ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 09, September 2026

Trust-Aware Artificial Intelligence Frameworks for Secure Intrusion Detection in Industrial IoT Networks: A Comprehensive Review

[1] Tejas Bajirao Tambe, [2] Dr. Saroj Kumar Nanda

[1] [2] Ajeenkya D. Y. Patil University, Pune, India
Email: [1] Tejastambe9@gmail.com, [2] sarojkumarnanda1979@gmail.com

Abstract*— The rapid growth of the Industrial Internet of Things (IIoT) systems has changed the landscape of automation and smart manufacturing and has presented intricate security issues. Traditional intrusion detection systems are unable to accommodate the dynamism, heterogeneity, trust sensitivity of IIoT systems. This review discusses the development of trust-sensitive artificial intelligence (AI) models that increase the detection accuracy using behavioral trust modeling, relationship analysis, and federated learning. The main goal is to analyse the potential to enhance AI-based detection models using trust metrics which are based on the reputation of the device, social relations, and interaction patterns to prevent malicious actions and data poisoning. A variety of benchmark datasets is tested to model a wide range of IIoT security scenarios: the SIoT Network is a social IoT interaction and topology-based trust construct; the WSN Secure and Energy-Efficient one focuses on the energy optimization and data integrity balance; Edge-IIoTset focuses on centralized and federated building designs; ToN-IoT provides a combination of protocol-rich traffic and multiple attack vectors; and BoT-IoT provides large-scale botnets as a way to benchmark. It was found in the analysis that the combination of trust evaluation mechanisms and AI-based intrusion detection does contribute to accuracy, energy efficiency, and interpretability significantly. These methods can be shown to be highly promising in edge deployments and federated deployments where scalability and privacy are of key concern. The authors further find that the dynamism, explainability, and trustfulness of AI security frameworks are vital to achieve robust security protection against the Industry 4.0 and next-generation IIoT ecosystems.

Index Terms: Trust-Aware AI, Industrial IoT Security, Intrusion Detection Systems, trust metrics, Federated Learning, Edge Computing, Social IoT Networks.

I. INTRODUCTION

Industrial ecosystems have been redefined by the fast development of the Industrial Internet of Things (IIoT) that connects machines, sensors, and cyber-physical systems to make industrial automation intelligent and predictive, as well as to create self-optimizing production in Industry 4.0 settings. Nevertheless, such hyperconnected character predisposes IIoT infrastructures to a range of cyber threats never seen before, such as distributed denial of service (DDoS) and insider threats and data manipulation, both of which can disrupt the continuity of operations and data integrity [1], [2]. The classical Intrusion Detection Systems (IDS) that were developed to operate on traditional IT-based architectures perform poorly in securing these dynamical, heterogeneous, and resource-constrained IIoT networks because they are not scalable, lack contextual awareness, and are not adaptable to changing attack surfaces [3], [4]. Recent studies emphasize that there is an urgent necessity to shift to trust-conscious artificial intelligence (AI) systems that are built on behavioral modeling, relationship analytics, and adaptive learning to provide an effective IIoT security [5]. Unlike the traditional models of IDS, which mainly pay attention to the traffic patterns and rule-based anomaly detection, trust-aware models include contextual intelligence through the measurement and analysis of the reliability of a device, its reputation, and interaction behaviour in distributed environments [6]. These type of trust assessment systems are essential in IIoT ecosystems where autonomous communication of devices is a reality, often at a different level of reliability and without a central authority and thus providing a place through which malicious nodes or corrupted devices can spread attacks without detection [7]. The simplicity of protecting IIoT has to do with its heterogeneous topology and multi-layered communication model where varied industrial protocols and types of data coexist, including wireless sensor networks (WSN) through edge-cloud models. Besides, with the development of IIoT systems toward decentralized frameworks based on edge computing and federated learning, it becomes even harder to maintain data privacy, energy efficiency, and cross-domain trust [8]. An example is federated intrusion detection systems like FedIn-NID and Edge-IIoTset, which can be used to illustrate how the use of distributed model training can increase the robustness of the detection mechanisms without compromising the local data confidentiality [9]. However, these systems are susceptible to threats of trust like malicious updates, poisoned edges, and inconsistent propagation of trust [10].

Trust-conscious artificial intelligence thus attempts to enhance conventional learning-based IDS by developing multi-dimensional trust measures such as behavioral similitude, topologies relations, communication integrity and past reputation in the detection pathway. The work by Ahmed

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 09, September 2026

et al. [1] and Bekri et al. [7] emphasize that federated learning with trust management enhances the accuracy of detection as well as reliability of nodes within distributed IIoT networks. Likewise, hybrid deep learning networks like GCN+LSTM [6] and fuzzy graph attention networks [11] have demonstrated prospective outcomes in modeling relational dependencies between the entities within the network, which reinforces the performance of intrusion detection in dynamic environments. Finally, the recent trend of trust-based AI-based IDS is a paradigm shift in IIoT cybersecurity. These frameworks use smart calculation of trust, co-ordination of models across a distributed system, and response mechanisms to counteract advanced attacks at real time. This paper presents these structures in their entirety, examining their architecture, datasets, challenges, and trends, to provide a unified framework of intrusion detection in Industry 4.0 settings, which is safe, trustworthy, and self-learning.

The main aim of this review is to give a detailed discussion of trust-conscious artificial intelligence (AI) systems designed to be used in secure intrusion detection in the Industrial Internet of Things (IIoT) systems. The proposed study will examine how a unified approach based on trust models, behavioral analytics, and intelligent decision-making engines, can enhance the precision of detection, versatility, and strength against advanced cyber threats in multi-faceted IIoT systems. This review focuses on the intersection of trust management and AI by noting how the reliability and transparency of intrusion detection frameworks can be improved with the help of dynamic trust measures, which are based on the behavior of the device, reputation, and relationships between the parties of interaction.

