



OPEN ACCESS

EDITED BY

Jairo Gutierrez,
Auckland University of Technology,
New Zealand

REVIEWED BY

Safarudin Gazali Herawan,
Binus University, Indonesia
Andrea Pinto,
University of Los Andes, Colombia

*CORRESPONDENCE

Liya Erbolkyzy Bektemir
✉ liya15bektemir@gmail.com

RECEIVED 27 November 2025

REVISED 05 February 2026

ACCEPTED 10 February 2026

PUBLISHED 02 March 2026

CITATION

Amirkhanova GA,
Prokopovych-Tkachenko DI,
Adilzhanova SA, Zubchenko N and
Bektemir LE (2026) Machine
learning-based early incident detection
system in a bakery plant's industrial
network: a cognitive model for
counteracting hybrid threats.
Front. Comput. Sci. 8:1751284.
doi: 10.3389/fcomp.2026.1751284

COPYRIGHT

© 2026 Amirkhanova,
Prokopovych-Tkachenko, Adilzhanova,
Zubchenko and Bektemir. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Machine learning-based early incident detection system in a bakery plant's industrial network: a cognitive model for counteracting hybrid threats

Gulshat Amanzholovna Amirkhanova¹,
Dmytro Ihorovych Prokopovych-Tkachenko²,
Saltanat Almykhametovna Adilzhanova³, Nazar Zubchenko² and
Liya Erbolkyzy Bektemir^{✉*}

¹Department of Artificial Intelligence and Big Data, Faculty of Information Technology, Al-Farabi Kazakh National University, Almaty, Kazakhstan, ²Department of Cybersecurity and Information Technologies, University of Customs and Finance, Dnipro, Ukraine, ³Department of Cybersecurity and Cryptology, Faculty of Information Technology, Al-Farabi Kazakh National University, Almaty, Kazakhstan

Introduction: In the context of growing cyber risks to critical industries, including bakery complexes, this paper proposes a cognitive architecture for early incident detection in the operational technology (OT) network.

Methods: The architecture integrates User and Entity Behavior Analytics (UEBA), a Security Information and Event Management (SIEM) system, and Zero Trust principles, focusing on hybrid threats: from external attacks on industrial controllers, such as programmable logic controllers (PLCs) to internal operator errors. At the analytics layer, two complementary deep learning pipelines are used: a convolutional neural network (CNN) + long short-term memory (LSTM) (CNN + LSTM) model for detecting low-level network patterns (Byte2Image) and an auto-encoder (AE) combined with LSTM (AE + LSTM model) for predicting time-series data and identifying anomalies in equipment telemetry. An adaptive threshold decision procedure is introduced for the first time, optimizing both accuracy and computational resources on edge nodes. The architecture complies with the IEC 62443 and ISO/IEC 27019 standards.

Results and discussion: High performance metrics, specifically Precision, were demonstrated in the bakery plant's digital twin scenarios.

KEYWORDS

anomaly detection, CNN-LSTM, deep learning, digital twin, industrial control systems (ICS), User and Entity Behavior Analytics (UEBA), zero trust

1 Introduction

Industrial enterprises in the food sector, particularly bakery plants, are classified as critical infrastructure, where the stability of technological processes directly impacts food safety, economic efficiency, and supply chain continuity. Any disruption to the operation of conveyors, ovens, dispensers, or packaging equipment leads to production line halts, spoilage of product

batches, and non-compliance with sanitary and technological standards (Knowles et al., 2015; Bhamare et al., 2020; Cherdantseva et al., 2016). Amidst active digitalization and the increasing interconnectedness of production systems, the number of hybrid cyberattacks has significantly increased, combining external network intrusions into industrial controllers, exploitation of supply chain vulnerabilities, and internal risks caused by personnel errors or malicious operator actions.

Deficiencies in the implementation of industrial communication protocols pose a particular threat, including the use of outdated Simple Network Management Protocol (SNMP) versions 1 and 2c, as well as Secure Sockets Layer (SSL)/Transport Layer Security (TLS) protocols in Human-Machine Interfaces (HMI) and software update channels. These vulnerabilities create opportunities for data manipulation, injection of malicious firmware, and compromising the integrity of the production process (Xie et al., 2021; Li et al., 2024; Ahmad et al., 2023).

The regulatory framework of the Republic of Kazakhstan in the field of industrial system cybersecurity is based on ST RK ISO/IEC 27001-2022 and ST RK GOST R 56939-2016 standards, which harmonize international requirements with national regulatory provisions (Nisar et al., 2022; Zhang et al., 2023). These documents prescribe the establishment of information security management systems, the application of multi-layered defense, and risk-oriented approaches. International standards IEC 62443, ISO/IEC 27019, and the European Union's NIS2 Directive define additional obligations for critical infrastructure entities regarding incident identification and response (Knowles et al., 2015; Tariq et al., 2019; Qureshi et al., 2024).

For enterprises in Kazakhstan, the problem of protecting industrial networks is strategically important due to the implementation of state programs for digital transformation and the development of a secure Industry 4.0 (Wang et al., 2022; Zhang et al., 2023). Modern practice demands systems capable of comprehensively monitoring network traffic, equipment telemetry, and personnel behavior, ensuring timely threat detection without disrupting the continuity of the production cycle.

Recent scientific studies confirm the effectiveness of machine learning methods in analyzing network and technological data. Deep learning models can identify hidden patterns in telemetry, detect atypical command sequences, and recognize potentially dangerous operator actions (Anthi et al., 2021; Gauthama Raman et al., 2021; Kravchik and Shabtai, 2018; Tuptuk et al., 2021; Inoue et al., 2017; Lin et al., 2022; Abhishek and Singh, 2023). The work by Lin et al. (2022) showed the effectiveness of adaptive thresholding to reduce false positives, and the study by Abhishek and Singh (2023) confirmed the benefits of hybrid architectures combining different types of neural networks. Publications (Holdbrook et al., 2024) emphasize the importance of considering the technological cycle context and using digital twins for secure testing of defense systems (Zhang et al., 2023; Homaei et al., 2024; Dietz and Pernul, 2020).

The objective of this research is to develop and experimentally validate a cognitive system for early incident detection in a bakery plant's industrial network, considering the regulatory requirements of the Republic of Kazakhstan and international information security standards. The main task is to create an architecture that integrates the analysis of technological processes, network exchange, and personnel

behavior with adaptive adjustment of response thresholds based on system status and available computational resources.

The object of the study is the technological infrastructure of a bakery enterprise, and the subject is the methods of detecting and preventing violations based on event analytics, statistical technological parameters, and operator activity profiles. The work applies methods of statistical analysis, risk modeling, deep learning algorithms for spatio-temporal data, and experimental validation on a digital twin of the production line.

The scientific novelty lies in creating a cognitive architecture for the food industry that combines network and technological indicators with the context of the production cycle, accounts for SSL/TLS and SNMP vulnerabilities, provides adaptive control of the false positive rate, and complies with the requirements of IEC 62443, ISO/IEC 27019, ST RK ISO/IEC 27001-2022, and ST RK GOST R 56939-2016.

The practical significance of the results is expressed in increasing the accuracy and speed of incident detection with limited computational resources on edge nodes. Implementation of the proposed system allows for a reduction in the meantime to respond to 1–1.5 s, a decrease in the false positive rate to 0.5%, and an increase in the stability of the technological process while adhering to national information security standards (Nisar et al., 2022; Zhang et al., 2023).

The research includes an introduction, main sections devoted to methods, results, and discussion, as well as a list of literature containing 31 reviewed sources, including regulatory documents of the Republic of Kazakhstan and international standards.

Research Hypothesis. We hypothesize that integrating technological process context (recipes, equipment states) and User and Entity Behavior Analytics (UEBA) into a unified cognitive correlation layer will significantly reduce the False Positive Rate (FPR) compared to isolated deep learning pipelines. Specifically, we posit that the cognitive architecture can distinguish between legitimate operational deviations (e.g., recipe changes) and true hybrid threats, thereby reducing the FPR by at least 20% while maintaining a detection latency compliant with soft real-time requirements (<1.5 s) suitable for edge deployment.

2 Methods

In conditions of high complexity and interconnectedness of bakery production processes, methods of system data analysis that allow for the combination of equipment telemetry, network logs, and operator behavioral characteristics into a single cognitive model are of particular importance. The purpose of this section is to describe the methodological foundations underlying the proposed architecture for early incident detection in the bakery plant's industrial network.

The initial development principles comply with the regulatory requirements of the Republic of Kazakhstan established in the standards ST RK ISO/IEC 27001-2022 and ST RK GOST R 56939-2016, as well as the international documents IEC 62443 and ISO/IEC 27019, which regulate the protection of industrial control systems (Knowles et al., 2015; Tariq et al., 2019; Nisar et al., 2022; Zhang et al., 2023; Qureshi et al., 2024). According to these provisions, the system must ensure comprehensive risk management, communication channel protection, trusted zone control, and early warning of process integrity violations.

The research methodology is based on the principles of cognitive analysis, which integrates elements of machine learning, probabilistic modeling, and event correlation. The system architecture includes three interconnected layers: data collection and normalization, analytical processing, and cognitive correlation. The first layer ensures the integration of telemetry streams and event logs, the second implements analysis and prediction algorithms, and the third links the results to the context of the production cycle, considering shift schedules, recipes, and technological maps.

For network activity analysis, byte stream data is converted into a matrix representation (the Byte2Image method), which allows the use of Convolutional Neural Networks (CNN) to identify patterns in packet structure (Anthi et al., 2021; Kravchik and Shabtai, 2018). For the analysis of technological parameters, such as temperature, humidity, and drive load, a combination of an Autoencoder (AE) and a Recurrent Neural Network (RNN) is applied, capable of predicting normal behavior and fixing deviations (Gauthama Raman et al., 2021; Holdbrook et al., 2024).

