Clustering for Adaptive Cybersecurity in Smart Infrastructure

Sonu Devi[1], Dharmender Kumar[2], Narender Kumar[3]

{ pooniasonu066@gmail.com[1], dharmindia24@gmail.com[2], narenderster@gjust.org[3] }

Department of Computer Science and Engineering,
Guru Jambheshwar University of Science and Technology (GJU S&T), Hisar, Haryana[1,3],
Department of Artificial Intelligence & Data Science
Guru Jambheshwar University of Science and Technology (GJU S&T), Hisar, Haryana[2],

Abstract. Smart infrastructure faces growing cyber threats that require intelligent and adaptive security. For detecting intrusions and anomalies in such environments this survey emphasizes clustering as a key technique. Clustering-based systems show better accuracy and resilience as compared to traditional methods. However, challenges such as latency, limited computing power, dataset dependency and the need for continuous threat intelligence still affect deployment. To address these issues, future research is moving toward neuromorphic and bio-inspired clustering, federated and privacy-preserving clustering, temporal–spatial clustering and energy-efficient quantum clustering. Strengthening clustering techniques with deep neural networks, hybrid approaches and improved frameworks can further enhance cybersecurity in smart infrastructure.

Keywords: Cyber-Physical Systems, IoT Security, Clustering, Cluster Validity Indices, Anomaly Detection, Adaptive Cybersecurity.

1 Introduction

Smart infrastructure systems bring substantial enhancements in efficiency, safety and flexibility because of its complex, interconnected and dynamic nature. On the other hand, it introduces the significant challenges in cybersecurity, such as intelligent transportation networks, power grids, industrial control systems, and Internet-of-Things (IoT) ecosystems. Due to this, there is a growing need for adaptive cybersecurity mechanisms in a distributed, resilient, and scalable manner to detect and respond to threat. Clustering algorithms can reveal previously unknown attack vectors, identify compromised devices, and support dynamic adaptation of security policies by grouping similar behaviors or traffic patterns. Clustering algorithms has emerged as powerful tools for uncovering hidden structure for driving anomaly detection, intrusion detection, and threat intelligence applications in high-volume, high-dimensional data. However, selecting appropriate algorithms, distance measures, and cluster validity indices that can automatically determine the optimal number of clusters, assess clustering quality and ensure robust performance under evolving network conditions depends on the effectiveness of clustering.

Deep Embedded Clustering (DEC) and its semantic variants uses neural networks to learn representations of feature and cluster assignments simultaneously, which enable detection of

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2362395

complex attack patterns in network traffic and log data. Variational autoencoder based methods and quantum assisted models further extend clustering to high-dimensional, time-series and cyber physical system data by capturing nonlinear relationships and enhance generalization.

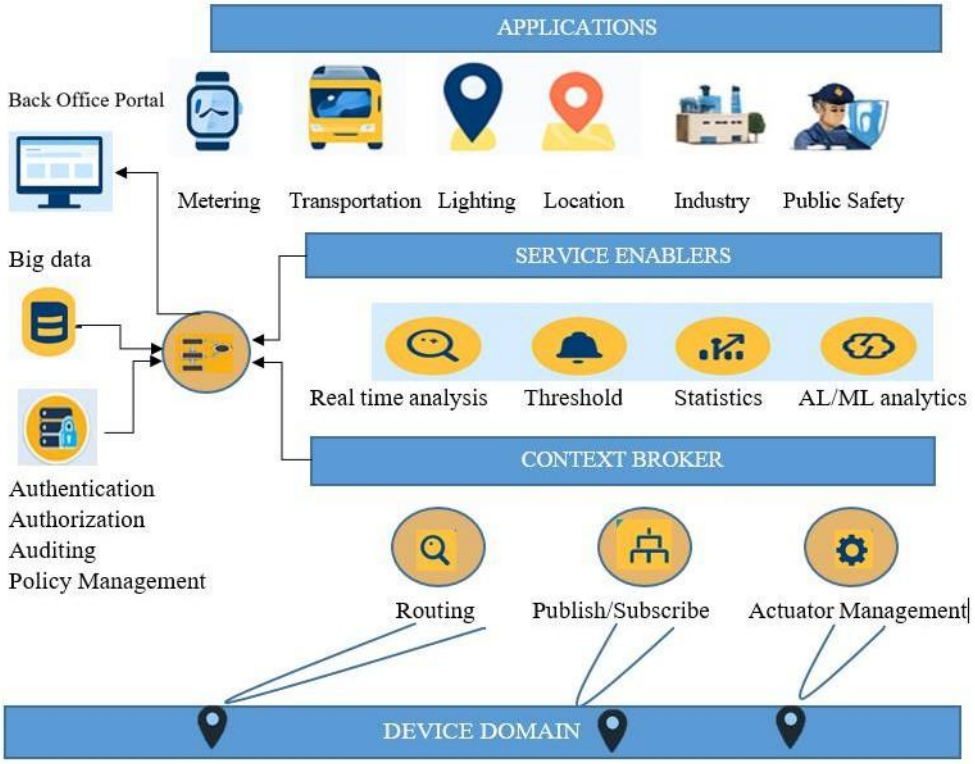


Fig. 1. Smart Infrastructure System [64]

In this review we study the key applications of clustering in distributed IoT architecture, deep learning & machine learning approach for cybersecurity, network security and anomaly detection systems. This paper aims to equip researchers and practitioners with a deep understanding of how clustering can drive adaptive cybersecurity in smart infrastructure. We conclude by outlining future research directions including neuromorphic and bio-inspired clustering system, privacy-preserving federated clustering architecture, temporal-spatial fusion clustering, energy-aware explainable quantum clustering.

2 Methodology

We have conducted a systematic review of the published work in great depth. We use the combined search terms such as clustering techniques: clustering, unsupervised learning, for cybersecurity: anomaly detection, intrusion detection, cybersecurity and for cluster measure assessment terms: cluster validity index, silhouette, Davies-Bouldin. We also added field names

as search terms like cyber-physical systems, IoT, smart grid, and industrial control systems so as to maintain focus for purposes of application.

We searched academic databases in order to have a firm foundation like: a leading repository in the engineering and computer research world, ACM Digital Library, Multidisciplinary Science, SpringerLink etc. along with Scopus and Web of Science major interdisciplinary indexing services.