The expanded nature of this review comes in formulating open issues and research gaps that can be utilized in the field. The main problems are the control of heterogeneous network structures, the avoidance of trust drift and adversarial manipulation, the collaboration of models that preserve privacy, as well as the deployment that can operate efficiently with limited resources on edges. Another aspect of explainable AI (XAI) examined in the paper is the need to enhance model transparency and user trust in decisions based on trust in intrusion detection models.

The major contributions of the Paper given as:

- ✓ Gives a detailed discussion on the role of trust modeling and behavioral intelligence and federated learning in improving intrusion detection in Industrial IoT (IIoT) settings.
- ✓ Compares six major IIoT datasets, namely, SIoT Network, WSN Secure and Energy-Efficient, Edge-IIoTset, ToN-IoT, X-IIoTID and BoT-IoT, systematically to evaluate their relevance, diversity and applicability to trust-based AI security evaluation.
- ✓ Identifies outstanding challenges, including scalability, preserving privacy, and explainability, as avenues of future research to create resilient industry 4.0 networks with energy-saving and trustworthy intrusion detection systems.

II. BACKGROUND AND THEORETICAL FOUNDATION

A. Industrial IoT Security Architecture and Threat Landscape

The Industrial Internet of Things (IIoT) is based on the complex multi-layer architecture providing the integration of sensing, communication, processing, and application layers into each other, which allows making intelligent automation and real-time decisions. Poor authentication and lack of trust control exposes the application layer to malwares, unauthorized access, and insider abuse, which offer services like tracking services, analytics, and decision automation [10], [11].

These vulnerabilities are increasing with the growing assault area of IIoT systems as millions of heterogeneous devices inter-network among production networks, supply chains and cloud systems. Conventional solutions to intrusion detection, which were initially developed to support the environment of static IT networks, are not able to manage the volume of data speed, device variety, and protocol heterogeneity of the IIoT environment [12]. Villafranca et al. [10] suggested a single AI-based intrusion detection framework using which IIoT anomaly detection and adaptive control methods are integrated and pointed out that IIoT systems should consider data diversity and contextual correlation. Benkhelifa et al. [14] also found out that traditional IoT intrusion detection systems are not sufficiently scalable and resilient to use in the industrial context, as they tend to be very slow to generalize in multi-protocol contexts and dynamic threat environments. The latest research has examined the concept of multi-layer security control, as well as AI-improved intrusion detection models to enhance the defensive position of IIoT systems. Kumar et al. [11] came up with a secure blockchain-based IDS which ensures data immutability and decentralized validation which reduces spoofing and insider attacks in distributed industrial systems. Deng [12] proposed a hybrid LightGBM and multi-layer perception network that is more effective at classifying anomalies by extracting lightweight but discriminatory features of IIoT traffic. In the same vein, Li et al. [13] introduced the CRSF framework that combines CNN2D-RNN and SVM classifiers to identify sophisticated spatiotemporal intrusion patterns in industrial control systems, which yields significant resistance to dynamic DoS and data manipulation threats.

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 09, September 2026

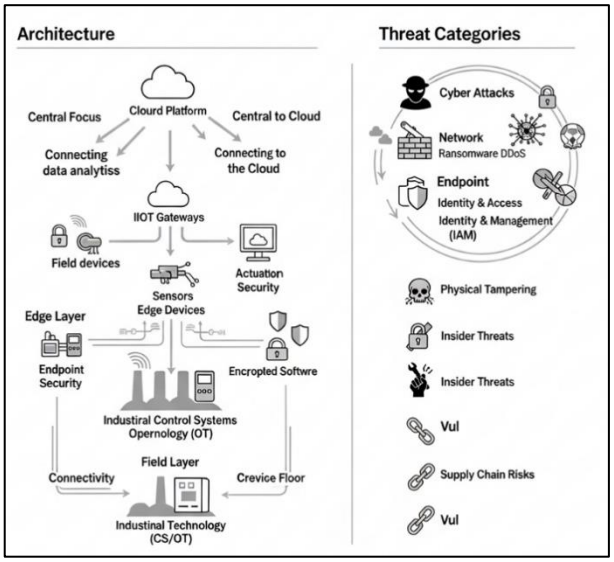


Figure 1. Industrial IoT Architecture and Threat Categories

This figure 1 provides the multi-layer structure of an Industrial IoT (IIoT) ecosystem and the discovered cyber threat environment. In the left part, hierarchical levels of the field devices and sensors on the edge, IIoT gateways and cloud platforms are displayed with a focus on data connectivity, endpoint security, and actuation control. Key industrial threats are classified into the right section and they include cyberattack (ransomware, DDoS, IAM breaches), physical intrusion, insider threat and supply chain vulnerabilities. Another interest, as shown in the literature, is growing in dataset realism and contextual modeling. Kumar et al. [15] introduced a security testbed, IoTForge Pro that facilitated the generation of realistic industrial intrusion datasets to overcome the absence of standardized testing environments of industrial IDS studies. These trends show that intrusion detection precision can be significantly enhanced using deep learning and hybrid AI architectures, which is contextual and trust-aware modeling without compromising energy-efficiency and scalability. However, there are still unresolved issues such as real-time detection latency, adaptive trust management, and cross-domain generalization, as well as other problems, to roll these systems out to a production-scale IIoT network.