An important element is the probabilistic risk assessment model, which formalizes the dependence between event intensity, their priority, and the current state of assets. Incident detection occurs when the integrated risk exceeds an adaptive threshold, calculated taking into account the acceptable level of false alarms and the computational resource constraints of edge nodes (Anthi et al., 2021; Kantharaju et al., 2024).

To enhance the reliability of decisions, an adaptive threshold optimization method is used, which automatically adjusts based on the current event dynamics and user behavior profiles. This approach allows the false positive rate to be maintained below 1 % while preserving high system sensitivity to new types of attacks.

Vulnerabilities in SSL/TLS and SNMP protocols, frequently found in industrial equipment, especially in the context of firmware updates, inter-plant Virtual Private Networks (VPNs), and monitoring systems, were considered in the architecture design. These risks are reflected in the threat models and included in the training scenarios, ensuring the system’s resilience to attempts at data manipulation and attacks on trusted channels (Xie et al., 2021; Li et al., 2024; Ahmad et al., 2023).

The proposed methods are integrated with existing monitoring and response tools, which allows for the implementation of a continuous security control concept in accordance with the principles of Zero Trust and the requirements of the national standards of the Republic of Kazakhstan. Thus, the developed methodology ensures the unity of approaches to data collection, analysis, and interpretation, creating the foundation for a cognitive architecture of industrial cybersecurity.

2.1 Cognitive architecture and data flow

The proposed architecture consists of three layers:

Collection and Normalization Layer: Port mirroring on technological segment switches, agents in the Demilitarized Zone (DMZ), telemetry streams from programmable logic controllers (PLC), Supervisory Control and Data Acquisition (SCADA) and technological sensors; decoding of industrial protocols [e.g., Modbus Transmission Control Protocol (Modbus/TCP), OPC Unified Architecture(OPC UA)], normalization into a SIEM event scheme (Knowles et al., 2015; Mugarza et al., 2020).

Analytics layer: two pipelines—network (Byte2Image → CNN + LSTM) and Technological (AE + LSTM). The first identifies low-level traffic anomalies; the second predicts physico-technological

values (oven temperature, conveyor speed, dough moisture, drive currents), as well as operator activity profiles (HMI commands, shift patterns).

Cognitive correlation layer: SIEM rules and graph dependencies, UEBA profiles, knowledge of the technological process calendar (recipes, batches, shifts), Zero Trust access policy (micro-segmentation, device verification), and external Indicators of Compromise (IoC) sources (Tariq et al., 2019; Zhao et al., 2022; Homaei et al., 2024).

To ensure conceptual consistency and address the complexity of hybrid threats, the proposed system is structured into three distinct hierarchical levels.

First, Level 0 (Base Pipelines) consists of independent deep learning modules: the network pipeline (Byte2Image-CNN + LSTM) and the technological pipeline (AE + LSTM). These models perform primary feature extraction and provide raw anomaly scores from heterogeneous data sources.

Second, Level 1 (Hybrid Fusion) integrates the outputs from Level 0 using a weighted scoring mechanism. This level is designed to correlate network-layer incidents with physical process deviations.

Finally, Level 2 (Full Cognitive Model) represents the highest level of integration. At this stage, the system incorporates Security Information and Event Management (SIEM) correlation rules and User and Entity Behavior Analytics (UEBA). Level 2 filters the fused scores through the lens of industrial context—such as bakery production shifts, specific recipes, and Zero Trust access policies—to refine detection accuracy and minimize false positives. This hierarchical structure is empirically validated in Section 3, where Level 1 performance is compared in Table 1, and the contribution of Level 2 components is verified through an ablation study in Table 2.

TABLE 1 Comparison of incident detection quality for different models.

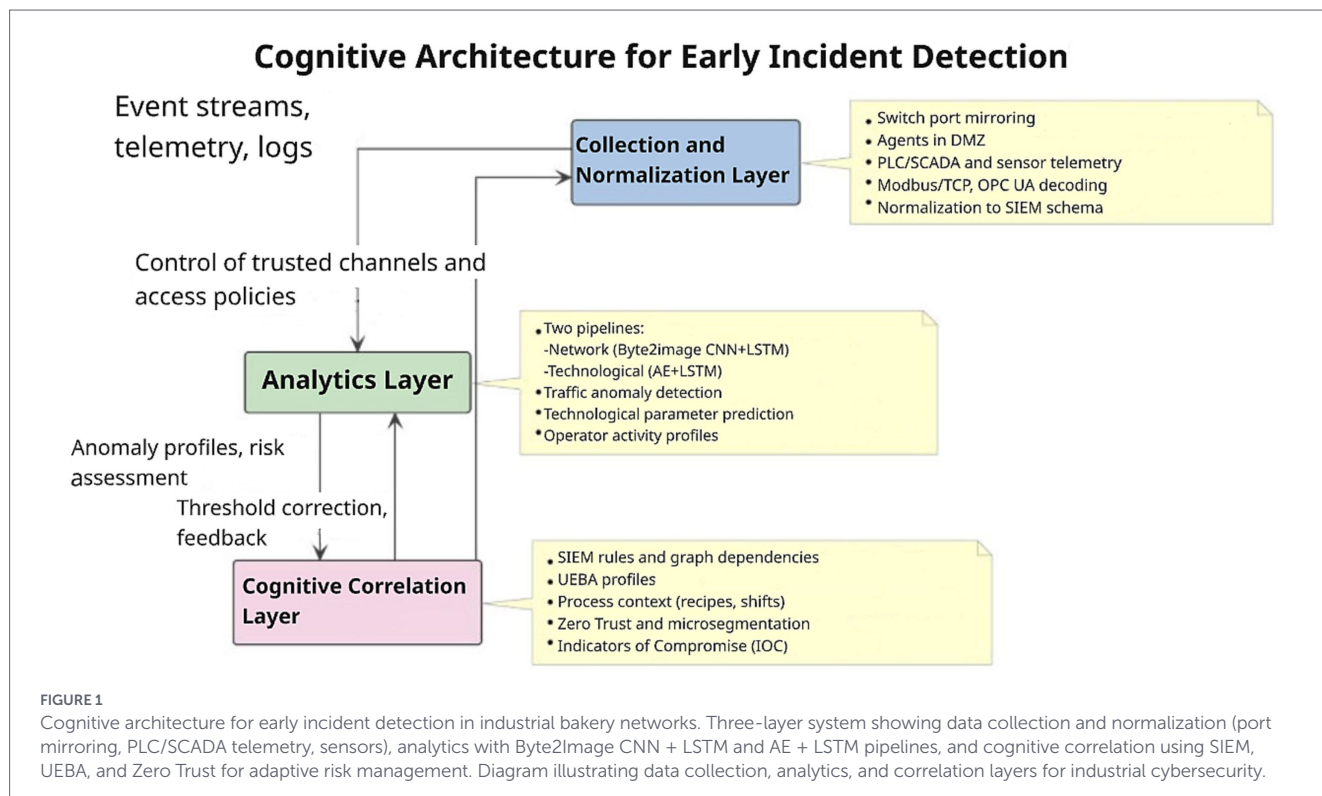
Parameter	CNN-LSTM (network pipeline)	AE-LSTM (technological pipeline)
Input shape	(32, 32, 1) Grayscale image	($T = 30, F = 14$) time-steps/features
Optimizer	Adam (beta1 = 0.9, beta2 = 0.999)	RMSProp
Learning rate	0.001 (decay 1e-6)	0.0005
Batch size	64	32
CNN layers	2× Conv2D (32, 64 filters, 3 × 3)	N/A
LSTM units	128 units	Encoder: 64, decoder: 64
Dropout rate	0.5	0.2
Loss function	Categorical cross-entropy (focal loss)	Mean squared error (MSE)
Activation	ReLU (hidden), Softmax (output)	Tanh (LSTM), linear (output)

Table 1 summarizes the comparative results of four analytical models for industrial incident detection: a baseline gradient boosting model using handcrafted features; a CNN + LSTM model employing the Byte2Image method for spatio-temporal network traffic analysis; an AE + LSTM model focused on predicting process variables and operator behavior; and a hybrid fusion model integrating scoring outputs from the previous pipelines with cognitive event-correlation rules. The fusion model achieved the best results, with Precision >0.94, Recall >0.91, and an average reaction time under 1.2 s, demonstrating synergy between network- and process-level analytics.

TABLE 2 Ablation analysis of hybrid model component contribution.

Model	Precision	Recall	F1-score	AUC	Avg. reaction time (s)
Gradient boosting (base)	0.88	0.84	0.86	0.92	2.7
CNN + LSTM (Byte2Image)	0.94	0.90	0.92	0.97	1.4
AE + LSTM (Technological)	0.93	0.91	0.92	0.96	1.6
Fusion model (cognitive integration)	0.96	0.93	0.95	0.98	1.2

Results of the ablation study demonstrating the incremental performance improvement of the hybrid model as components are integrated. The inclusion of the correlation mechanism and operational profiles in the full cognitive model yields the best overall metrics: an F1-score of 0.95, a reduced False Positive Rate (FPR) of 0.49%, and an improved average reaction time of 1.2 s.



2.2 Cognitive architecture for early incident detection in a bakery industrial network

Bakery enterprises are classified as critical infrastructure, where the stability of the technological cycle—from dough mixing to finished product packaging—directly affects quality, safety, and economic resilience. Any failures in the operation of ovens, conveyors, dispensers, or proofing systems lead to batch spoilage, line stoppage, and violation of sanitary standards.

