

Cyberattack Incident Response in Industrial Control Systems: A Critical Review of Detection-to-Response Approaches

Okafor, Chidozie Raymond Patrick^{1*}; Otuonye, Anthony Ifeanyi²;
Nwanga, Emeka Mathew³; Esomonu, Nkechi Faustina⁴

^{1;2;3;4}Department of Information Technology (IFT), School of Information & Communication Technology,
Federal University of Technology, Owerri, Nigeria.

Correspondence Author: Okafor, Chidozie Raymond Patrick^{1*}

Publication Date: 2026/10/26

Abstract: Industrial control systems (ICS) increasingly operate within interconnected information and operational technology environments, where cyber incidents can propagate from communication networks to control functions and physical processes. Conventional incident response remains dependent on analyst judgement, predefined rules and sequential manual activities, potentially causing delays and inconsistency during numerous, rapid as well as technically complex incidents. Although recent research has advanced machine-learning-based detection, anomaly monitoring, security orchestration and automated containment, existing literature remains fragmented across individual stages of the incident-response lifecycle. This critical review examined automated cyberattack incident response in ICS through four (4) computational functions: attack detection, attack-vector classification, severity assessment and response selection. Evidence from ICS-security surveys, machine-learning studies, automated-response research, threat-evaluation frameworks, process-aware monitoring and relevant standards were synthesized. The review revealed a principal research gap in effectively integrating detection outputs with appropriate and situation-specific response decisions. Hence, existing approaches frequently generated alerts and classifications without adequately incorporating asset criticality, physical-process consequences, response confidence, operational constraints and safety implications of autonomous actions. Accordingly, the review proposed an intelligent response pipeline that transforms heterogeneous security and process features into attack representations, classifies attacks, assigns sensitive severity levels as well as maps incident to appropriate response under explicit safety and human-oversight constraints. Priority research directions included: explainability, adaptive learning, data quality, benchmark limitations, response validation, adversarial robustness and safe autonomy. This synthesis, provided the foundation for developing intelligent computational models that progress from cyberattack detection, severity-aware to sensitive incident-response selection in ICS environments.

Keywords: Industrial Control Systems; Cyber-Attack; Incident Response; Machine Learning; Severity Assessment; Automated Response; Explainable AI; Operational Technology; Security Orchestration.

How to Cite: Okafor, Chidozie Raymond Patrick; Otuonye, Anthony Ifeanyi; Nwanga, Emeka Mathew; Esomonu, Nkechi Faustina (2026) Cyberattack Incident Response in Industrial Control Systems: A Critical Review of Detection-to-Response Approaches. *International Journal of Innovative Science and Research Technology*, 11(10), 1618-1626.
<https://doi.org/10.38124/ijisrt/26oct369>

I. INTRODUCTION

Industrial control systems (ICS) support the monitoring and control of physical processes in sectors such as electric power, manufacturing, oil and gas, water, transportation and other critical infrastructures. Modern ICS architectures increasingly combined operational technology (OT) with enterprise information technology, remote access, industrial networks, cloud services and interconnected digital components. This convergence improves operational visibility and flexibility but also increased the cyber-attack

surface. The National Institute of Standards and Technology (NIST), emphasized that operational technology (OT) environments, have distinctive performance, reliability and safety requirements that must be considered alongside cybersecurity controls (Stouffer et al., 2023). The security consequences of compromise in ICS differ materially from those of conventional information systems as a successful attack may affect confidentiality and integrity while altering process states, availability disruptions, equipment damage and other related safety consequences. Reviews of major ICS incidents, showed that attacks such as Stuxnet, BlackEnergy,

CrashOverride and Triton demonstrated how cyber operations can progress from initial compromise to manipulation of industrial processes and safety-relevant functions (Alladi et al., 2020; Firoozjaei et al., 2022). Consequently, incident response in ICS cannot be treated simply as an information technology (IT) containment problem; rather appropriate response actions must account for the state and criticality of the physical process.