B. Artificial Intelligence in Intrusion Detection

AI has taken over as a core element of the contemporary Intrusion Detection Systems (IDS) in Industrial Internet of Things (IIoT) environments, which is capable of automated learning, anomaly detection and pattern recognition within heterogeneous networks. Support Vector Machines (SVM), Decision Trees (DT) and Random Forests are machine learning (ML) algorithms that have been extensively used to recognize abnormal traffic behaviors through the use of past data to inform the decision. Nevertheless, all these classic paradigms tend to fail when dealing with high-dimensional, imbalanced, and dynamic IIoT data and thus less generalization performance is achieved during unknown attack situations [10]. To address these issues, deep learning (DL) algorithms have been embraced to identify hierarchical features and to identify intricate intrusion patterns through the use of Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), or hybrid CNN-LSTM applications. Experiments such as that of Li et al. [13], have shown that deep hybrid models can capture spatial and temporal correlated data in industrial network traffic and be more effective in detecting multi-vector attacks such as DoS and probe intrusions than the shallow ML models.

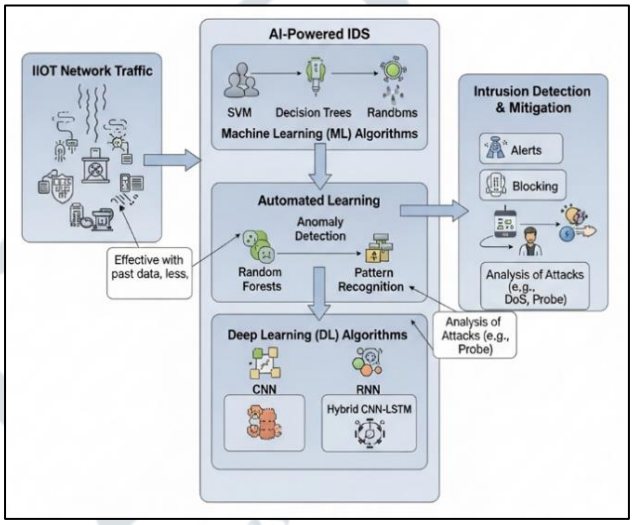


Figure 2. Artificial intelligence in intrusion detection for IIoT

This figure 2 represents an AI-based IDS architecture that takes traffic IIoT network traffic amounts to machine learning and deep learning algorithms to detect anomalies, analyze attacks and automatically block intrusion through automated detection alerts and blockers. In spite of such advances, AI single-only models have obvious limitations in IIoT applications. The majority of the DL models presuppose data independence and reliability, which is impossible in heterogeneous and semi-trusted industrial ecosystems [11]. Recently, federated learning (FL) is also a prospective approach to distributed intrusion detection that enables multiple IIoT devices to cooperatively learn global models without data exchange, ensuring better privacy and scalability. FedIn-NID framework is proposed by Mao et al. [25], and it is shown that intrusion detection in large-scale and heterogeneous IIoT systems can be significantly enhanced with the help of federated learning. Likewise, Hu et al. [23] proposed a privacy-sensitive FL method to SDN-based IIoT that has a high detection rate and a low communication load. Nevertheless, these systems are not yet trust-awarely coordinated and thus are prone to shipment of bad updates, malicious actors, and gradient exploitation. Lack of some form of trust evaluation means that AI-based IDS can spread incorrect knowledge among nodes, making the entire mechanism less reliable in detection [18]. This

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 09, September 2026

weakness highlights the dire necessity to implement trust models in AI-based IDS systems to provide secure, explainable, and robust decision-making in IIoT systems.

C. IIoT Systems Trust Modeling

Trust modeling is another IIoT security intelligence layer, which is intended to analyze the reliability, credibility, and actions of devices involved in industrial connections. The trust, in this case, is not limited to authentication but is rather dynamically determined in terms of its use when entities interrelate, cooperate, and sustainable performance under various conditions of operating. The trust measures are generally divided into three; behavioral trust which is a result of previous communication and consistency of behavior; social trust which is an outcome of relationships strength, reputation and peer opinion; and contextual trust which depends on environmental characteristics including network condition, latency and energy supply [7].

Padmavathi and Saminathan [18] suggested a federated edge intelligence platform based on trust-based access control, in which the calculation of trust scores is performed using historical interactions and energy values to address data sharing decisions between IIoT nodes. A federated machine learning based trust-aware authentication model was proposed by Ahmed et al. [1], revealing that false alarms are minimized and more robustness against malicious edge nodes are achieved by introducing trust weight in the learning processes. In that manner, Bekri et al. [7] have created a distributed trustful IoT architecture based on the Software-Defined Networking (SDN), AI, and blockchain, to provide safe communications and self-adaptive propagation of trust in industrial ecosystems. Previous trust models tended to use more fixed or centralized ways of evaluation thus they are not suited to a more dynamic and huge IIoT setting. As an example, in federated or multi-layer industrial networks, time-varying behaviors are challenging to traditional reputation-based trust systems, and context drift cannot be addressed. In contemporary models, there are dynamic trust computation methods which involve updating the trust score on a periodical basis with feedback, probabilistic models or reinforcement learning approaches [5]. This is to make trust dynamic to adjust to the behavior of the devices and the emerging threats. There are substantial advantages to combining trust computation with AI-based intrusion detection: it is possible to selectively aggregate the models in the process of federated learning, to filter irrelevant sources of data, and improve the interpretability of the systems by correlating the results of detecting problems with trust. Nevertheless, the standardized metrics of trust, cross-domain transfer mechanisms, and lightweight computation models to be adapted to resource-constrained IIoT nodes are yet to be defined through further research.

