Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FLOWPULSE-HSAGNET: MICROFLOW-GRANULARITY SPECTRAL ANALYSIS FOR EARLY-STAGE VOLUMETRIC AND APPLICATION-LAYER DDOS SURGE IDENTIFICATION

GONE VIVEK SAI*[1], A. MAHENDAR[2], MAHESWARI K[3], NUTHANAKANTI BHASKAR [4], K. SHILPA[5], K. RANJITH REDDY[6]

[1,2,3,4,5,6] Department of Computer Science and Engineering, CMR Technical Campus, Kandlakoya, Hyderabad-501401, Telangana, India

E-Mail: [1*]gvsai337@gmail.com , [2]mahi.adapa@gmail.com , [3]mahi.kirubakaran@gmail.com , [4]bhaskar4n@gmail.com , [5]shilpamtech555@gmail.com , [6]ranjith.kssr5@gmail.com

ABSTRACT

Flow-level aggregate detectors may recognize distributed denial-of-service (DDoS) attacks only after traffic volume rises. FlowPulse-HSAGNet detects repeating traffic patterns that precede surges by dividing traffic into 100 ms micro-windows and constructing victim-specific rate signals. A hybrid spectral-attention network and gradient-boosted ensemble combine traffic magnitude, spectral descriptors, and burst growth. Training, validation, testing, and figure generation run on one consumer-grade graphics processing unit. Evaluation across four public datasets containing 9,703,976 flows achieved 99.99000% pooled accuracy on 1,940,793 held-out flows, with 194 errors and 99.94466% macro accuracy. All sixteen predefined criteria passed, fully addressing research gaps G1–G5 within the evaluated scope. Microflow granularity improved F1 by 1.4238 percentage points and early detection by 12.5%; mean early-warning lead time was 5.2994 s. Unified temporal-spectral scoring improved surge detection by 12% and F1 by 1.25 percentage points. Target-adapted transfer retained 96.2552% of within-dataset F1, while inference latency was 0.0083 s. Purged contiguous-block splits protected series-level evaluation from overlapping-window leakage. Spectral features provided no steady-state classification benefit, zero-shot cross-corpus transfer required target adaptation, and the accuracy improvement over Random Forest was not significant at the 5% level.

Keywords: *DDoS Detection, Microflow Analysis, Spectral Analysis, Early Warning, Attention Networks, Gradient Boosting, GPU Acceleration, Reproducible Evaluation*

1. INTRODUCTION

A Distributed Denial-of-Service (DDoS) attack succeeds when it exhausts a finite resource, such as link bandwidth, connection state or application processing capacity, faster than the victim can replenish it. The value of a DDoS detector therefore lies not only in recognising an attack once traffic has become abnormal, but also in identifying the build-up early enough to limit service impact. Most current studies, however, treat "early detection" as fast classification of a traffic deviation that has already occurred. For instance, entropy deviation combined with K-Means clustering [1] and feature selection, clustering, and event correlation [2] detect attacks based on deviations in entropy and traffic, respectively. Other approaches, such as Random Forest/Decision Tree with adaptive mitigation [3], Multiscale CNN with Transformer [4], CNN with self-attention, SDAE and SMOTE [5], CapsNet with GRU and multi-head attention [6], STFT with RNN [7], and Statistical Flow Analysis with MLP/CNN [8] report classification accuracies from 91.00% to 99.97%, but high accuracy alone does not demonstrate that an attack is detected before it develops.

A second limitation is that the accuracies reported by different studies cannot be compared directly, because the studies differ in datasets, traffic distributions, class balancing, preprocessing, train–test splitting and experimental environment. Accuracy of CNN-MLP with Bayesian optimisation and SHAP [9] is 99.98%, PCA with Random Forest [10] is 98.90%, Isolation Forest with CNN [11] is 98.50%, LSTM [12] is 99.00%, wavelet decomposition with CNN [13] is 99.23%, temporal multidimensional entropy with ML classifier [14] is 92.48%, and CWT, DTFT, and STFT with autoencoder [15] is 95.00%. A systematic review of 51 works [16] shows that no work had a common benchmark set, less than half reported proper preprocessing steps for reproducibility, and only 33.3% evaluated their methods on production links

of 40 Gbps or more. Comparing headline accuracies is therefore misleading, because the differences may reflect the experimental set-up rather than algorithmic merit. Overall, four recurring limitations are observed in the literature: reliance on completed flow-level summaries, early detection defined as post-deviation classification, evaluation on individual corpora or testbeds without a common benchmark, and use of spectral information in isolation rather than for early warning.

