess the performance of the ML algorithms under consideration. We apply MI and ET in feature selection phase to rank and select the most relevant and informative features. The results of MI and ET feature selection are illustrated in fig. 4 (a) and (b). Based on the results of MI method, we removed the 6 low-importance features with MI score of $<= 0.1$, including 'Mean', 'Proto', 'sDSb', 'sTos', 'SrcJitAct', 'DstJitAct'. While ET method, which assigns importance scores to each feature, is used to rank and elect the 10 best performing features, including 'dTtl', 'pLoss', 'sTtl', 'Sport', 'RunTime', 'SrcRate', 'Rate', 'Sum', 'Max', and 'Min'.

The performance of the trained ML models is measured using classical evaluation metrics, including Acc, Precision, and Recall. Confusion matrix parameters are T_P, T_N, F_P, and F_N which are the number of true positives, true negatives, false positives, and false negatives, respectively. Acc, Precision, and Recall can be mathematically expressed by the following:

$$\text{Acc} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{2}$$

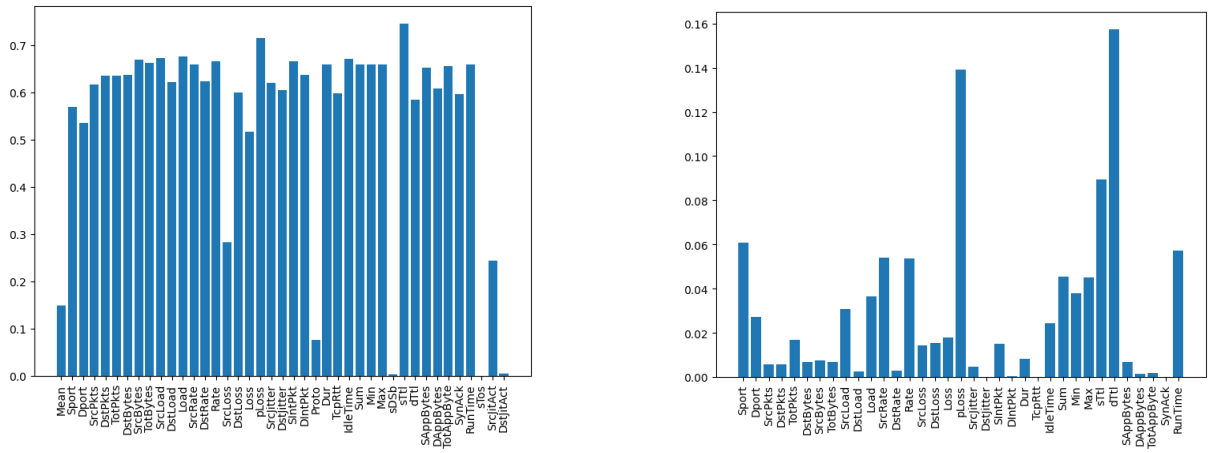
$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{3}$$

$$\text{Recall} = \frac{T_P}{T_P + F_N} \tag{4}$$

Acc tends to prioritize the common classes over rare classes, making it challenging for ML classifiers to perform well on minority classes [45]. For multi-class classification such as the case under study, both Micro F1-score and Macro F1-score can provide insights into the results which aggregates Precision and Recall across all classes. Micro F1-score uses the total T_P, T_N, F_P, and F_N, while Macro F1-score calculates Precision, Recall, and F1-score for each class separately, then average the values obtained for each class. Micro F1-score, and Micro F1-score can be

TABLE 2. Description of the corresponding datasets.