Incident response in cyber physical systems is traditionally characterized around preparation, detection, analysis, containment, eradication, recovery and lessons learned. In recent times due to emerging trends in cyber-attacks, contemporary guidance has increasingly treated incident responses as an integral component of cybersecurity risk management. Hence, NIST SP 800-61 Rev. 3 finalized in 2025, explicitly linked incident response to the Cybersecurity Framework 2.0 and emphasized improving the efficiency and effectiveness of detection, response and recovery activities (Nelson et al., 2025). Also, in parallel, NIST SP 800-82 Rev. 3 stressed that OT security measures must preserve operational performance in ICS, reliability and safety (Stouffer et al., 2023). Given the growing volume and complexity of cyber events, automated approaches have been considered as one of the most efficient means to cybersecurity given rise to machine learning (ML) and deep learning (DL) techniques, widely investigated for intrusion and anomaly detection in ICS (Umer et al., 2022). Process-aware security monitoring has also expanded, with recent research emphasizing on the value of modelling the physical process rather than only relying exclusively on network observations (Rehman & Bahşi, 2024). However, detection is only one component of incident response, and related reviews of automatic incident-response solutions have found that research frequently focused on what information is used as input and what defensive action is produced as output. Hence, asset inventories, platform monitoring and network-traffic analysis were common inputs, while network isolation was a common response output. Subsequently, many proposed solutions described their inputs and outputs only vaguely, indicating that the field still lacked a mature, well-specified end-to-end response architectures (Karlzén & Sommestad, 2023).

This observation exposes a central problem, as a system may accurately identify a malicious event without being able to determine how serious the event is, which asset is most affected, what operational consequences are likely, or which response is safest and most appropriate. Accordingly, this review treats automated ICS incident response as a sequence of related but distinct computational decisions involving detection, attack-vector classification, severity assessment, and response selection. Separating these functions would reveal where current approaches were sufficient, where they remained weak and where integration is required. The

contribution of this review is therefore not intended as a catalogue of cyberattacks, rather a critical integrative review, providing a critical synthesis of the computational pathway through an observed security event to actionable response by identifying the limitations in current approaches, evaluating the requirements for safe automation in ICS, as well as proposing a conceptual research architecture for intelligent, severity-aware and sensitive responses.

II. REVIEW METHODOLOGY

➤ Design

The research adopted a structured critical-review design due to the heterogeneous nature of the literature on ICS incident response, which would render a quantitative meta-analysis inappropriate without substantial harmonization of outcomes. The review reporting logic, strategically followed the transparency principles promoted by PRISMA 2020, stating the review objective, information source, eligibility logic and synthesis approach (Page et al., 2021).

➤ Questions

- What attack characteristics and ICS assets are most relevant to automated incident-response decision making?
- Which computational approaches have been used for attack detection and classification in ICS?
- To what extent do existing approaches support severity assessment and response selection over detection alone?
- What technical, operational and safety constraints are evident in autonomous ICS incident responses?
- What capabilities should an intelligent end-to-end ICS incident-response model provide to close the identified gaps?

➤ Information Source

The review was grounded in peer-reviewed and authoritative sources, identified through targeted searches of major scholarly publishers and standards repositories, including Elsevier/ScienceDirect, Association of Computing Machinery (ACM), Springer Nature and NIST. Search concepts combined terms related to industrial control systems, operational technology, cyber-attack, incident response, automated response, machine learning, severity assessment, process monitoring, security orchestration and explainable artificial intelligence. Particular attention was given to recent literature published from 2020 onward while retaining foundational studies relevant to establishing the development of ICS security and incident response.

➤ Inclusion and Exclusion Criteria

The criteria for the review inclusion and exclusion logic are presented in table 1 as follows:

Table 1 Inclusion & Exclusion Logic

SN	Inclusion	Exclusion
1.	Peer-reviewed studies addressing ICS/OT cybersecurity, attack detection, threat evaluation, incident response & automated response	Studies unrelated to ICS/OT, & cyber-physical operational environments.
2.	Standards & authoritative guidance directly relevant to ICS/OT security & incident response	Commercial promotional material used as evidence for technical performance
3.	Studies proposing, or evaluating ML, DL, fuzzy, rule-based, orchestration, & other computational response mechanisms	Papers that merely mention automation without substantive relevance to response
4.	Recent review & survey studies relevant for establishing state-of-the-art research gaps & lapses	Duplicated, clearly irrelevant, and bibliographically inconsistent sources
5.	Studies discussing safety, physical-process impact, explainability & operational constraints	Generic cybersecurity material that was irrelevant for ICS-specific analysis.