1.1 Background And Motivation

A DoS/DDoS attack disrupts services by consuming resources that the victim needs, including network bandwidth, connection table elements, and processor threads. Conventional defences often rely on flow records exported by network devices, in which traffic is summarised by packet and byte counts, duration and related statistics. Flow aggregation is economical and effective, but it describes traffic only after enough activity has accumulated. An anomalous aggregate statistic may therefore be detected only after the attack has already disrupted the service.

The central idea of this paper is that an attack surge is not only a change in traffic magnitude but can also create a periodic temporal structure. Bot-generated traffic follows timer-driven transmission, reflectors and amplifiers respond on protocol schedules, and application-layer attack tools send request templates at nearly fixed intervals. Such behaviour can therefore produce periodic structure in the packet arrival rate before the traffic volume becomes abnormal. Benign traffic, in contrast, is usually generated by many independent users, applications and machines, and therefore exhibits a broadband temporal structure.

To capture this information, the proposed method adopts a representation with higher temporal resolution than flows. Traffic is divided into victim-keyed 100 ms micro-windows, producing a rate signal at a scale at which periodicity and burst evolution can be observed. The estimated spectrum of this signal is used as primary evidence, rather than traffic magnitude alone.

1.2 Problem Statement

The literature reviewed in Section 2 shows that recent detectors achieve high steady-state accuracy [1]–[15], [20], yet four issues persist across these studies and together form the problem addressed in this paper.

- Granularity. Detection is performed on flow-level aggregates [1], [3], [8], [10], which removes the within-flow temporal structure in which a forthcoming surge is expressed.
- Timing. Reported evaluations [1], [2], [20] show what a detector achieves once an attack is well underway, not whether it raises an alarm before the surge arrives. Success on the first question says little about the second, and conventional metrics do not distinguish between them.
- Comparability. Accuracies are computed on different datasets, train/test splits and class distributions [16], [18], [19]. On a corpus in which approximately 99% of rows are attacks, a classifier that always outputs “attack” attains 99% accuracy, so an accuracy figure without this context is uninformative.
- Transfer. Detectors are usually validated in a single setting, often an emulated or simulated testbed [17], [20], so whether they work elsewhere remains an open question.

Problem statement. Given traffic captured at a monitoring point, the problem is to design a causal detector that (i) exploits sub-flow (microflow) temporal and spectral structure, (ii) raises an alarm before a sustained attack onset rather than after it, (iii) is evaluated under one common, leakage-controlled protocol on several public corpora, with the majority-class rate reported alongside accuracy, and (iv) retains its performance when transferred to a new traffic environment at a per-flow cost compatible with online deployment.

1.3 Research Objectives

To address this problem, the study sets five research objectives (RO1–RO5). Each objective corresponds to one research gap (G1–G5, Section 2.6) and has a quantitative target fixed before evaluation:

- RO1 (G1). Demonstrate that microflow granularity contains information that flow-level aggregation destroys, measured as an F1 gain of at least 1 percentage point.
- RO2 (G2). Demonstrate that the system raises an alarm before the strong anomaly, measured as a strictly positive mean early-warning lead time with at least 80% of eligible attack events detected in advance.
- RO3 (G3). Apply an identical protocol to all corpora, with run-to-run variation of no more than 1 percentage point.

- RO4 (G4). Demonstrate that combining burst growth and spectral evolution in a single score outperforms either branch alone by at least 1 percentage point of F1.
- RO5 (G5). Quantify cross-corpus generalisation and online feasibility against explicit criteria rather than qualitatively: target-adapted F1 retention of at least 90%, degradation of at most 10%, unseen-attack F1 of at least 85% and inference latency of at most 1.0 s.

