Hybrid Intrusion Detection with Edge-IIoTset Dataset

AMJAD JUMAAH FRHAN
Department of Education and Islamic Studies, Sunni Endowment
University of AL Mashreq, Department of Cybersecurity Engineering Technology
Baghdad
IRAQ

Abstract: - New security issues, such as how to protect networks and edge devices, have arisen with the introduction of the Industrial Internet of Things (IIoT). The ever-changing nature and complexity of IIoT communications makes traditional intrusion detection systems (IDS) inadequate. Using the Edge-IIoTset dataset, we present an IDS hybrid model in this study. Improving detection performance and reducing false alarms are achieved through the employment of ML-DL algorithms. When compared to standalone methods, experimental results demonstrate that the hybrid approach can provide excellent levels of accuracy, recall, and precision. The model's performance and validation in real IIoT contexts are illustrated by figures and statistics tables.

Key-Words: - Artificial Neural Networks, Cybersecurity, Internet of Things, Machine Learning, SVM and XGBoost.

Received: July 17, 2025. Revised: November 9, 2025. Accepted: March 4, 2025. Published: April 7, 2026.

1 Introduction

As The integration of IIoT technologies in industrial sectors has changed how efficiently operations run. However, it has also increased the risk of complex cyber threats. This is especially dangerous for edge devices, which typically have limited resources and are widely distributed geographically. Legacy IDS solutions are likely to be incompatible with scaling, speed and flexibility. To cover these problems, we suggest a hybrid IDS model to work jointly based on ML and DL approaches using the Edge-IIoTset dataset to verify our proposal with extensive analysis [1].

The present study addresses the following research questions:

- Is it possible for a hybrid IDS to surpass conventional single-model methodologies in IIoT contexts?
- How well can the Edge-IIoTset dataset be used to compare different hybrid IDS models?
- What are the trade-offs between accuracy, detection rate, and cost at the edge?

2 Literature Review

2.1 Intrusion Detection Systems in IIoT

Intrusion detection systems (IDSs) are frequently employed to keep an eye on network events or systems and identify potentially harmful activity that manages to get past security perimeters (like firewalls). Evaluating intrusion detection techniques is crucial, and assessing the precision and effectiveness of IoT security techniques requires the usage of IoT-related datasets that represent actual IoT applications. However, one of the biggest challenges to evaluating intrusion detection techniques specific to IoT/IIoT applications is the absence of real-world datasets for these applications. Since the empirical validation and evaluation of such systems should fulfil performance expectations, the lack of these datasets makes it difficult to build and develop IoT-based intrusion detection algorithms [2].

The use of Intrusion Detection Systems (IDS) as solutions that are necessary in the protection of network infrastructures is not a new phenomenon: their importance has only grown with the advent of the Industrial Internet of Things (IIoT). The IIoT environments are very dependent on the inter-relationship between the sensors, actuators, gateways, and distributed edge devices; all of which present new surfaces of cyber-attacks that cannot be easily secured by traditional IDS solutions. The main problem of developing an IDS to address IoT and IIoT is the lack of real-life datasets to represent the

operational and attack conditions in these settings. The lack of such datasets makes it hard to benchmark the performance of IDS and thus the lack of such datasets will lead to systems that do not scale to heterogeneous operations. The inability of researchers to create models that would identify current multi-stage intrusion is limited by insufficient data on current topics like the inability to find information that is both temporally diverse and realistic in terms of traffic load.

The IIoT is also defined by limitations concerning the computing capacity, storage, and energy supply. The limitations make conventional IDS architectures that are traditionally created to run on centralized and high-capacity servers incompatible with the implementation of the distributed industrial nodes. The heterogeneity of IIoT, the sensitivity of industrial data, and the importance of nonstop work has become the fundamental challenges in terms of ensuring the safety of this field. Their evaluation proves that traditional IDS system does not possess the flexibility needed when applying to IIoT contexts because of the dynamic character of industrial traffic and low scalability of signature-based IDS architectures. In order to address these limitations, studies have redirected to the use of machine learning (ML) and deep learning (DL) to improve anomaly detection and behavioral modeling. These techniques are useful in detecting nonlinear, nuanced attack patterns that are not identified by standard signature-based techniques [3].

