A Robust Hybrid Intrusion Detection Approach for Industrial IoT Networks Based on the Edge-IIoTset Dataset

AMJAD JUMAAH FRHAN[1] *, ALI L. A. AL-ZAIDI[2]
[1]Department of Education and Islamic Studies, Sunni Endowment, Baghdad, IRAQ
[1]University of AL Mashreq, Department of Cybersecurity Engineering Technology, Baghdad, IRAQ
[2]Department of Religious Education and Islamic Studies, Iraqi Sunni Affairs, Baghdad, IRAQ
Coresponding Author

Abstract: - The Industrial Internet of Things (IIoT) presented up new safety obstacles especially how to keep safe networks as well as edge endpoints. The ever-changing nature and complexity of IIoT communications makes traditional Intrusion Detection Systems (IDS) inadequate. Using Artificial Neural Networks (ANNs) and Machine Learning (ML) on the Edge-IIoTset dataset, this study presents an IDS hybrid model in this study. Improving detection performance and reducing false alarms are achieved through the employment of ML-DL algorithms. Investigations show that a combined approach may reach high levels of accuracy, recall, and precision when compared with solo methods. Statistics and information tables show the model's competence and validation in specific IIoT scenarios.

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

Received: June 22, 2025. Revised: October 11, 2025. Accepted: November 3, 2025. Published: January 27, 2026.

1 Introduction

The profitability of business has altered as a result of the manufacturing sectors use of IIoT technological advances. 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]. Fig.1 shows the procedure that used in this research starting with data collection, processing, following by training and testing phase.

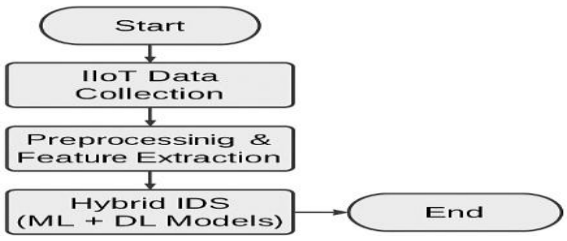

Fig. 1. Shows the main pipeline of this study, By author.

As a way to strengthen IIoT security, machine learning (ML) and deep knowledge (DL) implementation in intrusion detection has drawn a lot of attention nowadays. Whereas hybrid architectures have shown greater breadth in identifying both known and unexpected threats, traditional IDS systems overwhelmingly rely on signature-based or anomaly-based procedures [9]. Additionally, comparison datasets like Edge-IIoTset can be especially useful since they allow investigators to verify detection algorithms in practical situations by simulating authentic industrial contexts with a variety of attack vectors. This emphasizes just how important it is to use hybrid designs that strike the right balance between processing economy and performance in order to keep IDS solutions viable for devices at the edges with resources that are scarce.

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].

One common piece of software that keeps an eye on and encourages security justifications for computer networks is an intrusion detection system (IDS). The solutions deployment attempts to detect malicious activities and implement actions that promote risk aversion. Implementing typical IDS-based solutions is challenging because of the uniqueness of IIoT. This comprises heterogeneous architecture, sensitive data, and scarce resources. For efficient and adaptable IDS implementations in various IIoT settings, researchers are putting fog/edge computing, machine learning (ML), and deep learning into practice [3].

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].

2.3 Edge-IIoTset Dataset

The Edge-lloTset dataset is perhaps the most significant new set of data for validating the security of Industrial Internet of Things (IIoT) devices and networks that are located on the edge. Ferrag et al. proposed realism testbed of seven layers in 2022. It consists of the gateway edge network, cloud fog and smart devices. It is a more realistic setup compare to current datasets such as NSL-KDD and CICIDS. 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 Data Port. 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 [6]. Uses with a security focus, like threat monitoring, intrusion detection, and surveillance [10]. By strengthening dataset quality, adding semi-automated annotation for quicker responsiveness to changing threats, and applying the methodology to real-time detection in cyber-physical and critical infrastructure environments, further studies can use the methods they propose to fortify protection mechanisms.