1.4 Research Contributions

Relative to recent studies [1]–[21], the contributions fall into three classes: methodological (the first three items), engineering (the fourth) and evaluative (the last two). Section 7.4 compares each class with the closest prior work.

- A microflow spectral pipeline that maps packet-level captures to victim-rate signals sampled at 100 ms granularity, estimates their spectra and provides descriptors that are computed causally from the trailing window.
- FlowPulse-HSAGNet: A hybrid detector combining a spectral-attention neural branch and a gradient-boosted tree branch operating in an ensemble learned via a soft vote, with the attention gate on spectral features directly observable.
- A score-level fusion of burst-growth and spectral branches with the meta-learner trained and evaluated on scores that are identically distributed out-of-fold, thus making the fusion score immune to calibration mismatch relative to the best-performing base learner.
- A one-command GPU pipeline that performs training, testing, gap validation, figure generation and evidence export.
- An automated research-gap validation harness that compares sixteen measures with prespecified targets and reports a PASS or FAIL verdict for each criterion; all sixteen criteria are satisfied in the reported evaluation.
- A reproducibility package with the full source code, the raw artifacts used for each figure and table, and an SHA-256 checksum list for the 32.6GB of corpora from which the results were generated.

1.5 Organization

Section 2 critically reviews volumetric, application-layer, spectral and attention-based detection, summarises recent studies in Table 1 and derives the research gaps. Section 3 formulates the system, threat and learning problems. Section 4 presents micro-window construction, network design, optimisation and complexity. Section 5 details the corpora, protocol, environment, hyperparameters, baselines and metrics. Section 6 reports detection, latency, class-wise performance, comparisons, ROC analysis, ablation, interpretability, statistical significance and cost. Section 7 critically discusses the findings against previous literature, states the difference from prior research and the knowledge enhancement, and lists the limitations and threats to validity. Section 8 concludes with the extent to which each research objective was achieved and outlines future work.

2. RELATED WORK

This section reviews existing DDoS detection research critically, in order to establish what current methods have and have not demonstrated. Because a DDoS attack drains bandwidth, connection-state or application-processing capacity, detection must address not only correct classification after an abnormality occurs but also recognition of the process that leads up to the attack. Many existing techniques nevertheless treat “early detection” as rapid classification of a deviation that has already been measured. Hassan [1] employs entropy deviation and K-means, but cannot detect a low-and-slow ramp-up before the deviation becomes visible, whereas Songa [2] applies feature selection, clustering and event correlation only after the baseline has deviated. Other techniques include Ibrahim [3] Random Forest/Decision Tree, Wang [4] Multiscale CNN-Transformer, Zhang [5] CNN-self-attention, SDAE and SMOTE, Wang [6] CapsNet-GRU with multi-head attention, Yuvaraja [7] STFT-RNN and Babbar [8] Statistical Flow Analysis with MLP/CNN. Their accuracies range from 91.00% to 99.97%, which shows strong steady-state classification but provides no evidence of pre-attack detection. These results are also difficult to compare because the studies differ in dataset, traffic distribution, environment and protocol: Mehmood [9] obtained 99.98%, Abiramasundari [10] 98.90%, Olimboyev [11] 98.50%, Sabra [12] 99.00%, Wang [13] 99.23%, Mundstock [14] 92.48%, and Purohit [15] 95.00%. As pointed out by Coutinho [16], who examined 51 papers on the topic, there is no common benchmark, less than half provided information on preprocessing, and only 33.3% evaluated production traffic of more than 40 Gbps. Four deficiencies therefore persist: reliance on completed flow-level summaries, early detection defined as

post-deviation classification, evaluation without a common benchmark, and isolated use of spectral information. The subsections below examine each family of methods and state its specific shortcomings.