➤ Data Extraction

The review was synthesized against a common analytical matrix consisting of: attack type; targeted ICS assets; input data; computational technique; detection capability; classification capability; risk assessment; response capability; degree of automation; explainability; adaptability; safety awareness; evaluation metrics; and principal limitations. This matrix converted the descriptive review into a comparative assessment of what each class of approach actually accomplishes.

III. CYBERATTACK & INCIDENT-RESPONSE IN ICS

➤ ICS as a Cyber-Physical Environment

ICS combine sensing, communication, computation and actuation, with typical components including sensors and actuators, programmable logic controllers (PLCs), remote terminal units (RTUs), intelligent electronic devices (IEDs), supervisory control and data acquisition (SCADA) servers, human-machine interfaces (HMIs), engineering workstations and industrial communication networks. NIST SP 800-82

Rev.3, emphasized that OT systems directly interact with the physical environment and should be protected while preserving deterministic performance, reliability and safety (Stouffer et al., 2023). This architecture creates a fundamental distinction between cyber and physical consequences. A malicious network packet may be technically similar to traffic observed in an IT environment, yet its consequence depends on what it reaches and what industrial process it influences. A command directed at a PLC controlling a protection function is not equivalent in risk to an event affecting a non-critical workstation. Consequently, incident-response intelligence, requires necessary information about assets, process, dependencies and operational state.

➤ Attack Vectors & Required Automated Response

In automated ICS response, it is imperative to classify attack vectors by the operational pathway through which they affect the control environment in order to determine the best approach of automating its response. This classification and corresponding automated response actions are shown in table 2:

Table 2 Cyber-Attack Vectors in ICS

SN	Attack Class	ICS Manifestation	Response
1.	Network Disruption	DoS/DDoS; Protocol Flooding; Communication Interface.	Traffic Filtering; Network Segmentation; Controlled Isolation
2.	Integrity Manipulation	False Data Injection; Replay Attacks; Command Manipulation.	Process-Aware Validation.
3.	Control System Compromise	PLC/RTU/IED Compromise; Unauthorized Commands; Malicious Logic	Strong Safeguards.
4.	Credential Compromise	Phishing; Credential Theft; Privilege Escalation; Remote Access Abuse	Account Containment; Credential Revocation; Access Restriction
5.	Malware	Destructive Malware; Ransomware affecting OT Support Systems	Containment; Recovery; Preserve Critical Operations
6.	Man-in-the-Middle	Traffic Interception; Command Alteration; Session Manipulation	Correlation Network & Process Evidences
7	Process-Aware Attacks	Manipulation Designed to Remain within Plausible Process Ranges	Requires physical response rather than network signatures alone
8	APT Activity	Credential Access; Lateral Movement; Persistence & Process Impact	Event Correlation; Attack-Chain Reasoning, & Eventual Isolated.

➤ *Challenges in Cyber-Attack Incident Response Automation in ICS*

There are three (3) characteristics that makes automated response particularly difficult in industrial control systems:

- ICS often prioritizes availability and safety over rapid system modification, making a conventional IT response such as immediate shutdown mostly unacceptable;
- Industrial processes are interdependent, so isolating one component can cause cascading operational effects; and
- Attack evidence may be distributed across network traffic, device logs, control commands and physical-process measurements.

Hence, an automated response engine must integrate heterogeneous evidences rather than treating an intrusion alert as a complete representation of the incident.

IV. COMPUTATIONAL APPROACHES TO AUTOMATED ICS INCIDENT RESPONSE

➤ *Rule & Signature-Based*

Rule-based systems remain useful because they are transparent, computationally efficient and relatively easy to validate, as signatures, thresholds and predefined playbooks can identify known attack patterns and trigger deterministic actions. Their principal limitation is dependence on prior knowledge; hence they are less effective against previously unseen attacks, polymorphic behaviour, multi-stage campaigns and process-aware attacks that remain within expected network patterns.

➤ *Statistical and Anomaly-Based*

Statistical approaches identify deviations from a baseline. In ICS, deterministic or periodic process behaviour can provide strong baselines for anomaly detection. However, operational environments change because of maintenance, load variation, configuration changes and process transitions. A static threshold can therefore produce false alarms or fail to identify slowly evolving attacks. Effective anomaly detection requires context-aware baselines and mechanisms for handling legitimate process changes.