Table 1: Summary of AI-Based and Trust-Aware IDS Approaches in IIoT

Study / Year	Approach / Model	Key Features	Limitations	Scope for Future Work
Villafranca et al. [10]	AI-enabled unified IDS	Conceptual AI framework integrating anomaly detection and adaptive control	Lacks trust metrics and dynamic context awareness	Incorporate trust-based decision intelligence
Kumar et al. [11]	Blockchain-based IDS for IoT	Decentralized verification and data immutability	Limited scalability for large IIoT	Extend to hybrid AI–trust systems
Deng [12]	LightGBM + MLP Hybrid IDS	Efficient feature learning for IIoT traffic	Ignores relational trust dependencies	Integrate behavioral trust modeling
Li et al. [13]	CNN2D-RNN-SVM Hybrid Framework	Captures spatiotemporal attack patterns	High computational cost	Apply lightweight trust integration
Benkhelifa et al. [14]	Critical Review of IoT IDS Challenges	Identifies gaps in generalization and resilience	No trust-based validation layer	Introduce adaptive trust computation
Kumar et al. [15]	IoTForge Pro Dataset Generation Platform	Realistic IIoT intrusion dataset	Lacks trust-labeled data	Develop trust-annotated datasets
Padmavathi & Saminathan [18]	Federated Edge Intelligence with Trust	Trust-based access control and adaptive updates	Energy overhead in edge nodes	Optimize trust computation cost
Mao et al. [25]	FedIn-NID Federated IDS	Distributed learning for heterogeneous IIoT	Vulnerable to poisoned updates	Integrate trust-based participant validation

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 09, September 2026

III. TRUST-AWARE AI FRAMEWORKS FOR SECURE IIOT

A. Trust-Aware IDS Framework Taxonomy

The Industrial Internet of Things (IIoT) Trust-aware Intrusion Detection Systems (IDS) could be broadly classified as centralized, distributed and federated architectures with the ability to provide various advantages and trade-offs in terms of scalability, data privacy and real-time adaptability. Under centralized structures, intrusion detection and trust assessment is done on a central control node or a cloud server which gathers information on all the devices that are connected to it. An Online Network IDS (OI-SVDD + AS-ELM) which is designed by Gyamfi and Jurcut [16] is applicable in centralized industrial networks and is used to conduct real-time classification, and with high accuracy of detection, but low scalability because it heavily depends on central computing.

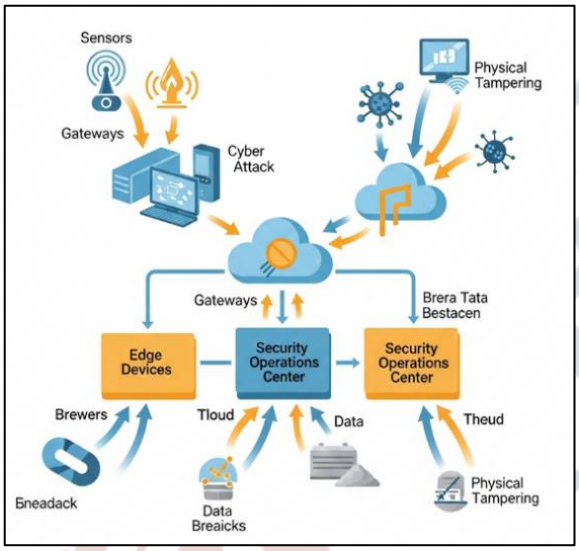


Figure 3. Security Operations Framework for IIoT Networks

This figure 3 provides a depiction of how edge devices, gateways and security operation centers communicate with each other in a system of Industrial IoT. It notes cyberattacks, data breaches, and physical tampering among the significant threats, and coordinated defense against them using centralized and edge-level security monitoring and response systems. In comparison, distributed frameworks devolve both the intrusion detection and trust assessment operations of numerous nodes, mitigating the threat of central compromise, and permitting more rapid local responses. Nizamudeen [17] proposed a deep learning framework as a swarm based model to operate in the multi-cloud IoT setting that enables individual nodes to self-learn and share detection outcomes, enhancing their resistance to dynamic network attacks. Padmavathi and Saminathan [18] proposed a federated edge intelligence method that uses trust-based access control, where the intrusion detectors are trained by the edge nodes and the trust scores are updated dynamically with the goal of blocking unreliable participants. This idea was further developed by Mao et al. [25], who introduced FedIn-NID, and the framework can be used to build a high-quality intrusion detection system in heterogeneous IIoT systems through the combination of global aggregation and node-level trust validation. Moreover, both Bekri et al. [7] and Hu et al. [23] have found that the edge intelligence with cloud orchestration is more scalable with complex tasks being offloaded to the cloud layers, yet trusted collaboration is maintained at the edge, leading to better latency characteristics and contextual awareness in the industrial realms.

B. Cooperation of Trust and AI Algorithms

A combination of AI-driven intrusion detection algorithms with the mechanism of trust computation is the main part of modern trust-aware IIoT systems. As an example, Gyamfi and Jurcut [16] used selective learning to minimize false positives in the classification of anomalies by using trust indicators. On the same note, Nizamudeen [17] was using swarm-based optimization to dynamically tune the trust weights, which allowed adaptive feature relevance to adapt to varying network settings.

Another critical dimension is model adaptation, and in this case, the information about trust has a direct impact on training and inference processes of AI algorithms. Padmavathi and Saminathan [18] incorporated the access control based on trust in federated edge nodes so that participants that can be trusted to contribute to global model only do so. The method reduces the chances of model poisoning and improves the accuracy of federated intrusion detection. Hu et al. [23] contributed more to this concept by introducing a privacy-saving federated learning model that uses trust scoring and secure aggregation, thereby sustaining the trade-off between detection and communication rate.

An increasing literature is based on hybrid models incorporating trust graphs with neural network architectures to use both structural and behavioral intelligence. Latif et al. [22] presented a Dense Random Neural Network (DRNN)-based IDS, which represents the propagation of trust among inter-connected IIoT nodes and enhances the ability to resist false data injection. As a form of trust-based attention network, Ben Atitallah et al. [19] introduced a Fuzzy Graph Attention Network (FGAT), in which the trust relationship between the nodes is captured as weighted-edges in the graph structures, enabling dynamic attention to trustworthy entities to be generated during learning.

