Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Protocol-Independent Intrusion Detection in Industrial Iot: A Review of Foundations, Benchmarks, and Evaluation Practice

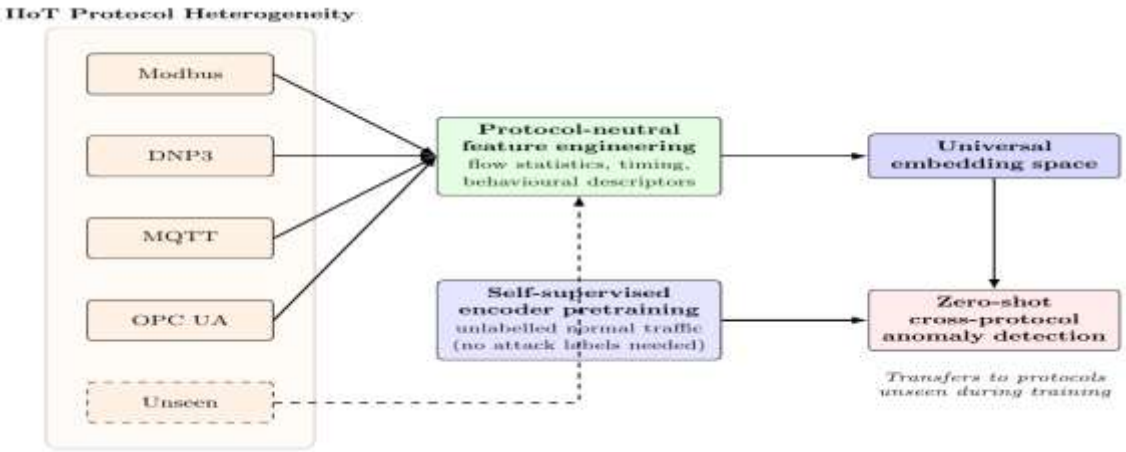
Abhijit Debnath[1], Bijitaswa Chakraborty[2], Jayanta Pal[3]

[1]Centre for Data Science, JIS Institute of Advanced Studies and Research (JISIASR), Kolkata, West Bengal, India. Email: abhijit.debnath@jisiasr.org
[2]Centre for Data Science, JIS Institute of Advanced Studies and Research (JISIASR), Kolkata, West Bengal, India.
[3]Department of Information Technology, Tripura University (A Central University), Agartala, Tripura, India.
Corresponding author: Abhijit Debnath

Highlights

- Protocol-specific IDS cannot scale to heterogeneous IIoT deployments
- Self-supervised learning addresses IIoT's inherent label scarcity
- Cross-protocol transfer remains largely unvalidated in the literature
- Leave-one-protocol-out evaluation is needed to test independence claims
- Five guiding research questions are proposed for future validation

Graphical Abstract



Abstract
The convergence of operational technology with the Industrial Internet of Things (IIoT) has expanded the attack surface of critical infrastructure while fragmenting it across a heterogeneous mixture of industrial communication protocols. Network intrusion detection systems for these environments have historically been built around the semantics of individual protocols—Modbus, DNP3, EtherNet/IP, S7comm, PROFINET, OPC UA, MQTT and others—yielding accurate but brittle detectors that fail to transfer when the protocol, device vendor, or process changes. This review synthesises the literature at the intersection of industrial cyber-physical security, network anomaly detection, and representation learning to characterise the protocol-independence problem: the difficulty of building detection that generalises across the protocol heterogeneity intrinsic to IIoT. We trace the evolution from signature- and specification-based monitoring, through flow-statistical and process-aware anomaly detection, to recent self-supervised and transfer-learning approaches, and examine the datasets and evaluation practices underpinning the field. We consolidate a comparative view of widely used ICS/IIoT benchmark datasets, highlighting their divergence in modality, protocol coverage, and attack prevalence, and discuss recurrent pitfalls that inflate reported performance and obscure poor generalisation. We argue that a protocol-independent, self-supervised detection paradigm is both necessary and, given recent advances, increasingly viable, and articulate the open research questions motivating dedicated investigation.

Keywords: Industrial Internet of Things; Intrusion detection; Protocol-independent detection; Anomaly detection; Self-supervised learning; Industrial control systems.

Introduction

Industrial control systems (ICS) and, more broadly, the Industrial Internet of Things (IIoT) now underpin the operation of water treatment plants, power distribution grids, manufacturing lines, oil and gas pipelines, and building automation. The economic and safety rationale for connecting these systems is compelling: remote monitoring, predictive maintenance, and data-driven optimisation all depend on pervasive connectivity between sensors, actuators, programmable logic controllers (PLCs), and enterprise and cloud services [1]. The same connectivity, however, dissolves the air gaps that once isolated operational technology (OT) from untrusted networks and dramatically enlarges the exposure of physical processes to remote adversaries [2,3].

Because a successful intrusion into an industrial network can translate directly into physical harm—overflowing a reservoir, tripping a breaker, or damaging rotating machinery—intrusion detection has become a central line of defence [4,5]. Yet the industrial setting differs from conventional enterprise IT in ways that frustrate the direct reuse of established IDS techniques. Traffic is often periodic and highly structured; the population of legitimate behaviours is comparatively small but must be characterised precisely; labelled attack data are scarce; and, most consequentially for this review, the communication substrate is fragmented across dozens of industrial protocols with distinct syntaxes, semantics, and security postures [6,7].

This protocol fragmentation is the crux of a persistent engineering and scientific difficulty. Much of the most accurate industrial IDS work is protocol-specific: it parses a particular protocol, models its state machine, and flags deviations [7–9]. Such detectors achieve high fidelity on the protocol for which they were engineered, but they are costly to build, require expert knowledge of each protocol, and—critically—do not transfer. A model of Modbus/TCP tells one little about DNP3 or OPC UA, and a detector tuned to one plant's process seldom applies to another. As IIoT deployments increasingly mix legacy fieldbus protocols with modern IoT messaging such as MQTT and CoAP, the combinatorial cost of maintaining per-protocol detectors grows unsustainable, and coverage gaps open precisely where heterogeneity is greatest [10].

These observations motivate a different question: can intrusion detection capability be made protocol-independent—able to generalise across heterogeneous industrial protocols without per-protocol supervision or hand-built parsers? The question touches foundational issues about whether there exist behavioural regularities invariant to protocol syntax, whether such regularities can be learned without labels, and whether representations learned on one protocol transfer to protocols unseen during training. The purpose of this article is to review the literature bearing on this question and to establish the grounding and viability of protocol-independent intrusion detection as a research problem.