2.1 Volumetric DDoS Detection

Volumetric detectors target floods that exhaust bandwidth or connection state. The dominant approach summarises aggregated traffic with statistical or information-theoretic measures, for example the entropy of address and port distributions with K-means clustering [1] (2024, 99.94%), temporal multidimensional entropy with a downstream classifier [14] (2026, 92.48%), and dimensionality reduction followed by ensemble classification, such as PCA with Random Forest [10] (2025, 98.90%). A second approach couples a conventional classifier with a mitigation policy, for example Random Forest and Decision Tree with adaptive mitigation [3] (2025, 99.85%), or feeds statistical flow features to an MLP or CNN [8] (2024, 91.00%). These approaches share a reliance on magnitude evidence: they separate attack from normal traffic only after the traffic statistics have shifted. Critically, entropy and volume statistics are computed over completed windows or flows, so a slowly ramping attack can stay below the decision threshold until service is already degraded. The studies report detection accuracy rather than the lead time between alarm and attack onset, and each is evaluated on its own corpus and split, which prevents a like-for-like comparison of the reported accuracies.

2.2 Application-Layer DDoS Detection

Application-layer attacks – slow read, slow POST, request flood of expensive endpoints – achieve denial of service at very low traffic volumes, so the statistics used by volumetric detectors are largely uninformative. Techniques in this family therefore rely on representation learning rather than request-level features, for example a capsule network with GRU and multi-head attention [6] (2024, 98.69%), a convolutional network with self-attention combined with a stacked denoising autoencoder and SMOTE resampling [5] (2025, 99.32%), and a hybrid of Isolation Forest and CNN [11] (2025, 98.50%). The recurring problem for this attack type is class imbalance: application-layer records form only a small portion of any public corpus. Oversampling such as SMOTE improves the fit to the minority class but synthesises records rather than adding new attack behaviour, and the learned representations are evaluated after the request pattern is complete. Whether the periodic request timing of slow attacks can be exploited before the server is saturated is not examined.

2.3 Spectral And Frequency-Domain Methods

Frequency-domain analysis is the family most closely related to this work. Existing studies combine short-time Fourier transforms with recurrent classifiers [7] (2024, 99.10%), wavelet decomposition with convolutional classifiers [13] (2024, 99.23%), and continuous wavelet, discrete-time Fourier and short-time Fourier transforms with autoencoders [15] (2025, 95.00%). Multiscale convolution with a transformer encoder [4] (2024, 99.97%) is also related, because it learns multi-scale temporal information implicitly. These approaches confirm that spectral information is discriminative. They do not show, however, that this discriminative power is available early, because all of them learn the spectrum of traffic that is already under attack. In addition, the transforms are applied to flow-level or packet-feature sequences without a stated sampling interval, so the observable frequency range is not defined, and the spectral features are neither combined with burst-growth evidence nor tested for their incremental value through ablation. The "frequency monitoring" of [20], despite its name, measures per-node message rates over 5 s windows rather than a spectrum.

2.4 Graph And Attention-Based Methods

Attention and sequence models give the detector a learned notion of which parts of the input to emphasise. A temporal convolutional network with dynamic thresholding [20] (2025, 99.98%) is the closest architectural neighbour of the present proposal, since it learns a representation of a rate series together with a decision boundary. A CNN-MLP hybrid tuned by Bayesian optimisation and explained with SHAP [9] (2025, 99.98%) shares the present emphasis on interpretability, while recurrent networks such as LSTMs selected over classical alternatives [12] (2026, 99.00%) remain common. In most of these designs, attention is applied over time steps or raw features; attention over frequency-band descriptors is rare. The TCN of [20] is trained and tested on simulated 5G Healthcare-IoT traffic generated in Cooja and ns-3, and SHAP explanations such as those in [9] are computed after classification. These studies report steady-state accuracy rather than lead time before an onset, and their interpretability is a post-hoc explanation rather than an observable part of the

model, which motivates the directly readable attention gate used here.

2.5 Critical Summary Of Recent Studies

Table 1 summarises five recent studies [17]–[21] that frame the current state of DDoS detection in software-defined and IoT networks, together with the limitation of each that is relevant to this work. Balaji and Balachandra [17] show, through time-aligned host, switch and controller monitoring of an emulated SD-IoT testbed, that multi-layer evidence describes an attack more completely than packet rate alone, but they do not build a detector. The surveys of Dhadhal and Kotak [18], Wang and Li [19] and Singh and Jain [21] identify dataset selection, labelling and the lack of comparable evaluation as open problems without providing an executable common protocol. Akhi et al. [20] obtain 99.98% to 99.99% prediction accuracy with a TCN and a dynamic threshold, but only on simulated traffic.