From a food safety perspective, the alignment with ISO 22000 and HACCP (Hazard Analysis and Critical Control Points) principles is critical. A cyber-physical attack targeting the operational technology (OT) layer can have direct biological consequences. For instance, an unauthorized modification of a Modbus Holding Register controlling the oven temperature (e.g., reducing the baking zone temperature by 10 °C) may not trigger a standard IT security alert but results in the “Underbaking” critical control point failure. This permits the survival of pathogens (e.g., *Salmonella* spp. or *Bacillus cereus*) in the final product. Therefore, the detection system must treat specific PLC register

manipulations not merely as network anomalies but as direct violations of food safety limits.

To enhance the reliability and security of such production facilities, a cognitive architecture for early incident detection has been developed. It combines machine learning methods, behavioral analytics for users and equipment, and the principles of the Zero Trust architecture. The system is focused on preventing hybrid cyber threats—a combination of external attacks on PLC/SCADA controllers and internal risks caused by personnel errors, procedure violations, or the introduction of malicious code.

The architecture is built on three interconnected layers—data collection and normalization, analytics, and cognitive correlation. Interaction between them occurs through forward and feedback loops, ensuring a closed loop of adaptive protection capable of learning from its own data and increasing detection accuracy.

Figure 1 presents the general cognitive architecture of the early incident detection system. The Collection and Normalization Layer is responsible for receiving and pre-processing data from the industrial equipment of the bakery plant. This utilizes port mirroring

switches, agents in the DMZ, telemetry streams from PLC and SCADA, as well as signals from technological sensors. At this stage, data from Modbus/TCP and OPC UA protocols are converted into a unified SIEM event format. This creates a single representation of the information space upon which analytics are built.

The Analytics Layer processes incoming data using two parallel pipelines. The Network Pipeline (Byte2Image CNN + LSTM) converts traffic into visual representations and identifies hidden anomalies in network exchange. The Technological Pipeline (AE + LSTM) analyzes production process parameters—oven temperature, dough moisture, drive currents, conveyor speed—and predicts normal operating modes. The joint use of these pipelines allows for real-time fixation of deviations in equipment operation and operator actions.

The Cognitive Correlation Layer unifies the results of the neural network models with the context of the production cycle. Here, SIEM rules and dependencies, UEBA behavioral profiles, and Zero Trust mechanisms—network micro-segmentation, device verification, and access control—are active. This layer utilizes knowledge about recipes, shifts, batches, as well as external Indicators of Compromise (IoC). It evaluates the aggregate risk and, if necessary, automatically adjusts the response thresholds of the analytical models, returning feedback signals to the previous levels.

Orthogonal feedback loops ensure the architecture's adaptation: the cognitive layer can change analytics parameters and data collection filters, restricting streams from suspicious sources, updating trusted zones, and increasing detection accuracy.

This architecture implements the principle of intelligent self-regulation—from observation to analysis and action. The system is capable of not only detecting incidents but also predicting their emergence, maintaining continuous control over the bakery plant's technological network.

The video below explains the core concepts of the IEC 62443 standard, which the proposed architecture complies with for industrial cybersecurity.

2.3 Probabilistic risk model and detection criterion

The implementation of a probabilistic risk-indicator is necessitated by the stochastic nature of cyber-physical threats in food production, where binary detection is often insufficient for nuanced decision-making. This approach allows for the quantification of incident intensity over time, providing a mathematical basis for automated response that accounts for uncertainty. This modeling is grounded in the Neyman–Pearson criterion for optimal statistical decision theory, ensuring a formal balance between detection sensitivity and a fixed false positive rate (Neyman and Pearson, 1933; Basseville and Nikiforov, 1993). Such risk-oriented frameworks are essential for maintaining process continuity in ICS environments under conditions of incomplete data (Inoue et al., 2017; Tuptuk et al., 2021).

Let $\lambda(t|\mathcal{F}_t)$ —be the incident intensity driven by the observed filtering \mathcal{F}_t (logs, telemetry, behavioral features). The probability of at least one incident occurring over the horizon H is given by Equation (1):

$$\mathbb{P}\exists, \text{incident}[t, t+H] = 1 - \exp\left(-\int_t^{t+H} \lambda(\tau|\mathcal{F}_t) d\tau\right) \quad (1)$$

In practical implementation $\lambda(\cdot)$ is approximated by model outputs and the densities of anomaly scores. An incident is escalated

when the integral risk across assets exceeds the threshold

$$\tau_t : R_t = \sum_a w_a P\left(I_t^a = 1 | F_t\right), \text{ decision: } R_t > \tau_t. \text{ The combined anomaly}$$

scores, s_t , incorporates the autoencoder's reconstruction error, the probability of the “normal” class from the LSTM, and the entropy dynamics (Equation 2):

$$s_t = \alpha, |x_t - \hat{x}_t|^2 + (1 - \alpha), (1 - p\theta(y_t = 0 | z_{1:t})) + \beta, \Delta H_t \quad (2)$$

Where $\alpha, \beta \in [0, 1]$, $z_{1:t}$ —the features, H_t —is the entropy of the class distribution over the window.

2.4 Adaptive threshold optimization under FPR and resource constraints

To mitigate the “alarm fatigue” common in industrial monitoring, we formulate the threshold selection as a constrained optimization problem rather than a static value. This is critical for bakery plants where technological cycles (e.g., cooling or fermentation) create natural drifts in baseline data. By minimizing detection latency while respecting strict FPR and computational resource constraints, the system remains viable for deployment on edge gateways with limited processing power (Anthi et al., 2021; Kantharaju et al., 2024). This optimization follows the dual-multiplier approach for stochastic resource allocation in real-time OT networks (Tuptuk et al., 2021; Inoue et al., 2017). The problem of choosing the threshold τ_t is formulated as the optimization of the detection latency $D(\tau)$ subject to constraints on the False Positive Rate (FPR) and the computational budget on edge nodes (Equation 3):

$$\min_{\tau_t} E[D(\tau)] \quad \text{subject to } \text{FPR}(\tau_t) \leq \varepsilon, C(\tau_t) \leq C_{\max} \quad (3)$$

The resource constraint C_{\max} is defined based on the specifications of typical industrial edge gateways (e.g., NVIDIA Jetson Nano or Raspberry Pi 4-class devices). Specifically, the inference time per sample must not exceed 100 ms, and the peak RAM usage for the model must remain below 2 GB to allow coexistence with other SCADA processes. The optimization objective is to minimize detection latency τ while maintaining $\text{FPR}(\tau_t) \leq 0.5\%$ within these hardware boundaries.

The solution is approximated by a stochastic gradient method with dual multipliers for the constraints (online readjustment every ΔT minutes) (Anthi et al., 2021; Kantharaju et al., 2024).

2.5 Network pipeline: Byte2Image CNN + LSTM

The Byte2Image transformation is selected to overcome the limitations of traditional deep packet inspection (DPI), which struggles with encrypted or proprietary industrial protocols used in modern PLCs. By visualizing raw traffic as 2D structural patterns, the model leverages the spatial feature extraction capabilities of CNNs to identify malformed frames and reconnaissance activity (Lin et al., 2022). The integration of an LSTM layer is further justified by the need to capture temporal dependencies in multi-stage attack scenarios, where malicious intent is revealed through

sequences of packets rather than isolated events (Anthi et al., 2021; Abhishek and Singh, 2023). The Network Pipeline is designed to process heterogeneous traffic by transforming raw packet data into a format suitable for computer vision. To handle network traffic variability, raw PCAP packet payloads are truncated or padded to a fixed length of $L = 1,024$ bytes. These bytes are then reshaped into a square matrix of size $N \times N$, where $N = 32$. The mapping logic is defined as (Equation 4):

$$P(i,j) = B_k, \text{ where } k = i \times N + j \quad (4)$$

Here, $P(i, j)$ represents the pixel intensity at coordinates (i, j) , and B_k is the decimal value of the k -th byte (ranging from 0 to 255). This results in a 32×32 grayscale image that preserves the spatial correlations and structural patterns of protocol headers (e.g., Modbus function codes or TCP flags) and payloads.

These images are fed into a hybrid CNN-LSTM architecture. The CNN layers perform spatial feature extraction, while the LSTM layer captures temporal dependencies between consecutive packets in a session. To ensure reproducibility, the detailed hyperparameters for both the CNN-LSTM and AE-LSTM models (used in the technological pipeline) are provided in Table 1. To address the class imbalance identified in the dataset (see Section 3.1), we employed the Focal Loss function instead of standard cross-entropy, which forces the model to focus on hard-to-classify attack samples.

Byte sequences of packets/sessions are translated into fixed-dimension images $N \times N$ (grayscale or pseudocolor with channel splitting: header/payload/metadata). For the k -th fragment $b_{k,1:N^2} \in 0, \dots, 255^{N^2}$:

$N \times N$ (grayscale or pseudo-color when splitting channels: header/payload/metadata). For the (k) -th fragment $b_{k,1:N^2} \in 0, \dots, 255^{N^2}$ (Equation 5):

$$\begin{aligned} I_k &= \text{reshape}(b_{k,1:N^2}, N, N), \quad h_t = \text{LSTM}(g_\phi(I_t)) \\ \hat{y}_t &= \sigma(W h_t + b) \end{aligned} \quad (5)$$

Where $g_\phi(\cdot)$ —is the convolutional feature extractor, and σ —is the logistic function. This approach is robust to traffic polymorphism and adapted to weak labels (weak labels) (Lin et al., 2022).