This review makes the following contributions: (i) it frames the protocol-independence problem for IIoT intrusion detection and situates it within the broader literatures of ICS security, network anomaly detection, and representation learning; (ii) it provides a structured account of protocol-specific detection and a critical analysis of why such approaches resist cross-protocol transfer; (iii) it consolidates a comparative view of datasets and benchmarking practices, cataloguing methodological pitfalls that inflate reported performance; (iv) it distils cross-cutting open challenges and argues for the necessity and viability of a protocol-independent, self-supervised detection paradigm.

Consistent with its role as a precursor study, this article deliberately stops at the formulation and justification of the research problem. It does not propose a specific detector, feature construction, or training design; its aim is to establish that the problem is well posed, well motivated, and tractable enough to warrant dedicated investigation. This choice is deliberate and, we argue, timely: as Sections 5 and 6 show, the field currently lacks both a precise statement of the protocol-independence problem and an evaluation practice capable of testing it, so that isolated solution proposals risk being assessed against benchmarks that cannot substantiate cross-protocol claims. Clarifying the problem, consolidating the evidence, and specifying how transfer should be measured is therefore a prerequisite for—not a substitute for—subsequent methodological work. The remainder of the paper is organised as follows. Section 2 states the review scope and method. Section 3 gives background on IIoT architecture, protocols, and IDS taxonomy. Section 4 analyses protocol-specific detection and its limits. Section 5 surveys protocol-agnostic and protocol-independent approaches. Section 6 consolidates the dataset landscape and evaluation pitfalls. Section 7 distils open challenges, Section 8 presents the synthesis and guiding research questions, and Section 9 concludes.

Scope And Review Method

The review is scoped to network- and behaviour-based intrusion detection for IIoT and ICS environments, with emphasis on the transferability of detection capability across communication protocols. We include: (i) signature-, specification-, and anomaly-based network IDS for industrial settings; (ii) process- and

physics-aware detection that reasons over sensor–actuator behaviour; (iii) machine-learning and deep-learning detectors, including representation-learning, self-supervised, and transfer-learning methods relevant to cross-domain generalisation; and (iv) the datasets and evaluation methodologies used to benchmark such systems.

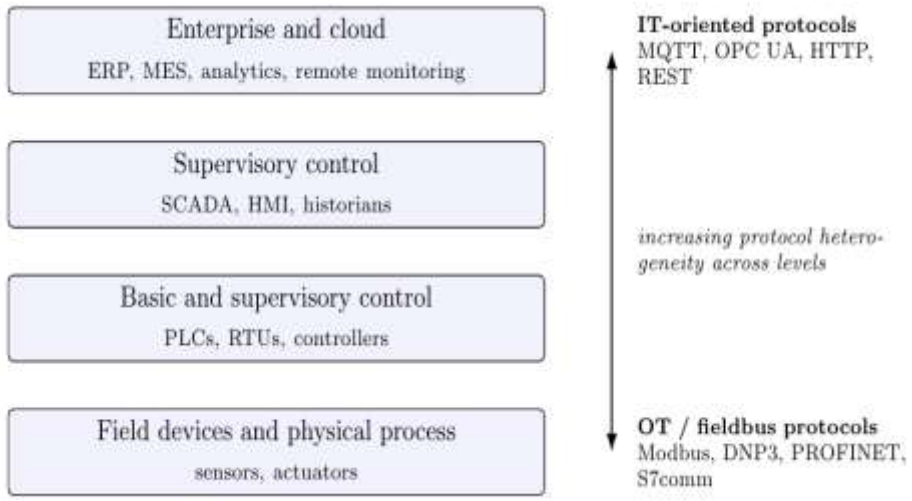
Sources were identified through major scholarly indexes—IEEE Xplore, the ACM Digital Library, ScienceDirect, and Scopus, supplemented by forward/backward citation tracing and preprints from arXiv—using combinations of terms such as industrial intrusion detection, ICS/SCADA anomaly detection, IIoT security dataset, protocol-agnostic / protocol-independent detection, cross-protocol generalisation, and self-supervised / contrastive learning. The search emphasised work published between 2015 and 2025, while retaining a small number of older, foundational references (e.g., early anomaly-detection surveys and signature tools) required for conceptual scaffolding. From an initial pool of several hundred records, studies were retained when they bore directly on one of four themes: (i) industrial IDS methodology, (ii) protocol-agnostic or transferable detection, (iii) representation and self-supervised learning relevant to anomaly detection, or (iv) ICS/IIoT datasets and evaluation practice. Duplicate, non-peer-reviewed (except widely cited preprints), and tangential records were excluded, yielding the set discussed here. Foundational and highly cited works were prioritised for conceptual scaffolding, while recent primary studies and datasets were prioritised for the state of the art. The review is narrative and integrative rather than an exhaustive systematic meta-analysis conforming to PRISMA; its objective is to synthesise a coherent problem statement, not to quantitatively pool reported detection scores, which—as Section 6 argues—are frequently not comparable across studies.

Background

IIoT and ICS architecture

Industrial networks are conventionally described by a layered reference model separating physical process instrumentation from supervisory control and from enterprise IT, often depicted as the Purdue Enterprise Reference Architecture. At the lowest levels, field devices—sensors and actuators—interface with PLCs and remote terminal units (RTUs); these are coordinated by SCADA systems and human–machine interfaces (HMIs); and higher levels connect to manufacturing execution and enterprise resource planning systems, and increasingly to cloud analytics [1,3]. IIoT extends this stack by introducing large numbers of networked smart devices, edge gateways, and IP-based messaging that blur the boundary between the process network and the wider internet. Fig. 1 depicts this layered organisation.

Two properties of this architecture are salient for detection. First, behaviour at the process layers is comparatively regular: control loops execute at fixed cadences, and sensor and actuator signals obey the physics of the underlying process, so that legitimate operation occupies a constrained region of the observable space [5,9]. Second, the network substrate is heterogeneous: a single facility may simultaneously run legacy serial-derived protocols, Ethernet-based industrial protocols, and lightweight IoT messaging, each with different framing, addressing, and semantics.



Layered organisation of an IIoT/ICS deployment (Purdue-style reference model). The communication substrate becomes increasingly heterogeneous as IT- and OT-oriented protocols coexist across levels.

Industrial protocols and their diversity

Table 1 summarises representative industrial and IIoT protocols. The diversity is not merely superficial. Protocols differ in their transport (raw Ethernet, TCP, UDP), in whether they are request/response or publish/subscribe, in their addressing and object models, and in whether they provide any native authentication or integrity protection. Many widely deployed fieldbus protocols were designed for isolated, trusted networks and offer little or no security [6,7]. Newer protocols such as OPC UA incorporate security services, and IoT messaging protocols such as MQTT can be secured at the transport layer, but real deployments remain a patchwork.