C. Comparative Examination of the Approaches Existing

Comparison of trust-aware AI frameworks in IIoT shows that various architectures have particular strengths, weaknesses and limitations. Centralized systems, including OI-SVDD + AS-ELM [16], are good at achieving high accuracy because of the full visibility of data but have high

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Volume 13, Issue 09, September 2026

latency and could have central bottlenecks. Swarm and distributed frameworks [21] offer autonomy and scalability and are subject to inconsistency in trust as well as communication overhead in synchronizing multiple nodes. Federated models including those suggested by Padmavathi and Saminathan [18] and Mao et al. [25] provide trade-offs in data privacy, accuracy and energy efficiency albeit may not perform well in the case of adversarial attacks or non-IID data.

Some of the models show significant improvement in terms of performance indicators. As an example, the FGAT model [20] is characterized by a high level of trust consistency and interpretability due to adaptive attention of unstructured graphs, whereas FedIn-NID [25] and SDN-enabled FL model by Hu et al. [23] have much lower latency and energy demands in contrast to centralized systems. Scalability and dynamic recalibration of trust are however an open problem particularly when it comes to large scale, multi-protocol IIoT networks that require real time scalability.

In general, the combination of trust measures and AI algorithms is a paradigm shift in the industrial cybersecurity. These frameworks do not only boost the accuracy of detection, but also ensure the provision of an explainable, energy-efficient, and privacy-preserving intrusion detection, which are the fundamental conditions of the security of future Industry 4.0 ecosystems [24] by embedding behavioral, relational, and contextual trust intelligence into learning-based IDS.

IV. BENCHMARK DATASETS AND EXPERIMENTAL ENVIRONMENTS

A. Overview of Benchmark Datasets

1. SIoT Network Dataset

The Social Internet of Things (SIoT) Network dataset is devoted to capturing relations and interaction between the IoT devices in order to study the trust-based patterns of communication in a big-scale social IoT setup.. The dataset has about 1.2 million interaction records between various categories of devices and relation types (ownership, co-location and functional). It contains more than 60 extracted features, including interaction duration, the rate of packet exchange, latency, and reputation score.

2. WSN Secure and Energy-Saving Dataset.

The WSN Secure and Energy-Efficient dataset is developed with the purpose of benchmarking lightweight intrusion detection and energy efficient security solutions in the Wireless Sensor Networks (WSN). Averages of the records include approximately 45 attributes, and they include node ID, transmission rate, energy consumption, trust value, packet integrity, and route stability [21].

3. Edge-IIoTset Dataset

Edge-IIoTset dataset, this is a realistic smart-factory IIoT dataset that aims to be used to test centralized and federated intrusion detection systems at the network edge [22]. It has a total of about 62 million records of traffic created with simulated industrial devices communicating with protocols like MQTT, Modbus/TCP, and HTTP. Records contain 83 features of network and system, such as the size of packets, flow time, payload entropy, and temporal features. The dataset has 14 types of attacks, such as DoS, MITM, reconnaissance, and data exfiltration, and both on the edge and cloud layer.

4. ToN_IoT Dataset

ToNIoT (Telemetry and Network-based IoT) dataset [23] is a multi-modal Industry 4.0 dataset, which consists of telemetry, operating system logs, and network traffic traces, to enable the analysis of holistic intrusion detections. The dataset consists of 7 data sources such as network flow logs, telemetry data and host based metrics, as well as 85 extracted features (CPU utilization, packet timing, frequency of connections, and anomaly flags). The ToN IoT provides a classification of 20 types of attacks including DDoS, backdoor, and data injection.

5. X-IIoTID Dataset

X-IIoTID (Extended Industrial IoT Intrusion Detection) data [24] This is a heterogeneous, protocol rich data set which is curated to model complex IIoT scenarios with diverse communication standards. It consists of some 30 million marked records of types of attacks in industry such as spoofing, replay, flood, and ransomware. Multiple IIoT protocols, such as BACnet, MQTT, CoAP and OPC-UA, are also covered in data set and application-layer and network-layer behaviors are modeled. It provides 92 attributes including time based, statistical as well as protocol specific attributes to ensure a realistic IIoT traffic model.

6. BoT-IoT Dataset

One of the most popular datasets used in the research of IoT and IIoT intrusion detection is the BoT-IoT (Botnet IoT) dataset [25]. It has to do with over 70 million records of traffic containing normal and botnet-infected traffic based on real-life network emulations. It has 46 extracted attributes, which include flow-based, packet-level, and statistical features being source IP, destination port, packet rate, and count of bytes. The simulated data represents the large categories of attacks such as DDoS, DoS, theft, scanning, and reconnaissance.