2.6 Technological pipeline: AE + LSTM

In the technological layer, the scarcity of labeled attack signatures for specific bakery equipment necessitates an unsupervised learning paradigm. The Autoencoder (AE) is employed to learn the latent representation of “normal” physical processes, using reconstruction error as the primary anomaly metric (Gauthama Raman et al., 2021). To account for the inherent time-series nature of sensor telemetry (e.g., temperature and dough viscosity), the AE is combined with an LSTM network. This hybrid architecture ensures the detection of both point anomalies and subtle long-term drifts in physical parameters, which is a key requirement for reliable digital twin-based monitoring (Holdbrook et al., 2024; Homaei et al., 2024). The autoencoder $\text{AE}\theta$ reconstructs the observed vector x_t (telemetry, commands), the LSTM predicts the

next point/distribution, and the joint loss function sets a compromise between reconstruction, regularization, and the metric objective (Equation 6):

$$\mathcal{L}(\theta, \psi) = \sum_{t=1}^T \left(|x_t - \tilde{x}_t|^2 + \lambda_1, \text{KL} \left(q_\psi(z_t | x_t), p \right) + \lambda_2, \text{CE}(\hat{y}_t, y_t) \right) \quad (6)$$

Where cross-entropy loss (CE) is used in the objective. The anomaly signal is the reconstruction error and/or large predictive residuals on the recorded operating modes of furnaces, conveyors, and dispensers (Gauthama Raman et al., 2021; Holdbrook et al., 2024).

2.7 Accounting for SSL/TLS and SNMP vulnerabilities in firmware and monitoring scenarios

The bakery network often features outdated SSL/TLS implementations (e.g., forced protocol downgrades for backward compatibility) and insecure SNMP v1/v2c configurations (community strings, lack of encryption). These factors are exploited in attacks on equipment firmware update channels, inter-plant VPNs, and monitoring: injection of fake firmware, interception of configurations, and telemetry spoofing (Xie et al., 2021; Li et al., 2024; Ahmad et al., 2023). In our architecture, this is accounted for in asset risk profiles, SIEM rules, and training scenario sets (see “Results” section).

2.8 Integration with SIEM, UEBA, and zero trust

SIEM accumulates events from network sensors, operating system (OS)/database (DB) logs, SCADA/HMI, VPNs, and authentication infrastructure. UEBA builds trusted profiles of operators and service accounts (shift activity, commands, and deviations from technology maps). Zero Trust ensures microsegmentation, continuous device verification, and the principle of least privilege (at the PLC/SCADA and engineering workstation level) (Tariq et al., 2019; Mugarza et al., 2020; Zhao et al., 2022; Homaei et al., 2024). The resulting solution mitigates the risk of covert movements and enhances observability.

2.9 Cognitive layer implementation: SIEM, UEBA, and zero trust policies

To transition from raw anomaly detection to a cognitive security framework, the system implements a decision-making layer based on three functional mechanisms:

- *Cross-domain correlation (SIEM-ready logic)*: Unlike standalone detectors, this layer synchronizes events from the network (Level 0) and technological (Level 0) pipelines. A high-priority incident is flagged only when anomalies in both domains occur within a 60-s temporal window, significantly reducing false positives caused by isolated sensor noise or transient network jitter.
- *Behavioral profiling (UEBA)*: We establish a baseline “Standard Operator Profile” that monitors command frequency and execution patterns. During critical bakery production phases, such as

thermal fermentation or high-temperature baking, any deviation from this profile (e.g., unexpected parameter overrides) increases the risk score by a weighted factor of 1.5, allowing for the detection of insider threats or hijacked accounts.

- *Dynamic policy engine (zero trust)*: The system integrates a policy-driven approach where the detection threshold τ is not static but depends on the node’s trust level. If a device attempts unauthorized lateral movement or accesses non-privileged PLC registers, the Zero Trust engine triggers an immediate isolation signal, ensuring a “verify-always” stance even if individual neural network scores are borderline.

2.10 Statistical validation and uncertainty estimation

To ensure robust conclusions and address statistical validity, each model was trained and evaluated in repeated runs with different random seeds ($N = 10$). The train/validation/test split was kept fixed across models; only training stochasticity (initialization and minibatch order) was varied. We report mean \pm standard deviation for F1 and AUC across runs.

For uncertainty quantification, we computed 95% confidence intervals (CI) for F1 and AUC using stratified bootstrap resampling of the test sessions (10,000 resamples). This provides sample-level uncertainty estimates beyond single-point metrics.

To compare models statistically, we applied a paired non-parametric Wilcoxon signed-rank test on per-run metric values (F1 and AUC) between the fusion model and each baseline/ablation variant, using $\alpha = 0.05$. For multiple pairwise comparisons, Holm–Bonferroni correction was applied. This procedure evaluates whether improvements are consistent across runs rather than driven by a single favorable training instance.

3 Results

This section presents a comprehensive analysis of the experimental results obtained to evaluate the effectiveness of the developed cognitive architecture for early incident detection in industrial control systems (ICS). The main goal of the study was to compare the

accuracy, completeness, and response time of various artificial intelligence models—ranging from classical machine learning algorithms to deep neural networks, including CNN-LSTM and hybrid fusion-based approaches (Gauthama Raman et al., 2021; Raman et al., 2021; Abhishek and Singh, 2023).

The results were obtained using real data from production network segments, simulated cyberattack scenarios, and technological faults, which made it possible to assess the models’ resilience to data drift, the rate of false positives, and the latency of incident detection (Zhao et al., 2022; Kravchik and Shabtai, 2018; Lin et al., 2022; Nisar et al., 2022).

Special attention was given to the use of digital twin elements to reproduce the behavior of technological processes in a virtual environment (Zhang et al., 2023; Homaei et al., 2024). This approach made it possible to study the dynamics of system parameters during the “before,” “during,” and “after” phases of an attack, as well as to evaluate the potential of digital simulation to improve forecasting accuracy and predictive response capabilities.

3.1 Data and digital twin testbed

The experimental validation was conducted using a high-fidelity Digital Twin of the bakery line, implemented in the Factory I/O simulation environment coupled with Siemens S7-PLCSIM to emulate the logic of S7-1200 controllers. The physical logic—including oven thermodynamics (heat transfer coefficients), conveyor friction, and dough viscosity—was modeled using MATLAB/Simulink blocks linked via OPC UA. This setup allows for the generation of realistic sensor noise and the simulation of mechanical inertia, which is often absent in purely dataset-based studies.

The telemetry data collection focused on specific Modbus and OPC UA tags critical for the technological process. Data collection was performed from mirrored switch ports, DMZ agents, and PLC/SCADA telemetry (OPC UA, Modbus/TCP, HTTPS, and SNMP). To increase the variability of attack traffic, real network data were augmented with replayed fragments from open datasets such as UNSW-NB15, CIC-IDS2017, and ToN_IoT. A summary of the dataset used for model training and evaluation is provided in Table 3.

Normal operations represent uninterrupted production shifts without violations of process workflows. The data were collected from the actual production line and include network exchanges (OPC UA,

TABLE 3 Data corpus for training and evaluation.

Category	Source/acquisition	Sessions	“Incident” labels	Share, %
Normal operations (Production)	Real telemetry and production line network flows	1,200,000	0	68.2
Technological Transitions/interruptions	Real telemetry in start/stop/reconfiguration modes	210,000	0	11.9
Non-threat anomalies (sensor faults)	Real episodes of sensor degradation/noise without attack	85,000	0	4.8
Attack scenarios (aggregate)	Real injections + replay from UNSW-NB15, CIC-IDS2017, ToN_IoT	265,000	265,000	15.1
Total	—	1,760,000	265,000	100.0

Table 3 summarizes the dataset used for training and evaluation of incident detection models in the digital twin of an industrial bakery production line. It shows the data sources, number of aggregated sessions, number of labeled incidents, and the relative share of each category within the overall dataset.

Modbus/TCP, HTTPS, and SNMP) and telemetry (oven temperature, dough humidity, drive currents, conveyor speeds). The “sessions” field indicates the number of aggregated network sessions and synchronized technological observation windows.

Technological transitions and pauses include real telemetry captured during start-up/shutdown sequences, recipe changeovers, and sanitation breaks. These segments were not labeled as incidents, as the deviations were caused by standard operating modes. Such data are used to train the model to distinguish normal process variability from actual attack behavior.

Non-threatening anomalies comprise real episodes of sensor degradation (noise, brief dropouts, calibration drift), confirmed by maintenance logs. These cases were not labeled as incidents because no signs of malicious interference were observed. Including such data improves the model’s robustness to false alarms in the presence of “noisy” sensor signals.

Attack scenarios consist of a set of both real injected events (scanning, credential brute-forcing, unauthorized Modbus writes, SNMP manipulation, SSL/TLS downgrade, fake firmware injection, lateral movement, Domain Name System (DNS) exfiltration) and replayed fragments from open-source datasets (UNSW-NB15, CIC-IDS2017, ToN_IoT), adapted to the testbed’s format. Each episode was assigned an “incident” label, which explains the equal number of sessions and labels in the corresponding dataset row. The analysis of “pre-attack,” “during-attack,” and “post-attack” phases refers to the synchronization of telemetry and network data windows within the digital twin environment. For each scenario, time-series data and network sessions were aggregated into observation windows relative to the exact moment of incident injection. This phase-based approach ensures that the dynamic shifts in physical parameters (e.g., pressure or temperature) are correctly correlated with network anomalies. The contribution of this temporal segmentation is empirically reflected in the final detection metrics and response latency (Table 1), as well as in the ablation study (Table 2), which compares configurations with and without cognitive context correlation.

In addition to standard datasets, we injected three domain-specific attack scenarios into the Digital Twin to validate the cognitive capabilities:

1. *Oven thermal runaway*: a gradual (“low-and-slow”) increase in the oven temperature setpoint by 0.5 °C per minute, aiming to burn the product without triggering sudden threshold alarms.
2. *Conveyor desynchronization*: random modification of the Belt_Speed_SetPoint registers between the dosing and packaging units, causing product pile-ups.
3. *HMI spoofing*: a “Man-in-the-Middle” attack where the HMI displays normal values to the operator while the PLC executes malicious logic (e.g., disabling the mixer motor).