As shown by the latest study, phishing email detection rate is much strengthened and false positives diminish when NLP-based discovery of features is used along with machine learning and combination approaches [4].

Furthermore, the latest literature points to the combination of fog and edge computing as one of the possible architectural improvements to implement an IDS. Edge-based IDS architectures minimize latency, enhance situational awareness, and support immediate response activities by pushing the capabilities of detection closer to the devices that produce the data. This is particularly important in industrial areas where delays may result in devastating effects of operations. Another crucial point made which is the presence of telemetry-oriented datasets, including TON_IoT, that should be used to estimate the real-time behavior of devices; nevertheless, even these are not as complex as the IIoT demands.

Overall, the reviewed literature shows that IDS in IIoT is still struggling with the issue of the lack of availability of datasets, system heterogeneity, and resource limitations. There is a great deal of consensus that successful IDS should be built on the combination of ML/DL techniques, edge-enabled systems, and realistic datasets that can be used to represent both benign and malicious activity. These findings have explained why hybrid IDS models need to be tested using sophisticated IIoT data like Edge-IIoTset.

2.2 Edge Computing and Security

The development of edge computing has been significantly accelerated in recent years by the quick development of the Internet of Things (IoT) and smart mobile devices. Though its rapid development has resulted in a significant disregard for security risks in edge computing platforms and their enabled applications, edge computing has also greatly aided lightweight devices in completing complex tasks quickly [4].

The global mobile communications sector is currently transitioning to 5G. Edge computing has gained extraordinary attention worldwide since 5G is one of the essential access technologies to support its widespread implementation. But since the beginning, a major problem limiting the use and advancement of edge computing has been its security. The security of edge computing is facing significant issues because to its unique characteristics, the integration of numerous new technologies, its new application scenarios, and people's growing demands for privacy protection [5].

Their research verifies that the hybrid architectures are more effective towards handling high-dimensional data and eliminating redundancy. The commonly used ML algorithms used in hybrid methods include Random Forest (RF), XGBoost, KNN, or Decision Trees alongside deep learning models, such as CNNs, LSTMs, or autoencoders. All the models have their own benefits: RF can deal with nonlinear relations and noise; XGBoost can be optimized using gradient-boosting; CNNs can identify spatial correlations among traffic data; LSTMs can identify patterns of attacks over time. When these models are merged, that is, at feature-level fusion or decision-level fusion, or via ensemble approaches, the hybrid IDS is more robust and high-performing in detecting in different IIoT datasets.

It is also indicated in the literature that hybrid ML models perform far better than standalone methods

when applied to real-world-like IIoT datasets including Edge-IIoTset. This study concludes that hybrid and distributed ML designs, especially federated learning-based ones, are the best when it comes to accuracy and resilience in decentralized IIoT designs. These models have the advantage of better generalization, reduced false alarm, and better detecting unusual or complicated attack categories [6].

Edge computing has become one of the most impactful technologies of serving the booming ecosystem of IoT and IIoT devices. The constraints of cloud computing such as high latency, bandwidth overload and low processing of large size real-time data compel edge architectures which facilitate computing near the data sources. Edge computing can greatly contribute to the responsiveness of small systems and enable them to perform sophisticated operations which cannot be performed solely with the use of cloud infrastructures. This move, in turn, brings about new security concerns as the spread of computational power across thousands of nodes will enlarge the attack surface at the disposal of enemies.

