CXG-DT: Confidence- and Explanation-Gated Digital Twin Response for IoT Intrusion Handling

1st Nauman Irshad Ali Shah
Department of Computer Science
University of Central Punjab
Lahore, Pakistan
naumanirshadalishah@gmail.com

2nd Ahmad Arsalan
Faculty of Information Technology and Computer Science
University of Central Punjab
Lahore, Pakistan
ahmad.arslan@ucp.edu.pk

3rd Muhammad Umer Amir
Department of Computer Science
University of Central Punjab
Lahore, Pakistan
umrramrr@gmail.com

4th Ali Ahmad
Department of Computer Science
University of Central Punjab
Lahore, Pakistan
aliahmaddev61@gmail.com

5th Danish Ali
Department of Computer Science
University of Central Punjab
Lahore, Pakistan
danish.ali02721@gmail.com

6th Syed Qarib Ali Naqvi
Department of Computer Science
University of Central Punjab
Lahore, Pakistan
Contactqaribnaqvi@gmail.com

***Abstract*—Internet of Things (IoT) intrusion detectors often report high accuracy yet still fail at the operational response stage, where uncertain predictions can cause false isolation or delayed mitigation. This paper presents CXG-DT, a confidence- and explanation-gated Digital Twin (DT) response framework for IoT intrusion handling. The pipeline detects attacks with lightweight models, estimates calibrated confidence, computes selective SHapley Additive exPlanations (SHAP) only when a risk gate opens, validates alerts against virtual device twin states, and permits monitor, throttle, isolate, or escalate actions only when an uncertainty score is acceptable. On a locked multiclass IoT evaluation (seed 42; 1800 test flows; 8 classes), CatBoost attains the best detection trade-off with 81.94% accuracy, 82.01% weighted F1-score, 98.18% receiver operating characteristic area under the curve (ROC-AUC), and 0.0037 ms/sample. Selective SHAP covers 57.2% of alerts with explanation stability S=0.998. The gated policy yields isolate/throttle/monitor rates of 18.67%/67.56%/13.78% with missed-escalation rate 0.000 and mean uncertainty 0.106. CXG-DT therefore shifts IoT intrusion detection system (IDS) evaluation from accuracy alone toward safe closed-loop response under uncertainty.**

***Index Terms*—IoT intrusion detection, Digital Twin, explainable AI, SHAP, edge AI, uncertainty gating, response policy, CatBoost**

I. Introduction

Internet of Things (IoT) networks face denial-of-service (DoS), distributed denial-of-service (DDoS), reconnaissance, spoofing, brute-force, web, and malware campaigns while operating under tight compute and latency budgets [1], [2]. Machine-learning intrusion detection systems (IDS) improve detection accuracy, but many studies stop at Accuracy and F1 and leave unresolved the operational question of *when* an automated action should be taken [3]. Over-confident false alarms can isolate healthy devices; under-confident true attacks can propagate through a topology and degrade service [4], [11].

Digital Twin (DT) layers can mirror device risk, confidence, and recommended actions in near real time [5]. Explainable artificial intelligence (XAI), especially SHAP, can attach feature-level evidence to alerts and support analyst trust [6], [7]. However, always-on SHAP is expensive on constrained edge gateways, and twin residuals alone do not define a safe action policy. Recent surveys on IoT security learning and deep cyber-defense emphasize both detection quality and deployability constraints [3], [4], [12].

CXG-DT addresses this gap through a closed-loop workflow that detects an intrusion, estimates prediction confidence, selectively generates a SHAP explanation, validates the evidence against the Digital Twin state, and finally selects an uncertainty-gated response action. Our contributions are: (i) a five-layer gated response architecture that couples detection with safe action; (ii) an uncertainty score combining confidence C, explanation stability S, and twin residual R; (iii) a comparative evaluation of six detectors with edge latency; and (iv) operational metrics beyond accuracy, including action mix and escalation safety under a locked seed-42 protocol.

