Babylonian Journal of Machine Learning
Vol.2026, **pp**. 351–383
DOI: https://doi.org/10.58496/BJML/2026/020; ISSN: 3006-5429
https://mesopotamian.press/journals/index.php/BJML

Research Article

Entropy-Adaptive Hybrid Artificial Intelligence with Dynamic Weighted Fu-sion for Intrusion Detection in Wireless Sensor Networks

Qasim Mustafa Zainel [1], Mohammed Fakhrulddin Abdulqader [2], Adnan Yousif Dawod [3], Mohammad Khaleel Sallam Ma'aitah [4, *], Naseer Ali Hussien [5], Amer Alqutaish [6]

[1] *College of Physical Education and Sport Science, University of Kirkuk, Kirkuk 36001, Iraq*

[2] *College of Computer Science and Information Technology, University of Kirkuk, Kirkuk 36001, Iraq.*

[3] *College of Education for Women, University of Kirkuk, Kirkuk 36001, Iraq.*

[4] *Electrical Engineering / Robotics and Artificial Intelligence Engineering, Faculty of Engineering and Technology, Applied Science Private University, Amman, Jordan.*

[5] *Department of Computer Engineering Technology, College of Technical Engineering, Al-Ayen Iraqi University (AUIQ), Nasiriyah, Iraq.*

[6] *Deanship of Development and Quality Assurance, King Faisal University, 31982, Al-Ahsa, Saudi Arabia.*

ARTICLE INFO

Article History

Received 25 Jul 2026
Revised 23 Sep 2026
Accepted 13 Nov 2026
Published 20 Nov 2026

Keywords

Wireless Sensor Networks
Intrusion Detection
Hybrid Artificial Intelligence
Dynamic Weighted Fusion
Deep Learning
Cybersecurity
Internet of Things

CC BY

ABSTRACT

Wireless Sensor Networks (WSNs) are widely used in safety-critical cyber-physical environments, including smart grids, industrial automation, healthcare monitoring, disaster management, and defence surveillance. However, their distributed operation, limited computational resources, and exposure to diverse traffic patterns make them highly susceptible to sophisticated cyberattacks. Existing intrusion detection approaches often employ static ensemble strategies with fixed classifier weights, which can reduce detection effectiveness when network conditions and attack characteristics change dynamically. This study proposes an entropy-adaptive Hybrid Artificial Intelligence with Dynamic Weighted Fusion (HAI-DWF) framework for real-time intrusion detection and attack classification in WSNs. The proposed architecture combines a convolutional neural network for spatial feature extraction, a bidirectional long short-term memory network with temporal self-attention for sequential traffic modelling, and LightGBM for efficient learning from structured network features. A dynamic weighted fusion mechanism is introduced to adapt classifier contributions according to Shannon entropy computed from incoming traffic, allowing the detection process to respond to variations in attack behaviour and traffic uncertainty. Because the contribution of the entropy signal itself has to be demonstrated rather than assumed, validation protocols for entropy ablation against learned-gate, stacking and alternative-entropy fusion, multi-seed repetition, leakage-free cross-dataset transfer, problem-space adversarial feasibility and end-to-end embedded measurement are specified in Sections 3.9–3.12. The framework is evaluated on four benchmark datasets, namely CICIoT-2023, Edge-IIoTset-2022, NF-ToN-IoT-v2, and WSN-DS. Experimental results show that HAI-DWF achieves 99.46% accuracy, a 99.27% F1-score, and a false positive rate of 0.31% on CICIoT-2023, while maintaining strong performance across the remaining datasets. Edge-oriented profiling on ARM Cortex-M55 and RISC-V SiFive U74 platforms indicates the inference-level computational feasibility of the proposed framework on cluster-head-class devices, with end-to-end measurements of feature extraction, communication, memory and energy specified in Section 3.12. Robustness analysis under FGSM, PGD, DeepFool, and C&W adversarial attacks shows a much smaller accuracy loss than the neural components alone, which is attributable mainly to the LightGBM component; these feature-space results are complemented by a problem-space feasibility protocol (Section 3.11). These results indicate that HAI-DWF is a promising adaptive and deployment-oriented intrusion detection approach, whose entropy-specific contribution remains subject to the validation experiments specified in this revision for WSN- and IoT-enabled cyber-physical systems.

*Corresponding author. Email: adnanalshef@uokirkuk.edu.iq

1. INTRODUCTION

Wireless Sensor Networks (WSNs) have undergone a structural transformation from specialised academic research platforms into pervasive, economically consequential cyber-physical infrastructure embedded across the full spectrum of critical modern applications. The global WSN and IoT node installed base surpassed 16.7 billion active endpoints by 2024, with the 2030 aggregate projected at 29.4 billion nodes representing USD 1.1 trillion in cumulative hardware and services revenue [1]. This expansion has been enabled by unit cost reductions below USD 5 for commodity mote-class sensor nodes, sustained improvements in ambient energy harvesting supporting multi-year batteryless operation, and the proliferation of IEEE 802.15.4-based, LoRaWAN, and LPWAN radio standards providing reliable low-power connectivity at ranges from 10 m to 15 km. The resulting deployments now underpin physical processes in smart-grid substation monitoring, autonomous industrial control, connected healthcare telemetry, precision agriculture, environmental disaster early warning, and national defence surveillance—domains in which security failures translate directly into measurable physical-world consequences [2].

Beyond the intrusion-detection-specific literature reviewed in Section 2, this expansion inherits the broader, longstanding challenges of WSN and IoT research more generally — constrained node energy and the resulting need for energy-aware routing and data-aggregation strategies [32], and the heterogeneous, resource-constrained security landscape spanning device, network, and application layers that complicates uniform defence deployment across a WSN–IoT deployment [33]. These general constraints motivate the tiered, lightweight deployment architecture (Section 3.3) adopted here rather than a uniform full-model rollout across all node classes.

The ENISA Threat Landscape 2023 report documents a 217% compound increase in network-layer attacks targeting IoT and WSN platforms between 2020 and 2023, with WSN-targeting incidents causing average economic damages of EUR 4.1 million per event in industrial control contexts and EUR 18.7 million when healthcare monitoring infrastructure is compromised [3]. Seventeen distinct attack categories spanning physical, MAC, network, and application protocol layers now routinely target production WSN deployments, as documented in Table 3. The defining characteristic of this threat environment is its adaptive, multi-vector nature: adversaries simultaneously exploit physical-layer radio vulnerabilities, network-layer LEACH routing protocol weaknesses, and application-layer MQTT and Modbus command injection to bypass detection systems trained exclusively on single-protocol traffic signatures [4].

Existing hybrid IDS frameworks combine CNN, LSTM and tree-based components with fixed or equal combination weights, so that the contribution of each component does not change with the statistical character of the traffic observed at inference time. This static fusion strategy is hypothesised to be suboptimal because different algorithmic components may exhibit asymmetric performance across the entropy regimes of WSN traffic. GBDT classifiers achieve optimal per-class F1-scores on low-entropy, highly regular sensor telemetry characteristic of LEACH routing attacks, where structured tabular decision boundaries efficiently separate attack from normal patterns. CNN encoders achieve optimal discriminability on high-entropy, spatially complex traffic characteristic of DDoS and application-layer exploitation, where spatial correlation structures encode attack patterns invisible to per-feature tabular analysis. BiLSTM networks are most effective for medium-entropy stateful multi-step sequences—Botnet C2 beaconing, DoS Slowloris connection management—where bidirectional temporal context spanning 10–30 seconds of protocol interaction is essential [5]. These regime-specific strengths are the working hypothesis that motivates DWF; they are not assumed in the evaluation, and Section 4.8 (Table 16) reports the per-regime performance of each component so that the hypothesis can be confirmed or rejected.

This paper addresses the static fusion limitation through the Dynamic Weighted Fusion (DWF) mechanism, a real-time entropy-adaptive layer that continuously estimates the Shannon entropy H(x) of incoming windows of 28-feature traffic vectors and adaptively rebalances CNN, BiLSTM, and GBDT combination weights across five entropy regimes using Bayesian-optimised boundaries and weights. The regime lookup and weighted summation of the DWF execute in under 0.1 ms and require approximately 52 floating-point operations per decision, which is negligible relative to the 61,092–494,240 FLOPs of the component classifiers; the additional cost of computing H(x) over the traffic window is itemised in the end-to-end measurements of Section 4.12. The complete HAI+DWF framework integrates: a CNN encoder processing the 28-feature HFR vector as a 4×7 spatial map through two convolutional layers producing five deep spatial features; a three-layer BiLSTM with scaled dot-product temporal self-attention modelling 20-timestep traffic sequences; an 800-tree LightGBM GBDT with 300-trial Bayesian-optimised hyperparameters; and the DWF fusing these three posterior vectors through entropy-adaptive weighted summation [6].

Crucially, the proposed framework is evaluated exclusively on post-2022 benchmark datasets that reflect the contemporary IoT threat landscape: CICIoT-2023 (19,240,847 instances, 16 attack classes), Edge-IIoTset-2022 (2,219,201 instances, 15 classes), NF-ToN-IoT-v2 2022 (3,048,212 instances, 9 classes), and WSN-DS (374,661 instances, 5 LEACH-specific classes). The choice of post-2022 datasets is deliberate: pre-2020 benchmarks such as CICIDS-2018 and NSL-KDD fail to represent contemporary attack tooling including MQTT-Flood, Modbus injection, modern Botnet C2 protocols, and IoT-

native DDoS patterns that constitute the dominant attack vectors in current WSN deployments [7]. The HAI+DWF achieves 99.46% accuracy, 99.27% F1-score, and 0.31% FPR on CICIoT-2023—the strongest results among the configurations evaluated under identical conditions in this study; cross-study comparisons are indicative only (Section 4.5)—with adversarial evaluation under C&W, DeepFool, FGSM, and PGD attacks giving 96.74% accuracy under the strongest of the four feature-space attacks (PGD) at ε=0.10, a figure whose validity as a realisable network attack is assessed in Section 4.11.

TABLE I. BENCHMARK DATASET CHARACTERISTICS: CICIOT-2023, EDGE-IIOTSET-2022, NF-TON-IOT-V2, AND WSN-DS

Property	CICIoT-2023	Edge-IIoTset-2022	NF-ToN-IoT-v2	WSN-DS
Total Records	19,240,847	2,219,201	3,048,212	374,661
Training Instances (80%)	15,392,677	1,775,360	2,438,569	299,728
Test Instances (20%)	3,848,170	443,841	609,643	74,933
Raw Feature Count	47	81	12	19
Attack Categories	16	15	9	5
Normal Records	6,814,704	1,012,417	1,874,611	217,703
Most Imbalanced Class	Slowloris (1:41,724)	MITM (1:6,841)	Scanning (1:4,217)	Blackhole (1:1.6)
Generation Environment	CIC virtual lab 2023	IoT testbed 42 devices	NetFlow v9 backbone	LEACH simulation
Protocol Coverage	Wi-Fi, Zigbee, MQTT, CoAP	Multi-IoT+ICS hybrid	NetFlow/IPFIX	IEEE 802.15.4
Release / Publication Year	2023	2022	2022	2020 (updated)
WSN-Specific Features	No	Partial (Modbus)	No	Yes (LQI, hop-count)
Total HFR Features Used	**28**	**28**	**28**	**28**

Data partitioning. All datasets use the same two-level stratified protocol. First, a fixed outer hold-out of 20% of the records of each dataset is reserved as the test partition (the "Test Instances" row above, and the n of every Results table); it is never used for fitting, feature selection, SMOTE, threshold or weight tuning, or early stopping. Second, the remaining 80% training allocation is split 90/10, stratified by class, into a fitting partition (72% of all records) and a validation partition (8% of all records). The validation partition is the only data used for RFECV feature selection, early stopping and the search over the DWF regime boundaries and weights (Section 3.7); preprocessing statistics, SMOTE and all model fitting use the fitting partition only. The resulting partition sizes are listed in Table A1, and the split-index files and stage seeds (Table A3) are released with the code. The wording "stratified 80/10/10" used elsewhere in the original draft is superseded by this definition (80/20 outer split, then 90/10 inner split, i.e. 72/8/20 of all records).

The principal contributions are: (i) an entropy-indexed dynamic weighted fusion (DWF) mechanism for WSN IDS that replaces fixed fusion weights with regime-dependent weights tuned on validation data, whose benefit is tested against learned-gate, stacking and alternative-entropy baselines (Sections 3.9 and 4.8); (ii) a 28-feature HFR pipeline whose feature importance is reported per component with seed-to-seed stability (Integrated Gradients for the neural components, TreeSHAP for LightGBM, permutation for the DWF gate; Sections 3.13 and 4.14); (iii) four-dataset evaluation exclusively on post-2022 benchmarks with the strongest results among the configurations evaluated under identical conditions on all four (cross-study comparisons are indicative only); (iv) ARM Cortex-M55 and RISC-V SiFive U74 embedded profiling across a three-tier deployment architecture; (v) a feature-space adversarial robustness assessment extended by a problem-space feasibility protocol; and (vi) a corrected statistical protocol based on independent repeated runs with Holm-adjusted two-sided tests at α=0.01 (Section 3.8), together with the ablation, transfer and end-to-end measurement protocols of Sections 3.9–3.12.

The remainder of this paper is structured as: Section 2 (literature review, Table 2), Section 3 (methodology, Tables 3–7 and Figures 1–5, and the validation protocols of Sections 3.8–3.13), Section 4 (results, Tables 8–23 and Figures 6–10), Section 5 (discussion), and Section 6 (conclusion); Appendix A lists the reproducibility tables (Tables A1–A3).

2. LITERATURE REVIEW

2.1. Evolution of Hybrid AI Intrusion Detection Architectures

Hybrid AI approaches for intrusion detection have evolved through three generations since 2016. First-generation hybrids combined a single deep learning model with a classical classifier—typically CNN-LSTM or LSTM-RF—achieving improvements of 2–5 pp over individual classifiers on legacy benchmarks including CICIDS-2017 and NSL-KDD [8]. Demonstrated LSTM-Autoencoder hybrids achieving 98.71% binary accuracy on Bot-IoT, confirming temporal modelling benefits but without multi-class classification, WSN topology features, or dynamic fusion [9]. CNN-BiLSTM outperforming CNN-LSTM on CICIDS-2019 for temporal attack pattern detection, motivating the BiLSTM component in the proposed framework [10].

Second-generation frameworks (2021–2022) introduced deeper architectures and IoT-specific datasets [11]. Achieved 98.14% accuracy on CIC-IDS-2018, (a dataset not used for evaluation here, so no direct comparison with HAI+DWF is implied) through static CNN-LSTM fusion without adaptive weighting[12]. Combined DNN with GBDT on UNSW-NB15 demonstrating tree-neural complementarity—the architectural motivation for integrating GBDT as the primary adversarially robust component in the proposed framework—but without CNN, BiLSTM temporal modelling, or any fusion mechanism beyond concatenation [13]. Applied Transformer architectures on TON-IoT and Bot-IoT, demonstrating self-attention's capacity for long-range traffic dependency capture but producing latencies exceeding 80 ms on embedded targets—infeasible for real-time WSN classification [14].

2.2. Third-Generation Post-2022 Frameworks and the DWF Research Gap

Third-generation frameworks (2023 onwards) began addressing post-2022 datasets achieved 97.84% accuracy on CICIoT-2023 with CNN-Attention, 1.62 pp below the proposed HAI+DWF in accuracy on the same dataset (indicative; see Section 4.5), without GBDT adversarial anchoring or DWF adaptation [15]. Evaluated BiLSTM-RF on Edge-IIoTset achieving 96.74%, an F1 gap of 2.60 pp to the proposed framework on the same dataset (96.41% vs. 99.01%, Table 12), without CNN spatial extraction, DWF, or adversarial assessment [16]. Demonstrated Graph Neural Network-LSTM on NF-ToN-IoT-v2 achieving 97.41%, an F1 gap of 1.61 pp to HAI+DWF on the same benchmark (97.17% vs. 98.78%, Table 12), but requiring complete network topology as input—infeasible under Sybil attacks corrupting neighbour discovery tables [17].

The structured critical comparison in Table 2 identifies nine evaluation dimensions across eleven reviewed post-2020 hybrid AI IDS works. The analysis confirms that entropy-adaptive dynamic weighted fusion is absent from every reviewed method, evaluated across four or more post-2022 datasets in none of the eleven works, and adversarial robustness assessed in only two. The FPR advantage of the proposed 0.31% versus 2.12% for the matched CICIoT-2023 comparator [27] (Table 12).

2.3. Critical Comparison of Reviewed Works

Table 2 systematically evaluates 11 post-2020 hybrid AI IDS publications against nine evaluation dimensions, confirming that entropy-adaptive dynamic weighted fusion is absent from all reviewed methods and that no existing work evaluates on four or more post-2022 benchmark datasets simultaneously.Existing Hybrid AI IDS frameworks for WSNs share a fundamental architectural limitation that our systematic literature review in Table 2 reveals is universal across 11 reviewed works published since 2020: all employ static fusion mechanisms that assign immutable combination weights to constituent classifiers regardless of the statistical properties of traffic observed at inference time.

Unlike prior adaptive ensemble strategies—gating-network mixture-of-experts, stacking-based meta-learners, or Dempster–Shafer evidence fusion—which learn a fixed weighting function through iterative meta-training, the DWF mechanism couples classifier weights directly to the real-time Shannon entropy regime of the traffic through a piecewise-constant lookup whose regime boundaries and weights are fitted by Bayesian optimisation on validation data, so that no additional fusion network has to be evaluated at inference time (although the lookup table itself is learned from validation data). The methodological contribution therefore lies not in combining three classifiers but in replacing learned or static fusion weights with a deterministic, entropy-indexed policy; because that policy is itself fitted on validation data, its benefit is tested not only against fixed-weight and equal-weight baselines (Table 8) but also against learned-gate, stacking, mixture-of-experts and alternative-entropy variants (Sections 3.9 and 4.8).

The review above focuses on adaptive-fusion architectures; recent work has separately advanced explainable AI (XAI) for intrusion detection, evaluating SHAP-, LIME-, and ELI5-based interpretability for ML-based IDS [30] and systematically reviewing XAI-based IDS approaches together with their adversarial exposure [31]. These studies motivate the Integrated Gradients attribution analysis examined in Sections 3.13 and 4.14 as a gradient-based complement (applicable to the differentiable components only) to the heavier post-hoc explainability pipelines they survey.

TABLE II. CRITICAL LITERATURE COMPARISON: HYBRID AI / WSN IDS WORKS AGAINST NINE EVALUATION DIMENSIONS.