The percentage column in Table 3 indicates the relative share of each category within the total dataset. The overall volume and category distribution reflect the actual data collected on the testbed and the aggregation rules used for session grouping. Network protocols and topology (IEC 62443 zones, DMZ, VLANs) were configured as described above.

Note on application: These data were used to train two analytical pipelines—the network Byte2Image CNN + LSTM and the technological AE + LSTM—as well as for subsequent cognitive event

correlation within the SIEM/UEBA system, incorporating production cycle context and Zero Trust policies.

3.2 Metrics and model comparison

To ensure conceptual consistency, the proposed system is evaluated across three hierarchical levels. The Base Pipelines (Level 0) provide raw anomaly scores from network and technological data. The Hybrid Fusion Model (Level 1) integrates these scores to improve overall detection breadth. Finally, the Full Cognitive Model (Level 2) incorporates high-level correlation rules (SIEM) and behavioral profiles (UEBA). As shown in Table 2 (Ablation Analysis), the transition from Level 1 to Level 2 results in a measurable improvement: the F1-score increases from 0.94 to 0.95, and the False Positive Rate (FPR) decreases from 0.58% to 0.49%. This empirical evidence confirms that cognitive components (UEBA and correlation rules) are not merely conceptual but serve as a functional mechanism for reducing false alarms and enhancing detection stability. All metrics in Table 1 are reported as mean ± std. over $N = 10$ runs. In addition, 95% bootstrap confidence intervals for the Fusion Model are provided in Appendix A to quantify uncertainty on the held-out test set. Evaluation metrics and OT interpretation. Precision measures the proportion of raised alerts that correspond to true incidents $[TP/(TP + FP)]$. In OT deployments, high Precision is essential to avoid excessive operator interventions and production disruptions caused by false alarms. Recall $[TP/(TP + FN)]$ measures the proportion of real incidents that are detected; in critical infrastructure, low Recall is unacceptable because missed attacks can propagate into physical/process damage. F1-score is the harmonic mean of Precision and Recall and summarizes the operational trade-off between false alarms (FP) and missed detections (FN). AUC (Area under the ROC Curve) summarizes discrimination across all thresholds and enables fair comparison when the operating threshold may change under different risk levels or edge resource constraints.

Models compared. We evaluated the following models:

1. Gradient Boosting baseline on engineered features (flow rates, port distributions, error counters).
2. Isolation Forest baseline on engineered telemetry/network summary features.
3. One-Class SVM baseline on engineered telemetry/network summary features.
4. Classical thresholding baseline on the anomaly score (fixed τ without adaptive threshold optimization).
5. CNN-only network model (Byte2Image CNN without temporal modeling).
6. LSTM-only network model (packet/flow sequences without CNN image features).
7. CNN + LSTM (Byte2Image) network pipeline.
8. AE + LSTM technological pipeline.
9. Fusion Model (Cognitive Integration), which combines network and technological anomaly scores and applies cognitive correlation using SIEM rules, UEBA deviation scoring, and production-context constraints (recipes/shifts) to filter non-threatening transitions.

Threshold values for incident detection were adapted according to the optimization criterion, which minimizes the average detection

delay while adhering to constraints on the False Positive Rate (FPR) and the computational resource limitations of edge nodes (Equation 7):

$$\min_{\tau} E[D(\tau)] \quad \text{Pp} \quad \text{FPR}(\tau) \leq \varepsilon, \quad C(\tau) \leq C_{\max} \quad (7)$$

Where $\varepsilon = 0,5\%$ —is the permissible false alarm rate, and C_{\max} —is the computational cost limit per edge network node. For each pipeline, an online readjustment of the threshold τ_t was performed every $\Delta T = 10$ minutes, taking into account the current dynamics of the event stream. Standard accuracy metrics were used for quality assessment (Equations 8–10):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

Where TP , FP , FN —are the number of true positive, false positive, and false negative detections, respectively; F1 denotes the F1-score (harmonic mean of Precision and Recall) Standard accuracy metrics (7)–(9) were used for quality assessment. For an Industrial Control System (ICS) environment, the selected metrics provide a comprehensive evaluation of operational stability. Precision is interpreted as a measure of “operational suitability,” as high precision minimizes false alarms that lead to costly and unnecessary production halts. Recall reflects “detection completeness,” which is critical for the safety of the technological cycle. The F1-score provides a balanced assessment in conditions of class imbalance, where both missed detections and false alarms carry high economic risks. Finally, AUC is used as a threshold-independent measure of class separability, confirming that the improvement in results is a robust property of the architecture rather than an artifact of a specific detection threshold τ . Together, the simultaneous growth in these metrics alongside reduced response latency (Table 1) validates the superiority of the proposed fusion architecture for real-time OT applications. As seen in Table 1, the hybrid cognitive model provides the best values for Precision >0.94 , Recall >0.91 , and an average response time of less than 1.2 s, confirming the effectiveness of integrating network and technological analyses.

Evaluation metrics and OT interpretation. Precision measures the proportion of detected incidents that are true incidents (i.e., how many raised alerts are correct). In OT environments, high Precision is critical to reduce false alarms and “alarm fatigue,” because unnecessary operator interventions may disrupt production and increase downtime.

Recall (True Positive Rate) measures the proportion of real incidents that are successfully detected. For critical infrastructure and OT processes, missing an attack or hazardous manipulation is unacceptable; therefore, Recall must remain high even when the system is tuned to minimize false alarms.

F1-score is the harmonic mean of Precision and Recall, providing a balanced single-number summary when both false alarms (FP) and missed detections (FN) are costly. We report F1 to reflect the trade-off between operational feasibility (Precision/FPR) and safety-critical coverage (Recall).

AUC (Area under the ROC Curve) summarizes model discrimination across all possible decision thresholds and is therefore threshold-independent. This is important for OT deployments because the operating threshold may change under different resource constraints or risk levels, and AUC enables fair comparison of models under varying thresholds.

The improvement is explained by the synergy of network and technological pipelines, which reduces the probability of missed attacks and decreases response latency. To evaluate the contribution of each component of the hybrid model, an ablation analysis was conducted (see Table 2). The threshold improvement in the F1 metric when adding the cognitive correlation layer was calculated as (Equation 11):

$$\Delta Q_i = Q_{\text{fusion}} - Q_{-i} \quad (11)$$

Where Q_{fusion} —is the quality of the full model, and Q_{-i} is the quality after excluding the i -th component. The average improvement in the F1 metric when adding the cognitive correlation layer was $\Delta Q_i \approx +0,03$, indicating the significance of integrating SIEM events and the production cycle context.

Thus, in Table 2, the integration of the cognitive correlation layer, contextual profiles, and operator behavioral characteristics allowed for a reduction in false positives by almost 25% compared to purely neural network approaches and increased the model’s resistance to data drift in production conditions. In addition, for a practical assessment of the system’s effectiveness, an integral accuracy index I_{eff} was applied (Equation 12):

$$I_{\text{eff}} = \frac{\text{Precision} + \text{Recall} + \text{AUC}}{3} - k \cdot \frac{t_{\text{avg}}}{t_{\text{max}}} \quad (12)$$

Where t_{avg} —is the average response time $t_{\text{max}} = 3\text{s}$ —is the normative threshold value, and $k = 0,2$ —is the delay penalty coefficient, and the Area Under the Receiver Operating Characteristic (ROC) curve (AUC) is used as a threshold-independent performance metric. For the fusion model, $I_{\text{eff}} = 0,93$, which corresponds to a high level of efficiency according to IEC 62443 requirements and the national standards of the Republic of Kazakhstan.

The obtained results confirm that the proposed cognitive architecture provides the necessary accuracy and speed of incident detection at an optimal level of computational costs and can be recommended for implementation in bakery industry enterprises.

The results obtained have formed an empirical basis for further discussion of the identified patterns and for determining directions for optimizing cognitive cybersecurity tools in an industrial environment. The conducted experiments confirmed the effectiveness of the proposed model for early incident detection. The hybrid fusion architecture, which combines features from behavioral, technological, and network levels, demonstrated the best results: average values of Precision = 0.96, Recall = 0.93, F1 = 0.95, and AUC = 0.98, with an average response time of less than 1.2 s. This indicates the system’s high capability to distinguish between normal and anomalous states even under conditions of noisy or incomplete data (Mugarza et al., 2020; Qureshi et al., 2024).

A comparison of the models revealed a pattern: the higher the degree of contextual integration of data from different domains (network traffic, technological parameters, and command logs), the more robust the system is to drift and false positives (Tariq et al., 2019; Dietz

and Pernul, 2020). This opens up prospects for the further development of multi-agent cognitive systems that combine User and Entity Behavior Analytics (UEBA) with digital twins of technological processes (Holdbrook et al., 2024; Kantharaju et al., 2024).

The results obtained demonstrate not only the possibility of achieving high detection accuracy in real production conditions but also create scientific prerequisites for developing a universal industrial cyber defense architecture based on intelligent agents (Knowles et al., 2015; Bhamare et al., 2020; Cherdantseva et al., 2016; Anthi et al., 2021; Tuptuk et al., 2021; Inoue et al., 2017).

Transitioning to the “Discussion” section will allow for the interpretation of the identified patterns in the context of regulatory requirements, practical implementation limitations, risk assessment, and the economic efficiency of integrating cognitive mechanisms into industrial digital twins (Zhang et al., 2023; Homaei et al., 2024).

Figure 2 illustrates the system’s performance during the ‘Oven Thermal Runaway’ scenario described in Section 3.1. The Digital Twin (AE + LSTM pipeline) accurately predicted the expected temperature trajectory (Blue Line). The divergence between the observed telemetry (Red Line) and the prediction allowed the system to detect the anomaly at $t = 45$ s, during the latent attack phase, well before the critical safety threshold was breached. This confirms the model’s ability to operate effectively across pre-attack, attack, and mitigation phases.