II. Related Work and Gap

Lightweight and explainable IDS methods optimize model size and SHAP/Local Interpretable Model-agnostic Explanations (LIME) transparency for edge or near-edge deployment [7], [8], [13]. Tree ensembles such as Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and CatBoost remain competitive for tabular IoT flow features because they offer strong multiclass discrimination with practical inference cost [14]–[16]. Neural baselines are often faster at inference but can lose accuracy on heterogeneous IoT attack taxonomies [12].

Digital Twin security studies focus on residuals, trust visualization, or federated twin coordination for cyber-physical and industrial IoT systems [5], [9], [17]. Autonomous response work emphasizes policy learning and automated mitigation, yet often omits twin-backed interlocks that block aggressive actions under high uncertainty [10], [18]. Dataset and benchmark efforts such as CICIoT2023 provide large-scale IoT attack coverage for reproducible evaluation [1].

The recurring gap is joint treatment of detection quality, selective explanation cost, twin consistency, and wrong-escalation risk under one reproducible protocol. Table I summarizes how CXG-DT differs from common practice. Evaluating these

TABLE I
CONCEPTUAL COMPARISON OF IDS/DT PRACTICE AND CXG-DT (NON-EMPIRICAL; N NOT APPLICABLE).

Dimension	Common practice	CXG-DT
Endpoint	Detection score	Gated response action
XAI use	Always / never	Selective SHAP gate
Twin role	Monitor / residual	Safety interlock input
Uncertainty	Often omitted	$U{=}f(C,S,R)$
Metrics	Acc/F1 only	+ action safety/latency

elements independently can hide important operational dependencies: a detector may achieve strong predictive performance while still producing uncertain alerts, an explanation may be informative but too costly to compute for every flow, and a Digital Twin residual may identify inconsistency without specifying a safe response. A unified evaluation is therefore needed to determine how confidence, explanation evidence, and device context jointly affect containment decisions. CXG-DT addresses this requirement by connecting the detector output to an explicit uncertainty interlock and by reporting response behavior alongside conventional detection metrics.

III. PROBLEM STATEMENT

Given flow features x, a detector returns class $\hat{y}$ and probability vector p. Let $C = \max_c p_c$ be confidence, A attack severity, S explanation stability, and R twin mismatch residual. The system must choose

$$a \in \{\text{monitor}, \text{throttle}, \text{isolate}, \text{escalate}\}$$

while minimizing missed severe attacks and wrong aggressive actions. Formal objective: characterize detection quality (Accuracy, weighted F1, ROC-AUC, latency) and enforce a hard interlock that forbids isolate when uncertainty $U > \tau$. The design therefore evaluates not only whether an attack is detected, but whether the recommended response remains safe under imperfect confidence and twin state.

a) Threat model.: The adversary can generate or replay malicious network flows representing DoS, DDoS, reconnaissance, spoofing, brute-force, web, and Mirai-style campaigns against monitored IoT devices. The attacker may vary traffic rate, timing, protocol use, and connection flags to evade the detector or provoke an unsafe automated response. The defender observes flow-level features at the gateway and maintains an authenticated Digital Twin state for each device. Compromise of the gateway, model parameters, twin service, orchestration channel, and training pipeline is outside the present scope; the evaluation therefore addresses evasion and response-safety risk at the traffic-classification layer rather than poisoning, model extraction, or control-plane takeover.

IV. PROPOSED METHOD: CXG-DT

A. Architecture

CXG-DT predicts an attack and confidence C, selectively obtains SHAP stability S, checks twin residual R, and selects a response. Isolation is permitted only for severe alerts with $U = 0.45(1-C) + 0.25(1-S) + 0.30R \le 0.45$; otherwise, the policy monitors, throttles, or escalates.

