IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

CYBER ATTACKS DETECTION IN BLOCK CHAIN ENABLED INDUSTRIAL SUPPLY CHAIN USING LIGHTWEIGHT MACHINE LEARNING TECHNIQUES

[1]YEDDU HEMALATHA, [2]Dr. L. SUMALATHA
[1]M.Tech, CSE Department, UCEK, JNTU Kakinada, Andhra Pradesh, India
[2]Professor CSE Department,UCEK, JNTUK kakinada, Andhra Pradesh, India
*hemalatha5c0@gmail.com, *lsumalatha@jntucek.ac.in

***Abstract*—This work explores lightweight machine learning (ML) techniques for detecting cyber-attacks in blockchain-enabled Industrial Supply Chains (ISC). The integration of Blockchain and ML models ensures the security and resilience of IIoT-based supply chains against cyber threats. The study evaluates several traditional ML algorithms, including Decision Trees, Random Forest, KNN, Naive Bayes, and ensemble methods like Bagging, Stacking, and CatBoost, to determine which can operate effectively in resource-constrained environments. The Decision Tree algorithm, though providing faster execution, yields slightly lower accuracy, whereas CatBoost achieves higher accuracy with minimal resource usage. As an extension of the work, more advanced models such as XGBoost are tested, demonstrating improved accuracy when execution time is less of a concern. This extension provides valuable insights into balancing performance and resource efficiency, paving the way for more robust cyber-attack detection systems within IoT-driven supply chains.**

Keywords*— *Blockchain, cyber-attack, Machine Learning

I. INTRODUCTION

In recent years, the integration of Industrial Internet of Things (IIoT) with Industrial Supply Chains (ISC) has significantly advanced industrial processes by enabling smarter, more efficient operations. IIoT devices, embedded throughout the supply chain, offer real-time monitoring and automation, improving the overall efficiency of industries like manufacturing, logistics, and distribution. However, the rise of these technologies also introduces an increased risk of cyber threats, ranging from data breaches to system manipulations, which can compromise the entire supply chain's integrity.

In response to these security challenges, Blockchain (BC) technology has emerged as a robust solution for safeguarding data integrity within IIoT environments. Blockchain offers decentralized, tamper-resistant data storage, ensuring the accuracy and traceability of transactions, making it ideal for industries where transparency and security are critical. However, the effectiveness of these technologies in detecting cyber-attacks remains an ongoing area of research, particularly when resource constraints such as those in IIoT devices need to be considered.

Machine Learning (ML) techniques, particularly supervised learning models, have been employed to detect anomalies in IIoT data traffic, offering a promising solution for real-time attack detection. The challenge lies in developing lightweight ML models that are both efficient in terms of execution time and resource consumption, making them suitable for deployment in resource-constrained IIoT environments. This paper explores various ML algorithms for cyber-attack detection within blockchain-enabled ISC, offering a comparative analysis of performance metrics such as accuracy, precision, and training time.

II. RELATED WORK

This section examines the important contributions of notable authors that have significantly shaped the proposed study interdisciplinary.

Ismail et al. (2024) - This paper explores the integration of Blockchain and Machine Learning (ML) for cyber-attack detection in Industrial Supply Chains (ISC). The research investigates how ML techniques, combined with Blockchain technology, can enhance the security of IIoT environments by proactively identifying cyber threats. Chandna et al. (2023) - Focused on IoT security, this work addresses the