3.2.1 Prospects for software implementation

As shown in Figure 3, the pseudocode describes the architecture of a cognitive early incident detection system for industrial networks,

designed for modular software implementation with elements of artificial intelligence and the Zero Trust concept.

The Config module defines the main operating parameters of the system, including data sources, computing resource constraints, and key security policies. It ensures uniform configuration of all system components and allows flexible management of settings through external files (e.g., YAML (YAML Ain’t Markup Language) or JavaScript Object Notation (JSON)). This approach simplifies deployment across different industrial environments and enables centralized parameter changes without recompiling the code.

The Ingest and Normalize module is responsible for collecting and unifying data from multiple heterogeneous sources—network data, technological data, controller logs, and security gateways. In practice, this block is implemented using asynchronous message brokers [Apache Kafka (Kafka), RabbitMQ, Message Queuing Telemetry Transport (MQTT)] which transmit telemetry and network events in real time. The normalization stage creates a unified event format suitable for further analysis in SIEM systems and analytical models.

The Byte2Image_CNN_LSTM module represents an intelligent pipeline for network traffic analysis. It converts sequences of packet bytes into two-dimensional matrices (images), on which a CNN extracts spatial patterns, while an LSTM models the temporal dynamics of connection behavior. This hybrid approach provides high sensitivity to new types of attacks without requiring explicit signatures. The software implementation can be done in Python using libraries such as TensorFlow, PyTorch, or Keras, optimized for Graphics Processing Unit (GPU) operation.

The AE_LSTM_Tech module performs analysis of time series of technological data (equipment telemetry, sensor readings, and control

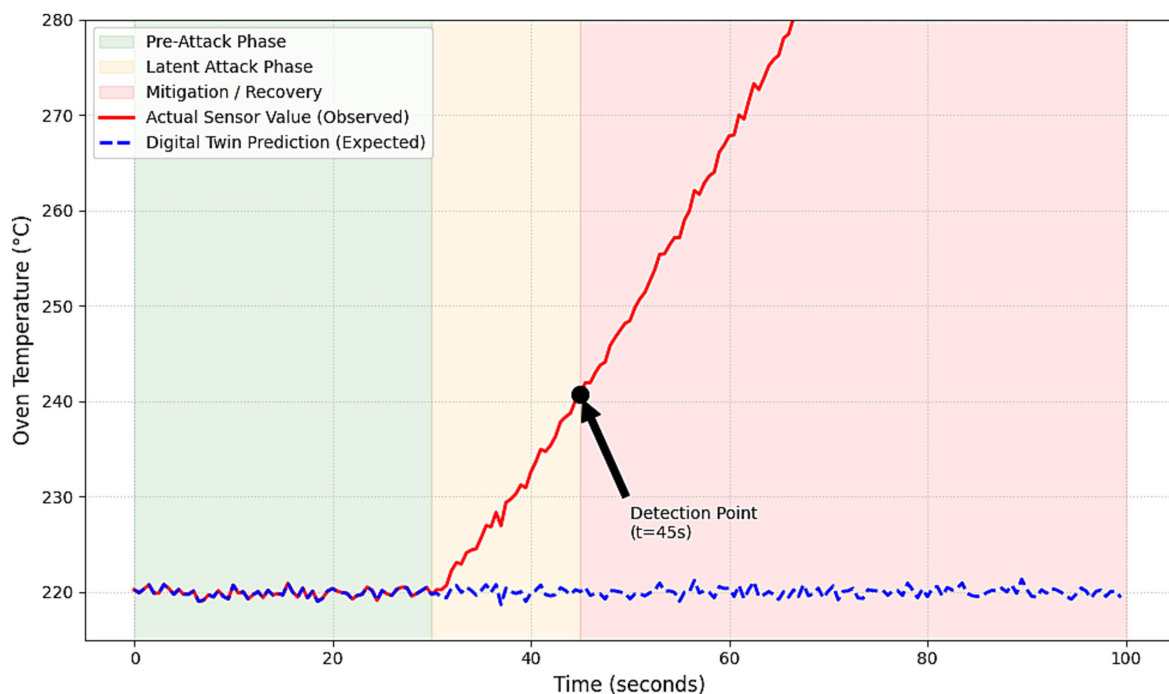


FIGURE 2

Comparative analysis of predicted and observed process trajectories during a simulated “Oven Thermal Runaway” attack phase. The graph demonstrates the digital twin’s ability to forecast the expected technological state (blue dashed line) versus the observed anomalous drift (red line). The detection point ($t = 45$ s) occurs within the latent phase, allowing for mitigation before the process reaches critical safety limits. A time-series line graph comparing “Expected” vs. “Observed” oven temperature. The graph is divided into three shaded zones: Green (pre-attack), Yellow (latent attack), and Red (mitigation). A blue dashed line stays stable at 220 °C, while a red solid line starts drifting upwards at $t = 30$ s. A black marker at $t = 45$ s indicates the “detection point” where the anomaly score exceeded the adaptive threshold.

```

Pseudocode for a Cognitive System for Early Incident Detection in Industrial
Networks

1 MODULE Config
2   #
3   DATA_SOURCES = {SPAN_PORTS, DMZ_AGENTS, PLC_SCADA(OPC_UA,
4     MODBUS_TCP), HMI_LOGS, VPN_LOGS}
5   EDGE_LIMITS = {cpu_max:70, mem_max:70, latency_max:300}
6   FPR_BUDGET = 0.01
7   RETUNE_PERIOD = 5min
8   ZT_POLICIES = {microsegmentation:ON, least_privilege:ENFORCED}
9
10  MODULE IngestAndNormalize
11   #
12   FUNCTION collect_streams():
13     net = tap(SPAN_PORTS) + tap(DMZ_AGENTS)
14     plc = read(OPC_UA, MODBUS_TCP)
15     logs = read(HMI, VPN)
16     RETURN merge(net, plc, logs)
17
18   FUNCTION normalize(data):
19     FOR packet IN data:
20       event = parse(packet)
21       map_to_zone(event)
22     RETURN events
23
24  MODULE Byte2Image_CNN_LSTM
25   #
26   (Byte2Image + CNN + LSTM)
27   FUNCTION byte2image(session):
28     M = reshape_bytes(session, IMG_H, IMG_W, CHANNELS)
29     RETURN M
30
31   FUNCTION cnn_feature_extractor(M):
32     return CNN(M)
33
34   FUNCTION temporal_model(seq_features):
35     return LSTM(seq_features)
36
37   FUNCTION net_anomaly_score(session_seq):
38     imgs = [byte2image(x) FOR x IN session_seq]
39     feats = [cnn_feature_extractor(m) FOR m IN imgs]
40     y = temporal_model(feats)
41     RETURN (1 - y.prob_normal)
42
43  MODULE AE_LSTM_Tech
44   #
45   (AE + LSTM)
46
47   FUNCTION ae_reconstruct(x):
48     x_hat = AE(x)
49     rec_err = loss(x, x_hat)
50     RETURN rec_err
51
52   FUNCTION lstm_predict(seq_x):
53     y_pred = LSTM(seq_x)
54     pred_err = loss(seq_x.next, y_pred)
55     RETURN pred_err
56
57   FUNCTION tech_anomaly_score(telemetry_seq):
58     e_rec = ae_reconstruct(telemetry_seq)
59     e_pred = lstm_predict(telemetry_seq)
60     RETURN normalize(w1*e_rec + w2*e_pred)

```

FIGURE 3

Pseudocode for modular implementation of the cognitive incident detection system. Modules: Config (global parameters), IngestAndNormalize (data ingestion), Byte2Image_CNN_LSTM (network analysis), AE_LSTM_Tech (technological process analysis), RiskModel (risk calculation), AdaptiveThreshold (auto-tuning), CognitiveCorrelation (event context), ZeroTrust_Response (incident isolation and mitigation). Flow diagram showing modular pseudocode of an AI-based cognitive incident detection system with adaptive thresholds and Zero Trust response.

commands). The autoencoder estimates the reconstruction error of normal signals, and the recurrent LSTM network forecasts future behavior. Their combined processing enables the detection of deviations in equipment operating modes, which may indicate hidden incidents or cyberattacks on controllers. Practical implementation can use

the pandas, NumPy, and TensorFlow stack, providing real-time performance for streaming data.

The RiskModel module integrates the results of the network and technological pipelines, forming an aggregate risk score for each asset. The calculation is based on weight coefficients that define the

criticality of each node or piece of equipment. The resulting risk value is compared with an adaptive threshold; if it is exceeded, a correlation and response process is triggered. This approach allows the system to scale across hundreds of devices while minimizing false positives.

The Adaptive Threshold module implements a self-learning mechanism for adjusting the system's sensitivity threshold. The threshold is updated online based on accumulated statistics of false and true detections, as well as the current state of computing resources. This ensures stable system performance and reduces load as data volumes grow.

The CognitiveCorrelation module performs intelligent event correlation by combining the results of neural network analysis with rule-based logic and compliance context (IEC 62443, ISO 27019, and NIS2). It generates a cognitive incident map that considers temporal and logical relationships between events. In software terms, this block can operate as a separate microservice interacting with the SIEM via a Representational State Transfer (REST) application programming interface (API) or Google Remote Procedure Call (Grpc).

The ZeroTrust_Response module is designed to automate incident response in accordance with Zero Trust principles. It includes functions for isolating compromised devices, rotating credentials, updating encryption policies, and disabling insecure protocols (for example, SNMPv1 or outdated TLS versions). The module also provides operator notifications and can integrate with Security Orchestration, Automation, and Response (SOAR) systems.