Table 1. Study Selection Criteria

Category	What We Included	What We Excluded
Publication Type	Journal articles and conference papers indexed in Scopus, Web of Science, and other well-known databases.	Non-peer-reviewed reports, editorials, thesis work
Language	English publications	Non-English publications
Time Frame	Studies published between 2010 and 2025	Older studies (before 2010)
Domain Focus	Cybersecurity research in CPS, IoT, smart grids, industrial systems	Studies unrelated to cybersecurity or outside our domains of interest
Method Requirements	Studies using clustering algorithms for threat detection	Works focusing only on supervised learning or purely theoretical clustering
Evaluation Standards	Studies discussing or applying cluster validity measures	Studies that didn't consider cluster quality evaluation
Advanced Approaches	Research combining clustering with modern AI methods	Studies limited to basic machine learning without clustering integration

3. Related Work

3.1 Distributed IoT Architectures and Security Challenges

R. Roman et al. [1] discussed about centralized and distributed architecture of IoT. This paper includes the various security challenges and privacy policy that need to be focused to solved but also include some strength and properties of distributed IoT. A new class of attack, false data injection attack in pivotal infrastructure such as electric power grid has been presented by Y. Liu et al. [3]. In this they show attacker introduced arbitrary error into certain state variable which can exploit the configuration of a power system by successfully demonstrate the process through simulation using IEEE test system. In context of social multimedia, a hybrid deep learning based anomaly detection scheme is proposed for suspicious flow detection to enhance the reliability of SDN [4]. A novel approach which is able to find well separated cluster was given by J.C. Bezdek et al. [13] as VAT i.e. a tool for visual assessment of cluster tendency and

a cluster separation measure which assumed to have data density is given by D.L. Davies at al. [14] to provide reliable clustering. To improve CSP security in industrial IoT, T. Zhukbayeva et al. [25] introduced an edge computing centered integrated solution which analyze network traffic and detect intrusion. In this researchers used K-Means and DBSCAN algorithms to analyze traffic pattern.

Gus et al. [39] emphasizing the importance of lightweight cryptographic algorithms for resource-constrained devices. To ensure anonymity and protection against traffic analysis their study highlighted critical trade-offs between performance and security, underlining the necessity of IoT-specific Anonymous Overlay Networks (AONs) and Anonymous Routing Protocols (ARPs). Almutairi et al. [40] identifying significant vulnerabilities arising from resource constraints and cloud complexities in the analysis of the IoT and cloud computing integration. They emphasized the need for standardized end-to-end security frameworks incorporating AI-driven anomaly detection, blockchain-enhanced authentication and post-quantum cryptographic techniques in order to mitigate threats such as API exploits, insider attacks and adversarial AI. Dauda et al. [41] categorized cloud, edge and fog paradigms according to their scalability, latency and privacy capabilities in their systematic survey of IoT application architectures while to fortify distributed deployments they also noted that multi-protocol edge gateways and improved interoperability were required. Fernandez-Carames [42] examined new quantum-resistant cryptosystems focused on post-quantum IoT security and highlighting the challenges of implementing such systems on resource-limited IoT devices. The idea of Zero-Trust Foundation Models (ZTFMs) was developed by Li et al. [43] that incorporate zero-trust characteristics such as data confidentiality, least privilege access and continuous verification into federated and blockchain-supported IoT architectures. Offering algorithms for robust decentralized analytics and adversary detection Chen et al. [44] examined secure distributed inference mechanisms in IoT which highlighting the necessity of adaptive countermeasures against compromised nodes.

3.2 Deep Learning and Machine Learning Approaches for Cybersecurity

A hierarchical spatial-temporal features based intrusion detection system was proposed as novel IDS by W. Wang [5] to reduce false alarm rate effectively. This proposed IDS first use deep convolutional neural network to learn low level spatial features of network traffic and after that use long short term memory to learn high level temporal features. With Donut, an unsupervised anomaly detection algorithm based on variational autoencoders (VAE) for seasonal Key Performance Indicators (KPIs) Xu et al. [9] outperformed with varying patterns and data quality challenges. F-scores ranged between 0.75 and 0.9 for KPIs from one major global Internet company. Their research gave a novel KDE (Kernel Density Estimator) interpretation of reconstruction for Donut, making Donut the first VAE-based unsupervised anomaly detection algorithm with theoretical explanation that is sound.

Kontagora et al. [11] presented a comprehensive review by collecting results from different studies on Random Forest, Support Vector Machine, k-nearest Neighbors and Gaussian Mixture Models of machine learning algorithms for anomaly detection in IoT networks. Their focus is on security against new forms of threat and proposed integration with blockchain and edge computing like new generation technology. Using the method of kingma and welling [19] (Auto-encoding Variational Bayes), one can find a continuous latent variable alternative to the above Gibbs sampler, it is efficient and practical within hidden Markov model. An optimized

DEC (Optimized Deep Embedded Clustering) model combines representation learning with clustering to achieve a 94% detection rate on data from the DARPA98 dataset without labels. Xie T. Bin and Qing Wen Min [18] used Deep Embedded Clustering (DEC) which achieves simultaneous learning of both representations and cluster assignments by optimizing for KL divergence in the learned feature space, so that attackers can be detected without any supervision. Hajj R. A. [19] introduces Stochastic Gradient Variational Bayes for efficient inference in continuous latent-variable models, providing a Deep Anomaly Detection architecture. SDEC (Stochastic Deep Embedded Clustering) integrates transformer embeddings and autoencoders which leads to up to 85.7% clustering accuracy on textual security logs [20].

A hybrid CNN-GRU IDS was developed by Jablaoui et al. [45] that successfully captures both spatial and temporal information and it achieved accuracy (98.6%) on the NF-UNSW-NB15 dataset and accuracy (97.95%) on the NF-CSE-CIC-IDS2018. A comprehensive survey on DL-based IDS for emerging technologies such as IoT, edge and cloud computing conducted by Neto et al. [46] that emphasizing challenges like explainability, operationalization and deployment in real-world environments. Kimanzi et al. [47] found a research gap in IoT-specific IDS in their evaluation of cutting-edge DL algorithms such as CNN, RNN, DBN and autoencoders. Awajan [48] developed a protocol independent DL-based intrusion detection system (IDS) for Internet of Things networks with an average accuracy of 93.74% using a four-layer Fully Connected (FC) architecture. Wu et al. [49] highlighting future trends like model compression and resource optimization by analyzing multiple DL models such as transformers and GPT-based architectures. Zhang et al. [50] examined DL applications to handle spatiotemporal feature extraction and data imbalance that use hybrid CNN-RNN and GAN-based techniques. Gueriani et al. [51] and Yang et al. [52] outlining key challenges in training efficiency, real-world deployment and adaptive learning explored Deep Reinforcement Learning (DRL) for IDS.

3.3 Network Security and Anomaly Detection Systems

Li and Huang [10] conducted a comprehensive survey of anomalies in wireless networks. They systematically categorized anomaly-detecting techniques for WSN architectures and their detection category such as Hierarchical/Flat. Their survey also covered statistical methods, rule-based approaches, data mining techniques, computational intelligence strategies, game theory, graph-based methods and hybrid methodologies. It emphasized key design principles for anomaly detection techniques of WSNs fit, which addressed the constraints limited energy, weak computation capacity, poor memory and short communication range. Moustafa et al. [12] presented a comprehensive review on Network Anomaly Detection systems (NADS). They covers all the aspect about ensemble learning and Deep Learning and his research includes extensive applications like Data Centers, Internet of Things (IoT), fog computing of NADS systems.