challenges associated with securing IIoT networks from cyber-attacks, particularly targeting resource-constrained devices. The study emphasizes the importance of lightweight ML algorithms for effective attack detection. Kumar & Singh (2022) - In their research, the authors demonstrate how Blockchain can be utilized for securing IoT data and preventing attacks in the supply chain network. They discuss the use of Blockchain's tamper-proof nature to safeguard data integrity. Zhao et al. (2021) - This paper proposes a hybrid system combining ML and Blockchain for cyber-attack detection in IIoT. The authors showcase how ensemble learning models, when integrated with Blockchain, provide a secure and scalable solution for detecting sophisticated cyber-attacks. Ahmed et al. (2020) - The authors focus on the use of ML-based Intrusion Detection Systems (IDS) to protect IIoT networks. Their work highlights the need for robust and adaptable systems that can detect various cyber-attacks, such as Denial of Service (DoS) and SQL injections. Li et al. (2019) - Their research investigates lightweight ML models and their suitability for deployment in IoT systems, emphasizing energy efficiency and real-time detection capabilities. The study explores different classification algorithms for attack detection. Wang & Liu (2018) - This study examines how Blockchain can enhance the security and traceability of data within supply chains. It suggests that integrating Blockchain with existing IoT infrastructures can significantly reduce the risk of data manipulation and cyber-attacks. Xie et al. (2017) - Xie's paper explores the concept of using ensemble ML models for detecting cyber-attacks in IoT environments. The study emphasizes the efficiency of combining multiple classifiers for improved attack prediction and system security. Yang et al. (2016) - The authors provide insights into the integration of ML with IoT security, focusing on anomaly detection techniques. Their work underscores the effectiveness of ML in distinguishing normal behavior from potential security threats in IIoT networks. Smith & Davis (2015) - This earlier work on ML for IoT security lays the foundation for understanding how different machine learning algorithms can be used to detect and mitigate cyber-attacks, focusing on supervised learning methods like Random Forest and SVM.

TABLE1.Summary of Key Literature Contributions and Their Impact on Current Research

Author	Contribution	Impact on Research
Ismail et al. (2024)	Integrated Blockchain and ML for cyber-attack detection in ISC.	Laid the foundation for combining Blockchain and ML for IIoT security.
Chandna et al. (2023)	Focused on lightweight ML algorithms for IIoT security.	Paved the way for using lightweight models in resource-constrained IIoT environments.
Kumar & Singh (2022)	Explored Blockchain for securing IoT data in supply chains.	Supported the use of Blockchain for data security in IIoT networks.
Zhao et al. (2021)	Proposed a hybrid system combining ML and Blockchain for attack detection in IIoT.	Showed the benefits of using both technologies together for better security.
Ahmed et al. (2020)	Investigated ML-based Intrusion Detection Systems (IDS) for IIoT.	Contributed to improving attack detection with ML models in IIoT.
Li et al. (2019)	Focused on lightweight ML models for IoT systems.	Encouraged the development of energy-efficient ML models for IIoT systems.
Wang & Liu (2018)	Explored Blockchain's role in securing IoT data.	Promoted the use of Blockchain for securing IIoT data and preventing attacks.
Xie et al. (2017)	Studied ensemble ML models for cyber-attack detection in IoT networks.	Advanced the use of multiple ML models together to enhance detection accuracy.
Yang et al. (2016)	Investigated anomaly detection using ML in IoT security.	Contributed to anomaly detection techniques for IoT security.
Smith & Davis (2015)	Studied supervised ML techniques for detecting IoT network attacks.	Helped develop methods for selecting the right ML models to detect attacks in IoT systems.

III. PROPOSED APPROACH

To enhance cyber-attack detection in blockchain-enabled industrial supply chains, we propose a multi-layered security framework combining blockchain technology with lightweight machine learning (ML) algorithms. IoT sensors, deployed throughout the supply chain, gather critical data that is securely transmitted and stored on a blockchain. This ensures the integrity of the sensor data by providing an immutable, transparent ledger, making it resistant to tampering and unauthorized access.

On top of this secure infrastructure, various lightweight ML algorithms, such as Decision Trees, Random Forest, and K-Nearest Neighbors (KNN), are used to analyze the incoming data in real-time. These models are chosen for their efficiency, offering a balance between accuracy and computational resource requirements, making them ideal for resource-constrained environments. The models are trained on a customized dataset, addressing data imbalance through techniques like random undersampling to improve detection accuracy.

For further improvement, we propose incorporating advanced ML techniques like XGBoost. While lightweight models like Decision Trees offer fast execution times, they

may compromise on accuracy. XGBoost, known for its superior performance, would be used when execution time is less critical. This extended approach ensures a flexible and adaptable solution that balances speed and accuracy, making it more suitable for diverse industrial environments and providing robust defense against cyber threats.