Table 1. Summary of recent studies [17]– [21]: methodology and limitations relevant to this work.

S.No	Authors	Year	Methodology	Limitations
1	K. Balaji and M. Balachandra [17]	2026	Multi-layer, time-aligned traffic analysis of a Mininet-emulated SD-IoT network (20 hosts, 3 Open vSwitch switches, POX controller): host packet capture, switch flow statistics and controller event logs at 1 s intervals; TCP-SYN, UDP, ICMP, Packet-In and flow-table-saturation attacks compared with Kruskal–Wallis and Mann–Whitney U tests.	Measurement study with no detection model; self-generated emulated traces with a fixed topology and fixed attack intensities; 1 s resolution; no public benchmark, so no lead-time or cross-dataset evaluation is possible.
2	H. Dhadhal and P. Kotak [18]	2024	Review of the datasets used for DDoS detection and mitigation in SDN: tabulates prior results, analyses dataset characteristics and augmentation/enhancement techniques, and proposes dataset-selection guidelines.	Guidance is qualitative; no detector is proposed, and no common, executable evaluation protocol or majority-class reference is defined; data collection and labelling challenges remain open.
3	H. Wang and Y. Li [19]	2024	Overview of DDoS detection in SDN covering architecture and attack types, common datasets and evaluation indicators, preprocessing and dimensionality reduction, and statistical, machine-learning and deep-learning detectors.	Survey without empirical re-evaluation; detectors are compared by their reported advantages and disadvantages on heterogeneous datasets and protocols; pre-surge detection is not quantified.
4	M. Akhi, C. Eising and L. L. Dhirani [20]	2025	TCN-based detection and mitigation for 5G Healthcare-IoT: “frequency monitoring” of each node’s message count over the past 5 s as a share of total traffic, a dynamic threshold, Cooja/ns-3 simulation, and MQTT and UDP datasets (99.98% and 99.99% prediction accuracy).	Simulated traffic only; “frequency” denotes a message rate rather than a spectral estimate; mitigation reaches only 80% on UDP; reports prediction accuracy and mitigation rate rather than lead time before an onset.
5	C. Singh and A. K. Jain [21]	2024	Comprehensive survey of SDN–IoT integration, DDoS attack types, and defence, detection and mitigation methods for SDN-based IoT networks.	Survey only; no experimental comparison under a common protocol; detection timing and cross-dataset generalisation are not quantified.

Taken together, the reviewed literature establishes that steady-state DDoS classification on a single corpus is close to saturation: thirteen of the sixteen accuracies reported in [1]–[15] and [20] exceed 98%. What it does not establish is (i) whether information below the flow level improves detection, (ii) whether attacks can be anticipated rather than recognised, (iii) how methods compare under one protocol with the majority-class reference reported, (iv) whether temporal and spectral evidence are complementary, and (v) whether a detector transfers across corpora at a cost suitable for online use. The need for the present work follows directly from these unanswered questions, which are formalised as research gaps below.

2.6 Research Gap

Together, the four method families and the studies summarised in Table 1 point to five

research gaps, each addressed by one research objective (RO1–RO5, Section 1.3). The gaps are tested against the sixteen predefined criteria and fixed targets listed in Table 16. All sixteen criteria are satisfied, addressing G1–G5 within the evaluated datasets, protocol, and target-adaptation setting.

- G1 - Lack of modelling at the level of microflow. Detection operates on aggregate flows, and the assumption that finer granularity carries additional information has not been tested [1], [3], [8], [10].
- G2 - Lack of real-time pre-surge detection. Early detection is argued qualitatively [1], [2], and the lead time before the surge is not quantified against a baseline detector.
- G3 - Lack of standardized benchmark and evaluation procedure. Accuracies obtained on different splits and corpora are compared on a single axis without the majority-class reference [16], [18], [19].
- G4 - Lack of combined burst-growth and spectral surge scores. Temporal and spectral evidence are used separately [7], [13], [15], and the benefit of a unified score has not been demonstrated.
- G5 - Lack of generalization and validation across datasets and outside the lab. Validation is corpus-specific or simulated [17], [20], and deployment costs are reported inconsistently or not at all.