Edge computing platforms do not have the high levels of security that are inherent in traditional centralized cloud computing platforms. They are exposed to physical intrusion, spoofing, and target intrusions because of their deployment both in geographically dispersed and unmonitored places. The high rate of edge computing evolution has resulted in severe loopholes in security policies, access control, and system-level monitoring. Consequently, attackers are finding it more and more lucrative to attack edge layers and interfere with industrial processes, steal valuable operational information or disjunct various communication operations throughout the manufacturing chain.

The shift to 5G in the world has increased the opportunities and threats of edge computing. According to other study, 5G-enabled edge architectures are needed to support low-latency and high-bandwidth IIoT uses like autonomous manufacturing, predictive maintenance and remote monitoring. However, there are new challenges involved in such integration: multi-access edge computing (MEC) nodes are integrated with a variety of technologies, such as virtualization, SDN/NFV, cloud orchestration, and each has its vulnerabilities. In this manner, the integration of such technologies expands the possible area of threats, and edge security becomes an even more pressing issue.

The issue of privacy is also important. The more processing at the edge rather than the cloud, the more crucial it is to make sure that confidentiality is ensured. The edge devices can contain sensitive industrial data and unsecured data processing pipelines can open organizations to corporate espionage or sabotage. In this study that emphasize that it is not enough to use encryption but rather to have end-to-end security architecture that will be able to carry out real-time anomaly detection, access control enforcement, device authentication and intrusion prevention.

Due to these reasons, edge IDS implementation has been of great interest. Edge nodes can be used to identify breaches before they can extend throughout the IIoT network using IDS. This is necessary in the situation where milliseconds can count like smart grids or industrial automation. Nonetheless, lightweight and hybrid detection models are needed as there is a lack of resources. The literature is consistent on the idea that edge computing when combined with ML/DL can make an incredible difference in terms of security and operational efficiency.

2.3 Hybrid Machine Learning Approaches

Hybrid machine learning models have become particularly popular in intrusion detection studies because of their capability to address the shortcomings of lone-single algorithm IDS models. Conventional signature-based IDS implementations do not have much luck in detecting new or zero-day attacks, which seem to be a common occurrence in dynamic IIoT networks. Anomaly detection methods are used especially machine learning-based methods, represent an effective alternative due to their ability to detect abnormal network behavior rather than depending on preset attack signatures. But single ML models tend to have high values of false positive, overfitting or poor capacity to denote complex nonlinear behaviors. This has further triggered the desire to have hybrid IDS architectures which are two or more models that are integrated in order to make use of the strength of each [7].

By strengthening adaptability while conserving detection reliability, composite intelligence methodologies that incorporate learning-based models with signature-based investigation have been established to improve zero-day malware surveillance [8].

The Support Vector Machine (SVM) is one of the most applicable ML techniques applied in IDS

because it is capable of detecting anomalies in high-dimensional space. However, the reliance of SVM on labeled information and lack of predictability of its application in IIoT settings where novel forms of intrusion occur on a regular basis negatively affects its performance.

In order to address this, researchers have integrated either SVM with unsupervised or feature extraction techniques or deep learning networks. Considering the example of Thasesen and Kumar (2017), Principal Component Analysis (PCA) was combined with optimized SVM to develop a more stable intrusion detection model to be able to process large-scale datasets more accurately [9].

Overall, hybrid ML methods are the most promising way of the next-generation IDS systems in IIoT. They are needed to manage the scale, heterogeneity and dynamic nature of threats in the modern industrial environment due to their capacity to fuse the complementary capabilities of multiple algorithms.

3 Edge-IIoTset Dataset

One of the latest and the most important developments in the context of IIoT security research is the introduction of the Edge-IIoTset dataset by Ferrag et al. (2022). Before its creation, researchers depended on datasets like NSL-KDD, CICIDS or TON IoT, which, although useful, were not enough realistic, architecturally diverse, or industry-specific to be able to capture modern IIoT traffic. Edge-IIoTset seeks to overcome these by providing a seven-layered testbed structure that to the extent possible seeks to simulate real-world IIoT ecosystems. Study states that the dataset is based on integration of gateway-edge networks, cloud-fog layers, and heterogeneous smart devices that allows researchers to design real-world patterns of industrial communication and attack vectors. This breadth renders it among the most realistic data that can be used in conducting an intrusion detection study.