Ali et al. [17] conducted a comparison on different techniques used for IoT intrusion detection like DBSCAN, K-Means and Agglomerative Clustering. Their results showed that DBSCAN used to real-time detections of known and novel threats significantly whereas Agglomerative clustering outperforms with higher accuracy in terms of both precision and recall, as well as increased F1 score overall. Althobaiti's [23] IDS framework employs Labeled Cyber-Physical Datasets and ML-based classifiers Random Forest (99.5% accuracy) and SGD (93.6% accuracy). Statistical validations confirmed RF's superiority over SGDs in IoT/CPS environments with constrained resources. Kumar et al. [25] propose an edge-based anomaly

detection system combining K-means and DBSCAN for data clustering and KNN, RF or LR as Rule-based classifiers. The result achieved near-perfect ROC methodology scores, precision/recall curves in the same high performance class while using edge processing reduces latencies significantly in all cases except where wire speed makes it unnecessary. To address evolving threats and traffic complexity, modern network security research increasingly leverages advanced deep learning architectures and novel learning paradigms. The self-supervised transformer contrastive learning framework that significantly improves generalization and detection of unseen attacks in packet-level representations is one example of a transformer-based methodology that has demonstrated promise result [58]. Self-supervised approaches like SAFE employ masked autoencoders to develop robust representations for anomaly detection and transform network data into image-like form, outperforming current state-of-the-art methods [59]. By achieving better detection and interpretability via SHAP and LIME, CSAGC-IDS couples a conditional GAN for synthetic data generation with a cost-sensitive, attention-enhanced CNN to tackle class imbalance and complex traffic patterns [60].

S. Narmadha and N. Balaji [61] proposed hybrid model to reduce false positive rates and better capture sequential dependencies in network traffic by including PSO, tuned optimal autoencoder, LSTM systems. A multi-agent adaptive deep learning framework leveraging federated sharing among distributed agents developed by M. Soltani et. al. [62] that accommodates concept drift and enables continual learning in real-time environments. In addition to these advancements, recent survey emphasizes deep learning applicability in SDN and emerging technologies, but they also highlight the key challenges such as explainability, dataset constraints and deployment in practical contexts [46] [63].

3.4 Critical Infrastructure Security and Attack Models

Liu, Ning and Reiter [3] outlining a significant attack technique such as injecting false data into the electric power grid's state estimator, in recent times some experts have suggested a novel way to identify state attacks. To achieve success and goes undetected by existing bad measurement detection techniques, new class of attacks breaches power system designs. Using a small test system study which do not just affect but change results itself, they showed how to produce vectors for the very same sequential estimation procedure. Two possible attack scenarios were discovered in their research, one was attackers who are impossible to find because they are physically protected and another one was attackers who cannot extend to compromise the number of meters necessary to reach a state estimator. The simulation demonstration with IEEE test systems made clear that when the power grid meets possible malicious attacks, its protective measures need a fundamental change. Bhosale et al. [21] introduced quantum classical hybrid modeling for monitoring vital infrastructure and achieve 95.2% accuracy using Quantum-Assisted VAE (QAVAE) integrates VAE architectures and parameterized quantum circuits for anomaly detection in high-dimensional time-series cyber-physical data.

Maynard et al. [53] providing a cohesive approach for malware defense by proposing integrating cyber intrusion models with incident response frameworks using Sequential AND Attack Trees. Bayesian and multi-objective decision-support frameworks have also been established to optimize incident response in order to balance safety, operational continuity, and resource allocation [54]. Systematic reviews point out shortcomings in current methods and suggest future research areas by emphasizing the need for dynamic, multi-layered security models for

cyber-physical systems [55]. The technology's significance in threat identification and mitigation highlighted in surveys on safeguarding critical infrastructure underscore the role of artificial intelligence, showing AI potential for enhancing real-time monitoring and response [56]. To further increase the resilience of critical systems, emerging techniques such as generative AI and large language models (LLMs) are being explored for anomaly detection, threat intelligence synthesis, and automated incident response [57].

Table 2. Key Studies on Critical Infrastructure and IoT Security

Researchers	Year	Clustering Method	Validation Approach	Application Domain	Key Achievement	Technical Limitations	Deployment & Operational Constraints
Mehdi et al. [17]	2024	DBSCAN	Silhouette, Davies-Bouldin	IoT Networks	Highest accuracy among clustering methods for IDS	Sensitive to parameter selection; struggles with varying densities	High memory consumption limits its scalability on large or streaming datasets
Xifeng. et al. [20]	2016	SDEC Framework	Semantic Loss, Distributional Loss	Text-based Security	85.7% clustering accuracy on AG News	Limited modality (text-only); semantic drift risk	Not applicable to network traffic or sensor-based IoT data
Bhosale et al. [21]	2025	QAVAE	Reconstruction Error Metrics	Industrial IoT	Enhanced anomaly detection in high-dimensional data	Model complexity; quantum noise sensitivity	Requires quantum hardware, limiting real-world adoption
Althobaiti [23]	2025	Random Forest + Clustering	Precision, Recall, F1	CPS Security	99.5% accuracy in cyberattack detection	Overfitting risk on static datasets	Performance degrades under unseen or evolving attack patterns
T. Zhukabayeva et al. [25]	2025	Edge-DBSCAN + ML	Silhouette Score	Industrial CPS	Real-time threat detection with edge computing	Reduced clustering quality under constrained resources	Latency and compute limitations at edge nodes

B. Gus et al. [39]	2025	Lightweight Cryptography, Anonymous Routing	Literature review and case studies	IoT Edge / CPS	Enhanced security at resource-constrained IoT edge	Limited cryptographic strength vs. advanced threats	Lack of large-scale real-world deployment validation
M. Almutairi, F. T. Sheldon [40]	2025	IoT-Cloud Integration Security Methods	Comparative analysis	IoT-Cloud Systems	Identified key integration vulnerabilities and mitigation strategies	Assumes static architectures	Heterogeneity and interoperability hinder standardized deployment
T. M. Fernandez-Carames [42]	2024	Quantum-Resistant Cryptosystems	Literature review	IoT Security	Surveyed post-quantum cryptography for IoT	Computational overhead on constrained devices	Feasibility challenges for low-power IoT hardware
K. Li et al. [43]	2025	Zero-Trust Foundation Models for AI	Simulation / Prototype evaluation	IoT / Collaborative AI	Introduced zero-trust paradigm for AI-enabled IoT	Limited empirical validation	Scalability and real-time enforcement unresolved
Y. Chen, S. Kar, J. M. F. Moura [44]	2025	Secure Distributed Inference	Theoretical and practical evaluation	IoT / Signal Processing	Secure collaborative inference for distributed IoT	Trade-off between accuracy and security	Increased computational overhead in distributed settings
R. Jablaoui et al. [45]	2025	CNN-GRU Hybrid Deep Learning	NF-UNSW-NB15 & NF-CSE-CIC-IDS2018 datasets	IoT Networks	High accuracy intrusion detection in IoT	Dataset imbalance affects minority attack detection	Generalization issues across heterogeneous IoT environments
E. C. P. Neto et al. [46]	2025	Survey of DL Methods in IDS	Comparative analysis of 59 papers, 39 datasets	Cloud, Edge, IoT, MEC, ICS	Identified limitations and opportunities for DL-IDS	Lack of unified benchmarks	Explainability and operationalization challenges