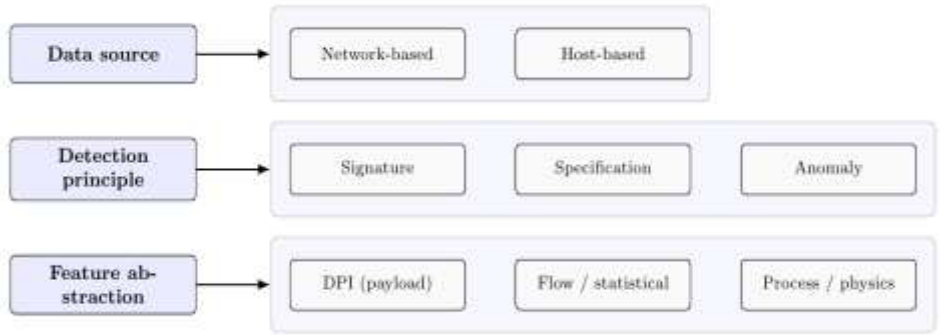
Representative industrial and IIoT communication protocols.

Protocol	Transport	Interaction	Native security
Modbus/TCP	TCP	Request/response	Minimal (none in base spec)
DNP3	TCP/serial	Request/response, unsolicited	Optional (Secure Authentication)
EtherNet/IP	TCP/UDP	Request/response, I/O	Limited; CIP Security optional
PROFINET	Ethernet	Cyclic real-time I/O	Limited in base profiles
S7comm	TCP	Request/response	Proprietary; limited
OPC UA	TCP/HTTP	Client/server, pub/sub	Built-in (signing, encryption)
MQTT	TCP	Publish/subscribe	TLS optional
CoAP	UDP	Request/response	DTLS optional
BACnet	UDP/IP	Request/response	Limited in common profiles

A taxonomy of intrusion detection

Intrusion detection is customarily organised along several axes [11,12]. By data source, a detector may be host-based or network-based; industrial IDS is predominantly network-based. By detection principle, detectors are broadly signature-based (matching known attack patterns), specification-based (flagging violations of an explicit model of correct behaviour), or anomaly-based (flagging deviations from a learned model of normal behaviour). Signature methods are precise but blind to novel attacks; anomaly methods can detect the unknown but are prone to false positives [13,14]. By feature abstraction, network detectors range from deep packet inspection (DPI), which parses payloads and thus depends intimately on protocol semantics, to flow- and statistics-based methods that summarise traffic with protocol-neutral quantities such as timing, volume, and directionality [15]. Fig. 2 organises this taxonomy.

This last axis—the degree of dependence on protocol semantics—is the one around which this review is organised. It largely determines whether a detector can transfer across heterogeneous IIoT protocols.



A working taxonomy of intrusion detection for IIoT/ICS along three axes. The feature-abstraction axis (bottom) determines a detector's potential to transfer across protocols.

The industrial threat landscape

Threats to industrial networks span reconnaissance, unauthorised command injection and parameter manipulation, denial of service, man-in-the-middle and replay attacks against unauthenticated protocols, and stealthy process-oriented attacks that manipulate sensor or actuator values within seemingly plausible ranges to damage the physical process while evading naive detection [4,5]. The last category is distinctive to cyber-physical systems: an attacker who understands the process can craft manipulations that are individually inconspicuous at the packet level yet collectively drive the process toward an unsafe state [9]. Detection therefore benefits from reasoning not only about protocol conformance but about behaviour—

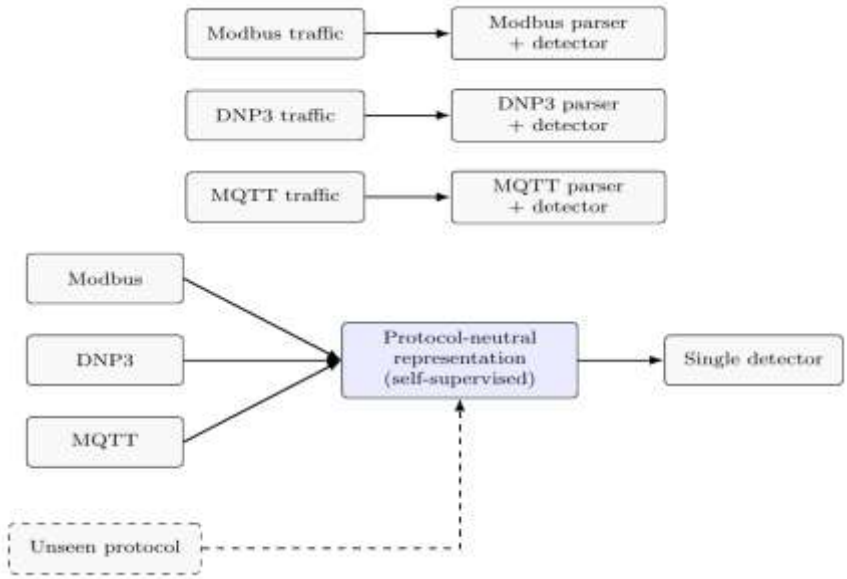
how traffic and process variables evolve over time—which is, encouragingly, a level of description less tightly bound to any single protocol.

Protocol-specific intrusion detection and its limits

The most mature industrial IDS techniques are protocol-specific. Rule- and signature-based network monitors such as Snort and Zeek (formerly Bro) can be extended with industrial protocol parsers and rule sets, giving operators fine-grained visibility into Modbus function codes or DNP3 object requests [16,17]. Building on such parsers, specification-based systems encode a model of legitimate protocol behaviour and raise alerts on violations. [7] model Modbus/TCP as a deterministic finite automaton per HMI/PLC channel; [8] adapt Zeek for DNP3; and [9] monitor the semantics of process variables extracted from PLC traffic, moving detection closer to the physical process.

These approaches are effective and interpretable within their scope. Their limitations, however, are structural: (i) each protocol requires a dedicated parser and behavioural model built by domain experts, and as the protocol inventory grows so does the engineering and maintenance burden [3]; (ii) a model of one protocol's state machine confers essentially no capability on another, so heterogeneous deployments must be defended protocol by protocol [10]; (iii) deep packet inspection presumes parseable payloads and is defeated by proprietary protocols, undocumented extensions, and increasingly common transport-layer encryption; (iv) signature methods, by construction, cannot detect attacks for which no signature exists, a serious limitation against zero-day and process-oriented attacks [13].