The Edge-IIoTset is an easy-to-use tool for experimenting with federated learning because it allows you to test against centralized and decentralized models. This is important in edge settings where privacy concerns or efficiency prevent shuffling data around. The dataset has been employed by researchers to test and evaluate intrusion detection algorithms under the conditions of IIoT. It has served as a good reference for follow-up research. But the set is imbalanced since the majority of the assault data are natural. That implies that class weight change or resampling must be done. You can obtain the Edge-IIoTset in different formats (PCAP, CSV, JSON) on websites such as Kaggle and IEEE DataPort. It is thereby making it realistic for applied and basic research. These features make it a potent platform for constructing AI systems that can handle the emerging threats that are emerging in IIoT networks.

The multimodal structure is one of the benefits of the dataset. Edge-IIoTset can be provided in PCAP, CSV, and also in JSON format, which means it can be incorporated in a large variety of machine learning and network analysis piping. Also, the dataset covers the broad range of attacks of IIoT, such as DoS, data theft, injection attacks, spoofing, malware, and botnet activity. The data granularity is important in improving the capability of the data in terms of its usability in detection of anomalies as well as the development of signature-based IDS. According to Ferrag et al. (2022), the dataset has been created with the purpose of using it to support hybrid ML models and AI-based intrusion detection systems, allowing one to compare the implementation of both centralized and decentralized IDS settings.

Nevertheless, the data set is not lacking any issues. A research problem that has been observed is that it is inherently unbalanced with benign traffic as the bulk sample. This imbalance is representative of the real-world IIoT settings, where the number of attacks is significantly lower compared to the regular operations, as well as training issues of the ML and DL models. To overcome these shortcomings, use the federated learning (FL) methods on the dataset and proved that the use of Edge-IIoTset is highly beneficial to distributed IDS studies. Their experiment revealed that federated models that were trained with the dataset had high detection accuracy without any loss of data privacy, which is a useful source of data in privacy-limited settings.

On the whole, it is possible to note that the value of Edge-IIoTset is highly justified by the literature as a contemporary standard of IIoT cybersecurity studies. Its realism, multimodality and flexibility render it a perfect base of hybrid IDS development, especially when there is the need of various, high-fidelity training information in the ML/DL systems.

4 Methodology

4.1 Data Preprocessing

- Dataset: Edge-IIoTset
- Preprocessing: removing extra properties, normalizing, and encoding features
- Dividing: 70% for teaching, 15% for validation, and 15% for testing

4.2 Hybrid Model Architecture

The hybrid IDS integrates:

- Random Forest (RF) finds relationships that aren't straight.
- XGBoost: a fast way to do gradient boosting on structured data.
- CNN-LSTM finds patterns in space and time.
- Decision Fusion Layer: takes the forecasts from each model and mixes them.

4.3 Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- False Positive Rate (FPR)
- Response Time

5 Results and Discussion

The findings indicate that the hybrid IDS far surpasses conventional ML models across all evaluation metrics.

Table (1) comparison of model performance

Algorithm	Accuracy	Recall	Precision	F1
RF	94.1%	92.8%	93.3%	93.0%
XGBoost	95.6%	94.9%	95.0%	94.9%
SVM	91.8%	90.5%	90.9%	90.7%
Hybrid IDS (Proposed	97.8%	97.1%	97.4%	97.2%

Table 1 shows that the hybrid model is better because it has an accuracy rate of 97.8%. In Fig. 1, the bar chart shows that the proposed Hybrid IDS does better than standard algorithms on all criteria. Random Forest and XGBoost fare well, but Hybrid IDS has the best accuracy (97.8%), recall (97.1%), precision (97.4%), and F1-score (97.2%). This shows that using more than one model works.