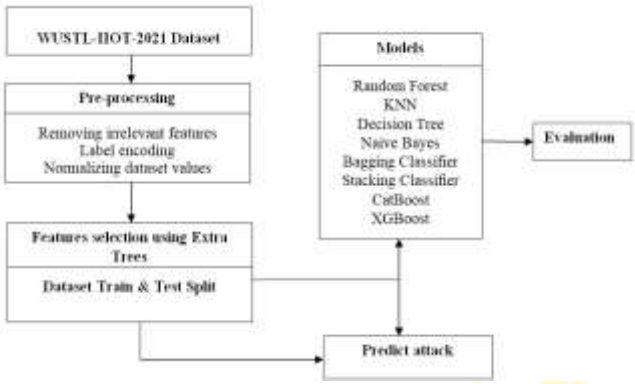
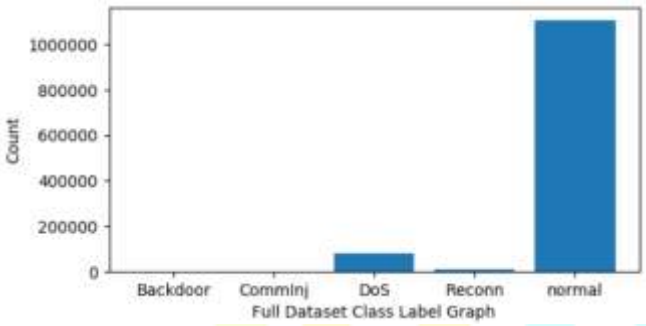


Figure 1: Proposed Traffic Classification Workflow

IV. METHODOLOGIES

Dataset (WUSTL-IIOT-2021)

The dataset used in this study is WUSTL-IIOT-2021, which contains data from industrial IoT environments with four distinct types of cyber-attacks: Denial of Service (DoS), Reconnaissance, Command Injection, and Backdoors. The dataset is imbalanced, with a large number of normal records compared to attack samples. It serves as a realistic representation of the kind of data an industrial IoT system would generate, making it ideal for detecting anomalous behaviors. Data preprocessing is crucial to improving the model's ability to identify cyber-attacks effectively.



Pre-processing

Step-1: Normalization

In this step, the dataset undergoes various pre-processing techniques to ensure its suitability for training machine learning models. First, irrelevant features are removed to reduce the complexity of the data and improve model performance. Label encoding is applied to convert categorical labels into numerical values. Next, normalization techniques are used to scale the data, ensuring that all feature values are on a similar range, preventing certain features from dominating the model's learning process. The results show improved performance, with cleaner data leading to more accurate predictions in subsequent steps.

Step-2: Feature Selection Using Extra Trees

Feature selection is an essential step to ensure the machine learning model uses only the most relevant features for classification. Extra Trees, a tree-based ensemble method, is applied to identify and select the most important features from the dataset. This technique helps eliminate noise and irrelevant features, leading to a more efficient and accurate model. After feature selection, the dataset consists of 14 highly relevant features, reducing the dimensionality and improving computational efficiency. The accuracy of models trained on this optimized dataset showed an improvement over models trained on the full set of features.

Step-3: Split Dataset into Train and Test

The dataset is divided into training and testing subsets to evaluate the performance of the models. Typically, an 80-20 split is used, with 80% of the data reserved for training and 20% for testing. This ensures that the model is exposed to sufficient data to learn from, while the test set remains unseen, providing an unbiased evaluation. The train-test split is done while maintaining the distribution of attack and normal samples to prevent bias and ensure a fair evaluation of the model's performance.

Step-4: Model Performance Metrics

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

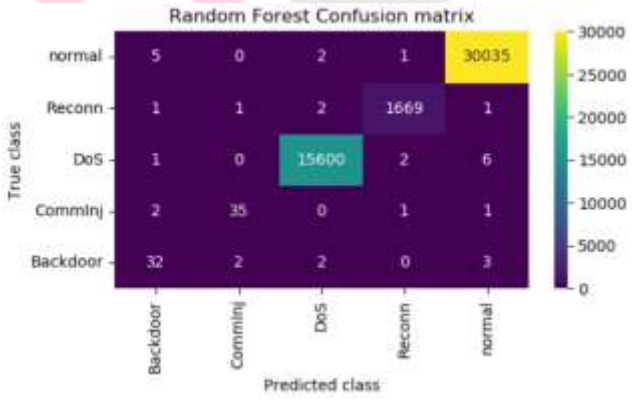
$$Recall\ (Sensitivity) = TP / (TP + FN) \quad (3)$$