Class Label	Original Imbalanced Dataset			Customized Dataset		
	Sample Size	Percentage	Total Size	Sample Size	Percentage	Total Size
Normal	1,107,448	92.72	1,194,464	150,000	63.28	237,016
DoS	78,305	6.5		78,305	33.03	
Reconnaissance	8,240	0.69		8,240	3.48	
Command Injection	259	0.021		259	0.11	
Backdoor	212	0.018		212	0.09	



(a) MI (b) ET

FIGURE 4. Feature importance using MI and ET.

expressed as follows:

$$\text{Micro-F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{Macro-Precision} = \frac{1}{N} \sum \text{Precision}_i \tag{6}$$

$$\text{Macro-Recall} = \frac{1}{N} \sum \text{Recall}_i \tag{7}$$

$$\text{Macro-F1} = 2 \times \frac{\text{Macro-Precision} \times \text{Macro-Recall}}{\text{Macro-Precision} + \text{Macro-Recall}} \tag{8}$$

In addition, we use MCC which provides a comprehensive measure of classification performance specifically when dealing with imbalanced dataset since it exhibited consistent performance across different classifiers [46]. This observation suggests that MCC is robust to data imbalance, as it takes into consideration all four components of the confusion matrix to provide a balanced assessment of classifier performance.

$$\text{MCC} = \frac{T_P * T_N - F_P * F_N}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}} \tag{9}$$

To find the optimal number of features for the simulation experiments, we examine the Acc of the ML models using D_c dataset. As we mentioned before, D_c is a customized dataset in which the testing set contains only the samples of the severely unrepresented classes: Command injection, and Backdoors. Figure 5 shows the Acc of all ML models versus the top 10 most important features, which are chosen after applying the ensemble feature selection using MI and ET techniques. We notice that the optimal number of features for most of the considered ML models is 5 features.

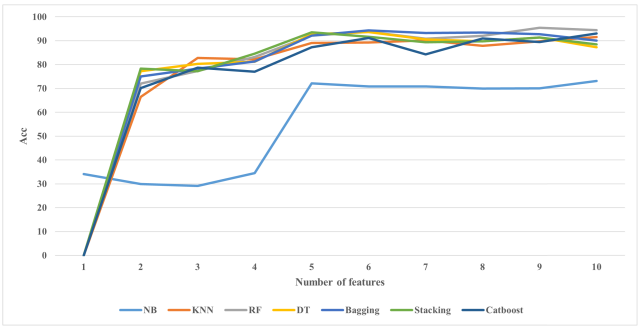


FIGURE 5. Acc vs number of features for D_c dataset.

Accordingly, we evaluate the ML models for the 5 high-importance features using predictions derived using both D_c and D_f datasets. Figure 6 presents the simulation results for the seven ML algorithms in terms of Acc, Precision, Recall, Micro F1-score, Micro F1-score, and MCC. After conducting numerous simulation experiments with various ML models