➤ *Machine Learning*

ML has become one of the dominant computational approaches to ICS intrusion and anomaly detection. Umer et al. (2022) categorized ML research into supervised, semi-supervised, unsupervised and reinforcement-learning approaches and emphasized that methods operate on both network-level and physical-process data. ML can improve classification and pattern recognition, but its practical utility depends on representative data, feature quality, class balance, generalization across sites and resistance to concept drift. Ensemble models such as Random Forest are particularly attractive when heterogeneous tabular features must be classified with reasonable interpretability and limited preprocessing. Their usefulness for an end-to-end response system, however, should not be overstated: a classifier answers the question of what an event resembles; it does not automatically determine the operationally safest response.

➤ *Deep Learning*

DL models can capture complex nonlinear and temporal relationships and have been widely investigated for ICS anomaly detection. Recent comparative studies continue to examine CNNs, LSTMs, autoencoders, GRUs and related architectures on benchmark datasets. Their principal weaknesses include computational cost, dependence on large or representative datasets, reduced transparency and susceptibility to distribution shifts. A high detection score on a benchmark dataset does not by itself establish suitability for autonomous response in a live plant.

➤ *Fuzzy Reasoning*

Fuzzy reasoning is useful where response decisions depend on imprecise concepts such as low, moderate and high severity. Instead of forcing every variable into a rigid threshold, fuzzy membership functions can represent gradual transitions and combine heterogeneous risk factors. This makes fuzzy reasoning a candidate mechanism for the severity-assessment stage, especially when expert knowledge and quantitative evidence must be integrated. Its limitation is that membership functions and rules require defensible design and validation.

➤ *Reinforcement Learning*

Reinforcement learning (RL) offers a natural formulation for sequential response decisions because an agent can learn actions based on state, reward and future consequences. However, direct exploration in a real ICS can be unsafe. RL is therefore more appropriate for controlled environments, digital twins, testbeds or offline policy learning unless strong safety constraints are imposed. The challenge is not merely learning an effective policy but demonstrating that the policy will not generate unsafe control actions under uncertainty.

➤ *Security Orchestration, Automated and Response (SOAR)*

Security orchestration, automation and response (SOAR) platforms demonstrate how alerts can be connected to predefined playbooks and technical controls. They can reduce repetitive analyst work and accelerate containment. Nevertheless, orchestration is not equivalent to intelligence. A playbook executes a decision; it does not necessarily determine whether the decision is appropriate for the current industrial state. The strongest architecture is therefore likely to combine orchestration with an analytical layer capable of classification, severity estimation and context-aware response selection.

➤ *Hybrid Intelligent Approaches*

Hybridization is increasingly attractive because no single computational method simultaneously provides strong detection, contextual reasoning, interpretability, adaptation and safe response. A hybrid architecture can use a supervised classifier for attack identification, fuzzy reasoning for severity assessment, rule or policy constraints for safety and orchestration for execution. Such a design separates responsibilities and makes validation more tractable.

V. CRITICAL COMPARATIVE ANALYSIS

Table 3 Comparative Analysis of Approaches to Automated Response

Approach	Detection	Classification	Severity	Response Selection	Explainability	ICS Suitability	Key Limitation
Rules	High for known patterns	Limited	Rule-defined	Strong for pre-defined cases	High	High for known scenarios	Poor adaptation to novel attacks
Statistical anomaly detection	High for deviations	Limited–moderate	Threshold-based	Usually, external	High	Moderate–high	Sensitive to changing baselines
Ensemble ML	High	High	Supports down-stream model	Usually, indirect	Moderate–high	High for tabular data	Depends on features and labelled data
Deep Learning	High	High	Possible	Limited in most studies	Low–moderate	Potentially high	Opacity, data and drift issues
Rules	High for known patterns	Limited	Rule-defined	Strong for pre-defined cases	High	High for known scenarios	Poor adaptation to novel attacks
Statistical anomaly detection	High for deviations	Limited–moderate	Threshold-based	Usually, external	High	Moderate–high	Sensitive to changing baselines
Ensemble ML	High	High	Supports down-stream model	Usually, indirect	Moderate–high	High for tabular data	Depends on features and labelled data
Deep Learning	High	High	Possible	Limited in most studies	Low–moderate	Potentially high	Opacity, data and drift issues
Fuzzy logic	Low as standalone detector	Low–moderate	Strong	Strong when rule base exists	High	High as decision layer	Rule/membership design
Reinforcement Learning	State dependent	Possible	Possible	Strong potential	Low–moderate	Conditional	Safety and exploration risk
SOAR	Depends on upstream tools	Rule-based	Rule-based	Strong execution	High	High with OT-aware controls	Does not inherently provide contextual intelligence
Hybrid Intelligent Model	Potentially high	High	High	High	Potentially high	High, if safety-aware	Integration and validation complexity