R. Kimanzi et al. [47]	2024	CNN, RNN, DBN, DNN, LSTM, AE, MLP, SNN	Dataset benchmarking and performance metrics	Network Security	Comprehensive review of DL algorithms for IDS	High computational cost	Limited adaptability to evolving threats
A. Awajan [48]	2024	Deep Fully Connected Network	Experimental evaluation with simulated & real intrusions	IoT Networks	Platform-independent intrusion detection	Performance depends on training data quality	Scalability across diverse IoT topologies remains unclear
Y. Wu et al. [49]	2024	DL Models: AE, DBN, DNN, CNN, RNN, GAN, Transformers	Comparative evaluation of models and datasets	Network Security	Overview of datasets, preprocessing, and DL models	Large model sizes	Resource constraints limit real-time deployment
Y. Zhang et al. [50]	2025	CNN-RNN Hybrid, GANs for imbalance	Comparative performance analysis	IDS in IoT / Distributed Networks	Improved spatiotemporal feature extraction and minority attack detection	Interpretability limitations	High computational overhead
A. Gueriani et al. [51]	2024	Deep Reinforcement Learning (DQN, Actor-Critic)	Survey of DRL methods	IoT Networks	Autonomous IDS for IoT environments	Sample inefficiency	Long training times impede real-time use

W. Yang et al. [52]	2024	DRL for Network Intrusion Detection	Survey and evaluation on public datasets	Network Security	Identified DRL methods outperforming traditional DL in some cases	Stability issues during training	Limited real-world validation
P. Maynard, Y. Cherdantseva, et. al.[53]	2025	Sequential & Attack Trees, Cyber Intrusion Modeling	Case studies with malware attacks	Critical Infrastructure	Integrated modeling of intrusions and incident response	Static attack assumptions	Scalability and real-time response limitations
S. Huang, C. M. Poskitt, L. K. Shar [54]	2025	Bayesian and Multi-Objective Decision Support	Multi-objective scenario analysis	Cyber-Physical Systems	Optimized incident mitigation balancing safety & operations	Model complexity	Difficult trade-off tuning in live systems
S. Huang, C. M. Poskitt, L. K. Shar [55]	2024	Security Modeling Framework	Systematic Literature Review	Cyber-Physical Systems	Comprehensive analysis of security modeling approaches	Lack of standardization	Limited tool support for practitioners
K. J. Raval [56]	2024	Survey-based AI Security Techniques	Literature synthesis and analysis	Critical Infrastructures	Identified AI-based defense opportunities	Conceptual focus	Deployment pathways remain unclear
Yigit [57]	2025	Generative AI / LLMs for Anomaly Detection	Case studies / Prototype evaluation	Critical Infrastructure	Explored LLMs for automated threat detection	Hallucination and reliability risks	Integration with real-time ICS environments unresolved

3.5 Datasets and Evaluation Frameworks

M. Ring et al. [6] focused on providing network based intrusion detection dataset with including 15 different features or properties and main contribution of their paper is to include 34 network based datasets. Moustafa and Slay [7][12] stressed how current collections of network traffic data lack any comprehensive sampling of modern data environments. Therefore, they took their establishment of the UNSW-NB15 dataset object as a chance to make a difference. This data set took real normal network traffic and modern modeled attack activities which were produced at the UNSW cyber security lab in a synthetic environment using the IXIA PerfectStorm tool. The UNSW-NB15 dataset is made up of packets and flows, with nine major families of attacks; analysis has already been done on it using Argus and the Bro-IDS tool provide similar output that you could extract to expand your scale.

Table 3. Cybersecurity datasets

Dataset	Environment	Attack Types	Key Strengths	Main Limitations	Reference
KDD Cup 99	Network IDS	DoS, Probe, R2L, U2R	Large-scale benchmark	Redundant & outdated	[65]
NSL-KDD	Network IDS	DoS, Probe, R2L, U2R	Reduced redundancy	Obsolete attack patterns	[66]
UNSW-NB15	Enterprise network	Fuzzers, Exploits, Backdoors	Modern attack diversity	Synthetic environment	[67]
CI-CIDS2017	Enterprise network	DDoS, Botnet, Brute Force	Realistic traffic	Severe class imbalance	[68]
CSE-CIC-IDS2018	Large-scale enterprise	Advanced DDoS, Botnet	Large and recent dataset	High computational cost	[69]
BoT-IoT	IoT network	DDoS, Data exfiltration	IoT-focused attacks	Simplistic normal traffic	[70]
TON_IoT	IoT & IIoT	Malware, Ransomware	Smart infrastructure realism	Limited attack variety	[71]
Edge-IIoTset	Industrial IoT	MITM, Injection, DoS	Industrial relevance	Limited benchmarks	[72]

3.6 Real-World Smart Infrastructure Incidents

Table 4 shows the mapping of real-world smart infrastrcutre incidents with AI-driven cybersecurity approaches that demonstarte the applicability of taxonomy across different domain such as healthcare, industrial, energy and transportation.

Table 4. Taxonomy of AI-Driven Cybersecurity Approaches for Real-World Smart Infrastructure Incidents

Taxonomy Category	Core Methods (Latest)	Smart Infrastructure Domain	Representative Real World Incident	Key Strength	Key Limitation	References

Clustering-Based Anomaly Detection	DBSCAN, HDBSCAN, Spectral clustering	Smart grids, ICS	Ukraine power grid cyberattack	Detects unknown attacks without labels	Sensitive to noise and concept drift	[73], [74]
Hybrid Learning Models	Clustering + Deep Learning	Industrial pipelines, IIoT	Colonial Pipeline ransomware	Combines interpretability and accuracy	Increased computational complexity	[75], [76]
Transformer-Based IDS	Self-attention, temporal transformers	Enterprise & IoT networks	Large-scale ransomware campaigns	Captures long-range dependencies	High training cost	[77], [78]
Graph Neural Network (GNN) Methods	GCN, GAT, temporal GNNs	Smart cities, transportation	V2I communication attacks	Models relational and coordinated attacks	Scalability challenges	[79], [80]
Self-Supervised Anomaly Detection	Autoencoders, contrastive learning	Smart grids, IIoT	Zero-day ICS exploits	Minimal labeled data needed	Threshold selection	[26], [67]
Adversarially Robust Models	Robust training, detection defenses	Critical infrastructure	Model evasion attacks	Improved resilience	Higher training overhead	[28], [30]

4. Cluster Validity Indices (CVIs)

Table 5 shows these indices that provide quantitative metrics of compactness and separation, which provide the automatic selection of the number of clusters and distinguish meaningful groupings from noise.