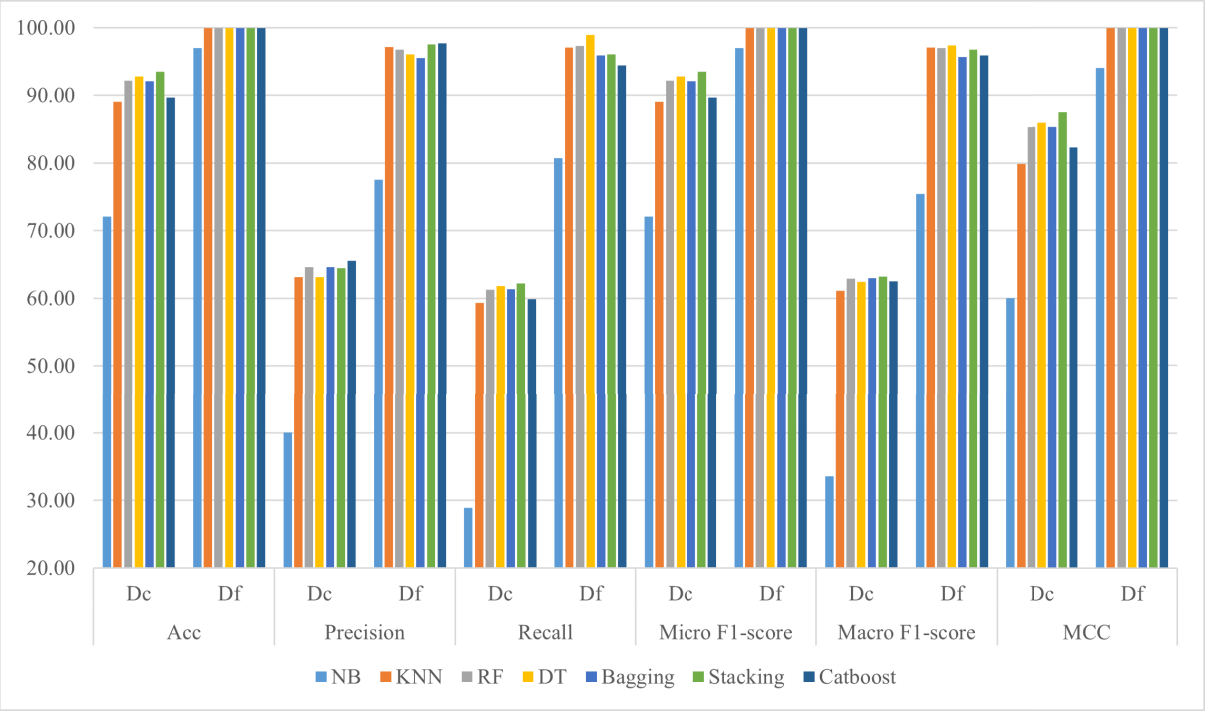


FIGURE 6. ML simulation results for both D_c and D_f datasets.

and feature selection techniques, it became obvious that achieving a high Acc, Precision, Recall, Micro F1-score, Micro F1-score, and MCC, on well-represented classes, including normal, DoS and Reconnaissance is straightforward. Further improvements do not exhibit significant gains when applying different methodologies for training and testing the ML models. Because of this, we focus instead on comparing the performance of the ML models based on their ability to classify the severely underrepresented classes: Command injection, and Backdoors, using D_c dataset as they are generally harder to predict. Figure 6 illustrates that RF, Bagging, Stacking, and Catboost are performing well in terms most of the metrics; however, Stacking model achieves the best Acc, Recall, Micro F1-score, Macro F1-score, and MCC, and Catboost has the highest Precision. It is worth mentioning that in our study, Acc evaluation is high on both training and testing datasets, indicating that overfitting is not an issue. In addition, our study focus on proposing a lightweight security system with simple ML algorithms using a customized dataset with more uniform class distribution, further reducing the risk of overfitting. The customized dataset with more uniform class distribution was achieved through RandomUnderSampler which could reduce the load of processing and the likeliness of overfitting [47].

Table 3 presents the simulation results for Mem, PT, and TT metrics. Mem of the seven models are 21, 22, 21, 21, 26, 27, and 27 Bytes for NB, KNN, RF, DT, Bagging, Stacking, and Catboost, respectively. We notice that the models' memory sizes are the same using both D_c and D_f datasets. In terms of TT, NB outperforms the other models with 0.032 and 0.028 seconds for D_c and D_f, respectively, followed by KNN, DT, Bagging, RF, and Catboost. Stacking has the highest processing time of with 39.682 and 39.36 seconds for D_c and D_f, respectively. Regarding PT, DT achieves the lowest prediction time of 0.002 and 0.004 seconds for D_c and D_f, respectively.We observe a significant increase in the prediction time of KNN model for D_f dataset. Notably, KNN typically stores the entire training dataset in memory during prediction, as it needs to calculate distances to all training points for each test point. Given that D_f is larger than D_c, this has the potential to impact the prediction time.

TABLE 3. ML results for Mem, PT, and TT metrics for both D_c and D_f datasets.

ML model	Mem	D_c		D_f	
		TT	PT	TT	PT
NB	21	0.032	0.007	0.028	0.01
KNN	22	0.14	0.007	0.137	4.597
RF	21	6.926	0.009	6.349	0.355
DT	21	0.155	0.002	0.159	0.004
Bagging	26	1.599	0.003	1.363	0.044
Stacking	27	39.682	0.016	39.36	3.265
Catboost	27	21.994	0.003	25.362	0.041

The main conclusions of this investigation can be summarized as follows:

- The WUSTL-IIOT-2021 dataset is characterized by high dimensionality and imbalance, closely resembling real-world scenarios for ISC systems. In this dataset, the proportion of attack traffic in comparison to normal traffic is notably low, reflecting the challenges typical of ISC systems in practical situations.