The comparison revealed an important distinction: computational performance at the detection stage does not imply end-to-end incident-response capability. Majority of the ICS literature, concentrated on whether an algorithm can distinguish normal from abnormal behaviour, or classify attack types. Fewer studies explicitly connected those outputs to a validated response policy, revealing a gap that is architectural, rather than purely algorithmic. Furthermore, attack type cannot be singlehandedly used to determine severity, as related network event may pose varying consequences depending on the targeted asset, its process role, current operating state and confidence in the classification. Existing threat-evaluation research has already

demonstrated the usefulness of combining attack sophistication and consequences into a threat score (Firoozjaei et al., 2022), supporting a broader response model where severity is treated as a dependent decision variable. Also, current research often evaluated models using classification metrics such as accuracy, precision, recall and F1-score. These metrics were necessary for detection and classification, but they were insufficient for evaluating an autonomous response system. Hence, a response model should additionally be assessed for response correctness, time to response, containment effectiveness, false-action rate, operational disruption, recovery impact and safety constraint violations.

VI. AN INTEGRATED CONCEPTUAL FRAMEWORK FOR INTELLIGENT ICS INCIDENT RESPONSE

The critical synthesis, supports an integrated architecture with seven (7) functional stages to include: data acquisition, preprocessing and normalization, attack detection, attack-vector classification, severity assessment, response selection and controlled execution with feedback. This architecture is deliberately modular in order to ensure that each stage is independently evaluated while the complete system is assessed as an incident-response pipeline (engine).

➤ Stage 1: Data Acquisition

Inputs should include network traffic features, control commands, device events, authentication events, process measurements, asset information, attack history and relevant threat intelligence. Asset identity and criticality are especially important because the meaning of an event depends on what the event can affect.

➤ Stage 2: Preprocessing & Feature Representation

The preprocessing layer should remove duplicate or corrupt observations, handle missing values, standardize representations and construct features that preserve attack-relevant information. Feature engineering should combine cyber indicators with contextual variables such as asset criticality and operational state. This prevents the response engine from treating a purely cyber signal as independent of the physical process.

➤ Stage 3: Attack Detection & Classification

The detection function determines whether the observed event is benign or malicious. If malicious, the classification function identifies the attack vector or attack family. A conceptual classifier can be represented as:

$$C(x) \rightarrow y, \quad (1)$$

Where:

x = the feature vector;

y = the predicted attack class.

Classification confidence should be retained rather than discarded because confidence can inform subsequent severity assessment.

➤ Stage 4: Severity Assessment

Severity should not be equated with attack class, rather a more conceptual formulation is generated as follows:

$$S = f(F, I, C, A, P) \quad (2)$$

Where:

S = incident severity;

F = attack frequency;

I = attack importance (estimated threat significance);

C = the criticality of the targeted asset or infrastructure;

A = anticipated operational / physical consequence;

P = classification/prediction confidence.

The variables can be normalized and mapped to Low, Medium and High severity through a fuzzy or other interpretable decision layer. These formulations were conceptual and requires empirical calibration before they can be put to operational use.

➤ Stage 5: Response Selection

The response engine should map the assessed incident state to a set of permissible actions. Candidate actions may include continued monitoring, alert escalation, blocking or filtering traffic, revoking credentials, isolating a network segment, isolating a non-critical endpoint, restoring a clean configuration, initiating recovery procedures or escalating to a human operator. The selected action should depend on severity, asset criticality, confidence, operational state and safety constraints.

A useful formulation is an action-selection function:

$$R = g(S, C, A, O, H) \quad (3)$$

Where:

S = severity,

C = asset criticality,

A = attack classification,

O = operational state,

H = human-oversight requirements.

The function should not be interpreted as a universal mathematical solution; but rather a guide that defined the necessary variables that a validated computational model must consider.