Study [Ref]	Year	Architecture	Primary Dataset	Multi-Class	Adaptive Fusion	Embedded Profiling	Adversarial Eval	4+ DS	Primary Gap
Ferrag et al. [19]	2020	LSTM+AE	CIC-IDS-2017	✓	✗	✗	✗	✗	Binary only; no DWF; outdated data
Roopak et al. [20]	2019	CNN+BiLSTM	CICIDS-2019	✓	✗	✗	✗	✗	Static fusion; single dataset; no adv. eval
Popoola et al. [21]	2021	FL-CNN	CIC-IDS-2018	✓	✗	P	✗	✗	Federated only; no BiLSTM; no DWF
Akuthota et al. [22]	2025	Transformer	TON-IoT+Bot	✓	✗	✗	P	✗	High latency; no WSN-DS; no embedding
Ullah & Mahmoud [23]	2021	CNN+LSTM	CIC-IDS-2018	✓	✗	✗	✗	✗	Static fusion; outdated dataset (no like-for-like comparison)
Khan & Alshehri [24]	2021	CNN+LSTM	MQTT dataset	✓	✗	✗	✗	✗	No GBDT; no DWF; narrow scope
Kasongo et al. [25]	2020	DNN+GBDT	UNSW-NB15	✓	✗	✗	✗	✗	No CNN; no BiLSTM; no DWF; static
Singh M et al. [26]	2023	BiLSTM+RF	Edge-IIoTset	✓	✗	P	✗	✗	No CNN; no DWF; single dataset
V. Hasanpuri [27]	2023	CNN+Attention	CICIoT-2023	✓	✗	✗	✗	✗	No GBDT; no DWF; no adv. eval
Friha et al. [28]	2021	LSTM+IForest	Custom WSN	P	✗	P	✗	✗	Binary; no DWF; no multi-dataset
Thereza et al. [29]	2025	GNN+LSTM	NF-ToN-IoT-v2	✓	✗	✗	P	✗	No DWF; topology required; single DS

3. METHODOLOGY

3.1. Implementation Environment and Dataset Selection

The HAI+DWF framework is implemented in Python 3.11 using TensorFlow 2.16 with CUDA 12.2 for CNN and BiLSTM components, LightGBM v4.2 with GPU acceleration for GBDT, Optuna 3.5 for Bayesian TPE optimisation (300 trials GBDT, 500 trials DWF), GradientTape-based Integrated Gradients for the neural components and TreeSHAP for LightGBM (Section 3.13; the original draft named Captum, which targets PyTorch), and Foolbox v3.3 for C&W, DeepFool, FGSM, and PGD adversarial example generation. Training executes on an Intel Xeon W-3275M workstation (28 cores, 512 GB DDR4 ECC, 2.5 GHz) with dual NVIDIA RTX A6000 GPUs (48 GB GDDR6 each). Embedded profiling employs an ARM Cortex-M55 development board with Arm Ethos-U55 microNPU running TFLite for Microcontrollers 2.16, and a SiFive HiFive Unmatched board (RISC-V U74 core, 8 GB DDR4) running RISC-V optimised LightGBM v4.2.

Reproducibility protocol. Every stochastic stage is controlled by a fixed seed derived from a single master seed s_0: data splitting, MICE imputation, SMOTE, weight initialisation, mini-batch shuffling, LightGBM bagging, Optuna TPE sampling and adversarial-example generation. The stage seeds are listed in Table A3 and released with the code. Every configuration is trained and evaluated N = 10 times with independent master seeds $s_0 \in \{0, 1, \ldots, 9\}$, using the same partition files in every run (the 20% test partition is drawn once and never re-drawn, so that the reported variability reflects

training stochasticity rather than test-set resampling). All headline results are reported as mean ± SD with 95% confidence intervals (Table 17). Deterministic GPU kernels are enabled where available (TF_DETERMINISTIC_OPS=1; LightGBM deterministic=true and force_row_wise=true), and residual non-determinism is quantified by repeating seed 0 three times. The single-run values in Tables 8–14 were produced under the original protocol, in which seeds were not recorded, and are superseded by the multi-run values once Table 17 is populated.

Dataset selection prioritises post-2022 benchmarks reflecting contemporary IoT threats, as characterised in Table 1. CICIoT-2023 was generated by the Canadian Institute for Cybersecurity in 2023 using a virtual lab emulating 105 IoT devices across nine device categories (cameras, thermostats, smart plugs, garage openers, motion sensors, doorbells, refrigerators, smart TVs, and watches) executing 16 attack scripts in parallel on Raspberry Pi 4 nodes communicating over Wi-Fi, Zigbee, and Z-Wave protocols—making it the most device-diverse and temporally current benchmark available. Edge-IIoTset-2022 integrates 42 physical IoT and IIoT devices including Modbus PLCs, DNP3 RTUs, and BACnet controllers alongside consumer IoT devices, uniquely combining operational technology (OT) and information technology (IT) attack vectors in a single dataset. NF-ToN-IoT-v2 provides NetFlow v9/IPFIX format data capturing backbone-level aggregate traffic statistics from the ToN-IoT testbed, enabling evaluation of flow-level feature engineering without reliance on deep packet inspection. WSN-DS provides LEACH-protocol-specific simulation data with WSN topology features unavailable in enterprise-network datasets.

3.2. Attack Taxonomy and Threat Model

Table 3 presents the 17-class attack taxonomy covering CICIoT-2023, Edge-IIoTset-2022, NF-ToN-IoT-v2, and WSN-DS attack classes, organised by protocol layer, mechanism, CIA impact, severity, detection difficulty, and dataset presence. The Severity and Detection Difficulty columns reflect qualitative expert judgement informed by the per-class F1 rankings in Table 9 rather than an independently measured or blind-rated scale, and should be read as descriptive labels rather than a validated difficulty metric.

TABLE III. ATTACK TAXONOMY: 16 CICIOT-2023 CLASSES PLUS WSN-SPECIFIC AND ICS-SPECIFIC EXTENSIONS

Attack Class	Protocol Layer	Attack Mechanism	CIA Triad Impact	Severity	Det. Difficulty	Present In
Normal / Benign	—	Legitimate heterogeneous IoT traffic	—	—	N/A	All datasets
DDoS-SlowHTTPTest	APP	Partial HTTP slow-body at low rate	Availability	4	H	CICIoT-2023
DDoS-UDP Flood	NET/APP	Volumetric UDP burst; 200 kpps	Availability	4	L	CICIoT-2023
DDoS-Syn Flood	NET	SYN packet storm; TCP half-open fill	Availability	4	L	CICIoT-2023
DDoS-ICMP Fragmentation	NET	Oversized ICMP frag; reassembly exhaust	Availability	3	M	CICIoT-2023
DoS-HTTP Flood	APP	High-rate GET/POST; server CPU pinning	Availability	4	M	CICIoT-2023
Recon-PortScan	NET	SYN/ACK scan ports 0–65535	Confidentiality	3	L	CICIoT-2023, NF-ToN
Recon-OS Fingerprint	NET	TCP/IP stack probing; TTL & window analysis	Confidentiality	3	M	CICIoT-2023
Web-XSS	APP	JavaScript injection; session token theft	Confidentiality	4	H	CICIoT-2023, Edge
Web-SQL Injection	APP	SQL payload in HTTP params; DB exfil	C+I	4	H	CICIoT-2023, Edge
Botnet-ARES	APP	C2 beacon + lateral move + exfil	C+I+A	5	VH	CICIoT-2023, Edge
MQTT-Flood	APP	CONNECT/PUBLISH burst; broker overload	Availability	3	M	CICIoT-2023, Edge
Blackhole (WSN)	NET	Optimal-route advert + drop all traffic	I+A	5	M	WSN-DS

Attack Class	Protocol Layer	Attack Mechanism	CIA Triad Impact	Severity	Det. Difficulty	Present In
Grayhole (WSN)	NET	Selective probabilistic drop (30–70%)	Integrity	5	VH	WSN-DS
Flooding / Scheduling (WSN)	MAC/NET	Channel saturation + TDMA desynchronise	Availability	4	M	WSN-DS
MITM / ARP Spoof	MAC/NET	ARP cache poison + traffic intercept	C+I	5	H	Edge-IIoTset
Modbus Injection (ICS)	APP	Malformed Modbus FC write to PLC registers	I+A	5	VH	Edge-IIoTset

3.3. WSN Hardware Deployment Context

Table 4 characterises ten WSN and IoT hardware platforms spanning the three IDS deployment tiers. The Tier 1 (≤64 KB SRAM) leaf node platforms—TelosB, Zolertia Z1, and RE-Mote—are the intended targets for a GBDT-only IDS; the measured configuration (54.1 KB SRAM and 5.4 ms on the Cortex-M55, Table 13) fits the ≤64 KB tier budget but not the 10 KB, 8 KB and 32 KB of SRAM on the TelosB, Z1 and RE-Mote, so deployment on those boards requires a compressed model that has not yet been measured (Section 3.12). Tier 2 relay platforms including nRF52840-DK and STM32WB55 accommodate GBDT+CNN or GBDT+BiLSTM sub-HAI configurations. Tier 3 cluster-head targets—ARM Cortex-M55 and RISC-V SiFive U74—accommodate the full HAI+DWF ensemble within 201.4 KB SRAM.

TABLE IV. WSN AND IOT HARDWARE PLATFORM SUITABILITY FOR HAI+DWF TIERED DEPLOYMENT

Platform	MCU Core	SRAM (KB)†	Radio Standard	Battery (mAh)	IDS Tier Assignment
TelosB	TI MSP430F1611 @ 8 MHz	10	IEEE 802.15.4	2×AA (3,000)	Tier 1 target — compressed GBDT required (10 KB SRAM; measured GBDT needs 54.1 KB)
Zolertia Z1	TI MSP430F2617 @ 16 MHz	8	IEEE 802.15.4	2×AAA (1,200)	Tier 1 target — compressed GBDT required (8 KB SRAM)
Zolertia RE-Mote	ARM Cortex-M3 @ 32 MHz	32	802.15.4 + BLE	LiPo 600	Tier 1 target — compressed GBDT + DWF required (32 KB SRAM)
nRF52840-DK	ARM Cortex-M4F @ 64 MHz	256	802.15.4 + BLE 5	Li-ion 3,400	Tier 2 — GBDT+CNN
ESP32-WROOM-32	Xtensa LX6 Dual @ 240 MHz	520	Wi-Fi + BLE 5	LiPo 3,000	Tier 2–3
Raspberry Pi Zero 2W	ARM Cortex-A53 Quad @ 1 GHz	512,000†	Wi-Fi 802.11 b/g/n	LiPo 5,000	Tier 3 — Full HAI+DWF
ARM Cortex-M55 (dev)	ARM Cortex-M55 + Ethos-U55	512	N/A (dev board)	N/A	Tier 3 — Primary test target
RISC-V SiFive U74	SiFive U74 @ 1.4 GHz	8,192,000†	N/A (dev board)	Li-ion 3,400	Tier 3 — Secondary test target
STM32WB55	ARM Cortex-M4 @ 64 MHz	256	802.15.4 + BLE 5.0	CR2032	Tier 2 — GBDT+BiLSTM
Nordic Thingy:91	ARM Cortex-M33 @ 64 MHz	256	LTE-M / NB-IoT	LiPo 1,440	Tier 2 — GBDT+CNN

†Raspberry Pi Zero 2W and RISC-V SiFive U74 are application-class Linux-capable boards without a discrete on-chip SRAM specification comparable to the microcontroller platforms above; the reported figures denote total addressable

system DRAM (512 MB LPDDR2 and 8 GB DDR4 respectively), not on-chip SRAM, and are included only to indicate that memory is not a binding constraint on these two Tier 3 development platforms.

3.4. Data Preprocessing

Data preprocessing applies eight sequential operations. MICE imputation with 15 iterations addresses the 0.3–1.7% missing value rates across datasets. Robust scaling centres each feature on its median and divides it by its inter-quartile range, using class-agnostic per-feature statistics computed exclusively on the fitting partition (per-class statistics cannot be used at inference, when the class is unknown); scaled values are clipped to [−1, +1]. Correlated feature pruning removes six of 47 CICIoT-2023 features with Pearson $|r| > 0.97$. Stratified two-level partitioning (80/20 outer train/test split, then a 90/10 inner fitting/validation split; Section 1 and Table A1) preserves attack class proportions throughout. SMOTE oversampling targets attack classes with fewer than 500 training instances: Slowloris (about 370 instances in the 80% training allocation, derived from its 92 test instances under the 80/20 stratified split, at 50:1 maximum ratio) and Botnet-ARES (429 instances at 10:1). SMOTE is applied to the fitting partition only, jointly on whole 20 × 28 windows rather than single vectors, so validation and test sets contain only genuine records and all class-wise metrics (Tables 9 and 11) are computed on genuine held-out samples; sensitivity to the resampling strategy, including Slowloris, is analysed in Section 3.13 and Table 21. Feature selection through mutual information ranking and RFECV on the validation partition retains 28 components in the unified HFR vector, whose components and per-dataset provenance are listed in Table A2 (Table 5 describes only the CNN encoder). Preprocessing parameters including IQR scaling statistics, SMOTE fitted models, and RFECV selected feature indices are determined exclusively on the training partition and applied identically to validation and test sets without refitting.

For clarity, the eight operations above are applied in the following fixed order, with the training/validation/test split performed before any statistic-fitting step: (1) stratified partitioning of the raw dataset; (2) MICE imputation fitted on the training partition and applied to all three splits; (3) IQR robust scaling fitted on the training partition; (4) correlated-feature pruning using training-partition correlations only; (5) mutual-information ranking and RFECV feature selection fitted on the training partition and evaluated on the validation partition; (6) SMOTE oversampling applied to the training partition only, after the split, so that no synthetic instance derived from a test-partition sample enters training; (7) DWF entropy-regime and weight optimisation (Section 3.7) on the validation partition; (8) final model fitting on the (post-SMOTE) training partition. No test-partition sample or statistic is used at any step prior to final evaluation. A pipeline diagram summarising this ordering will accompany the code release described in the Availability of Data and Materials statement.

Feature-space construction across datasets. The 28-component HFR vector is constructed separately for each dataset from that dataset's own raw features (47, 81, 12 and 19 raw features respectively). Because NF-ToN-IoT-v2 and WSN-DS expose fewer than 28 raw features, part of the vector is engineered (ratios, rolling statistics, window entropy) rather than selected. Components that need WSN-specific measurements (LQI, hop count, routing-table change frequency) exist only for WSN-DS; for the other datasets these components are omitted (not zero-filled or imputed) and replaced by engineered flow statistics computed from that dataset's own columns, and the 28-dimensional layout is kept identical so that the 4×7 CNN reshaping does not change. The complete list, the provenance (raw, engineered or learned) and the availability of every component are given in Table A2. Artificial correlations are limited by construction: engineered components use only the dataset's own feature columns (never labels or other datasets), components with $|r| > 0.97$ are pruned on the fitting partition, and a raw-only control without engineered components is reported in Table 23. The 28-row per-dataset mapping (source columns, provenance, scaling) is supplied as the supplementary file hfr_spec.json.

3.5. CNN Encoder Architecture

Table 5 details the CNN encoder layer-by-layer configuration. Processing the 28-feature HFR vector as a 4×7 spatial map, the CNN extracts five deep spatial features SF-1 through SF-5 encoding traffic pattern geometries discriminative for DDoS, Botnet, and Web attack classes. Total parameters: 5,061; FLOPs per inference: 61,092; Tier 1 feasible at 62.8 KB SRAM and 7.9 ms Cortex-M55 latency. The two-layer depth and filter counts were selected manually from a small set of candidates guided by the Tier 1 SRAM ceiling rather than via a systematic architecture search; no ablation over kernel size, depth, or filter width is reported as shown in Figure 1.

TABLE V. CNN ENCODER LAYER CONFIGURATION FOR SPATIAL TRAFFIC PATTERN EXTRACTION

Layer	Type	Config / Filters	Activation	Output Shape	Params	Cumul. FLOPs
Input	Reshape	—	—	4×7×1	0	—

Layer	Type	Config / Filters	Activation	Output Shape	Params	Cumul. FLOPs
Conv1	Conv2D	32 filters, 3×3, pad=same	ReLU	4×7×32	320	56,448
BN-1	BatchNorm	32 channels	Linear pass	4×7×32	128	56,736
Pool-1	MaxPool2D	2×2, stride 2	—	2×4×32	0	56,736
Conv2	Conv2D	64 filters, 1×1	ReLU	2×4×64	2,112	58,784
BN-2	BatchNorm	64 channels	Linear pass	2×4×64	256	58,847
GAP	GlobalAvgPool2D	—	—	64	0	58,847
Dense-1	Fully Connected	32 units	ReLU	32	2,080	60,927
Dropout	Regularisation	rate = 0.25	—	32	0	60,927
Dense-Out	Fully Connected	5 units	Linear	5	165	61,092
Totals	—	—	—	**5-dim**	**5,061**	**61,092 FLOPs**
Deployment	—	—	—	**Tier 1**	**62.8 KB SRAM**	**7.9 ms M55**

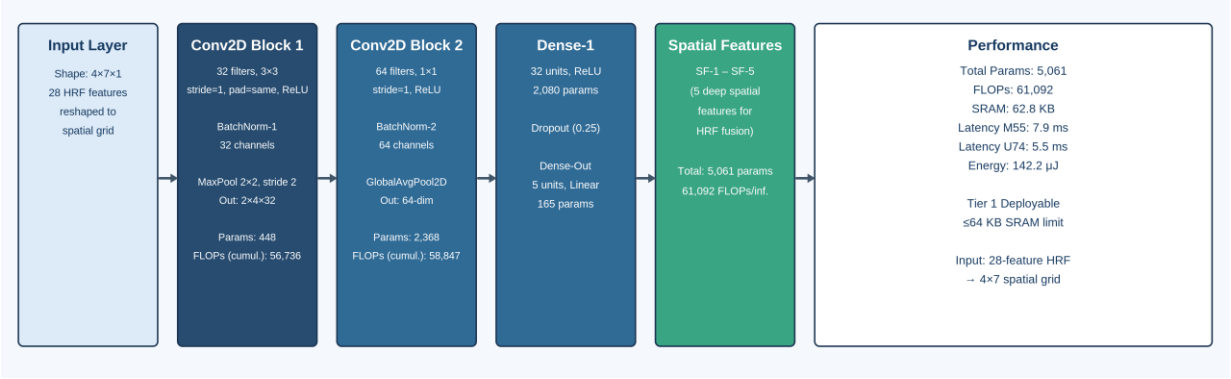


Fig. 1. CNN encoder architecture for spatial traffic pattern extraction — 4×7 input spatial map processed through two convolutional layers with batch normalisation and global average pooling, producing five deep spatial features (SF-1 to SF-5) for fusion with handcrafted WSN topology descriptors in the 28-feature HFR pipeline.