- The testing set plays a critical role in evaluating the performance of a ML model that has been fitted on a specific training set using an appropriate train-test split ratio and addressing underrepresented classes.
- The detection accuracy for low frequent attack classes is lower than the attacks with more instances is a common challenge in ML, particularly when dealing with imbalanced dataset, which requires proper methodology and evaluation metrics to enhance the detection process.
- Despite the dataset's imbalance, with less than 8% attack traffic, Precision, Recall, Micro F1-score, Micro F1-score, and MCC metrics are used to have a fair evaluation and understanding of the models performance across different classes.

IV. CONCLUSION

BC-enabled ISC systems integrated with constrained-resources IIoT devices and smart sensors face susceptibility to various significant cyber-attacks, posing potential threats to privacy, control, availability, and reliability. BC plays a crucial role in enhancing security and preventing various types of attacks in IIoT-ISC systems by providing an immutable and transparent ledger that securely store transaction records and data acquired from IIoT devices and smart sensors. However, with recent advances in cyber-security data science, ML emerges as an effective solution for the detection and mitigation of cyber-attacks within ISC. This paper presented an architectural security solution for BC-enabled ISC that embed a lightweight ML detection model to identify any malicious activities and mitigate their effects. A comparative analysis and performance assessment of several ML supervised techniques, namely, NB, KNN, RF, DT, Bagging, Stacking, and Catboost has been conducted. The WUSTL-IIOT-2021 imbalance dataset was used, which have been collected and tested from real scenarios of IoT devices in industrial settings. We applied lightweight methodology in terms of pre-processing and feature selection. This investigation aimed to elucidate the comparative advantages of these algorithms, considering their efficiency and resource utilization, with the goal of identifying the most suitable model for the detection of cyber-attacks in ISC.

REFERENCES

[1] M. Umair, M. A. Cheema, O. Cheema, H. Li, and H. Lu, "Impact of COVID-19 on IoT adoption in healthcare, smart homes, smart buildings, smart cities, transportation and industrial IoT," *Sensors*, vol. 21, no. 11, p. 3838, Jun. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/11/3838

[2] S. Ismail and H. Reza, "Security challenges of blockchain-based supply chain systems," in *Proc. IEEE 13th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2022, pp. 1–6.

[3] S. Ismail, H. Reza, K. Salameh, H. K. Zadeh, and F. Vasefi, "Toward an intelligent blockchain IoT-enabled fish supply chain: A review and conceptual framework," *Sensors*, vol. 23, no. 11, p. 5136, May 2023.

[4] M. Ohm, H. Plate, A. Sykosch, and M. Meier, "Backstabber's knife collection: A review of open source software supply chain attacks," in *Proc. Backstabber's Knife Collection, Rev. Open Source Softw. Supply Chain Attacks*, Lisbon, Portugal. Cham, Switzerland: Springer, Jun. 2020, pp. 23–43.

[5] M. Watney, "Cybersecurity threats to and cyberattacks on critical infrastructure: A legal perspective," in *Proc. Eur. Conf. Cyber Warfare Secur.*, vol. 21, no. 1, 2022, pp. 319–327.

[6] S. Ismail, M. Nouman, D. W. Dawoud, and H. Reza, "Towards a lightweight security framework using blockchain and machine learning," *Blockchain, Res. Appl.*, vol. 5, no. 1, Mar. 2024, Art. no. 100174.

[7] I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: An overview from machine learning perspective," *J. Big Data*, vol. 7, no. 1, pp. 1–29, Dec. 2020.

[8] A. Yeboah-Ofori, S. Islam, S. W. Lee, Z. U. Shamszaman, K. Muhammad, M. Altaf, and M. S. Al-Rakhami, "Cyber threat predictive analytics for improving cyber supply chain security," *IEEE Access*, vol. 9, pp. 94318–94337, 2021.

[9] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022.