➤ Stage 6: Human-Oversight Layer

Autonomy should be graduated rather than binary. Low-risk events may be handled automatically; moderate-risk events may be executed automatically subject to predefined safeguards; high-impact actions should require human approval or human-on-the-loop supervision unless a formally validated safety policy authorizes autonomous execution. This design acknowledges the operational reality of ICS and avoids the assumption that maximum automation is always optimal.

➤ Stage 7: Feedback

Post-response outcomes should feed back into the system. Successful responses, false positives, failed containment and changing operational conditions can provide

information for model recalibration. However, continuous learning should be controlled because unverified updates can introduce model drift or data-poisoning risks. Adaptation

should therefore be governed by change control, validation and rollback mechanisms.

VII. IDENTIFIED RESEARCH GAPS

The gaps identified from the review are outlined in table 4:

Table 4 Research Gaps

Gap	Evidence from Literature	Implication	Priority
Detection-Response Disconnection	Many studies focus on detection/anomaly classification; automated-response reviews identify vague input/output descriptions	Need an explicit mapping from classified incidents to validated response actions	Critical
Limited Severity Modelling	Threat evaluation frameworks exist, but severity is not consistently integrated into response pipelines	Severity should combine attack characteristics, asset criticality, consequence and confidence	Critical
Insufficient Physical-Process	Process-aware monitoring is emerging but remains distinct from response selection	Response models should integrate cyber and process evidence	Critical
Safety of Autonomous Action	Automation research often emphasizes speed; ICS guidance emphasizes safety and reliability	Autonomy requires safety constraints and human oversight	Critical
Explainability	Advanced ML/DL improves detection but may reduce transparency	Response decisions should be traceable to influential evidence and rules	High
Dataset	Benchmark datasets are useful but may not represent site-specific conditions	Cross-dataset and cross-domain validation is required	High
Concept Drift	Industrial processes change over time	Models need controlled adaptation and drift monitoring	High
Response-Level Evaluation	Accuracy/F1 are common detection metrics	Evaluation must include response time, action correctness and operational impact	High
Adversarial Robustness	AI-enabled ICS security introduces poisoning/evasion risks	Models and data pipelines require adversarial security controls	High
Lack of Standardized End-to-End Benchmarks	Studies use heterogeneous datasets and evaluation procedures	A common benchmark should evaluate detection through response	High

VIII. EVALUATION REQUIREMENTS FOR AUTOMATED ICS INCIDENT RESPONSE

A quality computational response model should not be validated using only accuracy, confusion matrices or F1-

score measurements. Those measures are appropriate for classification components but do not establish that a response policy is safe or effective. The review therefore recommends a layered evaluation framework as shown in table 5:

Table 5 Requirements for Automated Incident Response

Evaluation	Recommended Measures	Purpose
Detection	Precision, recall, F1-score, false-positive rate, false-negative rate, & ROC-AUC where appropriate	Determine ability to distinguish malicious from benign events
Classification	Macro/micro F1, per-class precision/recall, confusion matrix, balanced accuracy	Determine attack-vector recognition, especially under class imbalance
Severity	Agreement with expert labels, weighted error, calibration, sensitivity analysis	Determine whether risk/severity estimates are meaningful
Response	Response accuracy, containment success, inappropriate-action rate, time to response	Determine whether selected actions are correct and timely
Operational impact	Service disruption, process interruption, recovery time, critical-asset impact	Determine whether response causes unacceptable operational consequences
Explainability	Feature traceability, decision consistency, & explanation fidelity	Determine whether operators can understand and audit decisions
Robustness	Cross-dataset performance, drift sensitivity, adversarial tests	Determine generalization and resilience
Safety	Constraint violations, unsafe action rate, human-override frequency	Determine suitability for high-consequence environments

For a hybrid model, validation should be conducted separately for each functional component and then for the complete pipeline. An unseen test set should be used for final evaluation, and class imbalance should be explicitly addressed. Where possible, temporal splits are preferable to random splits because industrial attack data are time dependent and random splitting can produce overly optimistic estimates when adjacent observations are highly correlated.

IX. DISCUSSION

The central finding of this review is that the field has progressed further in automated observation than in automated decision making. Detection algorithms can identify suspicious behaviour with increasingly sophisticated models, but a security alert is not equivalent to a response decision. The operational meaning of an event emerges only after the system considers attack type, target, criticality, process state, confidence and potential consequence.