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 09, September 2026

Table 2: Summary of Benchmark IIoT Datasets for Trust-Aware Intrusion Detection

Dataset Name	Domain Focus	Total Records	Key Features (Count)	Attack Categories / Types
SIoT Network	Social IoT & Trust Topology	1.2 million	60 (interaction, reputation, latency)	10+ relationship-based attacks
WSN Secure & Energy-Efficient	Wireless Sensor Networks (WSN)	0.5 million	45 (energy, route stability, trust)	8+ (Sybil, sinkhole, selective forwarding)
Edge-IIoTset	Smart Factory & Edge IIoT	62 million	83 (network + system-level)	14 (DoS, MITM, reconnaissance, exfiltration)
ToN_IoT	Industry 4.0 Telemetry & Logs	25 million	85 (telemetry, network, host metrics)	20 (DDoS, backdoor, data injection, etc.)
X-IIoTID	Heterogeneous Industrial Protocols	30 million	92 (protocol & statistical)	24 (spoofing, replay, flooding, ransomware)
BoT-IoT	IoT Botnet & Traffic Analysis	70 million	46 (flow-based, packet-level)	5 major (DDoS, DoS, theft, scanning, recon)

V. ANALYSIS AND DISCUSSION

A. Performance Trends and Observations

Recent comparative analysis of trust-mindful AI systems shows that trust metrics offer considerable value in intrusion detection accuracy and a lower number of false positives in comparison to non-trust AI-based systems. Deep learning systems can be centralized only and are highly accurate, yet not always interpretable and resilient against dynamic attacks. Federated and edge trust-aware models (FedIn-NID and Fuzzy GAT) have similar detection rates and improve the level of transparency by using feature weighting by trust. The models can be more adaptable to heterogeneous IIoT traffic through learning contextual dependencies and relational behavioral patterns among nodes [26].

Table 3: Comparative Performance Metrics of Trust-Aware AI Frameworks [27]

Model / Framework	Architecture Type	Accuracy (%)	False Alarm Rate (%)	Interpretability Score (0–1)
Centralized CNN-RNN	Centralized	96.2	4.5	0.42
LightGBM-MLP Hybrid	Centralized	95.6	3.8	0.55
Federated FL-MU	Federated	97.4	2.9	0.73
Fuzzy GAT (FGAT)	Distributed	98.1	2.3	0.81
FedIn-NID + Trust Layer	Federated	98.6	2.0	0.88

The concept of trust-aware modeling has a significant contribution to the strength and ability to withstand failure of intrusion detection, because of the combination of behavioral and relational intelligence in AI algorithms. The system blocks malicious or unreliable data by weighting the model updates or the contribution made by features by the trustworthiness of the device to avoid model drift and false learning. This figure 4 is a comparison on the accuracy performance of different IDS models. FedIn-NID + Trust Layer has the highest accuracy (98.6%), closely trailed by Fuzzy GAT (98.1%), but the two have better detection reliability because of trust-centric integration and federated intelligence.

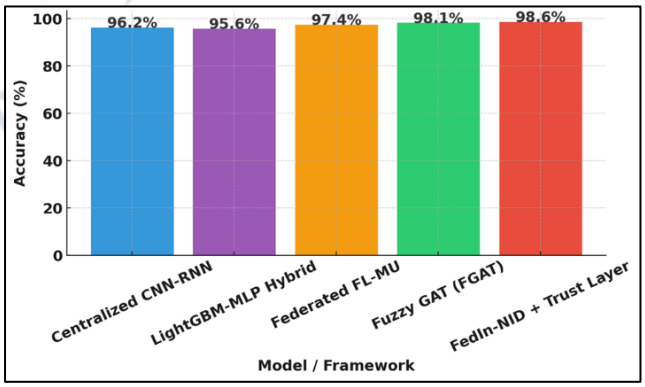


Figure 4. Model Accuracy Comparison for Trust-Aware IDS Frameworks

Models such as FedIn-NID and Fuzzy Graph Attention Networks are more resilient to adversarial and poisoning attacks and can be used to produce consistent performance despite 10-15 percent of nodes being adversarially or poisoned. Trust propagation guarantees the strengthening of trustworthy nodes and isolates the suspicious ones enhancing system continuity and adaptive learning.

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume 13, Issue 09, September 2026

B. Energy Efficiency and Edge Deployment Insights

Resource management in federated and edge environments is essential in ensuring efficiency in the operations without lowering the detection accuracy. The outcome is sustainable energy-performance trade-off which will increase device life and responsiveness of distributed intrusion detection [28].

Improved detection accuracy, latency, and computation cost have inherent tradeoffs in the deployment of trust-aware IDS in edge based IIoT environments. However, centralized systems, even though precise, have high latency and consumes a lot of bandwidth, whereas distributed and federated models have lesser latency, but more resources are needed in evaluation of the trust and synchronization.

C. Trust Dynamics and Model Explainability

Traditional deep learning systems, including CNN-RNN systems, are black boxes with low interpretability indices, thus are hard to reason decisions in industrial critical environments. The reason is the addition of SHAP and LIME-based explanations to increase the clarity of the decision by showing the features of trust that are involved in intrusion classification. Explainable AI (XAI) is the key to reducing the distance between human trust validation and industrial network automation with the help of AI. XAI allows the system administrator to have an insight into why certain intrusion decisions were made by the system, which enhances the operational transparency and adherence to the standards of industrial safety.

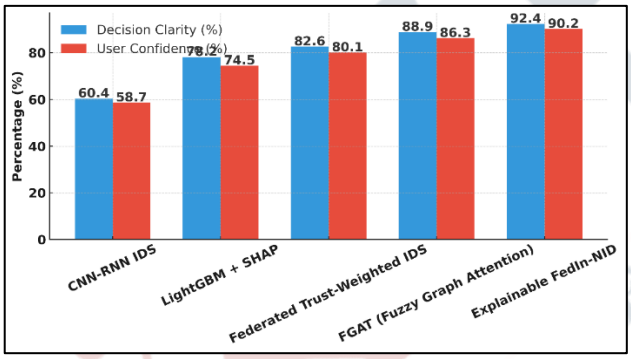


Figure 5. Comparative performance of IDS frameworks highlighting the increase in decision clarity and user confidence with explainable and trust-aware models

This figure 5 indicates that explainable and trust-conscious IDS such as FGAT and Explainable FedIn-NID are the ones with the most adequate decision clarity and user confidence, with better transparency, explainability, and trust in AI-based industrial intrusion detection systems. XAI methods are used in trust-aware schemes to visualize the flow of trust propagation, the weight of anomalies and credibility of the nodes, enabling the analyst to determine whether a decision is made on solid behavioral grounds. This minimizes the chances of incorrect classification of biased or poisoned inputs.