[10] Z. E. Huma, S. Latif, J. Ahmad, Z. Idrees, A. Ibrar, Z. Zou, F. Alqahtani, and F. Baothman, "A hybrid deep random neural network for cyberattack detection in the industrial Internet of Things," *IEEE Access*, vol. 9, pp. 55595–55605, 2021.

[11] M. Al-Hawawreh, E. Sitnikova, and N. Aboutorab, "X-IIoTID: A connectivity-agnostic and device-agnostic intrusion data set for industrial Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 5, pp. 3962–3977, Mar. 2022.

[12] J. G. Almaraz-Rivera, J. A. Perez-Diaz, J. A. Cantoral-Ceballos, J. F. Botero, and L. A. Trejo, "Toward the protection of IoT networks: Introducing the LATAM-DDoS-IoT dataset," *IEEE Access*, vol. 10, pp. 106909–106920, 2022.

[13] S. Ismail, D. Dawoud, and H. Reza, "Towards a lightweight identity management and secure authentication for IoT using blockchain," in *Proc. IEEE World AI IoT Congr. (AIIoT)*, Jun. 2022, pp. 077–083.

[14] S. Ismail, D. W. Dawoud, T. Al-Zyoud, and H. Reza, "Towards blockchain-based adaptive trust management in wireless sensor networks," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (eIT)*, May 2023, pp. 163–168.

[15] N. Malik, K. Alkhatib, Y. Sun, E. Knight, and Y. Jararweh, "A comprehensive review of blockchain applications in industrial Internet of Things and supply chain systems," *Appl. Stochastic Models Bus. Ind.*, vol. 37, no. 3, pp. 391–412, May 2021.

[16] S. Ismail, H. Reza, H. K. Zadeh, and F. Vasefi, "A blockchain-based IoT security solution using multichain," in *Proc. IEEE 13th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Mar. 2023, pp. 1105–1111.

[17] S. Ali, Q. Li, and A. Yousafzai, "Blockchain and federated learning-based intrusion detection approaches for edge-enabled industrial IoT networks: A survey," *Ad Hoc Netw.*, vol. 152, Jan. 2024, Art. no. 103320.

[18] R. K. Singh, R. Mishra, S. Gupta, and A. A. Mukherjee, "Blockchain applications for secured and resilient supply chains: A systematic literature review and future research agenda," *Comput. Ind. Eng.*, vol. 175, Jan. 2023, Art. no. 108854.

[19] S. Al-Farsi, M. M. Rathore, and S. Bakiras, "Security of blockchain-based supply chain management systems: Challenges and opportunities," *Appl. Sci.*, vol. 11, no. 12, p. 5585, Jun. 2021.

[20] S. S. Mathew, K. Hayawi, N. A. Dawit, I. Taleb, and Z. Trabelsi, "Integration of blockchain and collaborative intrusion detection for secure data transactions in industrial IoT: A survey," *Cluster Comput.*, vol. 25, no. 6, pp. 4129–4149, Dec. 2022.

[21] R. Machhan, R. Trehan, P. Singh, and K. S. Sangwan, "Blockchain, artificial intelligence, and big data: Advanced technologies for industry 4.0," in *Industry 4.0*. Boca Raton, FL, USA: CRC Press, pp. 141–160.

[22] M. Yang, M. K. Lim, Y. Qu, D. Ni, and Z. Xiao, "Supply chain risk management with machine learning technology: A literature review and future research directions," *Comput. Ind. Eng.*, vol. 175, Jan. 2023, Art. no. 108859.

[23] M. M. Salim, A. K. Comivi, T. Nurbek, H. Park, and J. H. Park, "A blockchain-enabled secure digital twin framework for early botnet detection in IIoT environment," *Sensors*, vol. 22, no. 16, p. 6133, Aug. 2022.

[24] S. Zhao, S. Li, and Y. Yao, "Blockchain enabled industrial Internet of Things technology," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 6, pp. 1442–1453, Dec. 2019.

[25] M. Zolanvari, M. A. Teixeira, and R. Jain, "Effect of imbalanced datasets on security of industrial IoT using machine learning," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Nov. 2018, pp. 112–117.