This distinction has direct implications for model design. A high-performing classifier may be an excellent component of an incident-response system without being an adequate incident-response system by itself. Similarly, fuzzy reasoning may be useful for severity assessment without being a detector. SOAR may execute actions effectively without determining which action should be selected. The appropriate research direction is therefore functional integration rather than the search for a single universally superior algorithm.

The review also challenges the assumption that fully automated response is inherently better than human-supported response. In IT environments, rapid isolation may often be beneficial. In ICS, however, isolation can affect control availability and physical processes. NIST's OT guidance explicitly recognizes performance, reliability and safety requirements. Consequently, automation should be optimized for safe and appropriate response, not simply minimum response time.

Explainability is similarly important. An operator should be able to determine why an incident was classified as high severity and why a particular action was selected. This is especially relevant where the model recommends disruptive actions. A transparent decision layer can also support auditing, model debugging and human override.

Finally, the review indicates that future research should move from isolated algorithm comparisons toward end-to-end evaluation. Research questions should therefore shift from 'Which classifier has the highest accuracy?' to 'Which architecture provides the most reliable, explainable and safe progression from heterogeneous cyber-physical evidence to an appropriate response?' This is a more consequential and scientifically meaningful problem for ICS cybersecurity.

X. CONCLUSION

This review critically examined the transition from conventional incident response toward intelligent and automated cyberattack incident response in industrial control systems. The literature demonstrated substantial progress in attack detection, anomaly monitoring, machine learning, process-aware security and security orchestration. However, the evidence also indicated that the research landscape remained fragmented as detection and classification were considerably more mature than severity-aware response selection, and many automated-response proposals do not adequately specify the relationship between their inputs, decision logic and defensive outputs.

The principal research gap is therefore the absence of sufficiently integrated, and situation-sensitive computational pipelines that can progress from cyberattack evidence to an appropriate response. A viable intelligent architecture should therefore distinguish attack detection, attack-vector classification, severity assessment and response selection while incorporating asset criticality, operational criticality, prediction confidence and safety constraints. Hybrid approaches are particularly promising because they can confidently combine data-driven classifications with interpretable severity reasoning to generate a controlled response orchestration.

The review further concludes that response automation in ICS should not be evaluated solely by speed or classification accuracy, rather an operationally meaningful system that demonstrates detection reliability, classification performance, severity validity, response correctness, explainability, adaptability, and operational resilience. These requirements provide a stronger foundation for future development of intelligent computational models for cyberattack incident response in power-grid and other critical-infrastructure in ICS environments.

RECOMMENDATIONS

Future research should develop integrated computational pipelines that systematically connect attack detection, classification, severity assessment, and response selection;

Response mechanisms should incorporate asset and operational criticality, prediction confidence, and safety constraints to ensure appropriate and controlled defensive actions against cyberattacks;

Future evaluations should assess response systems using detection reliability, classification performance, severity validity, response correctness, explainability, adaptability, and operational resilience.