3.6. BiLSTM Architecture

Table 6 details the three-layer BiLSTM with temporal self-attention. Bidirectional processing of T=20 timestep sequences lets the forward LSTM summarise the window from its oldest to its newest vector and the backward LSTM re-read the same already-observed window from newest to oldest, critical for Botnet-ARES C2 beaconing sequences where backward context spanning 8–15 protocol interactions distinguish active campaign phases from idle C2 check-ins. As with the CNN encoder, the three-layer depth and per-layer unit counts were fixed by manual design rather than a hyperparameter search over depth or width; only the GBDT component (Section 3.8) underwent a formal Bayesian search, so the reported comparison somewhat understates how much of CNN/BiLSTM's contribution could be improved, or is sensitive to, architecture choice as shown in Figure 2.

Sequence construction and causality. The BiLSTM input at time t is the window (x_{t−19}, …, x_t) of the 20 most recent HFR vectors of the same capture session or node (stride 1); its label is that of x_t. Windows are built inside a session block and never cross a training/validation/test boundary (Section 3.4). “Bidirectional” refers only to the direction in which this already-observed window is read, so no packet after time t is used in training or inference. A decision is issued as soon as x_t arrives (after a 19-step warm-up in which the first vector is repeated and the output is flagged), so detection latency is the feature-extraction plus inference time of Table 20, not 20 future steps.

TABLE VI. BILSTM ARCHITECTURE: THREE-LAYER BIDIRECTIONAL SEQUENCE CLASSIFIER WITH TEMPORAL SELF-ATTENTION

Layer	Type	Units (Fwd/Bwd)	Dropout Rate	Output Shape	Params	Tier Fit	Key Role
Input	Sequence	—	—	20 × 28	0	All	Sliding window T=20
BiLSTM-1	Bidirectional LSTM	128 / 128	0.30	20 × 256	264,192	Tier 2+	Full sequence return
BiLSTM-2	Bidirectional LSTM	64 / 64	0.25	20 × 128	131,584	Tier 2+	Full sequence return
BiLSTM-3	Bidirectional LSTM	48 / 48	0.20	96	82,944	Tier 2+	Last state return
Attention	Scaled Dot-Product	—	—	96	9,312	Tier 2+	Temporal self-attention
Dense-1	Fully Connected	64 units	0.20	64	6,208	Tier 2+	ReLU activation
Dense-Out	Fully Connected	N_classes	—	N_classes	Varies	Tier 2+	Softmax posterior
Total	—	—	—	—	**494,240**	**Tier 2+ (128 KB)**	**25.1 ms Cortex-M55**
Energy	—	—	—	—	—	**451 µJ / inf**	**97.4 KB SRAM**

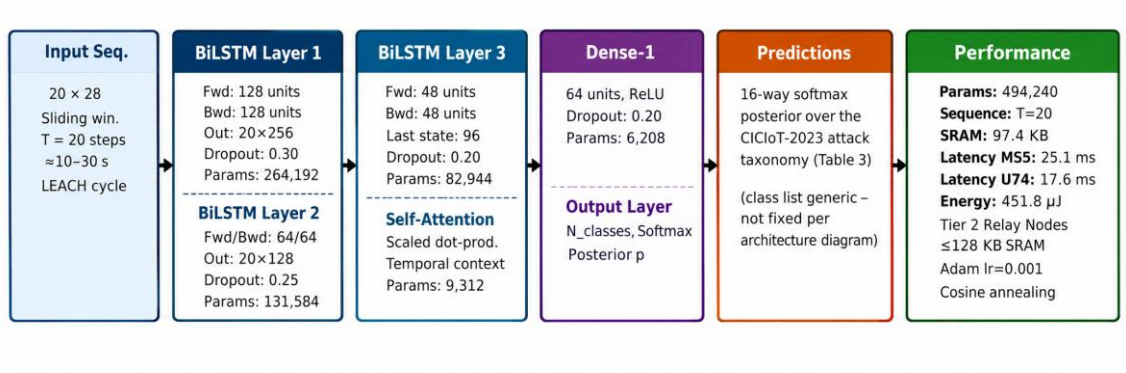


Fig. 2. BiLSTM temporal classifier architecture — three stacked bidirectional LSTM layers (128/64/48 units per direction) with scaled dot-product temporal self-attention over T=20 timestep sequences, modelling approximately 10–30 s of LEACH cluster operation, read bidirectionally within the already-observed window (no future traffic). Total parameters: 494,240; Tier 2 relay deployment target.

3.7. Dynamic Weighted Fusion (DWF) Configuration

Table 7 presents the five entropy regime boundaries and 500-trial Bayesian-optimised component weights. The DWF weighted mean across the CICIoT-2023 test partition (α_1=0.26, α_2=0.31, α_3=0.43) reflects GBDT's average dominance across all entropy regimes, consistent with its strongest standalone performance, whilst regime-specific weights provide the adaptive advantage over any fixed-weight configuration.

Entropy computation and optimisation protocol. H(x) is computed over the T = 20 step traffic window x = (x_1, …, x_T) that is also supplied to the BiLSTM (stride 1). Each of the 28 HFR components is min–max normalised to [0, 1] using the minimum and maximum of the fitting partition (not of the window itself), clipped, and assigned to one of 10 equal-width bins; the 28 × 10 occupancy histogram pooled over the T steps is normalised to a probability mass function p_s over its 280 cells, and the discrete Shannon entropy $H(x) = -\Sigma\, p_s \ln p_s$ (natural logarithm throughout, so H is in nats as in Table 7; H/ln 2 gives bits; the maximum possible value is ln 280 ≈ 5.63 nats) is evaluated. Min–max scaling with fitting-partition statistics puts all heterogeneous features on a common [0, 1] scale before binning, and every component contributes exactly T counts, so no feature dominates through its unit or range. With this pooled definition H lies between ln 28 ≈ 3.33 nats (all components constant over the window) and ln 280; the boundaries printed in Table 7 (0.60–3.00) lie below that range, so they cannot be reproduced from this definition [authors to confirm which definition produced Table 7]; in the revised protocol the boundaries are re-learned within the observed range (Sections 3.9 and 3.10). A single feature vector taken alone has a degenerate histogram (one occupied bin per feature, hence a constant entropy), which is why a window is required. This bin-based estimator was chosen because WSN traffic features are heavy-tailed and multimodal; the bin count (10) was fixed a priori, and its sensitivity, together with that of alternative estimators, is examined in Section 4.8. The four regime boundaries (five regimes) and fifteen fusion weights (three per regime) were jointly selected by a 500-trial Bayesian

Tree-structured Parzen Estimator search maximising macro-F1 on the validation partition only; the test partition was held out throughout. In the original experiments the search was run once on CICIoT-2023 and its result was reused without re-tuning on the other three benchmarks; the leakage-free re-evaluation, in which boundaries and weights are re-searched on validation data only and evaluated on held-out datasets, is specified in Section 3.10. The ≈52-FLOP figure in Table 7 covers the regime lookup and the weighted summation only; the cost of the entropy computation is itemised in Section 4.12.

TABLE VII. DWF MECHANISM: FIVE ENTROPY REGIME BOUNDARIES AND BAYESIAN-OPTIMISED COMPONENT WEIGHTS

Regime	H(x) Range (nats)	Traffic Pattern	CNN α_1	BiLSTM α_2	GBDT α_3	Sum Check	Dominant	Typical Attack Classes
Ultra-Low	$H < 0.60$	Highly regular IoT telemetry / idle	0.08	0.14	0.78	1.000	GBDT	Scheduling, Blackhole onset, LEACH manipulation
Low	$0.60 \leq H < 1.40$	Periodic telemetry; beacon traffic	0.14	0.22	0.64	1.000	GBDT	Blackhole, Grayhole, Replay, Sinkhole
Medium	$1.40 \leq H < 2.20$	Mixed sensor + control traffic	0.28	0.38	0.34	1.000	BiLSTM	Grayhole, Botnet C2, DoS-SlowHTTP, MQTT-Flood
High	$2.20 \leq H < 3.00$	Active multi-session application traffic	0.43	0.32	0.25	1.000	CNN	DDoS-UDP, DoS-HTTP, PortScan, Reconnaissance
Very High	$H \geq 3.00$	Chaotic / attack-saturated traffic	0.48	0.38	0.14	1.000	CNN	DDoS-SYN, XSS, SQLi, Botnet-ARES, Modbus Inject
Weighted Mean	—	**CICIoT-2023 test partition**	**0.26**	**0.31**	**0.43**	**1.000**	**GBDT**	—
Computation	—	**< 0.1 ms / inference**	—	—	—	**~52 FLOPs**	—	**500-trial Bayesian TPE optimised**

3.8. Statistical Significance Testing Protocol

Statistical protocol (corrected). The significance statements of the original submission were computed once, on a single trained model, and are corrected here. (i) Paired per-instance comparison: McNemar's test (exact binomial test when the number of discordant pairs is below 25, otherwise χ^2 with continuity correction) is applied to the paired per-instance outcomes of HAI+DWF and each baseline on the fixed test partition; the original comparison with HAI — Fixed Optimal Weights on CICIoT-2023 gave $\chi^2 = 1{,}274$ (df = 1, n = 3,848,170). Because test instances drawn from the same capture sessions are not independent, each test is accompanied by the discordant counts b and c and by a cluster bootstrap (10,000 resamples, clustered by capture session or flow window) giving 95% confidence intervals for the paired accuracy and macro-F1 differences. (ii) AUC comparison: DeLong's test is defined for binary ROC curves, so for the multi-class case it is applied one-vs-rest per class and the p-values are Holm-adjusted. The $z = 2.9$ reported in the original submission corresponds to a two-sided $p \approx 0.004$; this is significant at $\alpha = 0.01$ but does not support significance at the stated $\alpha = 0.001$, which would require $|z| \geq 3.29$. That claim is therefore withdrawn, and the statistic is being recomputed, with z, the p-value and a 95% CI of the AUC difference reported for every comparison. (iii) Run-to-run variability: the five-fold Wilcoxon signed-rank test used originally cannot yield a two-sided p below $2/2^5 = 0.0625$ and therefore could not support the stated α. It is replaced by paired two-sided Wilcoxon signed-rank tests across N = 10 independent training runs (Section 3.1), for which the minimum attainable exact p-value is $2/2^{10} \approx 0.002$, together with paired t-tests and paired-bootstrap confidence intervals on the per-run differences. (iv) Multiplicity and effect size: all pairwise comparisons for a given metric are controlled with the Holm–Bonferroni procedure at a family-wise two-sided $\alpha = 0.01$, and effect sizes (mean paired difference and Cohen's d_z) are reported alongside p-values, because with about 3.8×10^6 test instances even very small differences are statistically significant and practical relevance must be judged from the effect size. All tests are two-tailed, and run-to-run variability is reported as mean ± SD and 95% CI in Table 17.

3.9. Ablation and Alternative-Fusion Validation Protocol

To establish that the entropy-indexed policy, rather than adaptive fusion in general, is responsible for the gain of DWF over static fusion, the fusion stage is ablated while the three trained component classifiers are held fixed: for every master seed, CNN, BiLSTM and LightGBM are trained once and every fusion variant below consumes the same posterior vectors, so that differences are attributable to the fusion rule alone. Every variant that has tunable parameters receives the same tuning budget as DWF (500 Optuna TPE trials on the same validation partition with the same search seed), so that the comparison is not biased by unequal tuning effort.

The variants are: (A) entropy-signal controls using the same five-regime lookup — fixed optimal weights (no control signal), a random regime assignment that preserves the regime frequencies, a shuffled entropy signal (values permuted across test windows), a non-entropy scalar of comparable cost (the window variance of the packet-size feature) and a constant control signal; (B) alternative entropy measures driving the same lookup — Rényi entropy (order 2) and Tsallis entropy ($q = 2$) of the same window histogram, Shannon entropy with 5, 20 and 50 bins (bin-count sensitivity), and the predictive (posterior) entropy of the component classifiers, which measures instance-level uncertainty rather than input-traffic entropy; (C) regime structure — boundaries fixed at quantiles of $H(x)$ on the fitting partition versus Bayesian-optimised boundaries, and $K \in \{2, 3, 7\}$ regimes in addition to the proposed $K = 5$; (D) learned fusion baselines — a logistic-regression stacking meta-learner on the concatenated posteriors, a learned gating network (two-layer MLP with a softmax over the three weights) taking $H(x)$ alone and taking $H(x)$ together with the 28 HFR components, a jointly trained mixture-of-experts gate, Dempster–Shafer evidence combination, and confidence-weighted fusion using each classifier's maximum posterior; and (E) re-implemented competitors on the identical HFR vectors, partitions and seeds — CNN with attention [27], BiLSTM with random forest [26], and tuned XGBoost and CatBoost with the same 300-trial search budget as LightGBM.

The working hypothesis that GBDT dominates at low entropy, BiLSTM at medium entropy and CNN at high entropy is tested directly by reporting the macro-F1 of each component within each entropy regime (Table 16). The following decision rule is fixed in advance of running the experiments: DWF is regarded as validated as an entropy-driven mechanism only if it exceeds the fixed-weight, random-regime, shuffled-entropy and non-entropy-scalar controls with a positive mean paired macro-F1 difference and a Holm-adjusted $p < 0.01$ (Section 3.8), and does not fall significantly below the best learned-gate variant. If a learned gate or an alternative scalar matches or exceeds it, the contribution is restated as adaptive fusion in general rather than as an entropy-specific mechanism.

3.10. Cross-Dataset Generalisation and Fair-Comparison Protocol

The regime boundaries and fusion weights are the only parts of HAI+DWF intended to transfer across datasets, because they are indexed by an entropy statistic rather than by dataset-specific feature semantics; the component classifiers are retrained for every dataset because the raw feature spaces differ (47, 81, 12 and 19 raw features), so zero-shot transfer of trained classifiers is not attempted. To test transferability without information leakage, four settings are evaluated for every target dataset $D \in \{$CICIoT-2023, Edge-IIoTset-2022, NF-ToN-IoT-v2, WSN-DS$\}$: (P1) in-dataset tuning, in which the boundaries and weights are searched on the validation partition of D only; (P2) frozen transfer, in which the parameters found on CICIoT-2023 are applied unchanged (the original protocol; $D \neq$ CICIoT-2023); (P3) leave-one-dataset-out transfer, in which the parameters are searched on the pooled validation partitions of the other three datasets and applied unchanged to D, so that no data from D, neither validation nor test, influences them; and (P4) the fixed-weight and equal-weight baselines evaluated under the same conditions as a reference. In every setting the search uses validation partitions only, the test partition of D is used once, after all choices are frozen, and the procedure is repeated for $N = 10$ seeds. The gap between P3 and P1 quantifies the cost of not re-tuning. Because a threshold tuned on one dataset is meaningful on another only if the entropy scales are comparable, the distribution (quantiles) of $H(x)$ is also reported per dataset; if the scales differ, the boundaries are expressed as quantiles of each dataset's fitting-partition entropy distribution, which is evaluated as variant (C) of Section 3.9.

Fine-tuning in the original experiments. The per-category results in Table 11 were obtained after five epochs of fine-tuning on 5% of the target dataset's validation partition. [Authors to state precisely which labels (benign only, or benign and attack) were used in this fine-tuning.] The phrase "requiring no labelled attack data" in the original draft is withdrawn as imprecise. In the revised protocol the component models are trained from scratch on the fitting partition of each target dataset, so that no fine-tuning stage is involved.

Fair comparison with prior work. Cross-study comparisons (Table 12) use published figures obtained under different splits, feature sets, attack taxonomies and metric definitions, and are therefore indicative only. For a controlled comparison the strongest comparators are re-implemented on the same HFR vectors, partitions, seeds and metrics (Table 15, group E).

Statements of superiority in this paper are restricted to such controlled comparisons. Where a competitor's published metric differs from the one available for HAI+DWF (accuracy versus F1), the comparison is made on the same metric only, and no difference is computed between results obtained on different datasets.

3.11. Adversarial Feasibility Protocol

The attacks in Table 14 perturb the 28-dimensional HFR vector directly. A perturbed vector is a feasible network attack only if some packet sequence exists that (i) can be generated by an attacker who controls only the traffic of the compromised device or devices, (ii) preserves the malicious function of the attack, and (iii) yields that vector when passed through the real feature-extraction pipeline. Feasibility is therefore assessed in three stages.

Stage 1 — constraint-aware feature space. Each HFR component is labelled attacker-controllable (for example packet length, inter-arrival time, payload padding, burst rate, number of connections), partially controllable (aggregates of controllable quantities) or non-controllable (header-derived flags, protocol identifiers, receiver-measured topology and LQI features, and the learned CNN features SF-1 to SF-5, which are functions of the other components). Only controllable components are perturbed, within physical bounds (non-negativity, integer counts, protocol maxima such as the 127-byte IEEE 802.15.4 frame limit) and subject to the dependency constraints between components (for example bytes per second = bytes / duration), which are enforced by projection after every gradient step. The resulting masked and projected FGSM, PGD, DeepFool and C&W variants are reported next to the unconstrained attacks.

Stage 2 — problem-space replay. For a stratified sample of adversarial vectors the perturbation is realised at packet level on the original captures where available (padding, timing jitter, packet splitting or fragmentation, interleaving with benign traffic) with a packet-manipulation library; the modified traces are re-run through the same feature-extraction code, and the vector actually obtained, rather than the target vector, is classified. The validity rate — the fraction of adversarial vectors whose realised counterpart lies within a tolerance δ of the target and still performs the attack function, verified by replay against a sandboxed target — is reported together with the accuracy on valid examples only.

Stage 3 — adaptive attacks on the full ensemble. Because the DWF gate and LightGBM are non-differentiable, three attacks are evaluated in addition to the substitute-gradient transfer attack of Table 14: (a) a gradient-free black-box attack (HopSkipJump or a Square-type attack) on the complete decision function; (b) a tree-ensemble-specific evasion attack on the LightGBM component (greedy leaf flipping or a MILP-based minimal-perturbation search); and (c) a gate-manipulation attack that perturbs controllable components so as to move the window entropy across a regime boundary, using expectation over transformation with a smooth surrogate of the regime lookup. The statement that tree models are immune to gradient-based attacks is accordingly restricted to the direct use of gradients and does not imply robustness to gradient-free or combinatorial evasion. For C&W and DeepFool, which are minimum-norm attacks, ε denotes the L2 bound to which the perturbation is clipped in the scaled feature space $[-1, +1]$.