Table 5. Cluster Validity Indices (CVIs**)**

Internal CVIs	**Davies–Bouldin Index (DBI)** [14, 27]	$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} \left(\frac{S_j + S_j}{M_{ij}}\right)$	• $S_i = \frac{1}{\lvert C_i \rvert}\sum_{x \in C_i} \lVert x - C_i \rVert$ (inter- cluster scatter of cluster i) • $M_{ij} = \lVert C_i - C_j \rVert$ (distance between centroids)
	Silhouette Coefficient [15]	$S(x) = \frac{b(x) - a(x)}{max\{a(x), b(x)\}}$	• a(x) = average distance of x to points in same cluster • b(x) = minimum average distance of x to points in another cluster
	Dunn Index [29]	$DI = \frac{min_{1 \leq j \leq k} d(C_i, C_j)}{max_{1 \leq l k} diam(C_l)}$	• $d(C_i, C_j) = min_{x \in C_i,\ y \in C_j} \lVert x - y \rVert$ • $diam(C_l = max_{x,y \in C_l} \lVert x - y \rVert$

	Calinski–Harabasz (CH) Index [16]	CH = $\frac{Tr(B_k)/\ (k-1)}{Tr(W_k)/\ (n-k)}$	• $B_k = \sum_{i=1}^{k} \lvert C_i \rvert (c_i - c)(c_i - c)^T$ (between- cluster scatter) • $W_k = \sum_{i=1}^{k} \sum_{x \in C_i} (x - c_i)\ (x - c_i)^T$ (between- cluster scatter) • $c = global\ centroid$
	Xie–Beni (XB) [31]	XB = $\frac{\sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^{m} \lVert x_j - c_i \rVert^2}{n.min_{i \neq j} \lVert c_i - c_j \rVert^2}$	• u_{ij} = membership degree of x_i in cluster i • m = fuzzifier (typically m =2)
	PBM Index [32]	PBM(k) = $\frac{1}{k} . (\frac{E_1}{E_k} . D_k)^2$	• X = $\{x_1, x_2, \dots, x_n\}$ be the dataset • k = number of clusters • c_i = centroid of cluster i • c = overall data centroid (global mean) • $E_k = \sum_{i=1}^{k} \sum_{x \in C_i} \lVert x - c_i \rVert$ (within-cluster scatter for k clusters) • $E_1 = \sum_{x \in X} \lVert x - c \rVert$ (scatter when all data are in one cluster) • $D_k = max_{i \neq j} \lVert c_i - c_j \rVert$(maximum distance between any two cluster centroids)
	Stability-based internal validation [33]	Stability(k) = $\frac{2}{m(m-1)} \sum_{1 \leq i < j \leq m} Sim(P_i^{(k}$	• $D_1, D_2, \dots, D_m$= perturbed versions of dataset X • $P_i^{(k)}$= clustering partition of D_i with k clusters • Sim($P_i^{(k)}$, $P_j^{(k)}$) = similarity measure between two partitions (e.g. Adjusted Rand Index, Jaccard Index. Normalized Mutual Information)
Relative	**Gap Statistic** [34]		W$_k$ = $\sum_{r=1}^{k} \frac{1}{2\lvert C_r \rvert} \sum_{x_i, x_j \in C_r} \lVert x_i - x_j \rVert^2$ • C_r= cluster r

In-di-ces		Gap(k) = $\frac{1}{B}\sum_{b=1}^{B} \log(W^{*(b)}) - \log(W_k)$	• $\lvert C_r \rvert$= number of points in cluster r • $\lVert x_i - x_j \rVert$= Euclidean distance between points **Within – cluster dispersion for k clusters:** $W_k = \sum_{i=1}^{k} \sum_{x_j \in C_i} \lVert x_j - c_i \rVert^2$ Where: • $C_i = cluster\ i$ • c_i = centroid of cluster i **Compute difference of dispersion:** DIFF (k) = $(k-1)^{\frac{2}{p}} W_{k-1} - k^{\frac{2}{p}} W_k$ Where: • p = dimensionality of the dataset
	Krzanowski–Lai (KL) [35]	KL(k) = $\left\lvert \frac{DIFF(k)}{DIFF(k+1)} \right\rvert$	
	Information-criterion methods [36]	AIC (k) = -2 L_k + 2p_k BIC(k) = -2 L_k + 2p_k log(n)	• L_k = log=likelihood of the model with k clusters • p_k = number of free parameters in the model with k clusters • n = number of data points
	Fuzzy Partition Coefficient (FPC) [37]	FPC = $\frac{1}{n}\sum_{j=1}^{n} \sum_{i=1}^{k} u_{ij}^2$	• $u_{ij} \in [0,1]$ = membership degree of data point $x_j in\ cluster\ i$ • N = number of data points • K = number of clusters • Range : $\frac{1}{k} \leq FPC \leq 1$ • A value closer to 1 means a partition is more crisp (less fuzzy).
Fuzzy			

CVIs			
	Fuzzy Partition Entropy (FPE) [37]	FPE = $-\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{k} u_{ij}\log(u_{ij})$	• Same notations as above • Range: [0, log(k)] • Lower values indicate less fuzziness (crisper partitions).
	Fuzzy Silhouette [38]	FSI = $\frac{1}{n}\sum_{j=1}^{n} s_f(x_j)$ *Higher value indicate better clustering.	For a point x_j, $s_f(x_j) = \frac{b_f(x_j)-a_f(x_j)}{max\{a_f(x_j),b_f(x_j)\}}$ • $a_f(x_i) = \sum_{i=1}^{k} u_{ij}^{m} \cdot d(x_j, c_i)$ • $b_f(x_j) = \min_{l\neq i}\sum_{l=1}^{k} u_{lj}^{m} \cdot d(x_j, c_l)$ • m = fuzzifier (commonly m = 2) • $d(x_j, c_i)$ = distance between point x_j and centroid c_i
	Fuzzy XB variants [31]	XB = $\frac{\sum_{i=1}^{k}\sum_{j=1}^{n} u_{ij}^{m}\|x_j-c_i\|^2}{n \cdot \min_{i\neq j}\|c_i-c_j\|^2}$ *lower value indicates better clustering.	• u_{ij} = membership of point x_j in c • m = fuzzifier parameter • $\|x_j - c_i\|$ = distance from point x_j to ce • $\min_{i\neq j}\|c_i - c_j\|^2$ = minimum inter-cluster centroid distance

5. Result And Discussion

5.1 Security Challenges in Distributed IoT

Roman et al. [1] identify core issues of privacy, authentication, and key management across heterogeneous devices. Some critical features such as ubiquitous connectivity, autonomous behavior and scalability are also identified as foundational to IoT. During device mobility these same features introduce vulnerabilities such as weak authentication, inconsistent trust management and exposure. Rather than relying on traditional perimeter-based defenses, a major step was needed for dynamic, decentralized security models capable of adapting to distributed architectures. Due to limited computational resources privacy preservation in constrained devices is challenging which make many conventional cryptographic solutions impractical.

Liu et al. [3] demonstrate how false-data injection threatens critical infrastructures such as power grids, confirming the high vulnerability of distributed IoT systems. Malicious measurement data can be deliberately crafted by adversaries with knowledge of the power system model in a way that evades traditional bad data detection mechanisms it is demonstrated by their research. The key finding of their study was a mathematical formulation which illustrate how specific measurements are manipulate by attackers to alter the estimated system state without triggering alarm thresholds.