VI. FUTURE RESEARCH DIRECTIONS

A. Immutable Trust Management through the integration of Blockchain

The next generation IIoT intrusion detection systems will support the use of blockchain technology to provide unchangeable and transparent trust records among the devices. The blockchain guarantees decentralized verification of updates on trust which stops tampering and manipulation of the trust scores. Smart contracts can utilize data exchange through the evaluation of trust and improve the accountability among heterogeneous IIoT nodes.

B. Reinforcement Learning of Self-Learning Trust

Trust-aware AI frameworks which are combined with the reinforcement learning (RL) concept can help adapt to evolving threat scenarios independently. The self-learning agents are able to keep the values of trust updated with real-time feedback of the interactions with the devices and detection results.

C. Elucible and Ethical AI to Transparent Intrusion Decisions

With AI-based security becoming more autonomous, it is essential to be able to understand the explainability and ethical clarity. Trust-aware intrusion detection needs to be incorporated with Explainable AI (XAI) to explain how their decisions are made and the trust factors determine them.

D. Future-proof IIoT Security by Quantum-Safe Security Mechanisms

The advent of quantum computing is extremely threatening to traditional encryption and trust systems. Quantum-safe cryptography methods like lattice-based and hash-based encryption should be used in future IIoT security models where they are necessary to guarantee the safety of message transmission and confidence in transactions.

VII. CONCLUSION

This literature review has examined the development, issues, and opportunities of trust-conscious artificial intelligence (AI) models of secure intrusion detection in Industrial Internet of Things (IIoT) networks. The research points out that traditional Intrusion Detection Systems (IDS) can be useful in terms of baseline protection, but might not offer the flexibility, contextual awareness, and trust to the dynamic and heterogeneous characteristics of contemporary IIoT environments. The next-generation frameworks outperform the current ones by integrating AI-based anomaly detection with trust computation models, thereby guaranteeing higher accuracy, interpretability, and resistance to complex cyber threats of attacks such as insider and data-poisoning attacks. The benchmark dataset analysis, which involved SIoT Network, Edge-IIoTset, ToN-IoT, X-IIoTID and BoT-IoT demonstrated the role of trust relationships, protocol diversity, and topology variations in

intrusion detection models robustness and transferability. Trust-sensitive architecture, specifically federated and edge-based models, demonstrates impressive capabilities of improving the level of security without destroying privacy and energy savings. Combining behavioral, relational, and contextual trust metrics in the AI models enhances defense, and adaptive learning and real-time anomaly detection can be achieved. Also, the introduction of Explainable AI (XAI) offers additional transparency to industrial decisions, which will create human trust in the automated systems. The review however also highlights some serious missing links like data imbalance, scalability, privacy in trust sharing and absence of real time adaptive trust metrics which still need to be pursued. In general, the intersections between trust, AI, and industrial intelligence are an important move towards autonomous, secure and self-resilient IIoT ecosystems that will herald the intelligent and trustworthy Industry 4.0 and beyond.

REFERENCES

[1] K. Istiaque Ahmed, M. Tahir, S. Lun Lau, M. Hadi Habaebi, A. Ahad and A. Mughees, "Trust-Aware Authentication and Authorization for IoT: A Federated Machine Learning Approach," in IEEE Internet of Things Journal, vol. 12, no. 8, pp. 9889-9904, 15 April15, 2025, doi: 10.1109/JIOT.2024.3512657.

[2] Kumar, A., Sharma, B. & Noonia, A. Secure blockchain based intrusion detection for IoT networks. Discov Computing 28, 226 (2025). https://doi.org/10.1007/s10791-025-09754-4

[3] M. Ramaiah and M. Y. Rahamathulla, "Securing the Industrial IoT: A Novel Network Intrusion Detection Models," 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), Vellore, India, 2024, pp. 1-6, doi: 10.1109/AIIoT58432.2024.10574728.

[4] Y. Deng, "Design of Industrial IoT Intrusion Security Detection System Based on LightGBM Feature Algorithm and Multi-layer Perception Network," in Journal of Cyber Security and Mobility, vol. 13, no. 2, pp. 327-347, March 2024, doi: 10.13052/jcsm2245-1439.1327.

[5] Al Rawajbeh, M.; Maria Soosai, A.J.; Ramasamy, L.K.; Khan, F. Trustworthy Adaptive AI for Real-Time Intrusion Detection in Industrial IoT Security. IoT 2025, 6, 53. https://doi.org/10.3390/iot6030053

[6] M. Koca and I. Avci, "A Novel Hybrid Model Detection of Security Vulnerabilities in Industrial Control Systems and IoT Using GCN+LSTM," in IEEE Access, vol. 12, pp. 143343-143351, 2024, doi: 10.1109/ACCESS.2024.3466391.