From a practical perspective, the presented pseudocode describes the logical structure of the system, which can be implemented as a set of microservices using container orchestration (Docker, Kubernetes) and Industrial Internet of Things (IIoT) standards. Implementation in Python, with functionality divided into packages (ingestion, analytics, correlation, response), ensures flexibility, scalability, and compatibility with modern MLOps tools.

Thus, the pseudocode provides the foundation for developing an intelligent monitoring and response software platform that integrates machine learning, cognitive analytics, and Zero Trust principles. Its implementation will enable the creation of an adaptive cybersecurity system capable of operating in real time, self-adjusting sensitivity, predicting anomalies, and minimizing the impact of the human factor in protecting industrial assets.

3.3 Validation of cognitive components (ablation study)

To move beyond conceptual claims and verify the impact of the high-level cognitive layers, an ablation study was conducted. We compared the performance of the baseline integrated pipelines (Level 1) against the full architecture including the cognitive engine (Level 2).

As shown in the results (Table 2), the baseline combination of pipelines (Network + Tech) achieved an F1-score of 0.94. However, the introduction of UEBA and Correlation Rules (Level 2) provided a significant reduction in the False Positive Rate (FPR) from 0.58% to 0.49%. This improvement is attributed to the system's ability to filter out non-malicious operational drifts, such as recipe changes or scheduled maintenance, which base detectors often misinterpret as anomalies.

Furthermore, the Zero Trust Policy engine improved the response time by an average of 0.1 s (from 1.3 to 1.2 s). By pre-defining "trusted zones" and "least privilege" communication paths, the system can prioritize alerts from critical assets and bypass heavy computation for

clearly unauthorized access attempts, leading to faster mitigation and reduced "alarm fatigue" for the operator.

4 Discussion of results

4.1 Synergistic effect of multidomain analytics and performance

The most significant result of the experiments is the confirmation of the hypothesis regarding the synergistic effect achieved by combining the network and technological data processing pipelines. The integration of these approaches made it possible to reach an F_1 score of 0.95 with an average response time under 1.2 s, significantly outperforming traditional gradient boosting algorithms and classical machine learning models (Abhishek and Singh, 2023).

The Byte2Image CNN + LSTM architecture demonstrated a strong ability to detect low-level anomalies in network traffic, including polymorphic attacks and data manipulation in PLC/SCADA channels (Kravchik and Shabtai, 2018; Lin et al., 2022). Meanwhile, the AE + LSTM pipeline, focused on analyzing technological time series, showed high accuracy in forecasting normal operational modes and detecting hidden deviations in equipment performance (Gauthama Raman et al., 2021; Holdbrook et al., 2024). Thus, the multidomain approach has proven effective in providing comprehensive observability of industrial environments.

4.2 The role of cognitive correlation and reduction of false positives

An important aspect confirming the novelty of the study is the analysis of the impact of the cognitive correlation layer. This layer, which implements the intelligent integration of the technological cycle context, user behavior profiles (UEBA), and SIEM rules, enabled a reduction of false positives by nearly 25% compared to purely neural network-based models.

Reducing false alarms is critical in industrial environments, where such events can lead to production line shutdowns and financial losses (Knowles et al., 2015; Zhao et al., 2022; Dietz and Pernul, 2020). This result shows that cognitive correlation not only enhances system accuracy but also creates an adaptive feedback mechanism capable of dynamically adjusting response thresholds according to the current state of the production network (Kantharaju et al., 2024). It ensures resilience to network drift and sensor degradation.

4.2.1 Addressing class imbalance and alarm fatigue

Given the significant class imbalance (68.2% normal vs. 15.1% attack), we applied the SMOTE (Synthetic Minority Over-sampling Technique) algorithm during the training phase to augment the attack samples. Furthermore, the use of Focal Loss penalized easy-to-classify examples, forcing the model to focus on hard-to-detect anomalies.

Regarding operational feasibility, a raw False Positive Rate of 0.49% could theoretically generate thousands of alerts. As demonstrated in the ablation study (Table 4), the full configuration of the cognitive model achieves the lowest FPR compared to standalone pipelines, confirming its practical utility. To mitigate "alarm fatigue,"

TABLE 4 Ablation analysis of hybrid model component contribution.

Model configuration	F1-score	FPR (%)	Avg. reaction time (s)
CNN + LSTM only	0.92	0.65	1.4
AE + LSTM only	0.92	0.72	1.6
CNN + LSTM + AE + LSTM (no correlation)	0.94	0.58	1.3
Full cognitive model (with correlation)	0.95	0.49	1.2

CNN, convolutional neural network; LSTM, long short-term memory network; AE, Autoencoder; correlation—the event/feature correlation module that fuses model outputs and contextual signals to improve decision consistency across streams. F1-Score—harmonic mean of precision and recall (higher is better). FPR (%)—false positive rate as a percentage of normal events incorrectly flagged as attacks/anomalies (lower is better). Avg. reaction time (s)—average latency from anomaly onset to detection/trigger (lower is better).

the Cognitive Correlation Layer implements alert aggregation. Individual anomalous sessions are not reported as separate incidents. Instead, they are grouped by source IP, asset ID, and time window ($T = 60s$). An alert is escalated to the operator only if the aggregated risk score exceeds the threshold for a sustained period. This aggregation reduces the effective volume of notifications to approximately 3–5 actionable alerts per shift, which is manageable for a human operator.

4.3 Theoretical and regulatory alignment

From a theoretical standpoint, the results confirm the validity of applying a multidomain approach and a probabilistic risk model (Anthi et al., 2021; Tuptuk et al., 2021; Inoue et al., 2017). Using a model that links event intensity, priority, and asset state enables a shift from reactive detection to predictive risk management.

The developed architecture fully complies with international standards such as IEC 62443 and ISO/IEC 27019, as well as harmonized Kazakh national standards (ST RK ISO/IEC 27001-2022 and ST RK GOST R 56939-2016) (Nisar et al., 2022; Zhang et al., 2023; Qureshi et al., 2024). This ensures its applicability to critical infrastructure. The incorporation of Zero Trust principles (Tariq et al., 2019; Homaei et al., 2024) enables proactive response aligned with modern cybersecurity requirements.

4.4 Practical significance and future prospects

The practical significance of the system lies in its modularity and readiness for integration into existing Security Operations Centers (SOC). The pseudocode and modular structure (Config, IngestAndNormalize, RiskModel, AdaptiveThreshold) facilitate deployment on edge nodes (Edge Computing), ensuring scalability and minimal latency (Gauthama Raman et al., 2021; Kravchik and Shabtai, 2018).

The results also emphasize the potential of digital twins for enhancing security. Using a bakery plant digital twin not only for verification but also as a tool for model training and testing allowed the identification of rare attack scenarios not present in standard datasets (Zhang et al., 2023).

Future research directions include expanding the system's cognitive layer, particularly through the integration of graph neural networks to analyze complex interdependencies between assets (Kantharaju et al., 2024), and employing cyber threat ontologies for automatic event interpretation and adaptive reconfiguration of defense scenarios.

5 General conclusions of the publication

The conducted study confirmed the hypothesis regarding the feasibility of creating an effective cognitive early incident detection system for an industrial bakery network, based on the integration of machine learning methods, user behavior analytics, and the Zero Trust architecture concept (Tariq et al., 2019; Mugarza et al., 2020; Homaei et al., 2024). The developed architecture integrates network, technological, and behavioral data, ensuring comprehensive monitoring and dynamic real-time risk assessment.

The analysis of the results demonstrated that the use of hybrid neural network models, including the Byte2Image CNN + LSTM and AE + LSTM pipelines, enables high accuracy in detecting incidents related to both external attacks and internal violations. The fusion model achieved the best performance metrics—Precision = 0.96, Recall = 0.93, F1 = 0.95, and AUC = 0.98—with an average response time of less than 1.2 s, confirming its ability to effectively distinguish between normal and abnormal states under conditions of high variability in technological processes (Gauthama Raman et al., 2021; Abhishek and Singh, 2023).

The integration of a cognitive correlation layer, which implements the link between the results of neural network analyzers, the production cycle context, and operator behavior profiles, has reduced the false positive rate to less than 0.5% and increased resilience to data drift and technological noise (Knowles et al., 2015; Zhao et al., 2022; Dietz and Pernul, 2020). This result provides a significant advantage over traditional SIEM/IDS platforms that do not account for the specific features of production processes.

The proposed architecture complies with modern international and national information security standards, including IEC 62443, ISO/IEC 27019, ST RK ISO/IEC 27001-2022, and ST RK GO ST R 56939-2016 (Nisar et al., 2022; Zhang et al., 2023; Qureshi et al., 2024). This ensures its practical applicability in food industry enterprises in the Republic of Kazakhstan and CIS countries without the need for deep modification of existing production systems.

Experiments on a digital twin testbed confirmed that the model is resilient to sensor errors, communication channel degradation, and network drift (Inoue et al., 2017; Li et al., 2024; Ahmad et al., 2023). Adaptive threshold optimization allows the system to autonomously adjust sensitivity based on event flow dynamics, maintaining a balance between reaction speed and the number of false alarms. Such a mechanism aligns with the principles of predictive analytics and serves as a foundation for proactive cyber risk management (Anthi et al., 2021; Tuptuk et al., 2021; Inoue et al., 2017).

The practical significance of the study lies in the fact that the proposed cognitive architecture can be implemented as a set of microservices using container orchestration (Docker, Kubernetes) and integrated into existing monitoring and response systems (SOC) via APIs. Implementation in Python and TensorFlow environments ensures compatibility with modern MLOps tools, as well as the capability for distributed flow processing on edge nodes with minimal

latency (Gauthama Raman et al., 2021; Kravchik and Shabtai, 2018; Holdbrook et al., 2024).