Algorithm 1 CXG-DT detector training and online gated response

Require: Training set D_{tr}, flow stream $\mathcal{X}$, device twins $\mathcal{D}$, weights $w = (0.45, 0.25, 0.30)$, threshold $\tau = 0.45$
Train detector f on D_{tr}
for all incoming flow–device pairs $(x, d) \in \mathcal{X}$ **do**
  $(\hat{y}, p) \leftarrow f(x)$; $C \leftarrow \max_c p_c$; $A \leftarrow g(\hat{y})$
  Obtain S from the selective explanation gate and R from twin d
  $U \leftarrow 0.45(1-C) + 0.25(1-S) + 0.30R$
  if A is severe $\wedge$ $U \le \tau$ **then**
    $a \leftarrow \text{isolate}$
  else if C, S, and R provide conflicting evidence **then**
    $a \leftarrow \text{escalate}$
  else if A is medium or severe **then**
    $a \leftarrow \text{throttle}$
  else
    $a \leftarrow \text{monitor}$
  end if
  Execute or recommend a; update twin d with $(\hat{y}, C, R, U, a)$
end for

B. Detectors

Random Forest, XGBoost, LightGBM, CatBoost, multilayer perceptron (MLP), and Soft-Voting use the same 44-dimensional input and seed-42 split. Soft-Voting averages the component outputs.

C. Confidence and selective SHAP

Confidence uses calibrated probabilities, with margin and entropy supporting uncertainty diagnosis. SHAP is computed only for borderline or high-severity alerts, reducing explanation cost [6], [7].

D. Twin state and residual

Each of 20 virtual devices stores risk, recent prediction evidence, residual, and action. Residual R rises when a confident severe alert conflicts with a healthy twin state [5], [9]. The lightweight state update costs $O(1)$ per alert and $O(20)$ total storage.

E. Uncertainty gate and policy

$$U = w_1(1-C) + w_2(1-S) + w_3 R, \qquad w_1 + w_2 + w_3 = 1. \quad (1)$$

The locked run uses $(w_1, w_2, w_3) = (0.45, 0.25, 0.30)$ and $\tau = 0.45$. Severe, low-U alerts may be isolated; medium risk is throttled, conflicts are escalated, and low severity is monitored. The terms capture low confidence, unstable explanations, and twin disagreement, while $U > \tau$ blocks isolation.

F. CXG-DT Training and Online Response

Algorithm 1 presents the CXG-DT procedure in the same compact pattern as the detector, explanation, and twin components described above. Training is performed once, after which every incoming flow is evaluated by the confidence, selective-SHAP, twin-residual, uncertainty, and response stages.

The workflow is evaluated using a locked 8-class CICIoT-style feature space with 44 numeric features, balanced generation, a stratified 75/25 split (5400/1800), and seed 42 [1]. The classes are Benign, DDoS, DoS, Recon, Spoofing, BruteForce,

Web, and Mirai. Reported measures include accuracy, precision, recall, weighted F1, ROC-AUC, inference time, selective-SHAP coverage, action rates, wrong and missed escalation, and twin-risk summaries. All values in Section V come from this locked export.

CatBoost produces 1475 correct predictions from 1800 test flows, giving 81.94% accuracy and a 95% Wilson confidence interval of 80.10%–83.65%. Model differences are interpreted descriptively because only one split is retained and paired predictions or repeated-seed measurements are unavailable. Statistical model comparison therefore requires future paired resampling or repeated stratified runs.

V. Results

A. Detection leaderboard

Table II and Fig. 1 show CatBoost as the best F1 model (82.01%) with strong tree latency among top models (0.0037 ms/sample). MLP is fastest (0.0016 ms) but weakest (71.47% F1). Soft-Voting is competitive but slower than CatBoost.

The weighted F1 values in Table II jointly reflect precision and recall across the eight classes while accounting for class support. CatBoost achieves the highest value at 82.01%, followed closely by Random Forest at 81.52%; the difference is only 0.49 percentage points. XGBoost (80.50%), Soft-Voting (80.35%), and LightGBM (79.94%) form a second, tightly grouped tier. MLP falls to 71.47%, approximately 10.5 points below CatBoost, indicating that its very low latency does not compensate for its weaker multiclass discrimination. CatBoost is therefore selected for the confidence, SHAP, twin-validation, and action-gating stages.