[7] Wiem Bekri, Rihab Jmal, and Lamia Chaari Fourati. 2025. Distributed secured and trustworthy IoT framework based on SDN, AI, and blockchain. Cluster Computing 28, 16 (Nov 2025). https://doi.org/10.1007/s10586-025-05770-7

[8] K. -D. Lu, J. -C. Huang, G. -Q. Zeng, M. -R. Chen, G. -G. Geng and J. Weng, "Multi-Objective Discrete Extremal Optimization of Variable-Length Blocks-Based CNN by Joint NAS and HPO for Intrusion Detection in IIoT," in IEEE Transactions on Dependable and Secure Computing, vol. 22, no. 4, pp. 4266-4283, July-Aug. 2025, doi: 10.1109/TDSC.2025.3545363

[9] Samuel D. Okegbile, Ishaya P. Gambo, Artificial intelligence-driven security framework for internet of things-enhanced digital twin networks, Internet of Things, Volume 31, 2025, 101564, ISSN 2542-6605, https://doi.org/10.1016/j.iot.2025.101564

[10] Villafranca, A.; Thant, K.M.; Tasic, I.; Cano, M.-D. AI-Enabled IoT Intrusion Detection: Unified Conceptual Framework and Research Roadmap. Mach. Learn. Knowl. Extr. 2025, 7, 115. https://doi.org/10.3390/make7040115

[11] S. A. Abdulkareem, C. Heng Foh, M. Shojafar, F. Carrez and K. Moessner, "Network Intrusion Detection: An IoT and Non IoT-Related Survey," in IEEE Access, vol. 12, pp. 147167-147191, 2024, doi: 10.1109/ACCESS.2024.3473289.

[12] Vikas, Rajesh Prasad, Santosh Kumar Upadhyay, Amit Kumar. Zero Trust-Driven Anomaly Detection Framework for Wireless Sensor Networks [J]. Int J Performability Eng, 2025, 21(8): 463-471.

[13] S. Li et al., "CRSF: An Intrusion Detection Framework for Industrial Internet of Things Based on Pretrained CNN2D-RNN and SVM," in IEEE Access, vol. 11, pp. 92041-92054, 2023, doi: 10.1109/ACCESS.2023.3307429.

[14] Benkhelifa, E., Welsh, T., & Hamouda, W. (2018). A critical review of practices and challenges in intrusion detection systems for IoT: Toward universal and resilient systems. IEEE communications surveys & tutorials, 20(4), 3496-3509.

[15] P. Kumar, S. Mullick, R. Das, A. Nandi and I. Banerjee, "IoTForge Pro: A Security Testbed for Generating Intrusion Dataset for Industrial IoT," in IEEE Internet of Things Journal, vol. 12, no. 7, pp. 8453-8460, 1 April1, 2025, doi: 10.1109/JIOT.2024.3501017

[16] E. Gyamfi and A. D. Jurcut, "Novel Online Network Intrusion Detection System for Industrial IoT Based on OI-SVDD and AS-ELM," in IEEE Internet of Things Journal, vol. 10, no. 5, pp. 3827-3839, 1 March1, 2023, doi: 10.1109/JIOT.2022.3172393

[17] Nizamudeen, S.M.T. Intelligent intrusion detection framework for multi-clouds – IoT environment using swarm-based deep learning classifier. J Cloud Comp 12, 134 (2023). https://doi.org/10.1186/s13677-023-00509-4

[18] Padmavathi, V., Saminathan, R. A federated edge intelligence framework with trust based access control for secure and privacy preserving IoT systems. Sci Rep 15, 35832 (2025). https://doi.org/10.1038/s41598-025-19712-1

[19] S. Ben Atitallah, M. Driss, W. Boulila and A. Koubaa, "Securing Industrial IoT Environments: A Fuzzy Graph Attention Network for Robust Intrusion Detection," in IEEE Open Journal of the Computer Society, vol. 6, pp. 1065-1076, 2025, doi: 10.1109/OJCS.2025.3587486

[20] L. Takhellambam, U. Boby Clinton, N. Hoque, K. Robindro Singh and M. Bhuyan, "FL-MU: A Benchmark Dataset for Federated Intrusion Detection in IoT Networks," in IEEE Access, vol. 13, pp. 191037-191064, 2025, doi: 10.1109/ACCESS.2025.3628066.

[21] Y. Yang et al., "TGDCLNet: Teacher-Guided Denoising Contrastive Learning Network-Based IoT Network Intrusion Detection," in IEEE Internet of Things Journal, vol. 12, no. 18, pp. 38509-38525, 15 Sept.15, 2025, doi: 10.1109/JIOT.2025.3586290

[22] S. Latif et al., "Intrusion Detection Framework for the Internet of Things Using a Dense Random Neural Network," in IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 6435-6444, Sept. 2022, doi: 10.1109/TII.2021.3130248.

[23] T. Hu et al., "Efficient and Privacy-Preserving Network Intrusion Detection Based on Federated Learning in SDN-Enabled IIoT Network," in IEEE Internet of Things Journal, vol. 12, no. 20, pp. 41904-41923, 15 Oct.15, 2025, doi: 10.1109/JIOT.2025.3591598
[24] J. Long, W. Liang, K. -C. Li, Y. Wei and M. D. Marino, "A Regularized Cross-Layer Ladder Network for Intrusion Detection in Industrial Internet of Things," in IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 1747-1755, Feb. 2023, doi: 10.1109/TII.2022.3204034
[25] Wadibhasme, R. N., Chaudhari, A. U., Khobragade, P., Mehta, H. D., Agrawal, R., & Dhule, C. (2024, June). Detection and Prevention of Malicious Activities In Vulnerable Network Security Using Deep Learning. In 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET) (pp. 1-6). IEEE.
[26] J. Mao, Z. Wei, B. Li, R. Zhang and L. Song, "FedIn-NID: A Federated Learning Framework for Network Intrusion Detection in Large-Scale Heterogeneous Industrial IoT," in IEEE Transactions on Information Forensics and Security, vol. 20, pp. 9250-9262, 2025, doi: 10.1109/TIFS.2025. 3602226.
[27] J. Mao, Z. Wei, B. Li, R. Zhang and L. Song, "FedIn-NID: A Federated Learning Framework for Network Intrusion Detection in Large-Scale Heterogeneous Industrial IoT," in IEEE Transactions on Information Forensics and Security, vol. 20, pp. 9250-9262, 2025, doi: 10.1109/TIFS.2025. 3602226.