[26] M. M. Alani, "An explainable efficient flow-based industrial IoT intrusion detection system," *Comput. Electr. Eng.*, vol. 108, May 2023, Art. no. 108732.

[27] T. Gaber, J. B. Awotunde, S. O. Folorunso, S. A. Ajagbe, and E. Eldesouky, "Industrial Internet of Things intrusion detection method using machine learning and optimization techniques," *Wireless Commun. Mobile Comput.*, vol. 2023, pp. 1–15, Apr. 2023.

[28] M. Zolanvari, Z. Yang, K. Khan, R. Jain, and N. Meskin, "TRUST XAI: Model-agnostic explanations for AI with a case study on IIoT security," *IEEE Internet Things J.*, vol. 10, no. 4, pp. 2967–2978, Feb. 2023.

[29] A. M. Eid, A. B. Nassif, B. Soudan, and M. N. Injadat, "IIoT network intrusion detection using machine learning," in *Proc. 6th Int. Conf. Intell. Robot. Control Eng. (IRCE)*, Aug. 2023, pp. 196–201.

[30] H. Vargas, C. Lozano-Garzon, G. A. Montoya, and Y. Donoso, "Detection of security attacks in industrial IoT networks: A blockchain and machine learning approach," *Electronics*, vol. 10, no. 21, p. 2662, Oct. 2021.

[31] H. Mrabet, A. Alhomoud, A. Jemai, and D. Trentesaux, "A secured industrial Internet-of-Things architecture based on blockchain technology and machine learning for sensor access control systems in smart manufacturing," *Appl. Sci.*, vol. 12, no. 9, p. 4641, May 2022.

[32] S. Latif, Z. Idrees, J. Ahmad, L. Zheng, and Z. Zou, "A blockchain-based architecture for secure and trustworthy operations in the industrial Internet of Things," *J. Ind. Inf. Integr.*, vol. 21, Mar. 2021, Art. no. 100190.

[33] A. F. Author Surname, "Classification of cyber threat using machine learning models in supply chain management," *J. Cybersecur.*, vol. 10, no. 3, pp. 123–135, 2022.

[34] H. Wang, L. Sagbansua, and B. Alidaee, "Enhancing supply chain security with automated machine learning," Oct. 2023, *arXiv:2406.13166*.

[35] F. S. Cebeloglu and M. Karakose, "Comparative analysis of cyber security approaches using machine learning in Industry 4.0," in *Proc. IEEE Int. Symp. Syst. Eng. (ISSE)*, Oct. 2020, pp. 1–5.

[36] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine learning-based network vulnerability analysis of industrial Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6822–6834, Aug. 2019.

[37] V. Borgiani, P. Moratori, J. F. Kazienko, E. R. R. Tubino, and S. E. Quincozes, "Toward a distributed approach for detection and mitigation of denial-of-service attacks within industrial Internet of Things," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4569–4578, Mar. 2021.

[38] P. Arora, B. Kaur, and M. A. Teixeira, "Evaluation of machine learning algorithms used on attacks detection in industrial control systems," *J. Inst. Eng. (India), Ser. B*, vol. 102, no. 3, pp. 605–616, Jun. 2021.

[39] M. Zolanvari, L. Gupta, K. Khan, and R. Jain, *WUSTL-IIOT-2021 Dataset for IIoT Cybersecurity Research*. Washington, DC, USA: Washington Univ. St. Louis, 2021.

[40] S. Ismail, T. T. Khoei, R. Marsh, and N. Kaabouch, "A comparative study of machine learning models for cyber-attacks detection in wireless sensor networks," in *Proc. IEEE 12th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Dec. 2021, pp. 313–318.

[41] S. Ismail, D. W. Dawoud, and H. Reza, "A comparative study of datasets for cyber-attacks detection in wireless sensor networks," in *Proc. IEEE 3rd Int. Conf. Comput. Mach. Intell. (ICMI)*, Apr. 2024, pp. 1–6.

[42] I. Muraina, "Ideal dataset splitting ratios in machine learning algorithms: General concerns for data scientists and data analysts," in *Proc. 7th Int. Mardin Artuklu Sci. Res. Conf.*, 2022, pp. 1–9.