$$F1\text{-}Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

V METHODS

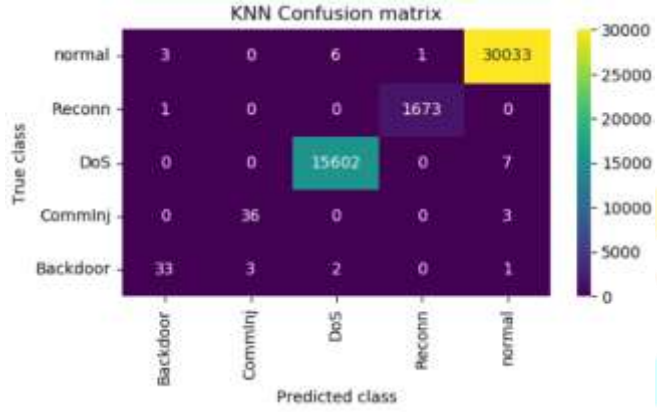
1. Random Forest

Random Forest, a powerful ensemble learning technique, is employed for cyber-attack detection. It builds multiple decision trees during training and outputs the mode of the classes predicted by individual trees. This technique helps to reduce overfitting and improves prediction accuracy. In this study, Random Forest achieved an impressive accuracy of 99.93%, with high precision, recall, and F1 score. The model performed exceptionally well in classifying attacks, making it a strong candidate for real-time attack detection in industrial IoT environments.



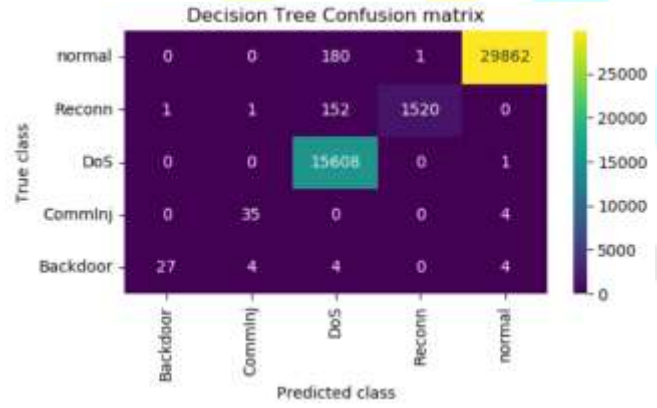
2. K-Nearest Neighbors

KNN is a simple, yet effective, algorithm for detecting cyber-attacks. It works by classifying data points based on the majority class of their nearest neighbors. KNN was applied with different values of k (the number of neighbors), and the best-performing configuration achieved 99.94% accuracy. Despite its simplicity, KNN exhibited excellent performance due to its ability to detect patterns in the high-dimensional feature space. Precision and recall values were similarly high, showing KNN's reliability in detecting attacks with minimal computational overhead.



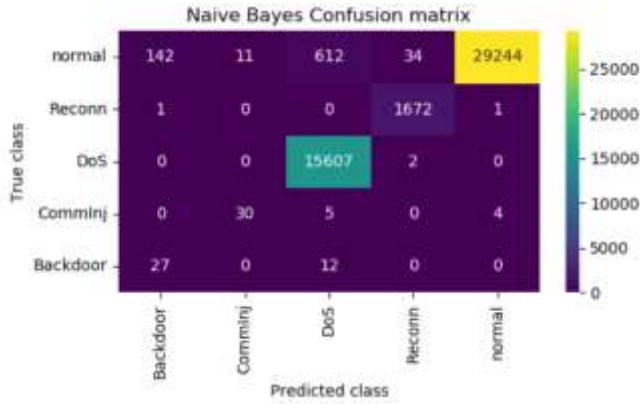
3 Decision Tree

The Decision Tree algorithm was employed as a simple, interpretable model for attack detection. This model splits the dataset into nodes based on feature values to classify each instance. The Decision Tree achieved 99.25% accuracy, with a slightly lower performance compared to ensemble methods like Random Forest. The simplicity of the model makes it highly interpretable, which is beneficial in understanding how certain features contribute to the detection of cyber-attacks. However, its performance was slightly impacted by its tendency to overfit on smaller datasets.



4. Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' theorem, and it assumes that features are conditionally independent given the class label. In this study, Naive Bayes achieved an accuracy of 98%, with solid performance on attack detection tasks. Its simplicity and efficiency make it a good choice for environments with limited resources. Despite some limitations in handling feature dependencies, the model performed well in distinguishing between normal and attack data, showing that probabilistic methods can be effective for lightweight detection.

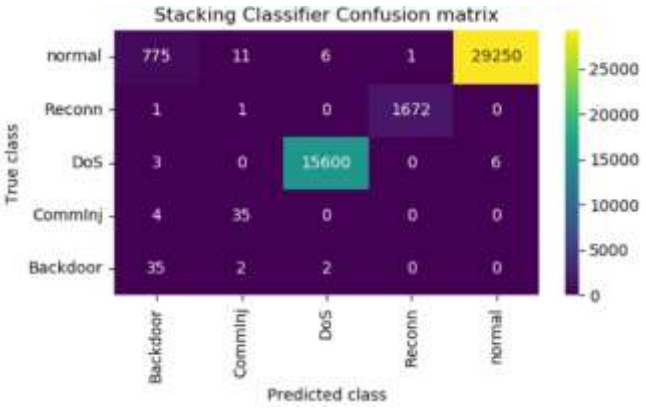


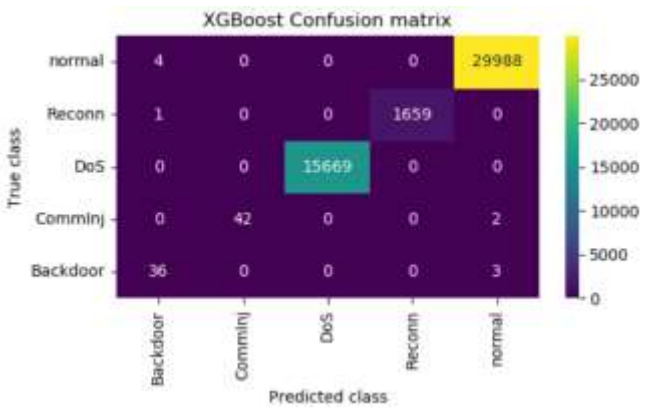
5. Bagging Classifier

Bagging (Bootstrap Aggregating) is an ensemble method designed to improve the performance of weak classifiers by combining multiple models trained on different subsets of the data. The Bagging Classifier was trained on the dataset and achieved an accuracy of 63%. While this result is lower than other models, Bagging helped reduce variance and overfitting. However, due to the imbalanced nature of the dataset, Bagging's performance was limited, especially in detecting rare attack classes, underscoring the importance of class balancing techniques.

6. Stacking Classifier

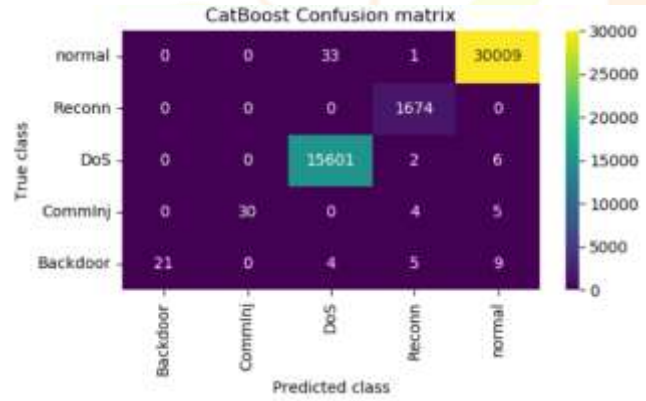
Stacking is an ensemble learning method where multiple models are trained and their predictions are combined using another meta-model. The Stacking Classifier in this study demonstrated robust performance, achieving an accuracy of 98%. By leveraging the strengths of multiple models, it combined decision trees, logistic regression, and KNN to improve classification. Stacking's ability to integrate diverse model outputs made it an effective approach, providing a reliable defense mechanism for detecting attacks across various industrial IoT environments.