5.2 AI-Based Intrusion Detection

A hybrid deep learning-based anomaly detection system proposed by Garg et al. [4] to extract spatial and temporal features from network traffic flows specifically for Software Defined Networks (SDNs), which leveraging both CNN and RNN models. In centralized SDN environments combining multiple deep learning architectures significantly enhances detection accuracy for suspicious flows is demonstrated. Improved adaptability to dynamic network conditions indicated by high precision and reduced false positive rates. However, their system required significant computational resources at the controller and exhibited dependency on training data volume and quality which pose challenge for large scale real time deployment. Wang et al. [5] introduced a hierarchical model which capable of learning both temporal- spatial correlations from network traffic.

5.3 Clustering Validation Methods

A novel approach which is able to find well separated cluster was given by J.C. Bezdek et al. [13] as VAT i.e. a tool for visual assessment of cluster tendency and a cluster separation measure which assumed to have data density is given by D.L. Davies at al. [14] to provide reliable clustering. To observe potential cluster boundaries before applying clustering algorithms, their result was demonstrated by rearranging pairwise dissimilarity matrices into an ordered, grayscale image. For unsupervised learning VAT reduce the risk of arbitrarily selecting clustering parameters which significantly improved preprocessing accuracy. However, their study is computationally expensive for large datasets, dependent on manual interpretation which limits its scalability on big data environments.

5.4 Edge-Computing Integration

An edge-computing intrusion framework [25] demonstrates scalability and real-time responsiveness which highlights the effectiveness of combining distributed processing with advanced clustering techniques for IoT protection. Their study designed to improve network traffic analysis and intrusion detection in Cyber-Physical Systems (CPS). Compared to traditional cloud-dependent IDS deployments their architecture strategically distributes detection capabilities between edge nodes and central control units which enables real-time anomaly detection with reduced latency.

6. Future Directions and Research Opportunities

6.1 Neuromorphic and Bio-Inspired Clustering Systems

A promising direction for future research is to develop adaptive cybersecurity which is based on spiking neural network based-clustering algorithms that mimic brain plasticity. In parallel

we will integrates genetic algorithm with inspired cluster evolution mechanisms that automatically adapt clustering parameters according to network threat landscapes. Building on these components, we will create swarm-based intelligence clustering systems using ant colony optimization principles without centralized coordination in distributed IoT devices for real-time threat pattern recognition.

6.2 Privacy-Preserving Federated Clustering Architectures

One of the important area of future research is to design clustering protocols based on homomorphic encryption which enable secure multi-party computation across the smart infrastructure domains. Additionally, we will develop differential privacy mechanisms for cluster validity indices that protect sensitive network patterns while maintaining detection accuracy. Moreover, creating a federated learning frameworks that are secured with block-chain which ensure GDPR compliance in cross-border smart city deployments where edge devices share clustering insights without exposing raw cybersecurity data.

6.3 Temporal-Spatial Fusion Clustering

It would be valuable to extend the research using four-dimensional clustering algorithms that utilizing time, place of occurrence, network topology and other information which leads to ultra-low latency threat detection. A promising direction is to develop dynamic clustering using cognitive radio inspired approach where clusters adapt to the availability of spectrum and interference pattern for real-time cyber-physical threat assessment.

6.4 Energy-Aware Explainable Quantum Clustering

We may use quantum annealing based clustering optimization for sustainable smart infrastructure to minimizes the total amount of energy consumed with achieving particularly good detection accuracy. Moreover, create a new framework that provide human-interpretable reasoning for quantum clustering decisions in cybersecurity contexts. Additionally, we will create hybrid quantum-classical clustering systems for battery-powered IoT devices with energy harvesting capabilities.

7. Conclusion

In the construction of adaptive cybersecurity in smart infrastructure this survey gives clustering as a central position. Traditional analytics are consistently outperformed, achieving high accuracy even against complex, emerging cyberattacks by hybrid architectures and edge-enhanced designs. Yet, the deployment of clustering-based solutions is still limited by real-time latency requirements, constrained computing resources, dataset dependency and the persistent need for dynamic and adaptive threat intelligence. Neuromorphic and bio-inspired clustering systems, privacy-preserving federated clustering architectures, temporal-spatial fusion clustering and energy-aware explainable quantum clustering represent prominent future research directions. Consequently, further innovation is required including clustering-driven network anomaly detection to mitigate false data injection attacks, the development of adaptable clustering frameworks consistent across diverse smart infrastructure applications, the integration of deep neural networks with clustering to enhance IDS performance and the

exploration of hybrid clustering algorithms for improved risk assessment and anomaly identification.

References

[1] Roman, R.: On the Features and Challenges of Security and Privacy in Distributed Internet of Things. Computer Networks. pp. 2266-2279 (2013) https://doi.org/10.1016/j.comnet.2012.12.018

[2] Chander, B.: AI-Based Internet of Things for Industry 5.0. Artificial Intelligence-based IoT. pp. 3-45 (2022) https://doi.org/10.1007/978-3-030-87059-1_1

[3] Liu, Y.: False data injection attacks against state estimation in electric power grids. in Proceedings of the 2009 ACM Conference on Computer and Communications Security, CCS 2009, Chicago, Illinois, USA. pp. 21-32 (2009) https://doi.org/10.1145/1653662.1653666

[4] Garg, S.: Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN. IEEE Transactions Multimedia. pp. 566-578 (2019) https://doi.org/10.1109/TMM.2019.2893549

[5] Wang, W.: HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. IEEE Access. pp. 1792-1806 (2018) https://doi.org/10.1109/ACCESS.2017.2780250

[6] Ring, M.: A survey of network-based intrusion detection data sets. Computers & Security. pp. 147-167 (2019) https://doi.org/10.1016/j.cose.2019.06.005

[7] Moustafa, N.: UNSW-NB15: A comprehensive data set for network intrusion detection systems. in MILCOM 2015, IEEE. pp. 1-6 (2015) https://doi.org/10.1109/MilCIS.2015.7348942

[8] Tang, T. A.: Deep learning approach for network intrusion detection in software defined networking. in Proc. IEEE ICCCN (2016) https://doi.org/10.1109/WINCOM.2016.7777224

[9] Xu, J.: Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. in Proc. WWW (2018) https://doi.org/10.1145/3178876.3185996

[10] Li, H.: Anomaly detection in wireless sensor networks: A survey. Journal of Network and Computer Applications. pp. 543-562 (2013) https://doi.org/10.1016/j.jnca.2011.03.004

[11] Kontagora, M. M.: Machine learning algorithms for anomaly detection in iot networks - A review. African Multidisciplinary Journal of Sciences and Artificial Intelligence. pp. 802-823 (2024) https://doi.org/10.58578/amjsai.v1i2.4014

[12] Moustafa, N.: A holistic review of network anomaly detection systems: A comprehensive survey. Journal of Network and Computer Applications. pp. 33-55 (2019) https://doi.org/10.1016/j.jnca.2018.12.006