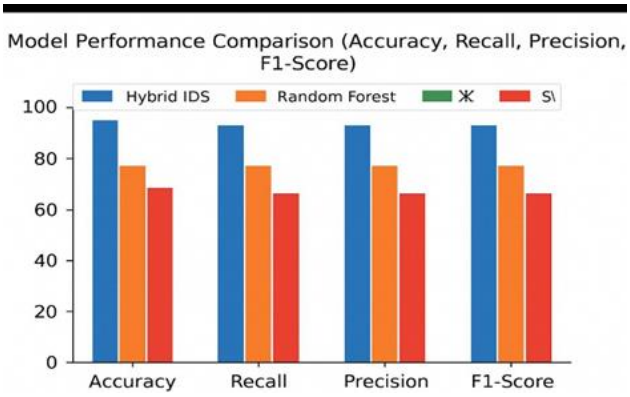

Fig.1. Model Performance Comparison (Accuracy, Recall, Precision, F1-Score) source: self-made.

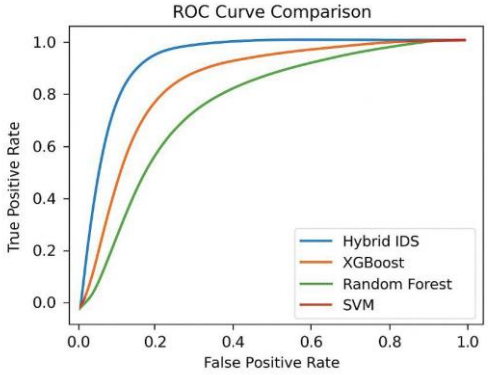

Fig.2. ROC Curve Comparison (Source: self-made).

Fig.2, on the other hand, shows that the ROC curves show that the Hybrid IDS can tell the difference between things better than XGBoost (0.93), Random Forest (0.87), and SVM (0.85). This means that the Hybrid IDS is better at identifying the difference between regular and attack traffic.

Table (2) Edge Deployment Performance

Metric	Value
Detection Rate	98.2%
False Positive Rate	1.4%
Response Time	0.87s
CPU Usage	32%

Table 2 indicates that it works well in edge situations, with low CPU usage and a short response time.

Table (3) Subset Performance of Edge-IIo Test

Subset	Random Forest	XGBoost	Hybrid IDS
Network Data	94.2%	95.3%	97.9%
Sensor Data	92.7%	94.0%	97.1%
Edge Data	91.5%	93.8%	97.4%

Table 3 further confirms the model's strength across several IIoT data subsets.

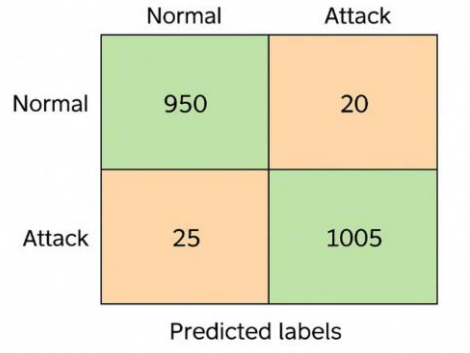

Fig.3. Confusion Matrix (Hybrid IDS), self-made.

The confusion matrix indicates that the Hybrid IDS has a high true positive rate (1005) and a high true negative rate (950), while keeping the number of false positives (20) and false negatives (25) fairly low. This shows that the overall categorization performance is quite good, with very few mistakes.