The latency column in Table II reports average inference time per test sample; a smaller value indicates lower detector-side delay. MLP is fastest at 0.0016 ms/sample but has the weakest F1. CatBoost is only slightly slower at 0.0037 ms/sample while delivering the best detection score; in this run it is roughly eleven times faster than XGBoost and more than thirty times faster than Random Forest. Soft-Voting inherits the cost of evaluating multiple members, while LightGBM is the slowest measured detector. The comparison therefore identifies CatBoost as the strongest accuracy–latency compromise. These values cover model inference only, not feature extraction, SHAP, twin updates, communication, or action execution.

Figure 1 presents the CatBoost confusion matrix, where rows are true classes and columns are predictions; diagonal cells are correct decisions and off-diagonal cells are errors. The diagonal contains 1475 correct predictions out of 1800, matching 81.94% accuracy. Recon is the clearest class, with 218 of 225 flows correctly identified, whereas Web is the most difficult, with 156 correct and leakage into Benign, BruteForce, and Spoofing. The matrix also shows DDoS and DoS confusion with Mirai, while some Mirai samples are labeled DDoS or DoS. These structured errors demonstrate why a predicted class should not directly trigger isolation: confidence, severity, SHAP evidence, and twin consistency are required before an aggressive response.

TABLE II
Model comparison for seed 42, 8 classes, 44 features, N_{train}=5400, and N_{test}=1800.

Model	Acc (%)	F1 (%)	AUC (%)	ms/sample
CatBoost	**81.94**	**82.01**	**98.18**	0.0037
RandomForest	81.33	81.52	98.02	0.1169
XGBoost	80.39	80.50	98.10	0.0405
SoftVoting	80.22	80.35	98.12	0.1047
LightGBM	79.78	79.94	97.88	0.1214
MLP	71.50	71.47	96.25	**0.0016**

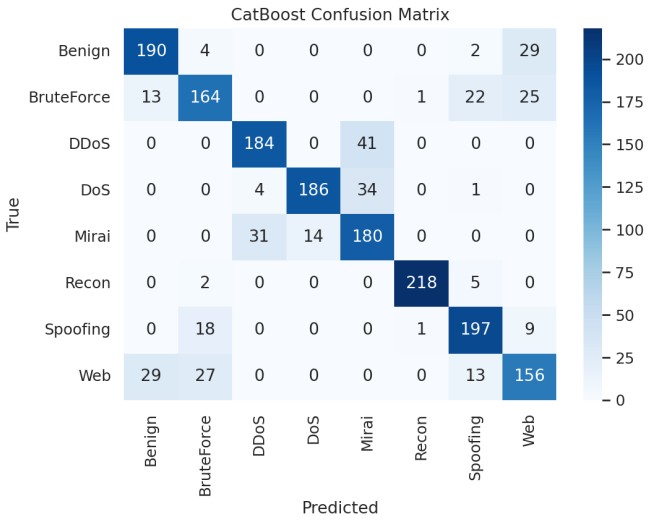

Fig. 1. CatBoost confusion matrix for 8 classes and $N_{\text{test}} = 1800$ (accuracy 81.94%).

B. Confidence behavior

Correct predictions show higher mean confidence (0.847) than incorrect ones (0.625), supporting the use of C inside the gate rather than raw labels alone.

Correct decisions have mean confidence 0.847, whereas incorrect decisions have mean 0.625. This 0.222 gap indicates that confidence is informative about reliability, although highly confident mistakes and uncertain correct predictions can still occur. A confidence-only threshold may therefore permit over-confident errors or suppress valid detections. CXG-DT instead combines C with explanation stability S and twin residual R.

C. Selective SHAP

SHAP was triggered for 57.2% of test alerts (46.26 ms/sample on gated subset) with stability S=0.998. Dominant features were inter-arrival time (IAT), Tot sum, syn_flag_number, Rate, Srate, Drate, Hypertext Transfer Protocol Secure (HTTPS), and User Datagram Protocol (UDP) (Fig. 2).