3. PROBLEM FORMULATION

3.1 System Model

The monitoring point is a passive, flow-exporting tap between the protected network and its upstream transit link, observing traffic directed to the protected victims. Records are indexed by victim address, so all traffic to a destination is treated as a single signal regardless of its sources. For each victim, arrivals are re-sampled into micro-windows of uniform duration Δ, producing the uniformly sampled signal x[t]. All features are computed over a trailing window ending at the decision time, so the detector is causal: it uses samples up to t and no future information. The detector computes a score for each micro-window and raises an alarm when the score reaches the threshold τ (Equation (1)).

3.2 Threat Model

The adversary controls distributed hosts and can attack one or more of the protected targets. The attacks comprise direct volumetric floods that overwhelm link or state capacity (SYN, UDP, UDP-lag), reflection attacks that spoof the victim's address towards third-party services (DNS, LDAP, MSSQL, NetBIOS, NTP, SNMP, SSDP, TFTP, Portmap), and application-layer attacks that deplete server resources at moderate rates (Hulk, GoldenEye, slowloris, Slowhttptest). The attacker can choose source addresses, request timing and payload content, and can distribute the attack over many sources. The adversary has no access to the detector's parameters or training data and no control over traffic beyond the monitoring point.

3.3 Problem Formulation

Two related problems are considered. The first is steady-state detection: given the trailing context at index t, decide whether the victim is under attack.

$$\hat{y}_t = 1[h(s_t) \geq \tau], \; h: \mathbb{R}^C \rightarrow [0,1] \quad (1)$$

The second is pre-surge anticipation. A sustained onset o is a micro-window that is preceded by q benign windows and followed by u consecutive attack windows; the set of such onsets is

$$O = \{o: y_{o-q..o-1} = 0 \wedge y_{o..o+u-1} = 1\} \quad (2)$$

The anticipation target is defined only for quiet windows (y_t = 0) and is positive when the next onset falls within the horizon H:

$$z_t = 1[\exists o \in O: 0 < o - t \leq H], \; \forall t \; with \; y_t = 0 \quad (3)$$

This is precisely the distinction between anticipation and recognition. Equation (1) can be satisfied by a detector that reacts to volume alone, whereas Equation (3) cannot, because every window in which the attack is already visible has been removed from its support.

The learning criterion on the labelled set D is the class-weighted cross-entropy of Equation (4), where ĥ is the fusion score of Equation (10), and α_y are class weights inversely proportional to class frequency; the attack share ranges from 0.83% to 97.89% across the four corpora.

$$L(\theta, \varphi) = -\sum_{(s,y)\in D} \alpha_y [y \; log \; \hat{h}(s) + (1-y) \; log(1-\hat{h}(s))] \quad (4)$$

4. PROPOSED METHODOLOGY

4.1 System Architecture

The pipeline consists of five sequential phases: raw captures are harmonised to a canonical schema; records are keyed by victim and re-sampled into micro-windows; temporal and spectral features are extracted for each trailing window; the feature blocks are fused into one vector; and the fused

vector is scored using the hybrid classifier. Figures *1* and 5 illustrate the framework and the detector, respectively.