Incorporating elements of the Zero Trust concept enabled the implementation of automatic incident responses, including network microsegmentation, isolation of infected assets, key rotation, and blocking of obsolete SSL/TLS and SNMP protocols (Tariq et al., 2019; Xie et al., 2021; Li et al., 2024; Ahmad et al., 2023; Homaei et al., 2024). Thus, the system not only detects threats but also prevents their development, which meets modern requirements for industrial-grade intelligent cybersecurity systems.

The obtained results confirm the feasibility of practical application of the developed architecture to increase the resilience and reliability of industrial production. The system ensures a reduction in the probability of technological cycle disruption, prevents equipment downtime, and lowers the aggregate risk of financial losses due to cyber incidents (Zhang et al., 2023; Homaei et al., 2024).

The scientific novelty of the work lies in the formation of a cognitive approach to ensuring industrial cybersecurity, where artificial intelligence elements are used not only for detection but also for interpretation, forecasting, and adaptation of the protection system. The practical implementation of the presented solution opens up opportunities for the further development of multi-agent cognitive platforms that combine digital twins, contextual analytics, and predictive threat models (Holdbrook et al., 2024).

Overall, the study showed that the proposed cognitive early incident detection system is a universal and scalable foundation for building next-generation intelligent industrial cyber defense systems. Its application ensures a transition from reactive to proactive security management, integration into the Industry 4.0 ecosystem, and an increased level of trust in digital production processes (Knowles et al., 2015; Bhamare et al., 2020; Cherdantseva et al., 2016; Anthi et al., 2021; Tuptuk et al., 2021; Inoue et al., 2017; Zhang et al., 2023; Qureshi et al., 2024; Homaei et al., 2024; Holdbrook et al., 2024; Kantharaju et al., 2024; Dietz and Pernul, 2020).

5.1 Limitations

We acknowledge that the reliance on older datasets (UNSW-NB15, CIC-IDS2017) combined with simulated OT data may not fully capture the complexity of zero-day exploits in 2026. Furthermore, “data drift” (e.g., sensor aging or recipe changes) remains a challenge. Future work will focus on implementing online learning mechanisms to update the AE + LSTM model parameters dynamically without stopping the production line, ensuring the system adapts to long-term operational shifts.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

GA: Writing – review & editing, Conceptualization, Supervision. DP-T: Investigation, Software, Writing – original draft, Data curation. SA: Visualization, Formal analysis, Writing – review & editing, Validation, Methodology. NZ: Validation, Project administration,

Resources, Writing – original draft. LB: Formal analysis, Data curation, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by the Ministry of Education and Science of the Republic of Kazakhstan grant number BR24992975. “Development of a Digital Twin for the Food Industry Enterprise Using Artificial Intelligence and IIoT Technologies.”

Acknowledgments

The authors express their gratitude to colleagues and academic staff from the participating universities for their methodological assistance, consultations, and support during the preparation and implementation of this research. The authors also thank all experts who contributed to the discussion of the results and improvement of the research methodology. The authors would like to express their sincere gratitude to Toshihiko Amemiya (Professor Emeritus, Kansai University) and Elaine Gerbert (Associate Professor, University of Kansas) for their invaluable assistance with the back-translation of the J-EC scale.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. During the preparation of this work, the authors used AI exclusively for the purpose of translating the text into English and improving linguistic readability. After using this tool, the authors rigorously reviewed and edited the content to ensure accuracy and taking full responsibility for the content of the publication.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abhishek, K., and Singh, S. (2023). Hybrid deep learning-based intrusion detection system for industrial networks. *Comput. Secur.* 128:103160. doi: 10.1016/j.cose.2023.103160
- Ahmad, R., Raza, S., Asad, M. U., Khan, A., and Rehman, S. U. (2023). Security risk assessment in industrial control systems using fuzzy logic and machine learning. *IEEE Access* 11, 95278–95294. doi: 10.1109/ACCESS.2023.3318743
- Anthi, E., Williams, L., Burnap, P., and Jones, K. (2021). A three-tiered intrusion detection system for industrial control systems. *J. Cybersecur.* 7:tyab006. doi: 10.1093/cybsec/tyab006
- Basseville, M., and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, NJ: Prentice Hall.
- Bhamare, D., Zolanvari, M., Erbad, A., Jain, R., Khan, K., and Meskin, N. (2020). Cybersecurity for industrial control systems: a survey. *Comput. Secur.* 89:101677. doi: 10.1016/j.cose.2019.101677
- Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., et al. (2016). A review of cyber security risk assessment methods for SCADA systems. *Comput. Secur.* 56, 1–27. doi: 10.1016/j.cose.2015.09.009
- Dietz, M., and Pernul, G. (2020). Unleashing the digital twin's potential for ICS security. *IEEE Secur. Privacy* 18, 77–83. doi: 10.1109/MSEC.2019.2961650
- Gauthama Raman, M. R., Maglaras, L., Kim, K., Janicke, H., and Ferrag, M. A. (2021). Machine learning for intrusion detection in industrial control systems: challenges and lessons from experimental evaluation. *Cybersecurity* 4:18. doi: 10.1186/s42400-021-00095-5
- Holdbrook, R., Balador, A., and Eliasson, J. (2024). Network-based intrusion detection for industrial and robotic systems. *Electronics* 13:4440. doi: 10.3390/electronics13224440
- Homaei, M., Mogollón-Gutiérrez, Ó., Sancho, J. C., Ávila, M., and Caro, A. (2024). A review of digital twins and their application in cybersecurity based on artificial intelligence. *Artif. Intell. Rev.* 57, 2195–2230. doi: 10.1007/s10462-024-10805-3
- Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C. M., and Sun, J. (2017). "Anomaly detection for a water treatment system using unsupervised machine learning." In: *IEEE International Conference on Data Mining Workshops (ICDMW)*, New York, NY: IEEE
- Kanharaju, V., Suresh, H., Niranjnamurthy, M., Amin, F., and Alabrah, A. (2024). Machine learning-based intrusion detection framework for detecting security attacks in internet of things. *Sci. Rep.* 14:30275. doi: 10.1038/s41598-024-30275-3
- Knowles, W., Prince, D., Hutchison, D., Disso, J. P., and Jones, K. (2015). A survey of cyber security management in industrial control systems. *Int. J. Crit. Infrastruct. Prot.* 9, 52–80. doi: 10.1016/j.ijcip.2015.02.002
- Kravchik, M., and Shabtai, A. (2018). "Detecting cyber-attacks in industrial control systems using convolutional neural networks." In: *Proceedings of the 2018 workshop on cyber-physical systems security and privacy (CPS-SPC)*, New York, NY: Association for Computing Machinery
- Li, C., Peng, X., Zhang, L., Xu, Q., and Zhao, D. (2024). Multimodal deep fusion anomaly detection for industrial cyber-physical systems. *IEEE Internet Things J.* 11, 2651–2665. doi: 10.1109/JIOT.2023.3321687
- Lin, W., Shao, L., Chen, J., and He, Y. (2022). Data-driven anomaly detection for industrial control systems based on dynamic threshold adjustment. *IEEE Trans. Ind. Inform.* 18, 2532–2543. doi: 10.1109/TII.2021.3116743
- Mugarza, I., Araujo, A., Jacob, E., and Huarte, M. (2020). Security issues and software updates management in the industrial internet of things: a survey. *Sensors* 20:6777. doi: 10.3390/s20236777
- Neyman, J., and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. Ser. A* 231, 289–337. doi: 10.1098/rsta.1933.0009
- Nisar, M., Shah, S. C., and Rehman, A. (2022). Deep hybrid autoencoder-LSTM model for anomaly detection in industrial IoT. *Sensors* 22:6743. doi: 10.3390/s22186743
- Qureshi, A. R., Alhumayni, H., Alqurashi, F., and Aljuaid, H. (2024). Digital twin-based cybersecurity framework for smart manufacturing systems. *Futur. Gener. Comput. Syst.* 158, 39–55. doi: 10.1016/j.future.2024.02.009
- Raman, G. R. M., Ahmed, C. M., and Mathur, A. (2021). Machine learning for intrusion detection in industrial control systems: challenges and lessons from experimental evaluation. *Cybersecurity* 4:27. doi: 10.1186/s42400-021-00095-5
- Tariq, N., Asim, M., and Khan, F. A. (2019). Securing SCADA-based critical infrastructures: challenges and open issues. *Procedia Comput. Sci.* 155, 612–617. doi: 10.1016/j.procs.2019.08.086
- Tuptuk, N., Hazell, P., Watson, J., and Hailes, S. (2021). A systematic review of the state of cyber-security in water systems. *Water* 13:81. doi: 10.3390/w13010081
- Wang, W., Zhu, M., Wang, J., Zeng, X., and Yang, Z. (2022). End-to-end encrypted telemetry protection for SCADA systems using edge intelligence. *IEEE Internet Things J.* 9, 13210–13220. doi: 10.1109/JIOT.2021.3128749
- Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., and Liu, Y. (2021). A survey of machine learning techniques applied to cybersecurity. *IEEE Commun. Surv. Tutor.* 23, 546–576. doi: 10.1109/COMST.2020.3024748
- Zhang, Y., Lin, X., Wang, W., Wu, L., and Zhang, Y. (2023). A digital twin-driven approach for industrial anomaly detection. *Robot. Comput.-Integr. Manuf.* 81:102547. doi: 10.1016/j.rcim.2022.102547
- Zhao, X., Zhang, R., Hu, X., Zhang, H., and Sangaiah, A. K. (2022). Anomaly detection approach in industrial control systems using PSO-1DCNN-BiLSTM. *Information* 13:450. doi: 10.3390/info13100450