As shown in Fig. 2, each horizontal bar is the mean absolute SHAP magnitude over selectively explained alerts; longer bars indicate greater average influence on model output. IAT is the dominant feature, showing the importance of packet inter-arrival behavior. Tot sum and syn_flag_number form the next tier, linking predictions to traffic volume and Transmission Control Protocol (TCP) connection initiation. Rate, Srate, and Drate

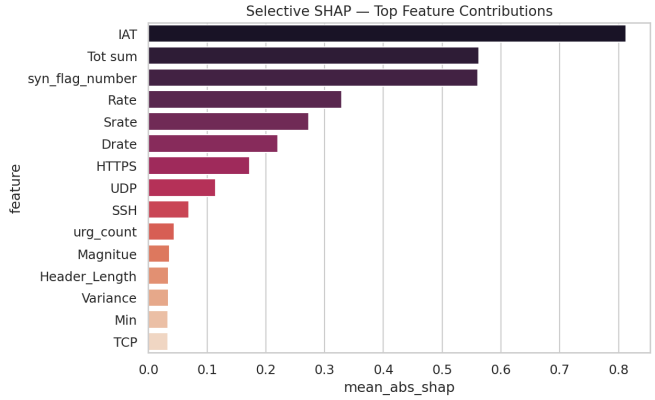

Fig. 2. Mean absolute SHAP contributions for the 57.2% selectively explained alerts ($S = 0.998$).

capture overall and directional intensity, while HTTPS and UDP represent protocol behavior. Absolute SHAP values measure influence strength, not whether a feature raises or lowers a specific class probability, and they do not prove causality. SHAP covers 57.2% of alerts, while stability S=0.998 indicates a consistent ranking in the locked evaluation.

D. Digital Twin states

All 20 virtual devices reached critical risk during the streamed replay.

Every twin finishes in the critical category, with no devices remaining at low, medium, or high risk. This shows that the twin accumulates repeated attack evidence instead of treating alerts as independent events. Sustained malicious activity therefore raises the operational priority of an affected device even when individual predictions vary in confidence. The fully saturated distribution also reveals a limitation: this replay does not demonstrate risk decay or recovery. Future evaluation should include benign intervals, device-specific behavior, and decay rules to confirm that a twin can leave the critical state after malicious activity stops.

E. Gated actions and safety

Action counts: throttle 1216, isolate 336, monitor 248 (Fig. 3). Mean confidence is highest for isolate (0.962) while mean U is lowest for isolate (0.045) (Fig. 4). Locked safety metrics are wrong-escalation 0.0161, missed-escalation 0.000, isolate rate 18.67%, throttle 67.56%, monitor 13.78%, detection accuracy 81.0%, and mean U=0.106.

Figure 3 converts all 1800 detector outputs into response decisions. Throttle is selected for 1216 alerts (67.56%), reflecting a preference for reversible rate limitation when evidence is meaningful but not strong enough for isolation. Isolation is applied to 336 alerts (18.67%), representing cases with high severity, strong confidence, twin agreement, and acceptable uncertainty. The remaining 248 alerts (13.78%) are monitored because their risk or evidence is too weak for active containment. No alert enters the analyst-escalation branch in this replay. Thus, CXG-DT does not equate every detected

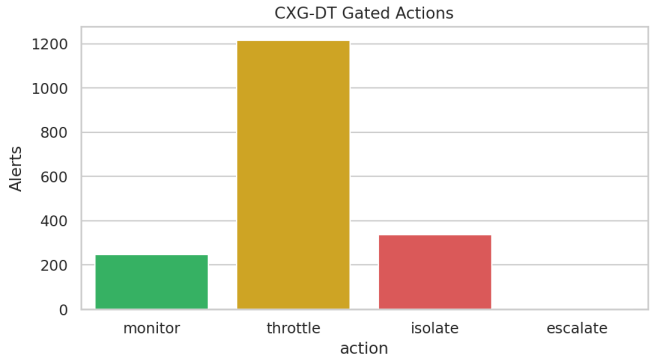

Fig. 3. Gated actions over 1800 alerts: 1216 throttle, 336 isolate, and 248 monitor.