[13] Bezdek, J. C.: VAT: A tool for visual assessment of (cluster) tendency. in IJCNN. pp. 2225-2230 (2002) https://doi.org/10.1109/IJCNN.2002.1007487

[14] Davies, D. L.: A cluster separation measure. IEEE Trans. Pattern Analysis and Machine Intelligence. pp. 224-227 (1979) https://doi.org/10.1109/TPAMI.1979.4766909

[15] Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. pp. 53-65 (1987) https://doi.org/10.1016/0377-0427(87)90125-7

[16] Calinski, T.: A dendrite method for cluster analysis. Communications in Statistics. pp. 1-27 (1974) https://doi.org/10.1080/03610927408827101

[17] Mehdi, S. A.: Clustering model for the first line of defense in IDS for IoT. African Journal of Biomedical Research. pp. 12228-12235 (2024) https://doi.org/10.53555/AJBR.v27i4S.6064

[18] Xie, J.: Unsupervised deep embedding for clustering analysis. in Proc. International Conference on Machine Learning. pp. 478-487 (2016) https://proceedings.mlr.press/v48/xieb16.html

[19] Kingma, D. P.: Auto-encoding variational bayes. in Proc. 2nd International Conference on Learning Representations (ICLR) (2014) https://doi.org/10.48550/arXiv.1312.6114
[20] Guo, X.: Deep embedded clustering with data augmentation. in Proc. 10th Asian Conference on Machine Learning (ACML). pp. 550-565 (2018) https://proceedings.mlr.press/v95/guo18b.html
[21] Bhosale, N.: Quantum-assisted variational deep learning for efficient anomaly detection in secure cyber-physical system infrastructures. International Journal of Advanced Computer Science and Applications (2025) https://doi.org/10.14569/IJACSA.2025.0160559.
[22] Tarekegn, A. N.: A new cluster validation index based on stability analysis. in Proc. ICPRAM (2025). https://doi.org/10.5220/0013309100003905
[23] Althobaiti, M. M.: Intelligent intrusion detection for IoT and cyber-physical systems using machine learning. International Journal of Advanced and Applied Sciences. pp. 92-105 (2025) https://doi.org/10.21833/ijaas.2025.06.009
[24] Hodge, V. J.: A survey of outlier detection methodologies. Artificial Intelligence Review. pp. 85-126 (2004) https://link.springer.com/article/10.1023/B:AIRE.0000045502.10941.a9
[25] Zhukabayeva, T.: An edge-computing-based integrated framework for network traffic analysis and intrusion detection to enhance cyber-physical system security in industrial IoT. Sensors. pp. 1-24 (2025) https://doi.org/10.3390/s25082395
[26] Zhang, X.: AOC-IDS: Autonomous online framework with contrastive learning for intrusion detection. In: Proceedings of IEEE INFOCOM 2024 (2024) https://doi.org/10.1109/INFOCOM52122.2024.10621346
[27] Halkidi, M.: On clustering validation techniques. Journal of Intelligent Information Systems. pp. 107-145 (2001) https://doi.org/10.1023/A:1012801612483
[28] He, K.: Adversarial machine learning for network intrusion detection systems: A comprehensive survey. IEEE Communications Surveys & Tutorials. pp. 538-566 (2023) https://doi.org/10.1109/COMST.2022.3233793
[29] Dunn, J. C.: Well-Separated clusters and optimal fuzzy partitions. Journal of Cybernetics. pp. 95-104 (1974) https://doi.org/10.1080/01969727408546059
[30] Singh, A.: Adversarial machine learning in cybersecurity. International Journal of Innovative Research in Technology. pp. 3347-3354 (2024) https://ijirt.org/article?manuscript=169990
[31] Xie, X. L.: A validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence. pp. 841-847 (1991) https://doi.org/10.1109/34.85677
[32] Pakhira, M. K.: A validity index for crisp and fuzzy clusters. Pattern Recognition. pp. 487-501 (2004) https://doi.org/10.1016/j.patcog.2003.06.005
[33] Ben-Hur, A.: A stability based method for discovering structure in clustered data. Pacific Symposium on Biocomputing (2002) https://doi.org/10.1142/9789812799623_0002
[34] Tibshirani, R.: Estimating the number of clusters in a dataset via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology). pp. 411-423 (2001) https://doi.org/10.1111/1467-9868.00293
[35] Krzanowski, W. J.: A criterion for determining the number of groups in a dataset using sum-of-squares clustering. Biometrics. pp. 23-34 (1988) https://doi.org/10.2307/2531893
[36] Schwarz, G.: Estimating the dimension of a model (bic). The Annals of Statistics. pp. 461-464 (1978) https://doi.org/10.1214/aos/1176344136
[37] Bezdek, J. C.: Pattern recognition with fuzzy objective function algorithms. Springer (1981) https://doi.org/10.1007/978-1-4757-0450-1
[38] Subbalakshmi, C: A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set. Procedia Computer Science, pp. 346-353, Vol. 46, (2015). https://doi.org/10.1016/j.procs.2015.02.030