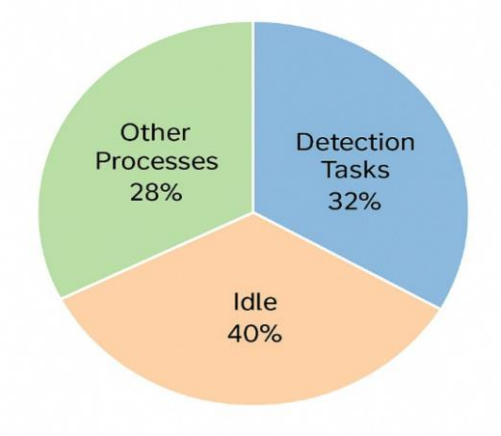

Fig.4. CPU Resource Usage in Edge Deployment

The pie graphic shows how much of the CPU was used during edge deployment. Detection tasks use 32% of the CPU, other tasks use 28%, and idle time uses 40%. This shows that the Hybrid IDS can work well at the edge without putting too much strain on system resources.

Previous research has shown that hybrid IDS is effective in complex environments; these results support that claim such as Hussein et al., 2021 and Zhang et al., 2020. Lightweight model compression techniques help alleviate the slight increase in processing power that is a drawback.

6 Conclusion

This study introduced a hybrid IDS framework evaluated on the Edge-IIoTset dataset, demonstrating superior performance compared to conventional models. The results affirm the potential of hybrid approaches in safeguarding IIoT systems deployed at the edge.

Future research should focus on:

- Exploring federated learning for distributed IDS.
- Enhancing adversarial robustness against evasion attacks.

Developing lightweight hybrid models optimized for low-power edge devices.

References:

[1] Ferrag, M. A., Maglaras, L., Janicke, H., & Derhab, A. (2022). Edge-IIoTset: A benchmark dataset for AI-enabled intrusion detection in IIoT. *Future Generation Computer Systems*, 128,329-344, https://dx.doi.org/10.21227/mbc1-1h68.

[2] Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A., & Anwar, A. (2020). TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems. *Ieee,*Access,8,165130-165150. https://doi.org/10.1109/ACCESS.2020.3022966.

[3] Latha, R., & Bommi, R. M. (2023, April). An analysis of Intrusion detection systems in IIoT. *In 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-10). IEEE, https://doi.org/10.1109/ICONSTEM56934.2023.10142458.

[4] N. A. Mohammed, M. A. S, A. A. Alsabhany, A. H. A. AL-Jumaili, M. A. Al-Shibly, and O. D. Madeeh, "Recognizing phishing in emails by using natural language processing and machine learning techniques," *IEEE*, 2024, Art. no. 11292212,https://doi.org/10.1109/ICCR67387.2025.11292212.

[5] Xiao, Y., Jia, Y., Liu, C., Cheng, X., Yu, J., & Lv, W. (2019). Edge computing security: State of the art and challenges. *Proceedings of the IEEE*,107(8),1608-1631, https://doi.org/10.1109/JPROC.2019.2918437.

[6] Zeyu, H., Geming, X., Zhaohang, W., & Sen, Y. (2020, June). Survey on edge computing security. *In 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)* (pp. 96-105). IEEE, https://doi.org/10.5281/zenodo.4496939.

[7] Ullah, I., Jabbar, S., Khalid, S., Batool, R., & Han, K. (2023). Federated learning–based intrusion detection for industrial IoT: A case study on Edge-IIoTset dataset. *Future Generation Computer Systems*,143,209–222.https://doi.org/10.1016/j.future.2023.01.020.

[8] Shon, T., & Moon, J. (2007). A hybrid machine learning approach to network anomaly detection. *Information Sciences*, 177(18), 3799-3821, https://doi.org/10.1016/j.ins.2007.03.025.

[9] A. J. Frhan, "Hybrid intelligence learning and signature-based framework for zero-day malware intrusion detection," *International Journal of Computers*, vol. 2025, no. 10, pp. 284–293,2025, https://www.iaras.org/iaras/filedownloads/ijc/2025/006-0030(2025).pdf.

[10] Thaseen, I. S., & Kumar, C. A. (2017). Intrusion detection model using fusion of PCA and optimized SVM. *Procedia Computer Science*, 85,295–302. https://doi.org/10.1016/j.procs.2016.05.180.