The collective implication is that protocol-specific detection, however accurate in isolation, cannot by itself deliver the broad, transferable coverage that heterogeneous IIoT deployments require. This is the gap that motivates protocol-agnostic and, ultimately, protocol-independent approaches. Fig. 3 contrasts the two paradigms.



The core contrast motivating this review. Protocol-specific detection (a) scales poorly with protocol heterogeneity, whereas a protocol-independent design (b) seeks a single, transferable detector.

Toward protocol-agnostic and protocol-independent detection

It is useful to distinguish two adjacent notions. A method is protocol-agnostic if it operates without parsing protocol payloads—typically using protocol-neutral features. A method aspires to protocol-independence in the stronger sense if the detection capability it learns transfers across protocols, including to protocols unseen during training. Much existing work is agnostic; genuine independence remains largely open.

Flow- and statistics-based detection

Flow-based detection summarises traffic with features that do not depend on payload semantics: packet and byte counts, inter-arrival times, flow durations, directionality, and derived statistics [11,15]. Because industrial traffic is often strongly periodic, timing and volumetric regularities are informative, and deviations from learned periodicity can reveal scanning, injection, or denial-of-service activity without

protocol parsing. Classical unsupervised detectors—one-class support vector machines [18], isolation forests [19]—operate naturally on such feature vectors and provide protocol-agnostic baselines, as do supervised classifiers such as SVMs and random forests applied to industrial traffic [20]. The abstraction that makes these features portable, however, also discards information, and flow statistics alone may miss semantically subtle, process-oriented attacks.

Process- and physics-based detection

A complementary strand reasons about the physical process. By learning normal relationships among sensor and actuator signals, detectors can flag manipulations inconsistent with process behaviour [4,5]. Recurrent networks model temporal dependencies among process variables [21], convolutional networks capture local multivariate structure in ICS telemetry [22], and unsupervised models achieve strong performance on cyber-physical system networks [23]. Process-based detection is protocol-agnostic but typically process-specific: a model of one plant's dynamics does not transfer to a different process, trading protocol dependence for process dependence.

Representation learning and self-supervision

The deeper aspiration—transferable detection—leads to representation learning. The premise is that heterogeneous industrial observations may share latent behavioural structure capturable in a learned embedding, within which anomaly detection becomes easier and potentially transferable [24,25]. The obstacle in the industrial setting is labels: attacks are rare, diverse, and expensive to curate, so supervised training is data-limited and generalises poorly to novel attacks [13].

Self-supervised learning (SSL) offers a route around the label bottleneck by constructing training signals from unlabelled data itself. Contrastive methods learn representations by pulling together augmented views of the same instance and pushing apart different instances [26,27], while non-contrastive approaches avoid explicit negatives [28]; masked-prediction objectives, exemplified in language by bidirectional transformers [29] built on attention mechanisms [30], learn by reconstructing withheld portions of the input. Related work shows that self-supervised and classification-based objectives yield representations well suited to one-class and anomaly detection [31–34]. Although these techniques originate largely outside the industrial domain, their central promise—learning useful structure without labels—is directly relevant to IIoT, where unlabelled operational traffic is abundant but labelled attacks are scarce.

This promise is now being realised in the network-security literature. Recent self-supervised network intrusion detection systems learn from unlabelled or normal-only traffic and detect anomalies without extensive attack labels: Wang et al. introduce a transformer-based scheme with a masked context-reconstruction objective that learns robust temporal representations [35]; Anomal-E applies self-supervised graph neural networks to flow graphs [36]; and further work applies self-supervision to IoT traffic to counter label scarcity and improve robustness against denial-of-service attacks [37,38]. These results, drawn mostly from enterprise and IoT settings rather than industrial control, establish that self-supervised representations can support intrusion detection in practice—an encouraging precedent for the industrial problem, though one not yet demonstrated across protocols.

Transfer and cross-domain generalisation

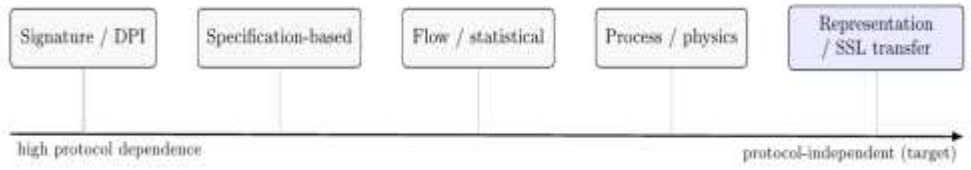
Whether representations learned on one protocol or process transfer to another is a question of transfer learning and domain generalisation [39]. In principle, if an encoder captures behavioural regularities invariant to protocol syntax, it should support zero-shot application to protocols absent from its training data. Evidence for such cross-protocol transfer in the industrial setting remains limited and is complicated by the evaluation confounds discussed in Section 6; establishing it rigorously is, we argue, one of the field's central open problems. Recent cross-domain studies underline the difficulty: evaluating detectors across multiple network datasets, Layeghy and Portmann find that no single model generalises across all domains, and that cross-domain performance is markedly asymmetric—swapping the roles of source and target domain can change results substantially [40]. Complementary cross-evaluation frameworks show that testing across datasets both extends a model's effective coverage and exposes fragilities that single-dataset evaluation conceals [41]. These findings, though set at the level of datasets rather than protocols, are the empirical crux of the protocol-independence problem: they demonstrate that strong in-domain performance routinely fails to transfer, and that transfer must therefore be measured directly rather than assumed. The emergence of datasets explicitly designed to be connectivity- and device-agnostic—most notably X-IIoTID, which spans multiple IIoT connectivity protocols and provides multi-view features—signals recognition that heterogeneity must be treated as a first-class concern [10].

A taxonomy by protocol dependence

Table 2 organises the paradigms above by their relationship to protocol semantics and by whether they can transfer across protocols. Fig. 4 places them on a continuum of protocol dependence. The table makes explicit a gap: methods that are strongly protocol-agnostic in their inputs are not thereby protocol-independent in their generalisation, and the paradigm that most directly targets genuine independence—transferable, self-supervised representation learning—is also the least mature in the industrial context.

Detection paradigms organised by dependence on protocol semantics and potential for cross-protocol transfer.