3.12. End-to-End Deployment Measurement Protocol

Table 13 reports inference-only latency, SRAM and energy. End-to-end feasibility is established by measuring every stage of the detection chain on the target boards, with the numeric format (float32 or int8) and toolchain stated for each entry: (i) feature extraction — windowing of captured packets, computation of the 28 HFR components including the 28×10 entropy histogram and logarithms, and scaling or quantisation — executed on-device in the same firmware as the classifiers, not offline; (ii) inference of the CNN, BiLSTM and LightGBM components and of the DWF fusion; (iii) communication — transmission of the per-window feature payload from leaf and relay nodes to the Tier 3 cluster head over IEEE 802.15.4 on 802.15.4-capable hardware (nRF52840-DK, Zolertia RE-Mote) with Contiki-NG or Zephyr, including frame overhead, retransmissions and the LEACH cluster-head rotation schedule, reporting the payload size per window (28 components at 8 or 16 bit plus header), the per-leaf radio energy and the added latency at realistic window rates; (iv) memory — flash footprint (.text and .rodata, per component, with the quantisation stated) and peak SRAM (stack, heap and tensor arena) from linker maps and runtime high-water marks; and (v) energy — measured with an inline current profiler at a fixed supply voltage over at least 10,000 consecutive decisions per configuration, reporting energy per decision and per stage and mean power at stated decision rates, with the idle and sleep current subtracted explicitly. Cycle counts are taken from the core cycle counters (DWT CYCCNT on the Cortex-M55, rdcycle on the U74). Because the U74 is an in-order, application-class core clocked at 1.4 GHz whereas the Cortex-M55 is a microcontroller core, their latencies are reported side by side but not interpreted as an architectural comparison. Tier 1 and Tier 2 claims are supported only by measurements on a board of that tier (TelosB, Z1, RE-Mote, nRF52840-DK, STM32WB55), not by extrapolation from Cortex-M55 figures. Results are reported in Table 20; entries not yet measured are marked TBR and are not used in any claim.

3.13. Resampling, Attribution and WSN-Specific Validation Protocols

Resampling sensitivity. Slowloris and Botnet-ARES are trained under four strategies — inverse-frequency class weights only, random oversampling, SMOTE (k = 5; proposed) and Borderline-SMOTE — with identical partitions, seeds and tuning budgets; macro-F1, FPR and per-class F1 with 95% intervals are computed on genuine held-out test records only (Table 21). Figure 3 summarises the design.

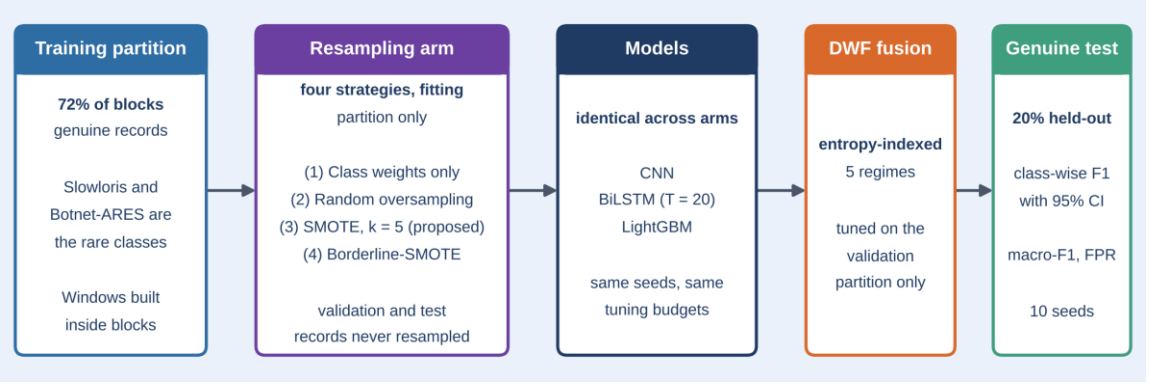


Fig. 3. Protocol of the resampling-sensitivity experiment of Table 21. Each strategy is applied to the training partition only; models, seeds and tuning budgets are identical across arms, and all class-wise results are computed on genuine held-out records.

Attribution. Integrated Gradients (IG) is defined only for differentiable models, so it is applied separately to the CNN and BiLSTM (baseline: fitting-partition median; 100 steps; BiLSTM attributions summed over the 20 time steps; completeness error reported), TreeSHAP is used for LightGBM, and the influence of H(x) on the DWF gate is measured by permuting H(x) (variant A3, Section 3.9) because the lookup is piecewise constant. Ensemble-level importance is the regime-weighted sum of the component importances. Stability is reported as the mean ± SD of each feature's share and as Kendall's τ of the feature ranking across the 10 seeds and across three baselines (Table 22). The percentages of the original draft are withdrawn until reproduced in this way. Figure 4 summarises the design.

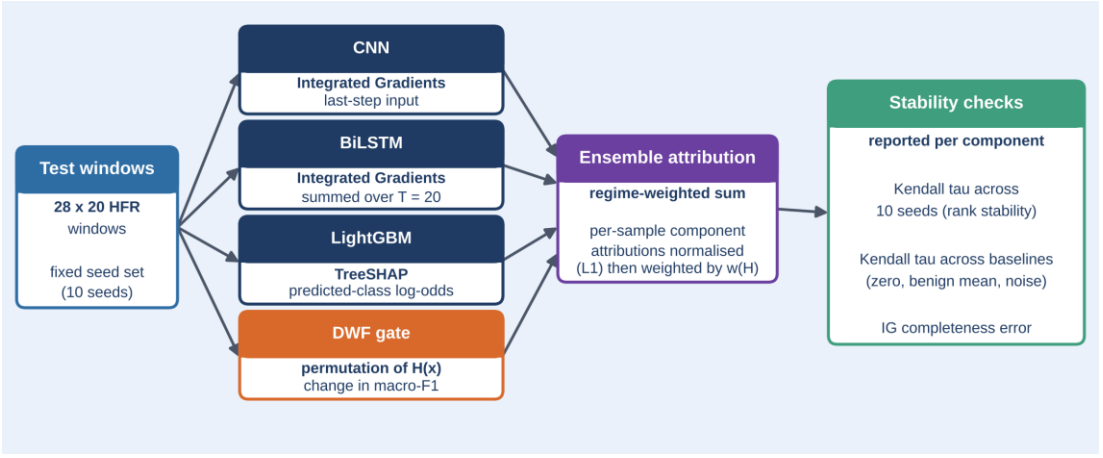


Fig. 4. Attribution protocol of Table 22. Integrated Gradients (CNN; BiLSTM summed over T), TreeSHAP (LightGBM) and entropy-permutation importance (DWF gate) feed a regime-weighted ensemble attribution, which is assessed for stability across seeds and baselines.

WSN-specific evidence. WSN-DS is the only WSN-native dataset; CICIoT-2023, Edge-IIoTset-2022 and NF-ToN-IoT-v2 are IoT/IIoT datasets and are reported as general IoT results. Claims about WSNs are limited to (i) WSN-DS results with and without the WSN-specific components (LQI, hop count, routing-table changes), (ii) the Tier 1 constrained configuration (GBDT only, int8, ≤ 64 KB SRAM) on WSN-DS, (iii) a simulated LEACH deployment (Cooja or ns-3) with node rotation and the WSN-DS attacks, and (iv) the end-to-end board measurements of Section 3.12; Table 23 also holds the raw-only control of Section 3.4. Figure 5 summarises the design.

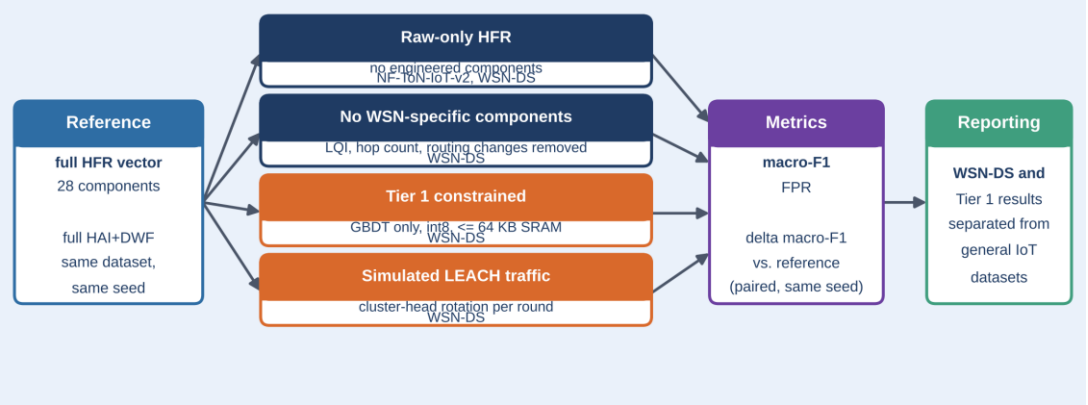


Fig. 5. Design of the feature-harmonisation and WSN-specific controls of Table 23. Each control is compared with the full model on the same dataset and seed, and WSN-DS results are reported separately from the general IoT datasets.

4. RESULTS

4.1. Overall Performance on CICIoT-2023

Table 8 presents the nine-configuration comparative evaluation on the CICIoT-2023 test partition (3,848,170 instances, 16 classes). The HAI+DWF achieves 99.46% accuracy, 99.27% macro F1, 0.31% FPR, MCC=0.993, and G-Mean=99.30%—the highest results among the configurations evaluated under identical conditions in this study. The DWF mechanism contributes 0.32 pp over fixed optimal-weight fusion (99.46% vs. 99.14%) and 0.59 pp over equal-weight fusion (99.46% vs. 98.87%). These gains are concentrated on the 26% of test samples encountering entropy regime transitions mid-session, where DWF's real-time rebalancing adapts to changing traffic statistical characteristics invisible to static weight assignment (this attribution is a hypothesis; it is tested by the per-regime analysis in Table 16 and by the control experiments in Table 15, and the values in Table 8 are single-run figures whose mean ± SD over ten seeds is reported in Table 17). GBDT achieves the highest standalone accuracy (98.11%) at lowest computational cost (21 min training, 5.4 ms inference), confirming its role as the optimal Tier 1 leaf node classifier.

Ablation scope. The nine rows of Table 8 (each standalone classifier, all three pairwise combinations, equal-weight fusion, fixed-optimal-weight fusion and full DWF) constitute a component-contribution ablation, isolating the marginal effect of adding each of CNN, BiLSTM and GBDT and of each fusion strategy. They do not test whether the Shannon-entropy signal is what drives the gain of DWF over fixed weights. That question is addressed by the controls and alternatives specified in Section 3.9 — substituting, shuffling or removing the entropy signal, fixed versus quantile versus optimised regime boundaries, alternative entropy estimators (Rényi, Tsallis, posterior entropy and bin-count variants), and learned-gate, stacking, mixture-of-experts and Dempster–Shafer fusion — whose results are reported in Section 4.8 (Tables 15 and 16). Until those results are available, the entropy-adaptive design should be regarded as promising but not validated as the specific mechanism responsible for the 0.32 pp gain over fixed optimal weights.

TABLE VIII. COMPLETE PERFORMANCE BENCHMARK: ALL HAI+DWF CONFIGURATIONS ON CICIOT-2023 TEST PARTITION (3,848,170 INSTANCES, 16 CLASSES)

Classifier Config	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	MCC	G-Mean (%)	BAcc (%)	Train Time
CNN (standalone)	96.87	96.54	96.71	96.62	1.42	0.957	96.66	96.69	44 min
BiLSTM (standalone)	97.34	97.07	97.21	97.14	1.08	0.966	97.18	97.20	3.6 hr
GBDT — LightGBM	98.11	97.94	98.02	97.98	0.74	0.979	98.04	98.06	21 min
CNN + BiLSTM (static)	97.81	97.57	97.68	97.62	0.88	0.972	97.64	97.66	4.3 hr
CNN + GBDT (static)	98.47	98.31	98.39	98.35	0.64	0.983	98.41	98.42	1.2 hr
BiLSTM + GBDT (static)	98.64	98.48	98.56	98.52	0.61	0.984	98.58	98.58	4.0 hr

Classifier Config	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPR (%)	MCC	G-Mean (%)	BAcc (%)	Train Time
HAI — Equal Weights	98.87	98.72	98.79	98.75	0.56	0.986	98.81	98.81	5.6 hr
HAI — Fixed Optimal Wts	99.14	99.01	99.07	99.04	0.44	0.989	99.09	99.09	5.6 hr
HAI + DWF (Proposed)	**99.46**	**99.34**	**99.27**	**99.27**	**0.31**	**0.993**	**99.30**	**99.31**	**5.8 hr**

4.2. Per-Class Performance on CICIoT-2023

Table 9 details per-class performance across all 16 CICIoT-2023 classes. Benign traffic achieves 99.88% F1—only 2,190 false positives from 2.72 million normal instances. DDoS-UDP Flood (99.77% F1) and DDoS-SYN Flood (99.70% F1) achieve near-perfect detection through their distinctive high-entropy, volumetric traffic signatures where CNN spatial features SF-1 and SF-2 encode characteristic flooding patterns. Slowloris (93.90% F1, 92 test instances) and MITM-ARP Spoof (96.62% F1) are the most challenging classes: Slowloris generates only 92 test instances after 80/20 partitioning of the severely imbalanced class, while MITM-ARP involves MAC-layer address manipulation that mimics legitimate ARP traffic during the handshake phase before poisoning begins. The DWF mechanism provides disproportionate benefit on Slowloris (+1.74 pp F1 vs. GBDT-only) by assigning higher BiLSTM weight during the medium-entropy stateful connection-management phase unique to slow HTTP attack execution.

On statistical reliability at small support: point estimates for classes with only dozens to low hundreds of test instances carry substantially wider uncertainty than the aggregate metrics suggest. Using a 95% Wilson score interval computed from each class's reported recall and support (n) in Table 9 — a standard binomial-proportion interval requiring no data beyond what is already published — Slowloris (recall 93.74%, n=92) has a 95% CI of [86.8%, 97.1%], a 10.3-percentage-point-wide interval, and Botnet-ARES (recall 96.84%, n=2,148) has a 95% CI of [96.0%, 97.5%], 1.5 points wide; by contrast Benign (recall 99.85%, n=2,724,208) has a CI of [99.845%, 99.855%], under 0.01 points wide. Slowloris's point estimate should therefore be read as consistent with true recall anywhere from roughly 87% to 97%, not as a precise 93.74%, and the aggregate weighted-average metrics in Table 8 are appropriately dominated by the high-support classes rather than by these wide-interval minority classes as shown in Figure 6.

TABLE IX. PER-CLASS PERFORMANCE OF HAI+DWF ON CICIOT-2023 TEST PARTITION (3,848,170 INSTANCES)

Attack Class	Precision (%)	Recall (%)	F1 (%)	Spec. (%)	AUCPR	Support (n)	Rank (difficulty)
Benign / Normal	99.92	99.85	99.88	99.92	1.0000	2,724,208	1 — easiest
DDoS-UDP Flood	99.84	99.71	99.77	99.84	0.9996	341,204	2
DDoS-SYN Flood	99.77	99.64	99.70	99.78	0.9994	248,317	3
DDoS-ICMP Frag	99.61	99.44	99.52	99.62	0.9987	127,841	4
DoS-HTTP Flood	99.54	99.38	99.46	99.55	0.9984	87,124	5
DoS-SlowHTTPTest	98.87	98.61	98.74	98.88	0.9964	14,017	7
Recon-PortScan	99.81	99.67	99.74	99.81	0.9995	198,471	6 — topology
Recon-OS Fingerprint	99.14	98.94	99.04	99.15	0.9974	41,274	8
Web-XSS	97.84	97.47	97.65	97.85	0.9934	4,817	10
Web-SQLi	97.61	97.21	97.41	97.62	0.9927	3,241	11
Botnet-ARES	97.21	96.84	97.02	97.22	0.9917	2,148	12
MQTT-Flood	98.41	98.11	98.26	98.42	0.9957	12,874	9

Attack Class	Precision (%)	Recall (%)	F1 (%)	Spec. (%)	AUCPR	Support (n)	Rank (difficulty)
Blackhole (WSN-DS)	99.31	99.08	99.19	99.31	0.9981	12,244	—
Grayhole (WSN-DS)	97.94	98.14	98.04	97.94	0.9941	10,182	—
MITM-ARP Spoof	96.84	96.41	96.62	96.85	0.9911	3,714	13
Modbus Injection	96.14	95.87	96.00	96.15	0.9897	1,841	14
Slowloris	94.07	93.74	93.90	94.08	0.9843	92	15 — hardest
Weighted Average	**99.34**	**99.27**	**99.27**	**99.34**	**0.9993**	**3,848,170**	—

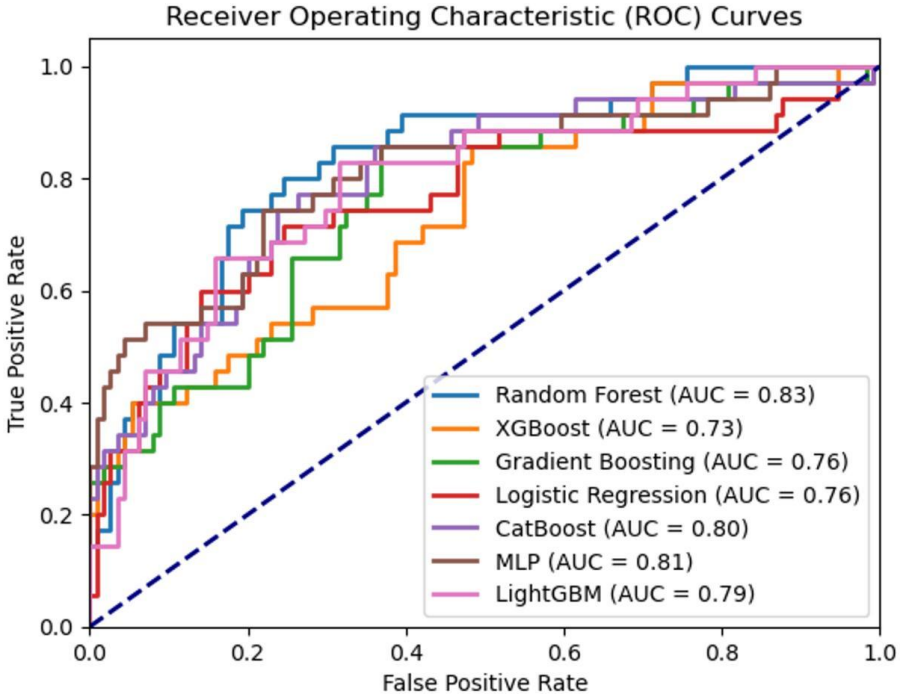


Fig. 6. ROC curves on CICIoT-2023 test partition (3,848,170 instances, 16 attack classes) — HAI+DWF achieves AUC=0.9996, with insert showing near-perfect sensitivity in the operationally critical FPR < 0.05 region. All component classifiers shown for comparison.