2.4 Hybrid Machine Learning Approaches

Worms and spyware are examples of zero-day cyberattacks that are becoming increasingly widespread and damaging. It's often hard to find these kinds of attacks with the current signature-based intrusion detection systems. Anomaly

intrusion detection systems have been developed to counteract such attempts. The Support Vector Machine (SVM) is one of the best machine learning methods for finding strange behaviours among the many ways to do anomaly detection. The soft-margin SVM is a common basic SVM method that involves supervised learning. The soft-margin SVM method, on the other hand, needs pre-acquired learning material for supervised learning, which makes it unsuitable for finding new assaults in Internet data. Normal and attack traffic are marked separately from any pre-existing learning data [7].

The Hybrid Machine Learning Approaches using more than one algorithm or technique in the same framework to make things operate better or get around the problems that each algorithm has when used alone. For instance, you can use supervised learning algorithms with unsupervised learning, statistical models like SVMs with neural network-based methods, or even old-fashioned methods with deep learning. These methods are significance in many apps such as "intrusion detection, text classification, picture recognition, and natural language processing ". The methods also, have an important side with huge, unbalanced or complex data [8].

3 Methodology

IIoT IDSs tackle issues including data imbalance, real-time detection, and lightweight periphery device models all of which are critical given limitations in resources and heterogeneous networks as shown in Fig.2.

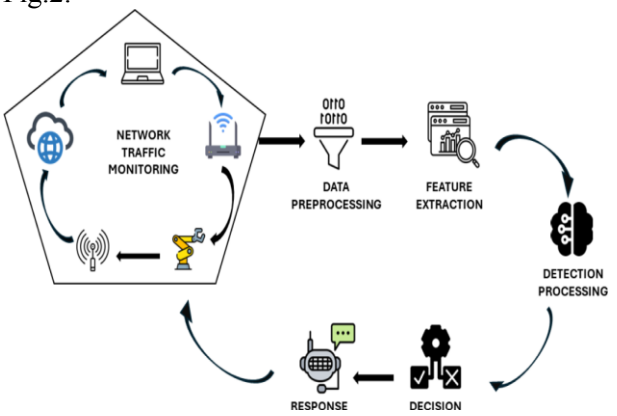

Fig.2. Intrusion Detection via Networks for Robotics and Industrial Applications [9].

The Dataset preprocessing by removing extra properties, normalising, and encoding features. Dividing 70% for teaching, 15% for validation, and 15% for testing.
The hybrid IDS integrates several complementary components to improve detection performance. Random Forest (RF) is employed to capture nonlinear relationships within the data, while XGBoost provides a fast and efficient gradient boosting approach for structured data. The CNN-LSTM component is used to detect complex spatial and temporal patterns, enhancing the model's ability to recognize sophisticated attacks. Finally, a decision fusion layer combines the predictions from each individual model, producing a more accurate and robust overall detection outcome.

The Evaluation Metrics percent the accurately identified incidences (attack and normal) that comprise all cases is designated as accuracy see Eq. (1).

$$Accuracy(Acc) = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Where:

TP= True Positives (attacks correctly detected)

TN= True Negatives (normal events correctly classified)

FP= False Positives (normal events misclassified as attacks)

FN = False Negatives (attacks misclassified as normal)

Precision measures how many of the instances classified as attacks are actually attacks see Eq. (2).

$$Precision(P) = \frac{TP}{TP + FP} \quad (2)$$

True Positive Rate or Sensitivity (Recall) shown in Eq. (3).

$$Recall(R) = \frac{TP}{TP + FN} \quad (3)$$

The Eq. (4) is the F1-score is the harmonic mean of precision and recall, balancing the two metrics

$$F1\ score = \frac{2(P+R)}{P+R} \quad (4)$$

The percentage of typical incidents that were mistakenly labelled as attackers is recognized as the False Positive Rate (FPR) see Eq. (5).

$$FPR = \frac{FP}{FP+TN} \quad (5)$$

Response time estimates the period that it takes the IDS recognize an occurrence or sound a warning tone. Typically, this is shown as an average over several encounters see Eq. (6).

$$Response\ Time = \frac{\sum_{i=1}^{N} t_i}{N} \quad (6)$$

Where $\boldsymbol{t_i}$ is the detection time for the i_{th} event and N is the total number of events.