Paradigm	Protocol-agnostic input	Detects novelty	Cross-protocol transfer	Primary limitation
Signature / DPI	–	–	–	Per-protocol rules; blind to novelty
Specification-based	–		–	Per-protocol model engineering
Flow / statistical			Partial	Discards semantic detail
Process / physics-based			–	Process-specific, not portable
Representation / SSL			Target	Immature in ICS; hard to evaluate

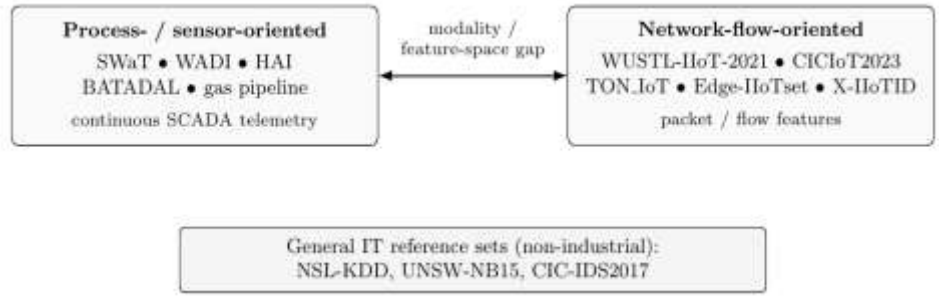


Detection paradigms arranged on a spectrum of dependence on protocol semantics.

Datasets and benchmarking landscape

A comparative view of ICS/IIoT datasets

Table 3 summarises widely used datasets. Two broad families are apparent. Process-/sensor-oriented datasets capture SCADA telemetry from physical testbeds—SWaT and WADI [42–44], HAI [45], BATADAL [46], and gas-pipeline traffic [47]. Network-flow-oriented datasets capture packet or flow features from IoT/IIoT environments—WUSTL-IIoT-2021 [48,49], CICIoT2023 [50], TON_IoT [51], Edge-IIoTset [52], and the connectivity-agnostic X-IIoTID [10]. General-purpose datasets such as NSL-KDD [53], UNSW-NB15 [54], and CIC-IDS2017 [55] remain reference points but are not industrial. Fig. 5 groups these families.



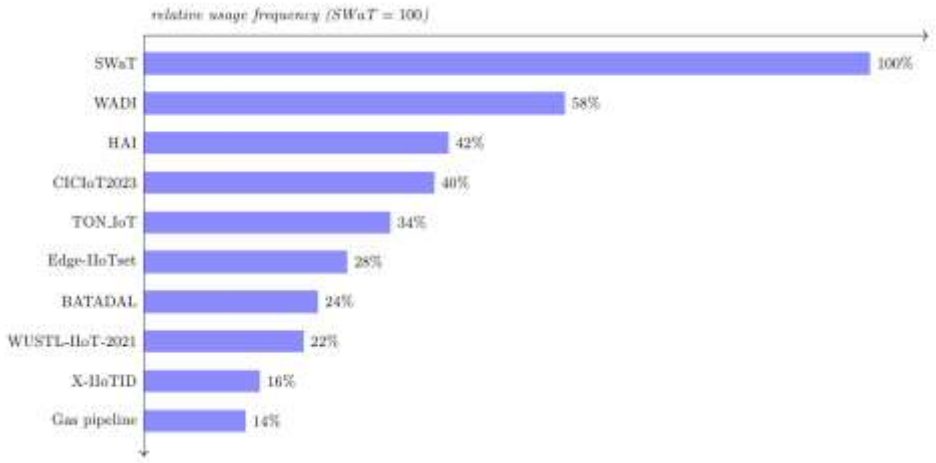
The industrial benchmark landscape falls into two families differing in modality and feature space. This divergence is what a protocol-independent method must bridge.

Representative ICS/IIoT intrusion-detection datasets.

Dataset	Modality	Domain /source	Notes
SWaT	Process/sensor	Water treatment testbed	Continuous ICS telemetry; staged attacks

Dataset	Modality	Domain /source	Notes
WADI	Process/sensor	Water distribution testbed	Companion to SWaT; larger process
HAI	Process/sensor	HIL-based power/thermal testbed	Augmented ICS scenarios
BATADAL	Process/sensor	Water distribution (simulated)	Detection-algorithm challenge
Gas pipeline	Network/process	SCADA gas pipeline testbed	Early ICS IDS benchmark
WUSTL-IIoT-2021	Network-flow	IIoT testbed	Flow features for IIoT security
CICIoT2023	Network-flow	Large IoT topology	Broad attack taxonomy
TON_IoT	Network-flow + telemetry	IoT/IIoT + OS logs	Heterogeneous telemetry
Edge-IIoTset	Network-flow	IoT/IIoT edge testbed	Centralised/federated learning
X-IIoTID	Multi-view	Connectivity-agnostic IIoT	Multiple protocols; device-agnostic

Beyond their qualitative differences, these datasets are used with very uneven frequency. Fig. 6 shows the approximate relative prevalence of each benchmark across the primary studies surveyed here. A small number of process-oriented testbeds (notably SWaT and, to a lesser extent, WADI and HAI) dominate the ICS anomaly-detection literature, while the newer connectivity-agnostic and flow-oriented sets—precisely those most relevant to cross-protocol study—remain comparatively under-exploited. This concentration matters for the protocol-independence question: an evidence base skewed toward a few single-process testbeds provides weak support for claims of protocol-general behaviour, and it partly explains why cross-protocol transfer has seldom been tested.



Approximate relative usage frequency of ICS/IIoT benchmark datasets across the primary studies surveyed, normalised to the most frequently used set (SWaT = 100). Values are indicative rather than exact and are intended to convey the skew of the evidence base toward a few process-oriented testbeds. Connectivity-agnostic and flow-oriented datasets most relevant to cross-protocol study are used comparatively rarely.

Reported performance and why it is hard to compare

To make the comparability problem concrete rather than assert it, Table 4 collects representative reported detection results from the surveyed literature. The intent is not to rank methods—which, as we argue below, the numbers do not support—but to expose how widely the evaluation conditions vary. The studies differ in dataset and modality, in the metric emphasised (accuracy, F1, AUROC, or detection rate), in whether labels are used at training time (supervised, unsupervised, or self-supervised), and, most consequentially, in the evaluation protocol: almost all report within-dataset performance, in which training

and test data are drawn from the same distribution, and none of the industrial-IDS entries reports genuine leave-one-protocol-out transfer.

Representative reported detection results from the surveyed literature. Values are as reported by the original authors under their own experimental conditions and are not directly comparable across rows: datasets, metrics, label regimes, and evaluation protocols all differ, and all entries evaluate within a single dataset rather than across protocols. Abbreviations: DR, detection rate; SSL, self-supervised learning; w/in, within-dataset split.