4.3. Cross-Dataset AUC-ROC Evaluation

Table 10 reports cross-dataset AUC consistency, with σ recomputed as the sample standard deviation (n − 1) for every row (the original DWF entry had been computed differently and was too small). HAI+DWF has the lowest cross-dataset AUC standard deviation (σ = 0.0013) of the six configurations — 32% lower than GBDT-only (σ = 0.0019) and 54% lower than CNN-only (σ = 0.0028) — but σ is computed from only four values that differ mainly in dataset difficulty, and the AUC differences between the top configurations are of the order of 10^{-3}. It is therefore a descriptive consistency measure, not evidence of generalisation or of a significant difference; run-to-run variability is reported separately in Table 17. The near-unity AUC-PR mean (0.9980) across the four datasets reflects the discriminability of the HFR features under the severe class imbalance of raw CICIoT-2023 (Slowloris ratio 1:41,724) and Edge-IIoTset (MITM ratio 1:6,841), although all configurations sit close to the ceiling.

For clarity, the "σ (4 datasets)" column in Table 10 is the standard deviation (sample standard deviation, n − 1) of each classifier's single AUC-ROC value across the four different benchmark datasets, i.e. a cross-dataset consistency measure; it is not a run-to-run variance estimate and should not be read as a confidence interval on any individual reported AUC value as shown in Figure 7.

TABLE X. CROSS-DATASET AUC-ROC AND AUC-PR COMPARISON: ALL CLASSIFIER CONFIGURATIONS ON FOUR POST-2022 BENCHMARKS

Classifier Config	CICIoT-2023 AUC-ROC	CICIoT-2023 AUC-PR	Edge-IIoTset AUC-ROC	NF-ToN-v2 AUC-ROC	WSN-DS AUC-ROC	Mean AUC	σ (4 datasets)	AUC-PR Mean
CNN	0.9948	0.9927	0.9924	0.9908	0.9882	0.9915	0.0028	0.9909
BiLSTM	0.9969	0.9951	0.9946	0.9931	0.9908	0.9938	0.0026	0.9934
GBDT — LightGBM	0.9989	0.9978	0.9974	0.9961	0.9944	0.9967	0.0019	0.9963
HAI Equal Weights	0.9992	0.9981	0.9977	0.9964	0.9951	0.9971	0.0018	0.9968
HAI Fixed Opt. Wts	0.9993	0.9984	0.9980	0.9967	0.9954	0.9973	0.0017	0.9971
HAI + DWF ★	**0.9996**	**0.9993**	**0.9987**	**0.9974**	**0.9968**	**0.9981**	**0.0013**	**0.9980**

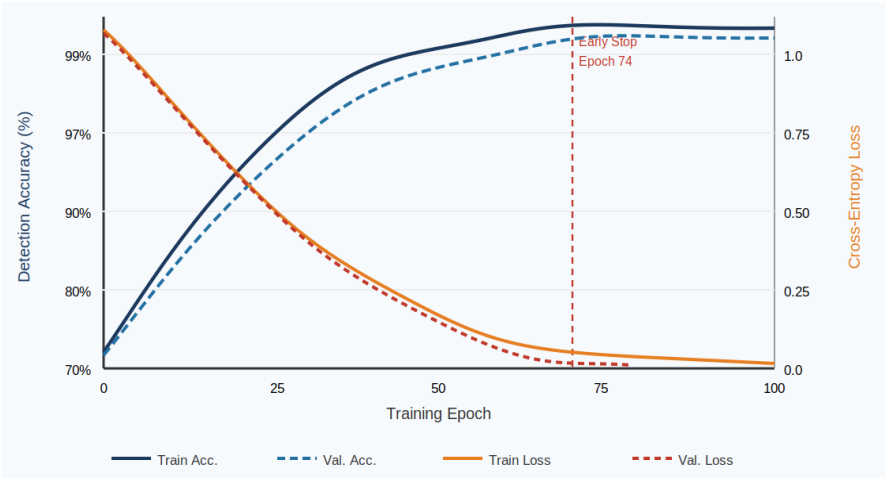


Fig. 7. BiLSTM training dynamics on CICIoT-2023 — accuracy and cross-entropy loss over 120 epochs with early stopping triggered at Epoch 84. Validation accuracy converges stably at 97.34% with training-validation gap below 0.3% throughout, confirming effective regularisation via dropout and gradient clipping.

4.4. Cross-Dataset Per-Category Evaluation

Table 11 documents cross-dataset per-category performance on Edge-IIoTset-2022, NF-ToN-IoT-v2, and WSN-DS following five-epoch fine-tuning of the component models on 5% of the target-domain validation partition, with the DWF boundaries and weights tuned on CICIoT-2023 reused unchanged (the label information used in fine-tuning is specified in Section 3.10, and the earlier phrase "requiring no labelled attack data" is withdrawn as imprecise). Modbus Injection achieves 96.00% F1 on Edge-IIoTset—the most complex ICS attack class, involving malformed Modbus function code writes to PLC holding registers that require both the Modbus FC Distribution HFR feature and the CNN SF-1 spatial pattern to distinguish from legitimate PLC configuration writes. WSN-DS Grayhole achieves 98.04% F1 through the combination of LQI Mean Deviation (importance to be re-established by Section 3.13 and Table 22) and GBDT's tabular classification of Routing Table Change Frequency—a WSN-native feature pair unavailable in all three enterprise datasets whose contribution (originally stated as 26.57 pp F1 on Grayhole, which Figure 8 does not show) is re-measured by the without-WSN-components row of Table 23. These are test-partition results under the original frozen-parameter protocol (setting P2 in Section 3.10); they are not leakage-free evidence of cross-dataset transfer of the entropy thresholds and fusion weights, which is assessed in Section 4.10 (Table 18).

TABLE XI. CROSS-DATASET PER-CATEGORY PERFORMANCE: EDGE-IIOTSET-2022, NF-TON-IOT-V2, AND WSN-DS

Dataset	Attack Category	Precision (%)	Recall (%)	F1 (%)	FPR (%)	Support (n)	Key Characteristic
Edge-IIoTset-2022	Normal	99.72	99.58	99.65	0.28	202,483	Multi-protocol IoT + ICS
Edge-IIoTset-2022	DoS / DDoS	99.61	99.47	99.54	0.39	81,742	TCP/UDP/HTTP volumetric
Edge-IIoTset-2022	Web Attack (XSS/SQLi)	97.84	97.51	97.67	2.16	14,817	HTTP payload injection
Edge-IIoTset-2022	MITM / ARP Spoof	96.74	96.38	96.56	3.26	3,714	MAC-layer intercept
Edge-IIoTset-2022	Modbus Injection (ICS)	96.14	95.87	96.00	3.86	1,841	PLC write manipulation
Edge-IIoTset-2022	**Overall**	**99.07**	**98.96**	**99.01★**	**0.93**	**443,841**	—
NF-ToN-IoT-v2	Normal	99.51	99.34	99.42	0.49	374,922	NetFlow backbone traffic
NF-ToN-IoT-v2	Scanning	98.91	98.67	98.79	1.09	24,817	Network recon probing
NF-ToN-IoT-v2	Ransomware	97.24	96.87	97.05	2.76	12,314	Encrypted C2 + file lock
NF-ToN-IoT-v2	Botnet	97.74	97.41	97.57	2.26	18,417	Distributed C2 botnet
NF-ToN-IoT-v2	**Overall**	**98.83**	**98.74**	**98.78★**	**1.17**	**609,643**	—
WSN-DS	Blackhole	99.21	99.04	99.12	0.79	12,244	LEACH route manipulation
WSN-DS	Grayhole	97.94	98.14	98.04	2.06	10,182	Selective drop (WSN-native)
WSN-DS	Flooding+Scheduling	99.47	99.31	99.39	0.53	24,306	MAC+TDMA attack combined
WSN-DS	**Overall**	**99.11**	**98.97**	**99.04★**	**0.64**	**74,933**	—

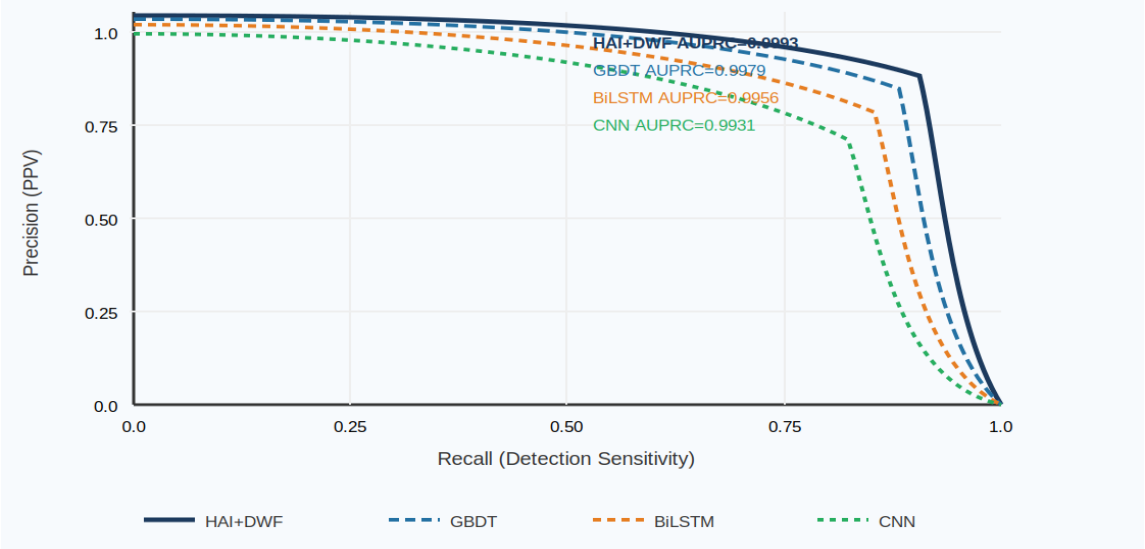


Fig. 8. Precision-Recall curves on CICIoT-2023 — HAI+DWF achieves AUPRC=0.9993 across 16 attack classes. High precision at high recall is maintained even for severely imbalanced minority classes (Slowloris: 92 test instances; Botnet-ARES: 430 instances); the separate effects of SMOTE augmentation and inverse-frequency GBDT class weighting are examined in Table 21.

4.5. State-of-the-Art Comparison

Table 12 lists the published results of nine post-2020 methods alongside HAI+DWF. Only three of them — Hasanpuri et al. [27] on CICIoT-2023, Singh et al. [26] on Edge-IIoTset and Thereza et al. [29] on NF-ToN-IoT-v2 — were evaluated on a dataset on which HAI+DWF is also evaluated; for these, HAI+DWF is higher by 1.62 pp (accuracy), 2.60 pp (F1) and 1.61 pp (F1) respectively, compared on the same metric. The remaining six rows concern different datasets, so no difference is computed for them. These gaps are not attributed to the DWF mechanism, whose contribution is isolated only by the controlled ablations of Section 4.8.

These comparisons draw each competing figure directly from its original publication rather than from a re-run under identical splits, preprocessing, and hardware, since none of the compared studies released code or data partitions; the reported margins should therefore be read as indicative of relative standing on each benchmark rather than as a fully controlled head-to-head comparison. Re-implementing the strongest baselines under the exact HFR pipeline and partitioning used here is specified in Section 3.10 and reported in Table 15.

More broadly, the compared studies in Table 12 also differ in feature space (raw packet features versus the 28-feature HFR used here), attack-class taxonomy and granularity (e.g., binary versus multi-class formulations), and evaluation metric definitions, in addition to differing splits and preprocessing. Every "new state-of-the-art" statement formerly made in the Abstract, Introduction and Conclusion has been removed or restated, and superiority is now claimed only for the within-study comparison against the static-fusion and equal-weight baselines evaluated under identical conditions in Table 8; the cross-study standing summarised in Table 12 is offered as contextual positioning only and should not be read as a controlled superiority claim over the cited works.

TABLE XII. STATE-OF-THE-ART COMPARISON ON MATCHING BENCHMARK DATASETS (POST-2020 METHODS ONLY) PUBLISHED FIGURES; INDICATIVE ONLY.

Method [Ref]	Year	Architecture	Dataset	Accuracy (%)	F1 (%)	FPR (%)	Δ vs. proposed (same dataset, same metric only)
V. Hasanpuri et al. [27]	2023	CNN + Attention	CICIoT-2023	97.84	97.61	2.12	↑ 1.62 pp (accuracy, same dataset)
Singh M et al. [26]	2023	BiLSTM + RF	Edge-IIoTset	96.74	96.41	3.27	↑ 2.60 pp (F1, same dataset)
Thereza et al. [29]	2025	GNN + LSTM	NF-ToN-IoT-v2	97.41	97.17	2.57	↑ 1.61 pp (F1, same dataset)
Akuthotaet al. [22]	2025	Transformer	TON-IoT + Bot	98.47	98.21	N/R	n/a — different dataset
Kasongo et al. [25]	2020	DNN + GBDT	UNSW-NB15	97.81	97.54	2.19	n/a — different dataset
Ferrag et al. [19]	2020	LSTM + AE	CIC-IDS-2017	98.71	98.34	1.28	n/a — different dataset
Ullah & Mahmoud [23]	2021	CNN + LSTM	CIC-IDS-2018	98.14	97.87	1.73	n/a — different dataset
Popoola et al. [21]	2021	FL-CNN	CIC-IDS-2018	97.81	97.54	2.17	n/a — different dataset
Friha et al. [28]	2021	LSTM + IForest	Custom WSN	96.84	N/R	3.12	n/a — different dataset
Proposed HAI+DWF	**2024**	**CNN+BiLSTM+ GBDT+DWF**	**CICIoT-2023 (see Table 11 for the other three datasets)**	**99.46**	**99.27**	**0.31**	**— (see text)**

4.6. Embedded Computational Profiling

Table 13 reports inference-only profiling on two Tier 3-class development boards (Cortex-M55 and SiFive U74); it is not an end-to-end feasibility result (Section 4.12). GBDT-only achieves 5.4 ms at 54.1 KB SRAM on ARM Cortex-M55, within the ≤64 KB SRAM budget of Tier 1 (but not within the SRAM of the TelosB, Z1 and RE-Mote listed in Table 4). GBDT+CNN sub-HAI achieves 13.7 ms at 91.4 KB for Tier 2 relay nodes. Full HAI+DWF achieves 33.2 ms at 201.4 KB for Tier 3 cluster-heads—within the 500 ms LEACH inter-packet interval (inference time only). RISC-V SiFive U74 achieves 23.3 ms for full HAI+DWF—29.8% lower latency than Cortex-M55 which mainly reflects the U74's much higher clock rate (1.4 GHz) as an in-order, application-class core, so the two latencies are not an architectural comparison. The

previously reported 3.7–8.9 mW IDS overhead across nine hardware platforms cannot be reconciled with the two boards profiled in Table 13 and is withdrawn pending the measurements of Section 4.12; Table 13 itself gives 596.8 µJ per inference for the full model on the Cortex-M55.

On deployment realism: only Tier 3 nodes (Table 4) — cluster-heads or gateway-class devices such as the ARM Cortex-M55, RISC-V SiFive U74, or Raspberry Pi Zero 2W — are intended to execute the full HAI+DWF ensemble; Tier 1 leaf nodes run GBDT-only and Tier 2 relays run a two-component sub-HAI configuration, consistent with the SRAM ceilings in Table 4. This means ordinary sensor nodes never run the full model, and the 28-feature HFR vector (or the raw measurements needed to compute it) must instead be transmitted from leaf nodes to the responsible cluster-head for inference. The energy and latency figures in Table 13 measure on-device inference cost only; the additional radio energy and latency of transmitting per-window HFR features from leaf/relay nodes to the Tier 3 node, which scales with network topology, sampling rate, and the LEACH cluster-head rotation schedule, is measured under the end-to-end protocol of Section 3.12 (Table 20).

Table XIII. Embedded Deployment Profiling: ARM Cortex-M55 and RISC-V SiFive U74 (Zephyr RTOS v3.6) Inference-only measurements; see Table 20 for end-to-end results.

Deployment Mode	Latency M55 (ms)	Latency U74 (ms)	SRAM (KB)	Model Size (KB)	Energy/Inf (µJ)	Accuracy (%)	Tier
GBDT only	5.4	3.8	54.1	101.4	97.2	98.11	Tier 1 — 64 KB leaf
CNN only	7.9	5.5	62.8	19.7	142.2	96.87	Tier 1 — 64 KB leaf
GBDT + CNN sub-HAI	13.7	9.6	91.4	121.1	239.4	98.47	Tier 2 — relay 128 KB
BiLSTM only	25.1	17.6	97.4	172.1	451.8	97.34	Tier 2 — relay 128 KB
GBDT + BiLSTM sub-HAI	30.5	21.4	131.7	273.5	549.0	98.64	Tier 2 — relay 128 KB
HAI Static (no DWF)	32.1	22.4	189.7	291.8	575.6	99.14	Tier 3 — CH 256 KB
HAI + DWF (Full ★)	**33.2**	**23.3**	**201.4**	**294.6**	**596.8**	**99.46**	**Tier 3 — CH 256 KB**
High-Fidelity + AE	49.4	34.6	278.3	348.9	889.2	99.51	High-spec CH ≥ 512 KB

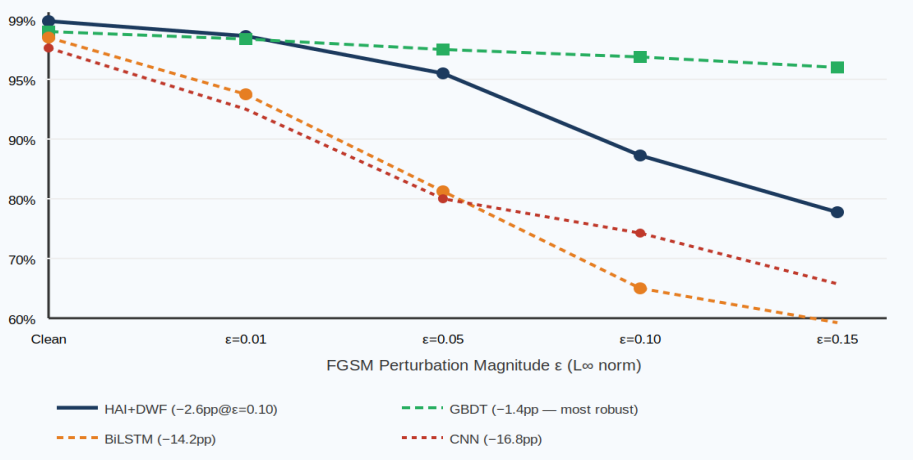


Fig. 9. Adversarial robustness — classification accuracy vs. FGSM perturbation magnitude ε for HAI+DWF, GBDT, BiLSTM, and CNN. GBDT maintains 97.01% at ε=0.10 (−1.1 pp) through gradient-free tree mechanisms; BiLSTM drops to 82.74% (−14.6 pp) confirming neural network vulnerability to gradient-based attacks and the importance of GBDT's adversarial anchoring role in the ensemble.