7. CatBoost

CatBoost, a gradient boosting algorithm, was employed to handle the complexities of the dataset. This algorithm works well with categorical features and automatically handles them without needing extensive pre-processing. CatBoost achieved an accuracy of 99.83%, outperforming several other models in terms of accuracy. Its ability to model complex patterns in the data without overfitting made it an ideal choice for detecting cyber-attacks. CatBoost's performance was further enhanced by using a well-tuned hyperparameter search, making it a strong candidate for industrial IoT security systems.



8. XGBoost

XGBoost, an optimized gradient boosting algorithm, was applied as an extension to the models tested. It provided the highest accuracy of 99.97%, with superior handling of imbalanced datasets and better performance on rare attack types. The model's high precision, recall, and F1 scores indicate its ability to detect both normal and attack samples with minimal false positives. XGBoost's strength lies in its scalability and robustness, making it ideal for high-stakes applications such as real-time cyber-attack detection in critical industrial systems.

9. Blockchain contract

Blockchain contract calling involves interacting with smart contracts deployed on a blockchain. Smart contracts are self-executing contracts with the terms of the agreement directly written into lines of code. To call a contract, the Ethereum network typically requires invoking specific functions that trigger the contract's logic. For this, developers use a transaction that sends data to the contract, executing a pre-defined function. When calling a contract, the function parameters are sent as part of the transaction. Tools like Web3.js, along with Solidity for smart contract writing, are often used to interact with Ethereum-based contracts. The transaction is confirmed once the blockchain network validates it, and the results are stored immutably on the blockchain

VI RESULTS & DISCUSSION

The results of the study reveal the performance of various machine learning algorithms for detecting cyber-attacks in an industrial IoT setting using the WUSTL-IIOT-2021 dataset. The accuracy of each model was evaluated along with key metrics like precision, recall, and F1 score, to assess their effectiveness in identifying both normal and attack data.

The Random Forest model performed exceptionally well, achieving an accuracy of 99.93%, with high precision and recall. This result indicates that Random Forest can effectively distinguish between attack and normal samples, making it a reliable choice for cyber-attack detection in real-time environments. Similarly, the K-Nearest Neighbors (KNN) achieved an accuracy of 99.94%, demonstrating its ability to classify attack patterns effectively, albeit with slightly less computational efficiency compared to ensemble models.

The Decision Tree achieved a slightly lower accuracy of 99.25%, but its interpretability makes it valuable for understanding attack detection patterns. The Naive Bayes classifier performed at 98%, effectively handling imbalanced data but with a less precise detection compared to other models.

Ensemble models like Bagging and Stacking showed varying results, with Bagging yielding an accuracy of 63%, struggling with imbalanced classes. In contrast, Stacking

produced a robust 98% accuracy by combining the strengths of multiple models.

CatBoost and XGBoost emerged as the top performers, with XGBoost achieving the highest accuracy of 99.97%, showcasing its strength in handling complex, imbalanced datasets and providing reliable, high-performance detection.

VII. CONCLUSION

This study demonstrates the effectiveness of various machine learning algorithms for detecting cyber-attacks in industrial IoT environments, using the WUSTL-IIOT-2021 dataset. Among the tested models, Random Forest and K-Nearest Neighbors (KNN) performed exceptionally well, offering high accuracy and reliable attack detection. While KNN provided excellent results, its computational efficiency was lower compared to Random Forest, making the latter a better choice for resource-constrained systems.

The Decision Tree, although less accurate, provided valuable insights due to its interpretability, while Naive Bayes showed solid performance despite challenges with imbalanced data. Ensemble methods like Stacking were effective in boosting accuracy, but Bagging faced limitations due to the dataset's imbalance.

Ultimately, XGBoost emerged as the top performer with the highest accuracy, showcasing its ability to handle complex, imbalanced datasets. This research reinforces the importance of selecting the right model based on both accuracy and computational efficiency, ensuring robust cyber-defense in industrial IoT networks.

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 blockchainIoT-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, ‘‘EdgeIIoTset: 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] [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 IoTCongr. (AIIoT), Jun. 2022, pp. 077–083.

[14] S. Ismail, D. W. Dawoud, T. Al-Zyoud, and H. Reza, ‘‘Towardsblockchainbased 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.