[43] B. Vrigazova, "The proportion for splitting data into training and test set for the bootstrap in classification problems," *Bus. Syst. Res. J.*, vol. 12, no. 1, pp. 228–242, May 2021.

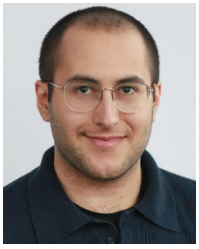
[44] S. Ismail, Z. El Mrabet, and H. Reza, "An ensemble-based machine learning approach for cyber-attacks detection in wireless sensor networks," *Appl. Sci.*, vol. 13, no. 1, p. 30, Dec. 2022.

[45] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review," *GESTS Int. Trans. Comput. Sci. Eng.*, vol. 30, no. 1, pp. 25–36, Nov. 2006.

[46] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using Matthews correlation coefficient metric," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, Art. no. e0177678.

[47] M. Saripuddin, A. Suliman, S. S. Sameon, and B. N. Jorgensen, "Random undersampling on imbalance time series data for anomaly detection," in *Proc. 4th Int. Conf. Mach. Learn. Mach. Intell.*, Sep. 2021, pp. 151–156.

SHEREEN ISMAIL (Member, IEEE) received the B.Sc. degree in computer engineering and the M.Sc. degree in computer engineering and networks from The University of Jordan, Amman, Jordan, and the Ph.D. degree in computer science from the University of North Dakota, USA. Throughout her Ph.D. studies, she was a Graduate Research Assistant with the School of Electrical Engineering and Computer Science. She is currently a Research Scientist in networking and cyber-security with Merit Network Inc. She taught undergraduate level at American University of Ras Al-Khaimah, Al-Zaytoonah University of Jordan, and Applied Science University of Jordan. Her research interests include wireless networks and cyber-security. Her contributions have been recognized through scholarship awards at prestigious conferences, including the 2023 IEEE International Symposium on Women in Services Computing (WISC 2023) and Women in CyberSecurity (WiCyS 2024).



SALAH DANDAN is currently pursuing the bachelor's degree in electrical engineering and computer science with the University of North Dakota. He is a Bachelor's Research Assistant, contributing to projects focused on GPU acceleration, high-performance computing, and the application of machine learning in cyber-security.

DIANA W. DAWOUD (Senior Member, IEEE) received the Ph.D. degree from the University of Surrey, and directs her current research towards the intricate domains of optical communication systems' physical layer, digital signal processing for communications, security systems, and the development of Internet of Things (IoT) systems. She is currently an Assistant Professor with the University of Dubai. She holds a patent in the field of energy-efficient communication systems, and her innovative unipolar transmission method has been featured in the esteemed IEEE TRANSACTIONS ON COMMUNICATION. She actively serves as a journal reviewer and contributes to conference committees. She co-chaired the inaugural IEEE Wireless Communications and Networking Conference (WCNC) Workshop on Optical Wireless Communication, in March 2023, laying the groundwork for a successful second edition, in 2024. She chairs the IEEE UAE Section ComSoc and SP joint chapter, and holds the position of the Vice-Chair at IEEE UAE Section WIE.

HASSAN REZA currently holds the position of a Professor of software engineering with the University of North Dakota, while also maintaining affiliations with North Dakota State University and UND Biomedical Engineering. He is a Technical Member of the Center of Excellence for Unmanned Aircraft Systems (UAS) since its establishment, the UND Research Institute for Autonomous Systems (RIAS), and the UND member of the National Academy of Inventors. His contributions extend to his role as Principal Investigator (PI) or Co-PI, resulting in the successful acquisition of funding totaling up to $15 million from prestigious entities, such as the Department of Defense (DOD), National Institutes of Health (NIH), National Aeronautics and Space Administration (NASA), National Science Foundation (NSF), NASA-EPSCore, and Rockwell Collins. His extensive expertise encompasses diverse domains, including architectural modeling of data/software intensive systems, model-based engineering of cyber-physical systems (with a focus on avionic and medical systems), software testing methodologies, the engineering of safety/security critical systems, and applied blockchained.