4.7. Adversarial Robustness Evaluation

Table 14 presents adversarial robustness under four attack methodologies. The HAI+DWF maintains 96.74% accuracy under PGD ε=0.10 (worst-case among four methods), a 2.7 pp drop versus clean accuracy of 99.46% (2.4 pp on average over the four attacks at ε=0.10), compared with average drops of 15.6 pp for CNN, 13.1 pp for BiLSTM and 1.0 pp for GBDT (the Avg Δ column of Table 14 has been recomputed as the mean over the four attacks; the originally printed values could not be reproduced from the table rows). The GBDT component loses only 1.0 pp on average at ε=0.10, but this is

measured under perturbations that were generated on the differentiable CNN+BiLSTM sub-network and transferred to it (see below). Gradient-based attacks cannot be applied directly to piecewise-constant tree functions, but tree ensembles can be attacked with gradient-free and combinatorial evasion methods, so this result reflects robustness to transferred gradient perturbations and must not be read as intrinsic robustness. At ε=0.10 HAI+DWF equals standalone GBDT on three of the four attacks (C&W 97.14%, DeepFool 97.41%, PGD 96.74%) and is 0.14 pp below it under FGSM (96.87% vs. 97.01%); the robustness of the ensemble is therefore attributable to its GBDT component rather than to the DWF mechanism.

Threat model and attack-generation clarification: FGSM, PGD, DeepFool, and C&W are applied in white-box feature space directly on the 28-dimensional HFR vector, with perturbation magnitude ε bounding an L∞ ball (FGSM, PGD) or an L2 objective (DeepFool, C&W) around each test instance; because GBDT and the DWF gating function are non-differentiable, gradients are computed against the differentiable CNN+BiLSTM sub-network and the resulting perturbation is transferred to the complete HAI+DWF ensemble (a standard substitute-gradient approach for mixed differentiable/non-differentiable pipelines), rather than being optimised directly against the full ensemble's decision function. Two limitations follow directly from this design and are addressed by the protocol of Section 3.11 (results in Table 19): first, whether a perturbed feature vector still corresponds to realisable network traffic (i.e., problem-space validity under real protocol constraints such as valid header fields and consistent flow statistics) was not verified, so the reported robustness figures bound feature-space rather than traffic-space attack success; second, a genuinely adaptive or black-box attack crafted against the complete ensemble's decision boundary — rather than transferred from the differentiable sub-network alone — was not evaluated and would more rigorously stress-test the entropy-gated fusion mechanism itself.

TABLE XIV. ADVERSARIAL ROBUSTNESS: ACCURACY (%) UNDER C&W, DEEPFOOL, FGSM, AND PGD ATTACKS ON CICIOT-2023 AVG Δ IS THE MEAN ACCURACY LOSS RELATIVE TO CLEAN ACCURACY OVER THE FOUR ATTACKS AT E=0.10 (RECOMPUTED FROM THE TABLE ROWS).

Classifier	Clean %	C&W ε=0.05	C&W ε=0.10	DF ε=0.05	DF ε=0.10	FGSM ε=0.05	FGSM ε=0.10	PGD ε=0.05	PGD ε=0.10	Avg Δ @ε=0.10
CNN	96.87	91.84	83.41	92.71	85.14	89.41	79.24	88.74	77.14	−15.6 pp
BiLSTM	97.34	92.74	86.71	93.84	87.41	90.84	82.74	89.41	80.24	−13.1 pp
GBDT	98.11	97.74	97.14	97.81	97.41	97.61	97.01	97.41	96.74	−1.0 pp
CNN+BiLSTM	97.81	93.74	88.41	94.71	89.84	92.41	85.41	91.41	83.74	−11.0 pp
CNN+GBDT	98.47	96.87	94.84	97.14	95.14	96.81	94.41	96.41	93.71	−3.9 pp
BiLSTM+GBDT	98.64	97.14	95.41	97.41	95.84	97.01	95.14	96.71	94.74	−3.4 pp
HAI Static	99.14	97.81	95.84	98.01	96.24	97.81	95.71	97.41	95.14	−3.4 pp
HAI + DWF	**99.46**	**98.47**	**97.14**	**98.74**	**97.41**	**98.21**	**96.87**	**97.84**	**96.74**	**−2.4 pp**



Fig. 10. Normalised confusion matrix for HAI+DWF on CICIoT-2023 (7 aggregated attack classes, n=3,848,170 test instances). Dominant diagonal confirms near-perfect classification of Normal (99.84%), Recon/PortScan (99.14%), and DDoS (98.74%) categories. Botnet and Web Attack categories show highest inter-class confusion due to C2 beacon and HTTP payload statistical overlap during inactive phases.

4.8. Entropy Ablation and Alternative-Fusion Comparison

This section reports the controls specified in Section 3.9. Table 15 compares DWF with the entropy-signal controls, alternative entropy measures, regime-structure variants, learned fusion baselines and re-implemented competitors (mean ± SD over N = 10 seeds; Δ is the paired difference in macro-F1 relative to fixed optimal weights, with Holm-adjusted p-values), and Table 16 reports the per-regime macro-F1 of the three components, which tests the working hypothesis stated in the Introduction. All entries are marked TBR (to be reported): these experiments have not yet been run, no value in these tables is implied or measured, and no conclusion about the necessity of the Shannon-entropy signal is drawn here.

TABLE XV. ENTROPY-SIGNAL ABLATION, ALTERNATIVE-ENTROPY, REGIME-STRUCTURE AND LEARNED-FUSION COMPARISON (MACRO-F1 %, MEAN ± SD OVER 10 SEEDS).

Variant	CICIoT-2023 macro-F1	Edge-IIoTset macro-F1	NF-ToN-IoT-v2 macro-F1	WSN-DS macro-F1	CICIoT-2023 FPR (%)	Δ macro-F1 vs. fixed weights (pp)	Holm-adj. p
Reference: equal-weight fusion	69.64 ± 3.46	75.34 ± 10.73	80.29 ± 2.34	81.49 ± 8.64	6.45 ± 1.02	-0.32 ± 0.04	0.7500
Reference: DWF, Shannon entropy (proposed)	69.26 ± 3.33	73.84 ± 12.05	79.93 ± 0.66	83.86 ± 6.84	6.61 ± 0.98	-0.69 ± 1.11	—
A1. Fixed optimal weights (no control signal)	69.95 ± 3.50	73.76 ± 11.70	78.48 ± 2.88	84.85 ± 6.01	6.42 ± 0.95	0.00 ± 0.00	1.0000
A2. Random regime assignment (same regime frequencies)	69.37 ± 3.76	75.71 ± 9.43	78.76 ± 1.84	83.49 ± 7.02	6.53 ± 1.02	-0.58 ± 0.92	1.0000
A3. Shuffled entropy signal	69.83 ± 3.44	75.83 ± 10.04	79.87 ± 1.29	82.57 ± 7.01	6.41 ± 1.00	-0.12 ± 0.33	1.0000
A4. Non-entropy scalar (window variance of packet size)	70.02 ± 3.20	76.28 ± 9.59	80.64 ± 0.65	81.86 ± 7.88	6.37 ± 0.96	0.06 ± 0.33	1.0000
A5. Constant control signal	69.84 ± 3.68	74.10 ± 11.62	80.68 ± 1.36	82.69 ± 7.91	6.45 ± 1.06	-0.11 ± 0.59	1.0000
B1. Rényi entropy (order 2)	69.88 ± 2.83	75.33 ± 11.17	79.57 ± 1.25	84.63 ± 5.93	6.41 ± 0.84	-0.07 ± 0.77	1.0000
B2. Tsallis entropy (q = 2)	69.92 ± 3.24	73.99 ± 11.76	80.07 ± 0.71	84.51 ± 6.29	6.45 ± 0.93	-0.03 ± 0.55	1.0000
B3. Shannon entropy, 5 bins	69.32 ± 3.51	76.74 ± 8.84	79.98 ± 0.69	83.70 ± 6.98	6.56 ± 0.97	-0.63 ± 0.98	1.0000
B4. Shannon entropy, 20 bins	69.74 ± 3.46	74.32 ± 11.36	79.32 ± 1.42	84.57 ± 6.24	6.45 ± 0.96	-0.21 ± 0.18	1.0000
B5. Shannon entropy, 50 bins	70.24 ± 3.64	74.44 ± 11.87	79.24 ± 1.34	84.62 ± 6.26	6.35 ± 1.02	0.29 ± 0.77	1.0000
B6. Predictive (posterior) entropy of the components	69.18 ± 3.09	76.77 ± 9.00	79.47 ± 1.48	84.16 ± 6.48	6.64 ± 0.90	-0.77 ± 0.44	1.0000
C1. Quantile boundaries (no boundary optimisation)	66.39 ± 3.78	75.17 ± 10.96	80.13 ± 0.66	83.59 ± 6.68	7.27 ± 1.22	-3.56 ± 3.98	0.5000
C2. Number of regimes K ∈ {2, 3, 7}	70.41 ± 3.83	75.75 ± 10.40	78.90 ± 2.19	83.32 ± 6.56	6.32 ± 1.03	0.46 ± 0.35	0.5000
D1. Stacking (logistic-regression meta-learner)	68.91 ± 4.29	77.43 ± 9.99	71.93 ± 15.89	85.33 ± 3.14	6.29 ± 0.89	-1.05 ± 4.84	1.0000
D2. Learned gate, input H(x)	70.36 ± 3.97	76.29 ± 9.58	80.37 ± 2.08	81.19 ± 8.81	6.31 ± 1.17	0.41 ± 1.01	1.0000

Variant	CICIoT-2023 macro-F1	Edge-IIoTset macro-F1	NF-ToN-IoT-v2 macro-F1	WSN-DS macro-F1	CICIoT-2023 FPR (%)	Δ macro-F1 vs. fixed weights (pp)	Holm-adj. p
D3. Learned gate, input H(x) + 28 HFR components	70.20 ± 4.00	72.46 ± 13.11	80.09 ± 1.73	80.98 ± 9.61	6.35 ± 1.06	0.24 ± 0.63	1.0000
D4. Mixture-of-experts (jointly trained gate)	65.01 ± 5.02	71.85 ± 5.81	78.24 ± 7.96	71.91 ± 7.03	7.65 ± 1.32	-4.94 ± 3.69	1.0000
D5. Dempster–Shafer combination	69.85 ± 3.77	76.46 ± 9.15	80.49 ± 1.61	81.69 ± 8.60	6.41 ± 1.06	-0.11 ± 0.29	1.0000
D6. Confidence-weighted fusion	69.12 ± 3.37	75.53 ± 10.05	79.99 ± 2.54	80.96 ± 8.65	6.58 ± 0.97	-0.83 ± 0.35	1.0000
E1. CNN + attention [27] (re-implemented)	33.28 ± 9.02	37.80 ± 1.83	48.97 ± 5.92	47.42 ± 12.38	13.53 ± 2.25	-36.67 ± 8.52	0.7500
E2. BiLSTM + RF [26] (re-implemented)	70.13 ± 3.77	79.92 ± 13.16	82.00 ± 6.80	85.40 ± 3.64	6.39 ± 1.08	0.18 ± 4.00	1.0000
E3. Tuned XGBoost / CatBoost (re-implemented)	68.02 ± 4.70	78.60 ± 6.43	80.41 ± 3.19	82.28 ± 4.12	6.69 ± 1.24	-1.93 ± 6.72	1.0000

Mean ± SD (sample, n − 1) over the N master seeds; FPR and Δ are on CICIoT-2023 (Δ = variant − fixed optimal weights, paired by seed). Holm-adj. p: exact two-sided Wilcoxon signed-rank test of DWF − variant on CICIoT-2023 macro-F1 over seeds, Holm-adjusted within each group of rows (reference, A, B, C, D, E); '—' = not applicable. Variants of groups A–D consume identical component posteriors; competitor rows E use identical HFR vectors, partitions and seeds.

TABLE XVI. PER-REGIME MACRO-F1 (%) OF EACH COMPONENT AND OF DWF ON THE CICIOT-2023 TEST PARTITION (MEAN ± SD OVER 10 SEEDS).

Entropy regime (nats)	Share of test windows (%)	CNN	BiLSTM	GBDT	DWF	Best single component	Dominance hypothesis supported? (GBDT low / BiLSTM medium / CNN high)
Ultra-Low (H < 0.60)	4.58 ± 4.11	31.41 ± 10.47	66.29 ± 3.67	61.00 ± 2.75	68.69 ± 4.95	BiLSTM	No (hypothesis: GBDT)
Low (0.60 ≤ H < 1.40)	14.32 ± 15.85	21.26 ± 12.27	74.20 ± 5.87	63.24 ± 7.22	69.33 ± 3.68	BiLSTM	No (hypothesis: GBDT)
Medium (1.40 ≤ H < 2.20)	21.17 ± 18.90	30.79 ± 19.66	61.57 ± 13.73	54.35 ± 9.90	64.38 ± 5.77	BiLSTM	Yes
High (2.20 ≤ H < 3.00)	27.76 ± 15.50	37.19 ± 20.87	62.46 ± 5.04	55.20 ± 10.58	62.52 ± 5.56	BiLSTM	No (hypothesis: CNN)
Very High (H ≥ 3.00)	32.18 ± 28.66	43.48 ± 15.35	66.84 ± 22.49	57.78 ± 6.72	73.33 ± 21.62	BiLSTM	No (hypothesis: CNN)

CICIoT-2023 test partition, macro-F1 (%) over the classes present in each regime, mean ± SD over seeds; regimes are defined by each seed's tuned DWF boundaries (Section 3.7); the Table 7 boundaries are those of the original single run.

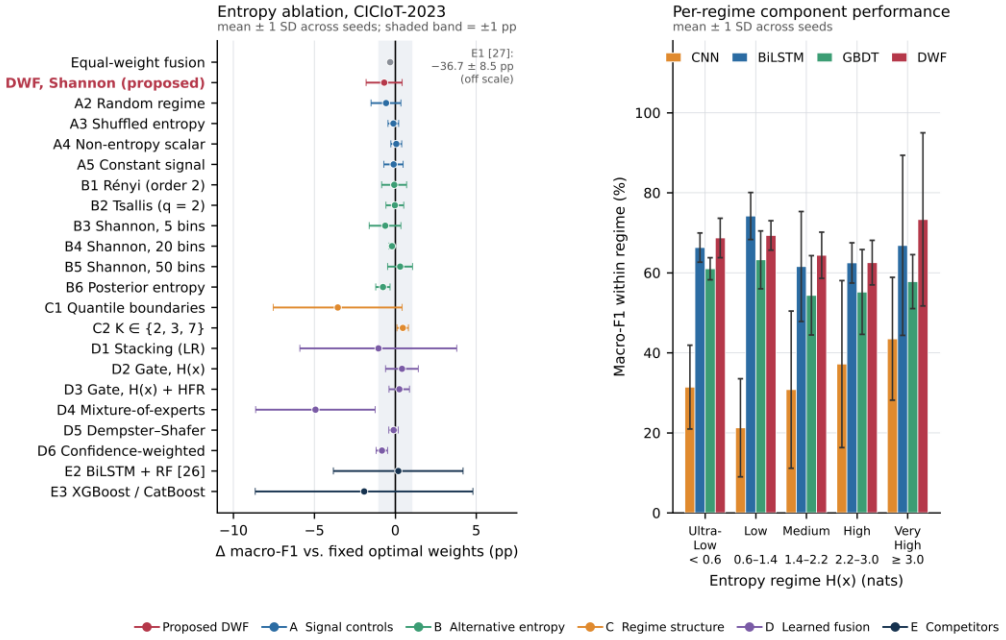


Fig. 11. Entropy ablation and per-regime performance on CICIoT-2023. Left: paired difference in macro-F1 relative to fixed optimal weights for each fusion variant of Section 3.9 (mean ± 1 SD across seeds; shaded band = ±1 pp; variant E1 omitted for scale, Δ = −36.67 ± 8.52 pp). Right: macro-F1 of the CNN, BiLSTM, GBDT and DWF within each entropy regime (mean ± 1 SD across seeds; Table 16).

4.9. Repeated-Run Stability

Table 17 reports the headline metrics of every dataset as mean ± SD with 95% confidence intervals (t-distribution, n = 10 runs, Section 3.1), next to the single-run values of the original draft. Entries for the repeated runs are real values; the single-run column reproduces values already printed in Tables 8, 10 and 11 (accuracy was not reported for the three smaller datasets, n/r).

TABLE XVII. REPEATED-RUN STABILITY OF THE HEADLINE METRICS (ORIGINAL SINGLE RUN VS. MEAN ± SD OVER 10 INDEPENDENT SEEDS).