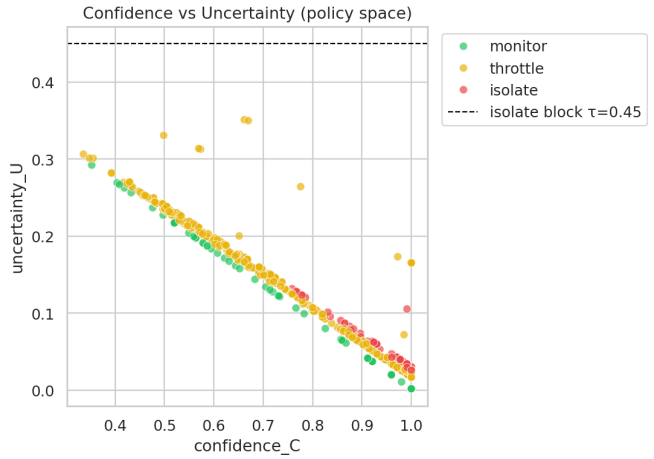

Fig. 4. Policy space with isolate mean $(C, U) = (0.962, 0.045)$ and interlock $\tau = 0.45$.

attack with isolation and avoids the most disruptive action for more than four-fifths of alerts.

Figure 4 positions each test alert by confidence C and combined uncertainty U, with color indicating the selected action. The downward trend is expected because $(1 - C)$ contributes directly to Eq. (1). Isolation points concentrate at high confidence and low uncertainty, with reported means $C = 0.962$ and $U = 0.045$. Monitor and throttle points overlap because response also depends on attack severity and twin evidence, not only the two displayed variables. Higher-U throttle points represent cases where residual uncertainty makes isolation unsafe. The dashed line at $U = \tau = 0.45$ is a hard interlock: any alert above it is forbidden from entering the isolate branch.

Wrong escalation is defined as applying aggressive containment to benign or low-severity ground truth. The confidence-only baseline records approximately 0.0100, while CXG-DT records 0.0161 (1.61%). The proposed policy is therefore higher by about 0.0061, or 0.61 percentage points, on this metric in the locked run. This must be interpreted together with the missed-escalation rate of 0.000: the current gate avoids missing severe cases but accepts more unnecessary aggressive actions. The comparison therefore shows a safety trade-off,

not an unconditional improvement. It motivates cost-sensitive calibration of the severity mapping, weights w_1, w_2, w_3, and threshold τ.

F. Available policy ablation

The locked export supports the partial ablation reported in Table III: confidence-only thresholding is compared with the full CXG-DT policy. The full policy records a wrong-escalation rate of 0.0161 versus 0.0100 for confidence only, an increase of 0.61 percentage points, while its missed-escalation rate is 0.000. Thus, the available comparison does not show an improvement in wrong escalation and is reported as a safety trade-off. Variants that independently remove SHAP stability or the Digital Twin residual were not retained in the export, so their separate contributions cannot be quantified without rerunning the experiment.

TABLE III
PARTIAL POLICY ABLATION FOR SEED 42 AND $N_{\text{test}} = 1800$.

Policy	Evidence used	Wrong esc.
Confidence only	Confidence C	0.0100
Full CXG-DT	C, S, R, severity A	0.0161

VI. Discussion

CatBoost offers the best accuracy–latency compromise for the detection stage among the six compared models. Selective SHAP keeps explanation cost off the critical path for easy benign decisions, which is essential for edge gateways with limited CPU budgets [7], [13]. The twin residual and uncertainty gate convert predictions into auditable actions rather than raw class labels [9].

In this locked run, missed escalation is zero, while wrong-escalation remains non-zero (0.0161), indicating that threshold and severity weights still need calibration on real traffic. Synthetic class-conditional generation enables paired reproducibility but is not a substitute for CICIoT2023 raw traces or live gateway telemetry [1]. Overall, CXG-DT reframes IDS reporting around safe closed-loop response under uncertainty.