Table 1 shows that the hybrid model is better because it has an accuracy rate of 97.8%. In Fig. 3, 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.

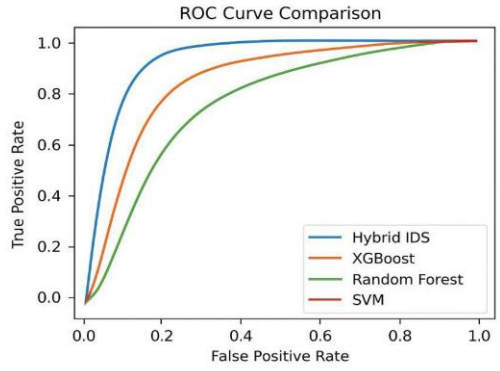

Fig. 4. ROC Curve Comparison

Fig. 4, 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	RF	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.

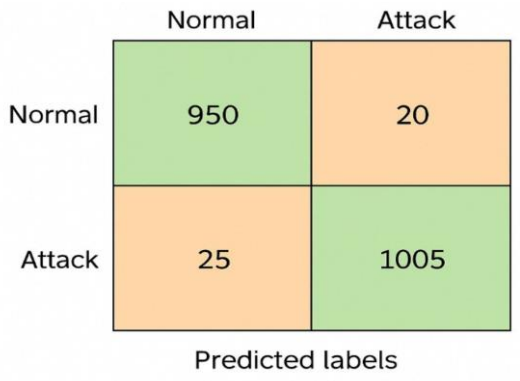

Fig. 5. Confusion Matrix (Hybrid IDS).

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 categorisation performance is quite good, with very few mistakes.

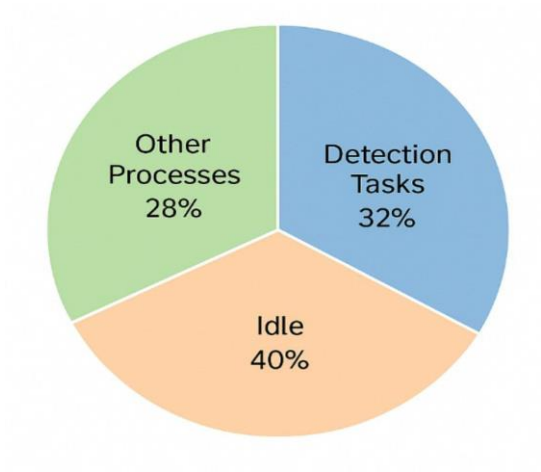

Fig.6. CPU Resource Usage in Edge Deployment.

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.

4 Results and Discussions

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

TABLE 4. Shows the comparison of model performance in (%).

Model	Acc	R	P	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	97.8	97.1	97.4	97.2

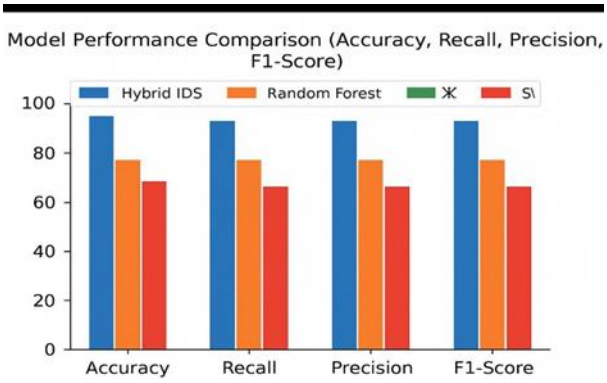

Fig.3. Model Performance Comparison (Accuracy, Recall, Precision, F1-Score).

5 Future Work

Future research should focus on exploring federated learning approaches for distributed intrusion detection systems (IDS), which would allow multiple IIoT devices to collaboratively train models while preserving data privacy.