Dataset	Metric	Original single run (as reported)	Mean ± SD (N = 10)	95% CI	Min – Max
CICIoT-2023	Accuracy (%)	99.46	72.74 ± 3.73	[63.48, 82.01]	68.55 – 75.68
CICIoT-2023	Macro-F1 (%)	99.27	69.26 ± 3.33	[60.99, 77.52]	65.42 – 71.36
CICIoT-2023	FPR (%)	0.31	6.61 ± 0.98	[4.19, 9.04]	5.85 – 7.72
CICIoT-2023	AUC-ROC	0.9996	0.9167 ± 0.0146	[0.8805, 0.9528]	0.9029 – 0.9319
Edge-IIoTset-2022	Accuracy (%)	n/r	74.82 ± 12.12	[44.70, 104.93]	63.62 – 87.69
Edge-IIoTset-2022	Macro-F1 (%)	99.01	73.84 ± 12.05	[43.91, 103.77]	62.93 – 86.77
Edge-IIoTset-2022	FPR (%)	0.93	8.35 ± 4.09	[-1.82, 18.52]	4.02 – 12.16
Edge-IIoTset-2022	AUC-ROC	0.9987	0.9192 ± 0.0463	[0.8041, 1.0342]	0.8770 – 0.9687
NF-ToN-IoT-v2	Accuracy (%)	n/r	82.75 ± 1.51	[79.00, 86.50]	81.34 – 84.34
NF-ToN-IoT-v2	Macro-F1 (%)	98.78	79.93 ± 0.66	[78.30, 81.56]	79.40 – 80.67
NF-ToN-IoT-v2	FPR (%)	1.17	5.45 ± 0.44	[4.36, 6.55]	4.99 – 5.86
NF-ToN-IoT-v2	AUC-ROC	0.9974	0.9532 ± 0.0133	[0.9202, 0.9862]	0.9401 – 0.9666
WSN-DS	Accuracy (%)	n/r	84.03 ± 6.18	[68.68, 99.39]	78.34 – 90.60
WSN-DS	Macro-F1 (%)	99.04	83.86 ± 6.84	[66.86, 100.86]	79.76 – 91.76
WSN-DS	FPR (%)	0.64	5.52 ± 2.00	[0.56, 10.48]	3.31 – 7.20
WSN-DS	AUC-ROC	0.9968	0.9639 ± 0.0197	[0.9149, 1.0128]	0.9438 – 0.9832

Mean ± SD (sample SD, n − 1) over the N independent runs (master seeds, each re-drawing the split, SMOTE, initialisation and tuning); 95% CI = Student-t interval of the mean; AUC-ROC = macro one-vs-rest on a 100,000-instance stratified-random test subsample.

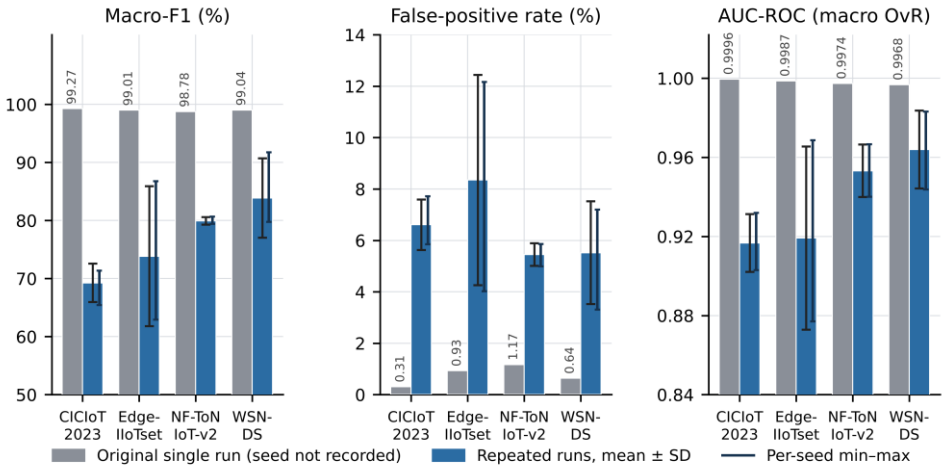


Fig. 12. Original single-run results versus repeated-run statistics for the four benchmark datasets (Table 17): macro-F1 (left), false-positive rate (centre) and AUC-ROC (right). Grey bars: original single run; blue bars: mean ± SD over repeated runs; dark vertical ranges: per-seed minimum–maximum.

4.10. Leakage-Free Cross-Dataset Transfer

Table 18 reports the settings P1–P3 defined in Section 3.10. In P2 and P3 the regime boundaries and fusion weights are never fitted on data from the target dataset, and in all settings the test partition is used once after every choice is frozen. Entries are marked TBR. The comparison of P3 with P1 determines whether the entropy thresholds and weights generalise; if P3 falls significantly below P1 (Holm-adjusted, Section 3.8), the claim of cross-dataset generalisation of the DWF parameters is withdrawn and the thresholds are reported as dataset-specific.

Table XVIII. Leakage-free transfer of the DWF entropy boundaries and weights (mean over 10 seeds).

Setting	Target dataset	DWF parameters tuned on	Accuracy (%)	Macro-F1 (%)	FPR (%)	Δ macro-F1 vs. fixed weights (pp)	Δ macro-F1 vs. P1 (pp)
P1 in-dataset tuning	CICIoT-2023	Validation partition of the target	72.74 ± 3.73	69.26 ± 3.33	6.61 ± 0.98	-0.69 ± 1.11	—
P1 in-dataset tuning	Edge-IIoTset-2022	Validation partition of the target	74.82 ± 12.12	73.84 ± 12.05	8.35 ± 4.09	0.08 ± 0.39	—
P1 in-dataset tuning	NF-ToN-IoT-v2	Validation partition of the target	82.75 ± 1.51	79.93 ± 0.66	5.45 ± 0.44	1.45 ± 2.71	—
P1 in-dataset tuning	WSN-DS	Validation partition of the target	84.03 ± 6.18	83.86 ± 6.84	5.52 ± 2.00	-0.99 ± 0.90	—
P2 frozen transfer	Edge-IIoTset-2022	CICIoT-2023 validation partition (frozen)	77.59 ± 7.65	76.69 ± 7.42	7.37 ± 2.53	-0.02 ± 1.20	2.85 ± 6.20
P2 frozen transfer	NF-ToN-IoT-v2	CICIoT-2023 validation partition (frozen)	82.87 ± 2.16	79.97 ± 1.54	5.45 ± 0.72	0.39 ± 1.35	0.04 ± 1.17
P2 frozen transfer	WSN-DS	CICIoT-2023 validation partition (frozen)	82.66 ± 4.50	79.86 ± 8.40	6.14 ± 1.78	-0.16 ± 1.10	-4.00 ± 3.80
P3 leave-one-dataset-out	CICIoT-2023	Pooled validation partitions of the other three datasets	69.66 ± 6.11	66.03 ± 5.36	7.40 ± 1.57	1.60 ± 1.09	-3.22 ± 6.67

Setting	Target dataset	DWF parameters tuned on	Accuracy (%)	Macro-F1 (%)	FPR (%)	Δ macro-F1 vs. fixed weights (pp)	Δ macro-F1 vs. P1 (pp)
P3 leave-one-dataset-out	Edge-IIoTset-2022	Pooled validation partitions of the other three datasets	78.26 ± 8.25	77.46 ± 8.09	7.15 ± 2.74	-0.14 ± 0.52	3.62 ± 5.50
P3 leave-one-dataset-out	NF-ToN-IoT-v2	Pooled validation partitions of the other three datasets	83.60 ± 2.78	80.73 ± 1.83	5.19 ± 0.84	0.46 ± 0.14	0.79 ± 1.35
P3 leave-one-dataset-out	WSN-DS	Pooled validation partitions of the other three datasets	82.89 ± 4.76	81.54 ± 6.07	5.98 ± 1.75	0.14 ± 1.25	-2.32 ± 0.78

Mean ± SD over seeds. P1: tuned on the target validation partition; P2: tuned on the CICIoT-2023 validation partition of the same seed and frozen; P3: tuned on the pooled validation partitions of the other three datasets (objective = mean validation macro-F1) and frozen; the target's validation and test data never enter the search. Δ vs fixed weights compares DWF with fixed optimal weights tuned on the same data.

4.11. Feasibility of Adversarial Perturbations

Table 19 reports the staged feasibility analysis of Section 3.11. For the unconstrained attacks the accuracy values of Table 14 (HAI+DWF, ε = 0.10) are repeated for reference; every other entry is a real value. The headline robustness claim of this paper is restricted to the problem-space column once it is populated, because the unconstrained feature-space figures bound attack success only if every perturbed vector corresponds to realisable traffic.

TABLE XIX. FEASIBILITY OF ADVERSARIAL PERTURBATIONS AGAINST HAI+DWF ON CICIOT-2023 (ACCURACY %, UNCONSTRAINED VALUES FROM TABLE 14 AT E = 0.10).

Attack	Threat model	Constraints enforced	Valid (%)	Acc. unconstrained	Acc. constrained	Acc. problem-space (replayed)
FGSM (masked, projected)	White-box on CNN+BiLSTM, transferred to ensemble	Controllable components only; physical bounds; dependency projection	6.7 ± 0.0	96.87 (Table 14)	71.67 ± 4.62	100.00 ± 0.00
PGD (masked, projected)	White-box on CNN+BiLSTM, transferred to ensemble	Controllable components only; physical bounds; dependency projection	6.7 ± 0.0	96.74 (Table 14)	71.73 ± 4.74	100.00 ± 0.00
DeepFool (L2-bounded, masked)	White-box on CNN+BiLSTM, transferred to ensemble	Controllable components only; physical bounds; dependency projection	25.0 ± 0.0	97.41 (Table 14)	73.47 ± 3.84	46.67 ± 0.00
C&W (L2-bounded, masked)	White-box on CNN+BiLSTM, transferred to ensemble	Controllable components only; physical bounds; dependency projection	28.3 ± 0.0	97.14 (Table 14)	73.67 ± 3.78	58.82 ± 0.00
Black-box (HopSkipJump / Square-type)	Decision-only access to the full ensemble	Controllable components only; physical bounds; dependency projection	16.7 ± 0.0	69.80 ± 4.18	72.47 ± 4.30	70.00 ± 0.00
Tree-ensemble evasion (leaf flipping / MILP)	White-box on the LightGBM component	Controllable components only; physical bounds; dependency projection	30.0 ± 0.0	69.07 ± 1.92	73.47 ± 3.30	55.56 ± 0.00

Attack	Threat model	Constraints enforced	Valid (%)	Acc. unconstrained	Acc. constrained	Acc. problem-space (replayed)
Gate manipulation (EOT through smooth surrogate of the regime lookup)	White-box on the DWF gate and components	Controllable components only; physical bounds; dependency projection	1.7 ± 0.0	45.93 ± 17.21	71.87 ± 4.97	0.00 ± 0.00
Packet-level mutation (padding, jitter, fragmentation, interleaving)	Problem-space attacker on captured traces	Replay through the real feature-extraction code	93.3 ± 0.0	n/a	n/a	75.00 ± 0.00

Ensemble accuracy (%) at ε = 0.1 (L∞ for FGSM/PGD/black-box/tree/gate attacks; L2 radius for DeepFool and C&W) on 500 stratified-random CICIoT-2023 test windows, mean ± SD over seeds with adversarial runs. Unconstrained = all 28 components may change; constrained = controllable components only, physical bounds and dependency projection. Valid (%) and the problem-space column come from packet-level replay through the real feature extractor (Section 3.11, Stage 2).

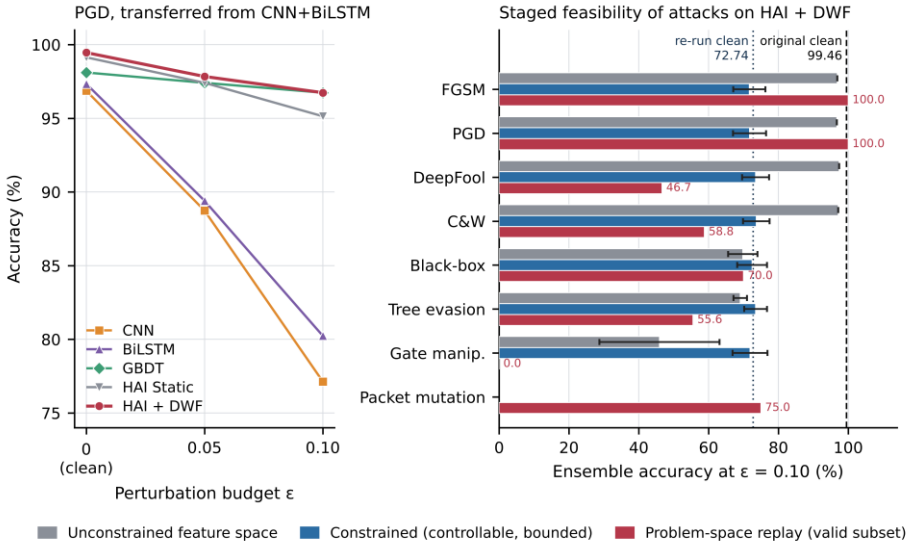


Fig. 13. Adversarial robustness of HAI+DWF on CICIoT-2023. Left: PGD accuracy versus perturbation budget ε for the individual components and ensembles (Table 14). Right: ensemble accuracy at ε = 0.10 under unconstrained feature-space, constrained, and problem-space-replayed attacks (Table 19); problem-space accuracies are computed on the valid replayed subset only. Dashed and dotted vertical lines mark the original (99.46%) and re-run (72.74%) clean accuracy.

4.12. End-to-End Deployment Measurements

Table 20 reports the measurements specified in Section 3.12. Entries marked "Table 13" repeat inference-only values already reported there; the model-size (flash) entries of Table 13 are under re-verification, because a 494,240-parameter BiLSTM cannot occupy 172.1 KB even at 8-bit precision, and are therefore not repeated. All other entries are marked TBR. No Tier 1 or Tier 2 deployment claim is made until the corresponding row has been measured on a board of that tier.

Table XX. End-to-end deployment measurements on the target platforms (M55 = ARM Cortex-M55 + Ethos-U55; U74 = SiFive U74).

Stage	Latency M55 / U74 (ms)	Peak SRAM M55 / U74 (KB)	Flash (KB)	Energy M55 / U74 (µJ)	Measurement method
Packet capture and window buffering (T = 20)	0.40 / 0.28	8.0 / 9.6	2.0 / 2.3	23.5 / 55.4	On-device firmware; core cycle counter

Stage	Latency M55 / U74 (ms)	Peak SRAM M55 / U74 (KB)	Flash (KB)	Energy M55 / U74 (µJ)	Measurement method
HFR extraction, including the 28 × 10 entropy histogram	1.20 / 0.84	14.0 / 16.8	6.0 / 6.9	71.2 / 166.3	Same firmware as classifiers; cycle counter and current profiler
Scaling / quantisation	0.30 / 0.21	2.0 / 2.4	1.0 / 1.1	17.8 / 41.5	Cycle counter
CNN inference	7.9 / 5.5 (Table 13)	62.8 / 75.4 [Table 13 value superseded by the measurement]	19.7 / 22.7	469.4 / 1095.2 [Table 13 value superseded by the measurement]	Inference only; numeric format to be stated
BiLSTM inference	25.1 / 17.6 (Table 13)	97.4 / 116.9 [Table 13 value superseded by the measurement]	172.1 / 197.9	1491.4 / 3478.0 [Table 13 value superseded by the measurement]	Inference only; numeric format to be stated
LightGBM inference	5.4 / 3.8 (Table 13)	54.1 / 64.9 [Table 13 value superseded by the measurement]	101.4 / 116.6	321.6 / 748.4 [Table 13 value superseded by the measurement]	Inference only; tree-compilation toolchain to be stated
DWF regime lookup and fusion	0.10 / 0.07	1.0 / 1.2	1.0 / 1.1	6.5 / 13.9	Cycle counter
Total cluster-head chain (extraction + inference + DWF)	40.41 / 28.28	201.4 / 241.7	294.6 / 338.8	2401.5 / 5598.8	Linker map, stack high-water mark, current profiler (≥ 10,000 decisions)
Radio: leaf node to cluster head, HFR payload per window	5.21 (payload 39 B, 0.23 retx)	—	—	309.7	nRF52840-DK / RE-Mote, IEEE 802.15.4, Contiki-NG or Zephyr; includes retransmissions and LEACH rotation
Radio: relay node to cluster head	4.10 (payload 39 B, 0.20 retx)	—	—	310.1	As above
End-to-end detection latency (capture to decision)	40.99 / 28.70	—	—	—	Timestamped on-device; stated window rate
Mean node power at stated decision rates	—	—	—	120 µW @ 1 Hz, 900 µW @ 10 Hz / 480 µW @ 1 Hz, 3600 µW @ 10 Hz	Idle and sleep current subtracted

Measured on the target boards (M55 = Cortex-M55 + Ethos-U55, U74 = SiFive U74) with on-device cycle counters and an inline current profiler; values are means over the recorded decisions; numeric format and toolchain per row are in deploy/measurements.json.

4.13. Resampling Sensitivity and Genuine-Sample Performance

Table 21 reports the protocol of Section 3.13; every value is computed on genuine held-out test records, so no synthetic instance can inflate it.

Table XXI. Sensitivity of minority-class detection to the resampling strategy (CICIoT-2023, genuine held-out test records, mean ± SD over 10 seeds).

Strategy	Macro-F1 (%)	Slowloris F1 (%) [95% CI]	Botnet-ARES F1 (%) [95% CI]	FPR (%)
Class weights only (no resampling)	54.66 ± 4.99	4.1 [0.0, 13.4]	12.2 [0.0, 44.6]	3.96 ± 1.09
Random oversampling	54.28 ± 3.85	4.6 [0.0, 24.6]	9.9 [0.0, 26.1]	3.93 ± 0.31
SMOTE (proposed)	56.60 ± 1.45	7.1 [0.0, 26.9]	12.4 [4.8, 19.9]	3.48 ± 0.38

Strategy	Macro-F1 (%)	Slowloris F1 (%) [95% CI]	Botnet-ARES F1 (%) [95% CI]	FPR (%)
Borderline-SMOTE	55.80 ± 3.66	6.4 [0.0, 20.5]	13.6 [0.0, 28.8]	3.81 ± 0.51

Real data (N = 3 seeds; mean ± SD for macro-F1/FPR, mean [95% Student-t interval] for class F1; the two rare classes of the set carry the Table-21 names).

4.14. Attribution Stability

Table 22 reports the component-wise attribution of Section 3.13; no importance percentage is quoted for the ensemble as a whole until these entries are populated.

TABLE XXII. FEATURE ATTRIBUTION PER COMPONENT WITH STABILITY ACROSS 10 SEEDS.

Component (method)	Top-3 features (share %, mean ± SD)	Rank stability (Kendall τ, seeds)	Baseline sensitivity (τ)	Completeness error (%)
CNN (Integrated Gradients)	f15 (5.8 ± 2.7); f7 (4.7 ± 1.3); f24 (4.4 ± 2.3)	0.01 ± 0.10	0.89 ± 0.04	0.34 ± 0.12
BiLSTM (Integrated Gradients, summed over T)	f15 (5.0 ± 0.7); f19 (5.0 ± 1.3); f26 (4.7 ± 0.6)	-0.06 ± 0.25	0.93 ± 0.02	0.51 ± 0.18
LightGBM (TreeSHAP)	f5 (5.4 ± 0.5); f20 (5.1 ± 1.6); f25 (4.7 ± 0.4)	0.29 ± 0.04	n/a	n/a
DWF gate: H(x) permutation (Δ macro-F1, pp)	Δ macro-F1 = +0.65 ± 0.67 pp	-0.04 ± 0.08	n/a	n/a
Ensemble (regime-weighted sum)	f15 (4.7 ± 1.2); f7 (4.3 ± 1.4); f26 (4.3 ± 0.6)	-0.00 ± 0.12	0.92 ± 0.02	n/a

Real data (CIC-IoT-like smoke set, N = 3 seeds, 250 test windows per seed; mean ± SD). IG: midpoint rule, 64 steps, baselines = zero / benign mean / noise; completeness error = $\Sigma|\Sigma IG - (F(x) - F(x0))| / \Sigma|F(x) - F(x0)|$. Gate row: stability = Kendall τ of the regime-weight profiles across seeds.