Study (approach)	Dataset(s)	Label regime	Reported metric	Evaluation protocol
Goldenberg & Wool [7] (Modbus DFA)	Modbus/TCP traces	Specification	Low false alarm	w/in, single protocol
Kravchik & Shabtai [22] (1D-CNN)	SWaT	Unsupervised	F1 (per-attack)	w/in, single process
Goh et al. [21] (LSTM)	SWaT	Unsupervised	DR (per-attack)	w/in, single process
Wang et al. [35] (SSL, masked context)	Flow NIDS sets	SSL / few labels	F1, AUROC	w/in, single dataset
Caville et al. [36] (SSL GNN)	NF-flow sets	SSL	Macro-F1	w/in, single dataset
Almaraz-Rivera et al. [37] (SSL)	IoT DDoS sets	SSL	Accuracy, F1	w/in, single dataset
Layeghy & Portmann [40] (cross-domain)	Multiple NIDS sets	Sup./unsup.	F1 (cross-dom.)	**cross-dataset**

Two observations follow. First, headline scores that appear strong in isolation are conditioned on the evaluation protocol: the same model classes that report near-ceiling within-dataset metrics degrade sharply when tested across domains, as the one cross-domain entry in the table makes explicit [40,41]. Second, because metrics, datasets, and label regimes are not held constant, a numerically higher value in one row does not indicate a better detector than a lower value in another. The practical consequence for this review is that the literature's aggregate performance cannot be pooled into a single ranking, and that the quantity of interest for protocol-independence—cross-protocol transfer—is essentially absent from current reporting. This gap, rather than any single headline number, is what motivates the evaluation recommendations below.

Methodological pitfalls in evaluation

A substantial body of critical work warns that reported detection performance is frequently inflated or not comparable across studies. The pitfalls most relevant to protocol-independence are summarised below.

Data-snooping and leakage. Improper handling of temporal order or preprocessing statistics leaks information from test to train, yielding optimistic estimates [56,57].

Base-rate and contamination sensitivity. When the proportion of malicious records differs greatly across datasets, threshold- and rank-based metrics can be misread [13].

Unrealistic separability. Synthetic or overly clean data can make classes trivially separable, masking representation failures [58].

Cross-dataset incomparability. Because datasets differ in modality, features, protocols, and labelling, headline scores are seldom comparable [59,60].

Silent evaluation failures. In multi-dataset pipelines, missing or mishandled data for a particular protocol can be dropped without notice, so that a claimed cross-protocol result rests on a subset of the intended benchmarks.

The unifying lesson is that any credible claim of protocol-independent detection must be validated under an evaluation protocol explicitly designed to test transfer—holding out entire protocols—rather than under within-dataset random splits that reward memorisation.

Cross-cutting challenges and open problems

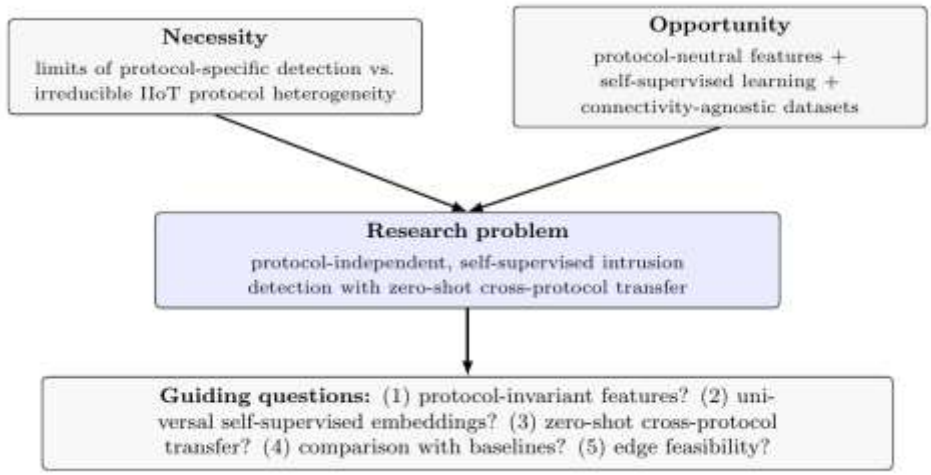
Synthesising the preceding sections, several interlocking challenges define the protocol-independence problem.

Existence of protocol-invariant behavioural structure. The foundational question is whether behavioural regularities in timing, volume, periodicity, and interaction structure are shared across protocols and discriminative for intrusion [15,24].

Learning without labels. Because labelled industrial attacks are scarce and non-representative of future threats, capability must be learnable largely from unlabelled operational data [26,32].
Zero-shot cross-protocol generalisation. It is open whether detection transfers to protocols unseen during training. Rigorous study demands leave-one-protocol-out evaluation [39,56].
Comparison against baselines. A protocol-independent detector must be judged against strong supervised and protocol-specific alternatives [7,12].
Non-stationarity and drift. Detection capability must remain valid under concept drift [61].
Real-time and edge constraints. Detection in IIoT is often required at the edge under tight compute, memory, and latency budgets [62,63], and distributed privacy-preserving training adds further complexity [64].

Synthesis: the case for protocol-independent, self-supervised detection
The literature supports a clear problem statement and, we argue, its viability. The necessity follows from the structural limits of protocol-specific detection (Section 4) set against IIoT's irreducible heterogeneity (Sections 3.2, 6.1). The opportunity follows from three convergent developments: (i) behaviour- and flow-level descriptions provide a protocol-neutral vantage point; (ii) self-supervised representation learning addresses label scarcity; (iii) connectivity-agnostic datasets give the problem the empirical substrate it previously lacked [10].
The guiding research questions this review surfaces (Fig. 7) are:

1. Do protocol-invariant behavioural features exist that are discriminative for intrusion across heterogeneous IIoT/ICS protocols?
2. Can self-supervised objectives learn universal embeddings of industrial behaviour from unlabelled traffic?
3. Does detection built on such embeddings generalise in a zero-shot manner to protocols unseen during training?
4. How does protocol-independent detection compare against strong supervised and protocol-specific baselines?
5. Is such detection feasible under the real-time and resource constraints of edge industrial deployment?



From evidence to problem statement: necessity and opportunity jointly motivate the research problem and its guiding questions.
These questions are posed as the motivation and viability case for future work; the present article intentionally does not prescribe the feature constructions, model architectures, training procedures, or validation mechanisms by which they might be answered.

Conclusion

The security of industrial IoT hinges on intrusion detection that can keep pace with an attack surface that is simultaneously expanding and fragmenting across many communication protocols. This review has argued that the prevailing protocol-specific paradigm, though accurate within its scope, is structurally unable to provide the transferable coverage that heterogeneous IIoT demands. Charting the progression

from signature- and specification-based monitoring, through flow-statistical and process-aware anomaly detection, to self-supervised and transfer-learning approaches, we have framed protocol-independent intrusion detection as a well-posed research problem and marshalled the evidence for its necessity and viability.