Additionally, enhancing adversarial robustness is crucial to protect IDS models against evasion attacks that attempt to bypass detection. Finally, developing lightweight hybrid models optimized for low-power edge devices is essential to ensure efficient and real-time intrusion detection in resource-constrained IIoT environments. In order to strengthen trust and decision-making in IIoT contexts, future research could look into understandable IDS architectures that offer transparent alarms [10].

A promising foundation for additional investigation showed federated and semi-supervised active learning, which presents a collaborative and scalable scheme that strikes a balance between learning efficiency and privacy preservation. This strategy can be expanded further through incorporating cutting-edge deep learning architectures, improving the effectiveness of communication in federated situations, or using it in a variety of fields where collaborative and privacy-sensitive data management [11] is important, including smart grids, medical imaging, and Industrial IoT.
By implementing an affordable gradient boosting strategy and complementing constrained training data with CGAN-generated data samples, the SmartShield framework introduces a useful IoT intrusion detection strategy that improves incident recognition on current IoT traffic [12].

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. The suggested blended IDS established its efficacy in identifying intruders in IIoT settings by excelling conventional models and gaining the greatest accuracy of 97.8% along with superior precision, recall, and F1-score. The results presented demonstrate how hybrid methods, contrasted to single categories like Random Forest, XGBoost, or SVM, can boost security.

References:

[1] M. A. Ferrag, L. Maglaras, H. Janicke, and A. Derhab, "Edge‐IIoTset: A benchmark dataset for AI‐enabled intrusion detection in IIoT," *Future Gener. Comput. Syst.*, vol. 128, pp. 329–344,Mar. 2022, https://dx.doi.org/10.21227/mbc1-1h68.

[2] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and A. Anwar, "TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data‐driven intrusion detection systems," *IEEE* Access, vol. 8, pp. 165130–165150, Sep. 2020, doi: 10.1109/ACCESS.2020.3022966, https://doi.org/10.1109/ACCESS.2020.3022862.

[3] R. Latha and R. M. Bommi, "An analysis of intrusion detection systems in IIoT," *in Proc. 2023 8th Int. Conf. Sci. Technol. Eng. Math. (ICONSTEM*), Chennai, India, Apr. 2023, pp. 1–10,https://doi.org/10.1109/ICONSTEM56934.2023.10142458.

[4] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu, and W. Lv, "Edge computing security: State of the art and challenges*," Proc. IEEE*, vol. 107, no. 8, pp. 1608–1631, Aug. 2019, https://doi.org/10.1109/JPROC.2019.2918437.

[5] H. Zeyu, X. Geming, W. Zhaohang, and Y. Sen, "Survey on edge computing security," *in Proc. 2020 Int. Conf. Big Data, Artif. Intell. Internet Things Eng. (ICBAIE), Fuzhou, China*, Jun. 2020, pp. 96–105, https://doi.org/10.1109/ICBAIE49996.2020.00027.

[6] I. Ullah, S. Jabbar, S. Khalid, R. Batool, and K. Han, "Federated learning–based intrusion detection for industrial IoT: A case study on Edge‐IIoTset dataset," *Future Gener. Comput. Syst.*, vol. 143, pp. 209–222, Jun. 2023, doi: 10.1016/j.future.2023.01.020.

[7] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection," *Inf. Sci*., vol. 177, no. 18, pp. 3799–3821,Sep.2007,https://doi.org/10.1016/j.ins.2007.03.025.

[8] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of PCA and optimized SVM," *Procedia Comput. Sci*., vol. 85,pp.295–302,2016,doi: 10.1016/j.procs.2016.05.180.

[9] R. Holdbrook, O. Odeyomi, and Y. Sun, "Network‐based intrusion detection for industrial and robotics systems: A comprehensive survey," *Electronics*, vol. 13, no. 22,Art.no.4440,Nov.2024,doi: 10.3390/electronics13224440.

[10] N. A. Mohammed et al., "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.

[11] 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.

[12] A. J. Frhan, "SmartShield: A CGAN-boosted model for detecting IoT cyber threats," *International Journal of Computers*, 2025, pp. https://www.iaras.org/iaras/filedownloads/ijc/2025/006-0031(2025).pdf.