REFERENCES

- [1]. Alladi, T., Chamola, V., & Zeadally, S. (2020). Industrial Control Systems: Cyberattack trends and countermeasures. *Computer Communications*, 155, 1–8. <https://doi.org/10.1016/j.comcom.2020.03.007>
- [2]. Aslam, M. M., Tufail, A., & Gul, H. (2025). Artificial intelligence for secure and sustainable industrial control systems: A survey of challenges and solutions. *Artificial Intelligence Review*, 58, 349. <https://doi.org/10.1007/s10462-025-11320-9>
- [3]. Ding, D., Han, Q.-L., Xiang, Y., Ge, X., & Zhang, X.-M. (2018). A survey on security control and attack detection for industrial cyber-physical systems. *Neurocomputing*, 275, 1674–1683. <https://doi.org/10.1016/j.neucom.2017.10.009>
- [4]. Ferdous, J., Islam, R., Mahboubi, A., & Islam, M. Z. (2023). A review of state-of-the-art malware attack trends and defense mechanisms. *IEEE Access*, 11. <https://doi.org/10.1109/ACCESS.2023.3328351>
- [5]. Firoozjaei, M. D., Mahmoudiyar, N., Baseri, Y., & Ghorbani, A. A. (2022). An evaluation framework for industrial control system cyber incidents. *International Journal of Critical Infrastructure Protection*, 36, 100487. <https://doi.org/10.1016/j.ijcip.2021.100487>
- [6]. Goncharov, E., Kruglov, K., & Dashchenko, Y. (2019). Five ICS cybersecurity myths based on Kaspersky Lab ICS CERT experience. *at-Automatisierungstechnik*, 67(5). <https://doi.org/10.1515/auto-2019-0016>
- [7]. Guembe, B., Azeta, A., Misra, S., Osamor, V. C., Fernandez-Sanz, L., & Pospelova, V. (2022). The emerging threat of AI-driven cyber attacks: A review. *Applied Artificial Intelligence*, 36(1), 2037254. <https://doi.org/10.1080/08839514.2022.2037254>
- [8]. Jalalvand, F., Baruwat Chhetri, M., Nepal, S., & Paris, C. (2025). Alert prioritization in security operations centers: A systematic survey on criteria and methods. *ACM Computing Surveys*, 57(2). <https://doi.org/10.1145/3695462>
- [9]. Karlzén, H., & Sommestad, T. (2023). Automatic incident response solutions: A review of proposed solutions' input and output. *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES 2023)*, 37:1–37:9. <https://doi.org/10.1145/3600160.3605066>
- [10]. Kim, H. M., & Lee, K. H. (2022). IIoT malware detection using edge computing and deep learning for cybersecurity in smart factories. *Applied Sciences*, 12(15), 7679. <https://doi.org/10.3390/app12157679>
- [11]. Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security and Applications*, 22, 113–122. <https://doi.org/10.1016/j.jisa.2014.09.005>
- [12]. Mazini, M., Shirazi, B., & Mahdavi, I. (2019). Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms. *Journal of King Saud University – Computer and Information Sciences*, 31(4), 477–484. <https://doi.org/10.1016/j.jksuci.2018.03.011>
- [13]. McLaughlin, S., Konstantinou, C., Wang, X., Davi, L., Sadeghi, A.-R., Maniatakos, M., & Karri, R. (2016). The cybersecurity landscape in industrial control systems. *Proceedings of the IEEE*, 104(5), 1039–1057. <https://doi.org/10.1109/JPROC.2015.2512235>
- [14]. Nelson, A., Rekhi, S., Souppaya, M., & Scarfone, K. (2025). Incident Response Recommendations and Considerations for Cybersecurity Risk Management: A CSF 2.0 Community Profile. NIST Special Publication 800-61 Rev. 3. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-61r3>
- [15]. Ni, C., & Li, S. C. (2024). Machine learning enabled Industrial IoT Security: Challenges, Trends and Solutions. *Journal of Industrial Information Integration*, 38, 100549. <https://doi.org/10.1016/j.jii.2023.100549>
- [16]. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- [17]. Rehman, M. ur, & Bahşi, H. (2024). Process-aware security monitoring in industrial control systems: A systematic review and future directions. *International Journal of Critical Infrastructure Protection*, 47, 100719. <https://doi.org/10.1016/j.ijcip.2024.100719>
- [18]. Salahdine, F., & Kaabouch, N. (2019). Social engineering attacks: A survey. *Future Internet*, 11(4), 89. <https://doi.org/10.3390/fi11040089>
- [19]. Stouffer, K., Pease, M., Tang, C., Zimmerman, T., Pillitteri, V., Lightman, S., Hahn, A., Saravia, S., Sherule, A., & Thompson, M. (2023). Guide to Operational Technology (OT) Security. NIST Special Publication 800-82 Rev. 3. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-82r3>
- [20]. Stouffer, K., Pillitteri, V., Lightman, S., Abrams, M., & Hahn, A. (2015). Guide to Industrial Control Systems (ICS) Security. NIST Special Publication 800-82 Rev. 2. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.SP.800-82r2>
- [21]. Umer, M. A., Junejo, K. N., Jilani, M. T., & Mathur, A. P. (2022). Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations. *International Journal of Critical Infrastructure Protection*, 38, 100516. <https://doi.org/10.1016/j.ijcip.2022.100516>
- [22]. Wang, Z., Sun, L., & Zhu, H. (2020). Defining social engineering in cybersecurity. *IEEE Access*, 8, 85069–85078. <https://doi.org/10.1109/ACCESS.2020.2992807>