VII. Limitations and Future Work

Limitations include synthetic data generation, global rather than per-alert stability in the locked export, and lack of hardware energy measurements. Future work will validate on full CICIoT2023 [1], learn (w_i, τ) from cost-sensitive objectives, add neighbor-aware propagation scoring in the twin graph [17], and deploy the gate on constrained edge hardware with measured joules and end-to-end action latency [18].

VIII. Conclusion

CXG-DT couples lightweight IoT intrusion detection with selective SHAP, Digital Twin validation, and an uncertainty-gated response policy. Locked results identify CatBoost as the strongest detector (81.94% Acc / 82.01% F1 / 0.0037 ms) and demonstrate an operational closed loop with transparent action decisions. The framework reframes IoT IDS evaluation around safe response under uncertainty rather than accuracy alone.

Acknowledgment

The authors would like to thank the Center for Emerging Networks & Technologies (CENT) Lab at the University of Central Punjab (UCP), Lahore, Pakistan, for technical support and constructive feedback.

Code Availability

The complete implementation of CXG-DT, including the locked experimental notebook, pretrained models, and evaluation scripts, will be made publicly available upon publication at: https://github.com/Nauman-Irshad/CXG_DT_IoT_Response_Implementation

References

[1] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, Art. no. 5941, 2023. Available: https://doi.org/10.3390/s23135941

[2] N. Chaabouni, M. Mosbah, A. Zemmari, C. Sauvignac, and P. Faruki, "Network intrusion detection for IoT security based on learning techniques," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2671–2701, 2019. Available: https://doi.org/10.1109/COMST.2019.2896380

[3] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Sayed, and M. Guizani, "A survey of machine and deep learning methods for Internet of Things (IoT) security," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1646–1685, 2020. Available: https://doi.org/10.1109/COMST.2020.2988293

[4] D. S. Berman, A. L. Buczak, J. S. Chavis, and C. L. Corbett, "A survey of deep learning methods for cyber security," *Information*, vol. 10, no. 4, Art. no. 122, 2019. Available: https://doi.org/10.3390/info10040122

[5] M. Eckhart and A. Ekelhart, "Digital twins for cyber-physical systems security: State of the art and outlook," in *Security and Quality in Cyber-Physical Systems Engineering*. Springer, 2019, pp. 383–412. Available: https://doi.org/10.1007/978-3-030-25312-7_14

[6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. Available: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[7] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022. Available: https://doi.org/10.1109/ACCESS.2022.3165809

[8] M. A. Yagiz and P. Goktas, "LENS-XAI: Layered explainable security for IoT intrusion detection," arXiv preprint arXiv:2501.00790, 2025. Available: https://arxiv.org/abs/2501.00790

[9] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21980–22012, 2020. Available: https://doi.org/10.1109/ACCESS.2020.2970143

[10] S. M. Sajjad, M. Yousaf, H. Afzal, and M. R. Mufti, "SIUV: A smart intrusion detection system for unmanned vehicles," *IEEE Access*, 2024. Available: https://doi.org/10.1109/ACCESS.2024.3361325

[11] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breitenbacher, and Y. Elovici, "N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12–22, 2018. Available: https://doi.org/10.1109/MPRV.2018.03367731

[12] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018. Available: https://doi.org/10.1109/TETCI.2017.2772792

[13] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *Journal of Information Security and Applications*, vol. 50, Art. no. 102419, 2020. Available: https://doi.org/10.1016/j.jisa.2019.102419

[14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD*, 2016, pp. 785–794. Available: https://doi.org/10.1145/2939672.2939785

[15] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. Available: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

[16] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018. Available: https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

[17] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019. Available: https://doi.org/10.1109/TII.2018.2873186

[18] I. Ahmad, S. Shahabuddin, T. Kumar, J. Okwuibe, A. Gurtov, and M. Ylianttila, "Security for 5G and beyond," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3682–3722, 2019. Available: https://doi.org/10.1109/COMST.2019.2916180