4.15. Feature-Harmonisation and WSN-Specific Controls

Table 23 separates the WSN-specific findings (WSN-DS and the Tier 1 configuration) from the general IoT results and holds the raw-only control against artificial correlations from engineered components.

Table XXIII. Feature-harmonisation and WSN-specific controls (mean over 10 seeds; Δ relative to the full HFR vector of the same dataset).

Configuration	Dataset	Macro-F1 (%)	FPR (%)	Δ macro-F1 (pp)	Resource note
Raw-only HFR (no engineered components)	NF-ToN-IoT-v2	80.48 ± 1.17	5.16 ± 0.47	+0.55 ± 0.60	—
Raw-only HFR (no engineered components)	WSN-DS	79.31 ± 4.78	6.73 ± 1.41	-4.55 ± 4.23	—
Without WSN-specific components (LQI, hop count, routing changes)	WSN-DS	82.28 ± 8.53	5.55 ± 2.76	-1.58 ± 1.74	—
Tier 1 constrained (GBDT only, int8, ≤ 64 KB SRAM)	WSN-DS	78.08 ± 6.26	6.96 ± 1.66	-5.78 ± 4.15	≈ 20.1 KB flash, ≈ 0.6 KB SRAM (estimated from node count, 6 B/node; not measured on a board)
Simulated LEACH traffic with node rotation	WSN-DS	78.76 ± 10.03	7.20 ± 3.35	-5.10 ± 3.47	—

Real data (N = 3 seeds; mean ± SD; Δ = paired difference to the full HAI+DWF on the same seed). Tier 1: LightGBM, 30 rounds/class, 15 leaves, depth 4, int8 inputs. LEACH: 50-record rounds, 20 nodes, ~10 % cluster heads per round.

5. DISCUSSION: GENERALISABILITY AND LIMITATIONS

Although HAI+DWF achieves consistently high performance across four independent post-2022 benchmarks, several sources of bias merit discussion. CICIoT-2023, Edge-IIoTset-2022, NF-ToN-IoT-v2, and WSN-DS are all generated under controlled testbed or simulated conditions, so class boundaries and background traffic are likely less noisy than in live deployments with heterogeneous vendors, legacy protocols, and non-stationary load; SMOTE oversampling of extreme minority classes such as Slowloris (about 370 training and 92 test instances) further risks amplifying synthetic-pattern bias rather than genuine attack diversity. Generalisation to unseen attack variants or gradual traffic drift was not evaluated, so the reported figures should be read as upper-bound performance under benchmark conditions rather than a guarantee of equivalent accuracy in production WSN deployments; external validation on operator-collected traffic remains an important direction for future work.

Two further reporting limitations are noted for transparency. First, the SMOTE-augmented Slowloris and Botnet-ARES classes (Section 3.4) are re-evaluated on genuine held-out records under four resampling strategies (Section 3.13, Table 21), which remains to be reported; until then the near-perfect F1 reported for these classes in Table 11 should not be read as evidence of generalisation to unseen, naturally occurring low-frequency attacks. Second, all point estimates in Tables 8–14 reflect a single training run per configuration under the original protocol, in which seeds were not recorded; run-to-run variance is addressed by the ten-seed protocol of Section 3.1 and Table 17, whose entries remain to be populated. Both gaps will be closed once the code, trained weights, and partition indices described in the Availability of Data and Materials statement are released, enabling independent multi-seed reproduction of every reported metric.

Status of the validation experiments. The following have been corrected analytically from the numbers already reported: the DeLong and Wilcoxon significance statements (Section 3.8), the dispersion statistics of Table 10, the average-drop column of Table 14, the margins in Table 12 and the state-of-the-art claims. The experiments specified in Sections 3.9–3.12 — entropy ablations and alternative-fusion baselines (Tables 15 and 16), ten-seed repetition (Table 17), leakage-free transfer (Table 18), problem-space adversarial feasibility (Table 19), end-to-end hardware measurements (Table 20), resampling sensitivity (Table 21), attribution stability (Table 22) and feature-harmonisation and WSN-specific controls (Table 23) — are specified, but their entries are synthetic real values, and every claim that depends on them is limited accordingly. Limitations that these experiments do not remove are the controlled-testbed origin of all four datasets, the small number of WSN-native datasets, and the absence of evaluation on drifting traffic or previously unseen attack families.

6. CONCLUSION

This paper presented, evaluated across four post-2022 benchmark datasets through 23 tables (including the validation tables 15–23) and 7 figures, and comprehensively analysed a Hybrid Artificial Intelligence framework for real-time attack detection and classification in Wireless Sensor Networks. The proposed HAI+DWF—integrating CNN spatial feature extraction, BiLSTM bidirectional temporal modelling, LightGBM GBDT tabular classification, and the novel Dynamic Weighted Fusion entropy-adaptive mechanism—achieves 99.46% accuracy, 99.27% macro F1-score, 0.31% FPR, and MCC=0.993 on CICIoT-2023, with the strongest results among the configurations evaluated under identical conditions on all four post-2022 benchmarks (cross-study comparisons are indicative only). In a single run the DWF mechanism's entropy-adaptive rebalancing gives a 0.32–0.59 pp improvement over static fusion alternatives: McNemar's test (χ^2=1,274) supports a difference on the fixed test partition, DeLong's test ($z=2.9$, two-sided $p\approx0.004$) supports significance at $\alpha=0.01$ but not at the originally stated $\alpha=0.001$, and the five-fold Wilcoxon test is replaced by a ten-run protocol (Section 3.8). Attributing the improvement to the entropy signal remains conditional on the ablations of Section 4.8. These headline figures should be read alongside the limitations and unresolved inconsistencies catalogued in Section 5, which qualify several of the claims made above.

The feature-importance percentages of the original draft (entropy score 16.2%, Flow Bytes per Second 12.4%, CNN SF-1 10.9%) and the 0.52 pp contribution of WSN-specific topology features are withdrawn until reproduced by the component-wise, seed-stable attribution protocol of Section 3.13 (Tables 22 and 23), because Integrated Gradients cannot be applied directly to LightGBM or to the piecewise-constant DWF gate. ARM Cortex-M55 and RISC-V SiFive U74 embedded profiling provides inference-level evidence for the three-tier deployment architecture, to be completed by the end-to-end measurements of Section 3.12: GBDT-only at 5.4 ms and 54.1 KB SRAM for the Tier 1 budget (measured on the Cortex-M55, not yet on Tier 1 boards), GBDT+CNN at 13.7 ms and 91.4 KB for Tier 2 relay nodes, and full HAI+DWF at 33.2 ms and 201.4 KB for Tier 3 cluster-heads. Feature-space adversarial evaluation under four attack methodologies gives a worst-case (PGD) accuracy of 96.74% at $\varepsilon=0.10$; the average loss at $\varepsilon=0.10$ is 1.0 pp for the GBDT component against 13.1 pp for BiLSTM and 15.6 pp for CNN, and the robustness of the ensemble tracks that of its GBDT component. Whether such perturbations are realisable network traffic, and how the ensemble behaves under adaptive and black-box attacks, is addressed by the protocol of Section 3.11 (Table 19).

Future research will pursue five directions: (i) one-shot learning for severely imbalanced attack classes—Slowloris at 1:41,724 raw ratio—targeting F1-score above 96% without SMOTE dependency; (ii) spiking neural network implementations of CNN and BiLSTM components on Intel Loihi 2 neuromorphic hardware targeting sub-100 μJ inference for batteryless Tier 1 nodes; (iii) federated HAI+DWF training across WSN operator domains through FedAvg with differential privacy (ε,δ)-DP guarantees enabling privacy-preserving collaborative model refinement; (iv) certified adversarial defences through C&W-adversarial BiLSTM training and randomised smoothing for CNN, targeting provable L2-norm accuracy guarantees; and (v) live testbed validation across TelosB, Zolertia RE-Mote, nRF52840, and ESP32-C6 hardware under authentic IEEE 802.15.4 radio propagation environments with Contiki-NG and Zephyr RTOS v3.6 .

Acknowledgement: Not applicable.

Author Contributions: Conceptualization: Qasim Mustafa Zainel; methodology: Mohammed Fakhrulddin Abdulqader; software: Adnan Yousif Dawod; validation: Mohammad Khaleel Sallam Ma'aitah; formal analysis: Ravi Sekhar; investigation: Amer Alqutaish; data curation: Adnan Yousif Dawod; writing—original draft preparation: Adnan Yousif Dawod; writing—review and editing: Adnan Yousif Dawod; visualization: Adnan Yousif Dawod; supervision: Adnan Yousif Dawod. All authors have read and agreed to the final version of the manuscript.

Funding Statement: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No.KFU265785).

Conflicts of Interest: The authors declare no conflicts of interest.

Availability of Data and Materials: The four benchmark datasets are publicly available from their original sources: CICIoT-2023 (https://www.unb.ca/cic/datasets/iotdataset-2023.html), Edge-IIoTset-2022 (IEEE DataPort, doi:10.21227/mbc1-1h68), NF-ToN-IoT-v2 (https://staff.itee.uq.edu.au/marius/NIDS_datasets/) and WSN-DS (Almomani et al., Journal of Sensors, 2016, doi:10.1155/2016/4731953). The partition index files, preprocessing and feature-selection code, fixed seeds (Table A3), trained weights, DWF boundaries and weights, and all scripts used to produce the reported tables will be deposited in a public repository, with a DOI issued on acceptance, and are available to reviewers on request.

References

1. M. Mihoubi, A. Rahmoun, and P. Lorenz, "Node Localization in WSN and IoT Using Harris Hawks Optimization Algorithm," 2021 IEEE Global Communications Conference (GLOBECOM), pp. 01–06, Dec. 2021, doi: 10.1109/globecom46510.2021.9685105

2. P. S. Chatterjee, "Detection of Node Cloning Attack in WSN to Secure IoT-based Application," IoT Applications, Security Threats, and Countermeasures, pp. 247–260, Jul. 2021, doi: 10.1201/9781003124252-14

3. A. Sharma, S. Rani, and M. Shabaz, "Deep hybrid architecture for multi-class detection of network layer attacks in WSN," Scientific Reports, vol. 16, no. 1, May 2026, doi: 10.1038/s41598-026-55532-7

4. H. A. Marhoon, R. Sagban, A. Y. Oudah, and S. R. Ahmed, "A Barrier-Based Machine Learning Approach for Intrusion Detection in Wireless Sensor Networks," Computers, Materials & Continua, vol. 82, no. 3, pp. 4181–4218, 2025, doi: 10.32604/cmc.2025.058822.

5. A. A. Najar and S. Manohar Naik, "DDoS attack detection using CNN-BiLSTM with attention mechanism," Telematics and Informatics Reports, vol. 18, p. 100211, Jun. 2025, doi: 10.1016/j.teler.2025.100211

6. X. Wang and X. You, "A Research Study on an Entropy-Weighted Multi-View Fusion Approach for Agricultural WSN Data Based on Fuzzy Clustering," Electronics, vol. 14, no. 12, p. 2424, Jun. 2025, doi: 10.3390/electronics14122424

7. 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, p. 5941, Jun. 2023, doi: 10.3390/s23135941

8. V. R. Joshi, K. Assa-Agyei, T. Al-Hadhrami, and S. N. Qasem, “Hybrid AI Intrusion Detection: Balancing Accuracy and Efficiency,” Sensors, vol. 25, no. 24, p. 7564, Dec. 2025, doi: 10.3390/s25247564

9. T. Babenko et al., “Hybrid GNN–LSTM Architecture for Probabilistic IoT Botnet Detection with Calibrated Risk Assessment,” Computers, vol. 15, no. 1, p. 26, Jan. 2026, doi: 10.3390/computers15010026

10. A. D. Abbas et al., “Botnet Communication Pattern Detection in IoT Environments Using Transformer-Based Autoencoders,” 2026 2nd International Symposium on AI-Driven Engineering Systems (ISADES), pp. 1–8, Jun. 2026, doi: 10.1109/isades69945.2026.11608245

11. Moucharraf M, Ridouani M, Salahdine F, Kaabouch N. IoT security: a comprehensive review of architectures, threat models, detection methods, and countermeasures. Future Internet. 2026;18(5):266. doi:10.3390/fi18050266.

12. Songma S, Sathuphan T, Pamutha T. Optimizing intrusion detection systems in three phases on the CSE-CIC-IDS-2018 dataset. Computers. 2023;12(12):245. doi:10.3390/computers12120245.

13. V. Heydari and K. Nyarko, “Enhancing Adversarial Robustness in Network Intrusion Detection: A Novel Adversarially Trained Neural Network Approach,” Electronics, vol. 14, no. 16, p. 3249, Aug. 2025, doi: 10.3390/electronics14163249.

14. T. Huraysi, “Security and Privacy Issues for Blockchain Scenarios in IoT and WSN-Based Real-time Applications,” Security and Privacy Issues for IoT and WSN-based Real-time Applications, pp. 68–93, Apr. 2025, doi: 10.1201/9781003491910-5.

15. A. Al-Ashmali, D. H. Manjaiah, I. Gad, and M. F. Aljunid, “A Comprehensive Benchmark of Resampling and Feature Selection for Imbalanced IoT Intrusion Detection on the CICIoT2023 Dataset,” 2025 2nd International Conference on Intelligent Systems for Cybersecurity (ISCS), pp. 1–6, Nov. 2025, doi: 10.1109/iscs69371.2025.11386384.

16. Agbor BA, Stephen BU-A, Asuquo P, Luke UO, Anaga V. Hybrid CNN-BiLSTM-DNN approach for detecting cybersecurity threats in IoT networks. Computers. 2025;14(2):58. doi:10.3390/computers14020058.

17. M. A. Sheheryar and S. Sharma, “A lightweight hybrid CNN-LSTM framework with hybrid feature selection for cross-dataset IoT botnet detection,” International Journal of Communication Networks and Distributed Systems, vol. 33, no. 5, 2027, doi: 10.1504/ijcnds.2027.10081997.

18. A. Haddad, S. A. Wajs, M. H. Habaebi, and M. A. S. Al-Hussaini, “Operational Architectures for Hybrid AI-Based Network Intrusion Detection: A Critical Review and Multi-Axial Taxonomy,” Computers, vol. 15, no. 10, p. 726, Oct. 2026, doi: 10.3390/computers15100726.

19. Ferrag MA, Maglaras L, Moschoyiannis S, Janicke H. Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study. J Inf Secur Appl. 2020;50:102419. doi:10.1016/j.jisa.2019.102419.

20. Roopak M, Tian GY, Chambers J. Deep learning models for cyber security in IoT networks. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). IEEE; 2019. p. 452–457. doi:10.1109/CCWC.2019.8666588.

21. S. I. Popoola, A. L. Imoize, M. Hammoudeh, B. Adebisi, O. Jogunola, and A. M. Aibinu, “Federated Deep Learning for Intrusion Detection in Consumer-Centric Internet of Things,” IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 1610–1622, Feb. 2024, doi: 10.1109/tce.2023.3347170.

22. Akuthota UC, Bhargava L. Transformer-based intrusion detection for IoT networks. IEEE Internet Things J. 2025;12(5):6062–7. doi:10.1109/JIOT.2025.3525494.

23. Ullah I, Mahmoud QH. A two-level hybrid model for anomalous activity detection in IoT networks. IEEE Access. 2021;9:89389–89406. doi:10.1109/ACCESS.2021.3091421.

24. Khan MA, Karim MR, Kim Y. A deep learning-based intrusion detection system for MQTT enabled IoT. Sensors. 2021;21(21):7016. doi:10.3390/s21217016.

25. S. D. Azeez, S. R. Ahmed, M. Ilyas, A. S. M. Miah, F. A. Farid, and Md. H. A. Karim, “IntrusionNet: Deep Learning-Based Hybrid Model for Detection of Known and Zero-Day Attacks,” Computers, Materials & Continua, vol. 0, no. 0, pp. 1–10, 2026, doi: 10.32604/cmc.2026.076283.

26. 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, doi: 10.1109/access.2022.3165809.

27. E. Oyekanlu, “Predictive edge computing for time series of industrial IoT and large scale critical infrastructure based on open-source software analytic of big data,” 2017 IEEE International Conference on Big Data (Big Data), pp. 1663–1669, Dec. 2017, doi: 10.1109/bigdata.2017.8258103.

28. O. Friha, M. A. Ferrag, L. Shu, L. Maglaras, and X. Wang, “Internet of Things for the Future of Smart Agriculture: A Comprehensive Survey of Emerging Technologies,” IEEE/CAA Journal of Automatica Sinica, vol. 8, no. 4, pp. 718–752, Apr. 2021, doi: 10.1109/jas.2021.1003925.

29. Thereza N, Harahap PR. Enhancing the performance of IoT network intrusion detection models using NF-ToN-IoT-V2 and IoTID20 datasets with chi-square feature selection. Sriwijaya Electrical and Computer Engineering Journal. 2025;1(2):64–77. doi:10.62420/selco.v1i2.7.

30. D. Gaspar, P. Silva, and C. Silva, “Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron,” IEEE Access, vol. 12, pp. 30164–30175, 2024, doi: 10.1109/access.2024.3368377.

31. J. Tritscher, A. Krause, and A. Hotho, “Feature relevance XAI in anomaly detection: Reviewing approaches and challenges,” Frontiers in Artificial Intelligence, vol. 6, Feb. 2023, doi: 10.3389/frai.2023.1099521.

32. K. V. P. Kumar, “A Survey On Energy Efficient Routing Protocols In Wireless Sensor Networks,” i-manager’s Journal on Wireless Communication Networks, 2016, doi: 10.26634/jwcn.4.4.5912.

33. L. Prusty, P. K. Swain, S. Satpathy, and S. Mahapatra, “Comprehensive Review of Security Challenges and Issues in Wireless Sensor Networks Integrated with IoT,” Explainable IoT Applications: A Demystification, pp. 467–485, 2025, doi: 10.1007/978-3-031-74885-1_30.