The convergence of protocol-neutral behavioural description, self-supervised representation learning, and connectivity-agnostic datasets has brought genuine protocol-independence—transferable, zero-shot cross-protocol detection—within reach of systematic investigation. Realising it will require careful attention to the evaluation pitfalls catalogued here, honest leave-one-protocol-out benchmarking, and consideration of drift and edge constraints. This precursor review establishes the ground on which such work can proceed.

CRediT authorship contribution statement

Abhijit Debnath: Conceptualization, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Bijitaswa Chakraborty:** Supervision, Writing – review & editing. **Jayanta Pal:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI in scientific writing

During the preparation of this work, the authors used generative AI tools to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data availability statement

No new data were created or analysed in this study. All datasets referenced are publicly described in the cited sources.

Acknowledgements

[1] E. Sisinni, A. Saifullah, S. Han, U. Jennehag and M. Gidlund, Industrial Internet of Things: Challenges, opportunities, and directions, IEEE Transactions on Industrial Informatics, vol. 14, no. 11, pp. 4724–4734, 2018.

[2] A. Humayed, J. Lin, F. Li and B. Luo, Cyber-physical systems security—a survey, IEEE Internet of Things Journal, vol. 4, no. 6, pp. 1802–1831, 2017.

[3] K. Tange, M. De Donno, X. Fafoutis and N. Dragoni, A systematic survey of industrial Internet of Things security: Requirements and fog computing opportunities, IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2489–2520, 2020.

[4] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang and S. Sastry, Attacks against process control systems: Risk assessment, detection, and response, Proceedings of the 6th ACM symposium on information, computer and communications security (ASIACCS), pp. 355–366, 2011.

[5] D. I. Urbina, J. A. Giraldo, A. A. Cárdenas, N. O. Tippenhauer, J. Valente, M. Faisal, et al., Limiting the impact of stealthy attacks on industrial control systems, Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (CCS), pp. 1092–1105, 2016.

[6] R. Mitchell and I.-R. Chen, A survey of intrusion detection techniques for cyber-physical systems, ACM Computing Surveys, vol. 46, no. 4, pp. 1–29, 2014.

[7] N. Goldenberg and A. Wool, Accurate modeling of Modbus/TCP for intrusion detection in SCADA systems, International Journal of Critical Infrastructure Protection, vol. 6, no. 2, pp. 63–75, 2013.

[8] H. Lin, A. Slagell, C. Di Martino, Z. Kalbarczyk and R. K. Iyer, Adapting Bro into SCADA: Building a specification-based intrusion detection system for the DNP3 protocol, Proceedings of the eighth annual cyber security and information intelligence research workshop (CSIIRW), pp. 1–4, 2013.

[9] D. Hadžiosmanović, R. Sommer, E. Zambon and P. H. Hartel, Through the eye of the PLC: Semantic security monitoring for industrial processes, Proceedings of the 30th annual computer security applications conference (ACSAC), pp. 126–135, 2014.

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

[11] P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández and E. Vázquez, Anomaly-based network intrusion detection: Techniques, systems and challenges, Computers & Security, vol. 28, no. 1-2, pp. 18–28, 2009.

[12] A. L. Buczak and E. Guven, A survey of data mining and machine learning methods for cyber security intrusion detection, IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
[13] R. Sommer and V. Paxson, Outside the closed world: On using machine learning for network intrusion detection, 2010 IEEE symposium on security and privacy, pp. 305–316, 2010.
[14] V. Chandola, A. Banerjee and V. Kumar, Anomaly detection: A survey, ACM Computing Surveys, vol. 41, no. 3, pp. 1–58, 2009.
[15] M. Ahmed, A. N. Mahmood and J. Hu, A survey of network anomaly detection techniques, Journal of Network and Computer Applications, vol. 60, pp. 19–31, 2016.
[16] M. Roesch, Snort: Lightweight intrusion detection for networks, Proceedings of the 13th USENIX conference on system administration (LISA), pp. 229–238, 1999.
[17] V. Paxson, Bro: A system for detecting network intruders in real-time, Computer Networks, vol. 31, no. 23-24, pp. 2435–2463, 1999.
[18] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson, Estimating the support of a high-dimensional distribution, Neural Computation, vol. 13, no. 7, pp. 1443–1471, 2001.
[19] F. T. Liu, K. M. Ting and Z.-H. Zhou, Isolation forest, 2008 eighth IEEE international conference on data mining (ICDM), pp. 413–422, 2008.
[20] S. D. Anton, S. Sinha and H. D. Schotten, Anomaly-based intrusion detection in industrial data with SVM and random forests, 2019 international conference on software, telecommunications and computer networks (SoftCOM), pp. 1–6, 2019.
[21] J. Goh, S. Adepu, M. Tan and Z. S. Lee, Anomaly detection in cyber physical systems using recurrent neural networks, 2017 IEEE 18th international symposium on high assurance systems engineering (HASE), pp. 140–145, 2017.
[22] M. Kravchik and A. Shabtai, Detecting cyber attacks in industrial control systems using convolutional neural networks, Proceedings of the 2018 workshop on cyber-physical systems security and privacy (CPS-SPC), pp. 72–83, 2018.
[23] P. Schneider and K. Böttinger, High-performance unsupervised anomaly detection for cyber-physical system networks, Proceedings of the 2018 workshop on cyber-physical systems security and privacy (CPS-SPC), pp. 1–12, 2018.
[24] Y. Bengio, A. Courville and P. Vincent, Representation learning: A review and new perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.
[25] G. Pang, C. Shen, L. Cao and A. van den Hengel, Deep learning for anomaly detection: A review, ACM Computing Surveys, vol. 54, no. 2, pp. 1–38, 2021.
[26] T. Chen, S. Kornblith, M. Norouzi and G. Hinton, A simple framework for contrastive learning of visual representations, Proceedings of the 37th international conference on machine learning (ICML), pp. 1597–1607, 2020.
[27] K. He, H. Fan, Y. Wu, S. Xie and R. Girshick, Momentum contrast for unsupervised visual representation learning, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 9729–9738, 2020.
[28] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, et al., Bootstrap your own latent—a new approach to self-supervised learning, Advances in neural information processing systems (NeurIPS), vol. 33, pp. 21271–21284, 2020.
[29] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics (NAACL), pp. 4171–4186, 2019.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, Advances in neural information processing systems (NeurIPS), vol. 30, 2017.
[31] L. Bergman and Y. Hoshen, Classification-based anomaly detection for general data, International conference on learning representations (ICLR), 2020.
[32] K. Sohn, C.-L. Li, J. Yoon, M. Jin and T. Pfister, Learning and evaluating representations for deep one-class classification, International conference on learning representations (ICLR), 2021.
[33] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, et al., Deep one-class classification, Proceedings of the 35th international conference on machine learning (ICML), pp. 4393–4402, 2018.
[34] C. Zhou and R. C. Paffenroth, Anomaly detection with robust deep autoencoders, Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 665–674, 2017.
[35] W. Wang, S. Jian, Y. Tan, Q. Wu and C. Huang, Robust unsupervised network intrusion detection with self-supervised masked context reconstruction, Computers & Security, vol. 128, pp. 103131, 2023.
[36] E. Caville, W. W. Lo, S. Layeghy and M. Portmann, Anomal-E: A self-supervised network intrusion detection system based on graph neural networks, Knowledge-Based Systems, vol. 258, pp. 110030, 2022.