[39] Gusita, B.: Securing IoT edge: A survey on lightweight cryptography, anonymous routing and communication protocol enhancements. International Journal of Information Security (2025) https://doi.org/10.1007/s10207-025-01071-7
[40] Almutairi, M.: IoT-Cloud integration security: A survey of challenges, solutions, and directions. Electronics. pp. 1394-1415 (2025) https://doi.org/10.3390/electronics14071394
[41] Dauda, A.: A survey on IoT application architectures. Sensors. pp. 5320-5345 (2024) https://doi.org/10.3390/s24165320
[42] Fernandez-Carames, T. M.: From pre-quantum to post-quantum IoT security: A survey on quantum-resistant cryptosystems for the internet of things. IEEE Internet of Things Journal. pp. 6457-6480 (2020) https://doi.org/10.1109/JIOT.2019.2958788
[43] Li, K.: Zero-Trust foundation models: A new paradigm for secure and collaborative artificial intelligence for internet of things. IEEE Internet of Things Journal. pp. 1-21 (2025). https://doi.org/10.1109/JIOT.2025.3603957
[44] Chen, Y.: The internet of things: Secure distributed inference. IEEE Signal Processing Magazine. pp. 64-75 (2018) https://doi.org/10.1109/MSP.2018.2846059
[45] Jablaoui, R.: Deep learning enabled intrusion detection system for IoT security. EURASIP Journal on Wireless Communications and Networking (2025) https://doi.org/10.1186/s13638-025-02477-6
[46] Neto, E. C. P.: Deep learning for intrusion detection in emerging technologies: A comprehensive survey and new perspectives. Artificial Intelligence Review (2025). https://doi.org/10.1007/s10462-025-11346-z
[47] Kimanzi, R.: Deep learning algorithms used in intrusion detection systems – A review. arXiv preprint, arXiv:2402.17020. pp. 1-31 (2024) https://doi.org/10.48550/arXiv.2402.17020
[48] Awajan, A.: A novel deep learning-based intrusion detection system for IoT networks. Computers. pp. 1-17 (2023) https://doi.org/10.3390/computers12020034
[49] Wu, Y.: Current status, challenges and future trends of deep learning-based intrusion detection models. Journal of Imaging. pp. 1-25 (2024) https://doi.org/10.3390/jimaging10100254
[50] Zhang, Y.: A review of deep learning applications in intrusion detection systems: Overcoming challenges in spatiotemporal feature extraction and data imbalance. Applied Sciences. pp. 1-20 (2025) https://doi.org/10.3390/app15031552
[51] Gueriani, A.: Deep reinforcement learning for intrusion detection in IoT: A survey. in 2023 2nd International Conference on Electronics, Energy and Measurement (IC2EM). pp. 1-7 (2023) https://doi.org/10.1109/IC2EM59347.2023.10419560
[52] Yang, W.: A survey for deep reinforcement learning based network intrusion detection. arXiv preprint, arXiv:2410.07612 (2024) https://doi.org/10.48550/arXiv.2410.07612
[53] Maynard, P.: Consistent and compatible modelling of cyber intrusions and incident response demonstrated in the context of malware attacks on critical infrastructure. arXiv preprint, arXiv:2505.16398 (2025). https://doi.org/10.48550/arXiv.2505.16398
[54] Huang, S.: Bayesian and multi-objective decision support for real-time cyber-physical incident mitigation. arXiv preprint, arXiv:2509.00770 (2025) https://doi.org/10.48550/arXiv.2509.00770
[55] Huang, S.: Security modelling for cyber-physical systems: A systematic literature review. arXiv preprint, arXiv:2404.07527 (2024) https://doi.org/10.48550/arXiv.2404.07527.
[56] Raval, K. J.: A survey on safeguarding critical infrastructures: Attacks, AI security, and future directions. International Journal of Critical Infrastructure Protection (2023) https://doi.org/10.1016/j.ijcip.2023.100647
[57] Yigit, Y.: Generative AI and LLMs for critical infrastructure protection: Evaluation benchmarks, agentic AI, challenges, and opportunities. Sensors. pp. 1-40 (2025) https://doi.org/10.3390/s25061666

[58] Koukoulis, I.: Self-supervised transformer-based contrastive learning for intrusion detection systems. In: Proceedings of IFIP Networking 2025. pp. 1-7 (2025) https://doi.org/10.23919/IFIPNetworking64979.2025.11076319
[59] Li, E.: SAFE: Self-supervised anomaly detection framework for intrusion detection. arXiv preprint, (2025) https://doi.org/10.48550/arXiv.2502.07119
[60] Zeng, Y.: CSAGC-IDS: A dual-module deep learning network intrusion detection model for complex and imbalanced data. arXiv preprint, (2025) https://doi.org/10.48550/arXiv.2505.14027
[61] Narmadha, S.: Improved network anomaly detection system using optimized autoencoder–LSTM. Expert Systems with Applications (2025) https://doi.org/10.1016/j.eswa.2025.126854
[62] Soltani, M.: A multi-agent adaptive deep learning framework for online intrusion detection. Cybersecurity. pp. 1-25 (2024) https://doi.org/10.1186/s42400-023-00199-0
[63] Ruffo, S.: Anomaly and intrusion detection using deep learning for software-defined networks: A survey. Expert Systems with Applications (2024). https://doi.org/10.1016/j.eswa.2024.124982
[64] Bohli, J. M.: SMARTIE Project: Secure IoT data management for smart cities. in Proc. 2015 Int. Conf. Recent Advances in Internet of Things (RIoT), Singapore. pp. 1-6 (2015) https://doi.org/10.1109/RIOT.2015.7104906
[65] Stolfo, S.: Cost-based modeling for fraud and intrusion detection: Results from the JAM project. Proc. DARPA Information Survivability Conf. and Exposition. pp. 130-144 (2000) https://doi.org/10.1109/DISCEX.2000.821515
[66] Tavallaee, M.: A detailed analysis of the KDD CUP 99 data set. Proc. IEEE Symp. Computational Intelligence for Security and Defense Applications. pp. 1-6 (2009) https://doi.org/10.1109/CISDA.2009.5356528
[67] Sattar, S.: Anomaly detection in encrypted network traffic using self-supervised learning. Scientific Reports. pp. 1-15 (2025). https://doi.org/10.1038/s41598-025-08568-0
[68] Sharafaldin, I.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proc. Int. Conf. Information Systems Security and Privacy (ICISSP). pp. 108-116 (2018) https://doi.org/10.5220/0006639801080116
[69] Sharafaldin, I.: A detailed analysis of the CICIDS2017 data set. Proc. Int. Conf. Information Systems Security and Privacy (ICISSP), In: Information Systems Security and Privacy (ICISSP 2018), Revised Selected Papers, Communications in Computer and Information Science. pp. 172-188 (2019) https://doi.org/10.1007/978-3-030-25109-3_9
[70] Koroniotis, N.: Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics. Future Generation Computer Systems. pp. 779-796 (2019) https://doi.org/10.1016/j.future.2019.05.041
[71] Moustafa, N.: An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things. IEEE Internet of Things Journal. pp. 4815-4830 (2019) https://doi.org/10.1109/JIOT.2018.2871719
[72] Ferrag, M. A.: Edge-IIoT set: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for Centralized and Federated Learning. IEEE Access. pp. 40281-40306 (2022) https://doi.org/10.1109/ACCESS.2022.3165809
[73] Presekal, A.: Intrusion detection system for digital substations using semi-supervised learning and traffic distance similarity clustering. IEEE Transactions on Smart Grid. pp. 576-589 (2025) https://doi.org/10.1109/TSG.2025.3611345
[74] Gu, H.: DEIDS: A novel intrusion detection system for industrial control systems. Neural Computing and Applications. pp. 9793-9811 (2022) https://doi.org/10.1007/s00521-022-06965-4

[75] Gopi, R.: Intelligent intrusion detection system for industrial internet of things environment. Computer Systems Science and Engineering. pp. 1567-1582 (2023) https://doi.org/10.32604/csse.2023.025216
[76] Aldhaheri, A.: Deep learning for cyber threat detection in IoT networks: A review. Internet of Things and Cyber-Physical Systems. pp. 110-128 (2024) https://doi.org/10.1016/j.iotcps.2023.09.003
[77] Long, Z.: A transformer-based network intrusion detection approach for cloud security. Journal of Cloud Computing (2024) https://doi.org/10.1186/s13677-023-00574-9
[78] Xi, C.: A novel multi-scale network intrusion detection model with transformer. Scientific Reports (2024) https://doi.org/10.1038/s41598-024-74214-w
[79] Bilot, T.: Graph neural networks for intrusion detection: A survey. IEEE Access. pp. 49114-49139 (2023) https://doi.org/10.1109/ACCESS.2023.3275789
[80] Guo, H.: EGNN: Energy-efficient anomaly detection for IoT multivariate time series data using graph neural network. Future Generation Computer Systems. pp. 45-56 (2024) https://doi.org/10.1016/j.future.2023.09.028