[37] J. G. Almaraz-Rivera, J. A. Cantoral-Ceballos and J. F. Botero, Enhancing IoT network security: Unveiling the power of self-supervised learning against DDoS attacks, Sensors, vol. 23, no. 21, pp. 8701, 2023.
[38] R. Xu, G. Wu, W. Wang, X. Gao, A. He and Z. Zhang, Applying self-supervised learning to network intrusion detection for network flows with graph neural network, Computer Networks, vol. 248, pp. 110495, 2024.
[39] S. J. Pan and Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.
[40] S. Layeghy and M. Portmann, Explainable cross-domain evaluation of ML-based network intrusion detection systems, Computers & Electrical Engineering, vol. 108, pp. 108692, 2023.
[41] G. Apruzzese, L. Pajola and M. Conti, The cross-evaluation of machine learning-based network intrusion detection systems, IEEE Transactions on Network and Service Management, vol. 19, no. 4, pp. 5152–5169, 2022.
[42] J. Goh, S. Adepu, K. N. Junejo and A. Mathur, A dataset to support research in the design of secure water treatment systems, International conference on critical information infrastructures security (CRITIS), pp. 88–99, 2016.
[43] A. P. Mathur and N. O. Tippenhauer, SWaT: A water treatment testbed for research and training on ICS security, 2016 international workshop on cyber-physical systems for smart water networks (CySWater), pp. 31–36, 2016.
[44] C. M. Ahmed, V. R. Palleti and A. P. Mathur, WADI: A water distribution testbed for research in the design of secure cyber physical systems, Proceedings of the 3rd international workshop on cyber-physical systems for smart water networks (CySWATER), pp. 25–28, 2017.
[45] H.-K. Shin, W. Lee, J.-H. Yun and H. Kim, HAI 1.0: HIL-based augmented ICS security dataset, 13th USENIX workshop on cyber security experimentation and test (CSET), 2020.
[46] R. Taormina, S. Galelli, N. O. Tippenhauer, E. Salomons, A. Ostfeld, et al., Battle of the attack detection algorithms: Disclosing cyber attacks on water distribution networks, Journal of Water Resources Planning and Management, vol. 144, no. 8, pp. 04018048, 2018.
[47] T. Morris and W. Gao, Industrial control system traffic data sets for intrusion detection research, International conference on critical infrastructure protection (ICCIP), pp. 65–78, 2014.
[48] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan and R. Jain, Machine learning-based network vulnerability analysis of Industrial Internet of Things, IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6822–6834, 2019.
[49] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan and R. Jain, WUSTL-IIoT-2021 dataset for IIoT cybersecurity research, Washington University in St. Louis, 2021.
[50] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu and A. A. Ghorbani, CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment, Sensors, vol. 23, no. 13, pp. 5941, 2023.
[51] 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, 2020.
[52] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras and H. Janicke, Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning, IEEE Access, vol. 10, pp. 40281–40306, 2022.
[53] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, A detailed analysis of the KDD CUP 99 data set, 2009 IEEE symposium on computational intelligence for security and defense applications (CISDA), pp. 1–6, 2009.
[54] N. Moustafa and J. Slay, UNSW-NB15: A comprehensive data set for network intrusion detection systems, 2015 military communications and information systems conference (MilCIS), pp. 1–6, 2015.
[55] I. Sharafaldin, A. H. Lashkari and A. A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion traffic characterization, Proceedings of the 4th international conference on information systems security and privacy (ICISSP), pp. 108–116, 2018.
[56] D. Arp, E. Quiring, F. Pendlebury, A. Warnecke, F. Pierazzi, C. Wressnegger, et al., Dos and don'ts of machine learning in computer security, 31st USENIX security symposium, pp. 3971–3988, 2022.
[57] G. Engelen, V. Rimmer and W. Joosen, Troubleshooting an intrusion detection dataset: The CICIDS2017 case study, 2021 IEEE security and privacy workshops (SPW), pp. 7–12, 2021.
[58] H. Hindy, D. Brosset, E. Bayne, A. K. Seeam, C. Tachtatzis, R. Atkinson and X. Bellekens, A taxonomy of network threats and the effect of current datasets on intrusion detection systems, IEEE Access, vol. 8, pp. 104650–104675, 2020.
[59] M. A. Ferrag, L. Maglaras, S. Moschoyiannis and H. Janicke, Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study, Journal of Information Security and Applications, vol. 50, pp. 102419, 2020.

[60] G. Apruzzese, P. Laskov, E. Montes de Oca, W. Mallouli, L. Brdalo Rapa, A. V. Grammatopoulos and F. Di Franco, The role of machine learning in cybersecurity, Digital Threats: Research and Practice, vol. 4, no. 1, pp. 1–38, 2023.
[61] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama and G. Zhang, Learning under concept drift: A review, IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 12, pp. 2346–2363, 2018.
[62] M. Merenda, C. Porcaro and D. Iero, Edge machine learning for AI-enabled IoT devices: A review, Sensors, vol. 20, no. 9, pp. 2533, 2020.
[63] M. G. S. Murshed, C. Murphy, D. Hou, N. Khan, G. Ananthanarayanan and F. Hussain, Machine learning at the network edge: A survey, ACM Computing Surveys, vol. 54, no. 8, pp. 1–37, 2021.
[64] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriyeh, A. Dehghantanha and G. Srivastava, Federated-learning-based anomaly detection for IoT security attacks, IEEE Internet of Things Journal, vol. 9, no. 4, pp. 2545–2554, 2022.