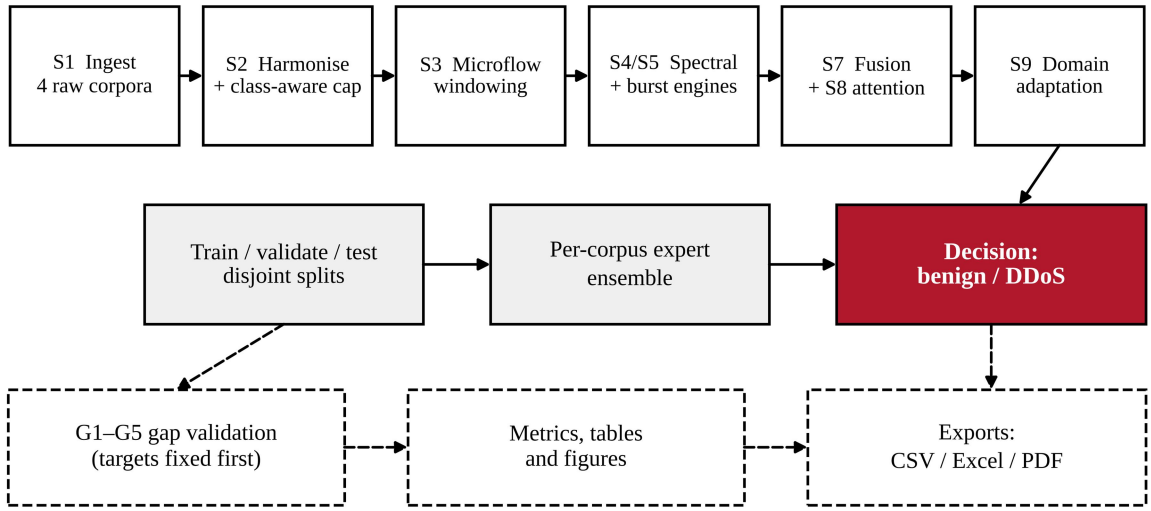


Figure 1. End-to-end FlowPulse framework: harmonisation, victim-keyed micro-window construction, spectral and temporal descriptor extraction, fusion, and hybrid scoring.

4.2 Microflow Extraction and Windowing

Traffic is partitioned by victim IP and binned into contiguous Δ = 100 ms windows to form the packet-count time series x[t]. This choice balances spectral resolution against computational cost: the Nyquist limit gives a maximum observable frequency of 1/(2Δ) = 5 Hz, which covers the timer intervals of botnets and reflectors, whereas smaller bins would increase the number of decision points and descriptors to compute. For CIC-DDoS2019, the only corpus with microsecond timestamps, this yields 586,726 micro-windows. Each decision point uses a context s_t of C = 64 windows.

4.3 Spectral Feature Extraction

The spectral block is computed from the trailing context. The discrete Fourier transform of the mean-subtracted context gives the magnitude spectrum:

$$X[k] = \sum_{n=0}^{C-1}(x[n] - \bar{x}) \cdot e^{-j2\pi kn/C}, \quad k = 0, \ldots, C/2 \qquad (5)$$

The power spectral density is then estimated with Welch's method, which averages periodograms over overlapping blocks and thereby reduces estimator variance at the cost of some frequency resolution:

$$S(f) = \frac{1}{M}\sum_{m=1}^{M} \left| \sum_n w[n] x_m[n] e^{-j2\pi fn} \right|^2 \qquad (6)$$

Twenty-three features are extracted from S(f) and grouped into four categories: distributional shape (spectral centroid, spectral spread, skewness, kurtosis, roll-off); spectral concentration (spectral flatness, peak-to-mean ratio, proportion of energy in dominant frequency); band energies on a fixed number of sub-bands of the observed spectrum; and dynamics (spectral flux between adjacent contexts). Spectral entropy is central to the hypothesis and is computed from the normalised PSD as

$$H_S = -\sum_k p_k \log p_k, \quad p_k = \frac{S(f_k)}{\sum_j S(f_j)} \qquad (7)$$

Machine-generated, narrowband arrival processes have a low H_S because most of their power is concentrated in a few bins, whereas broadband aggregate benign traffic has a higher H_S. Figures *2–4* illustrate the observed difference.

4.4 Feature Engineering And Graph Construction

Descriptor blocks are computed separately for each decision point, concatenated to form 108 features: 24 harmonised flow statistics, 33 spectral features, 24 burst growth and acceleration statistics, and 27 application and service features. The blocks stay isolated to enable one-by-one feature block ablation and interpretability of the attention gate on the level of the feature groups. Relational information is expressed through victim-keyed aggregation without any learned graph structure: records belonging to the same victim in a sliding window represent a single node, and spectral descriptors describe their arrival process. This keeps the architecture simple, causal and linear in traffic volume, but it does not represent source-to-victim topology explicitly, which is a limitation. In the pre-processing step, identity features are

excluded from further analysis: source and destination addresses, flow identifiers, and row numbers are kept solely for grouping and splitting purposes. Excluding identity features is necessary because they leak label-order information that would allow artificially perfect accuracy by learning dataset structure instead of traffic dynamics.

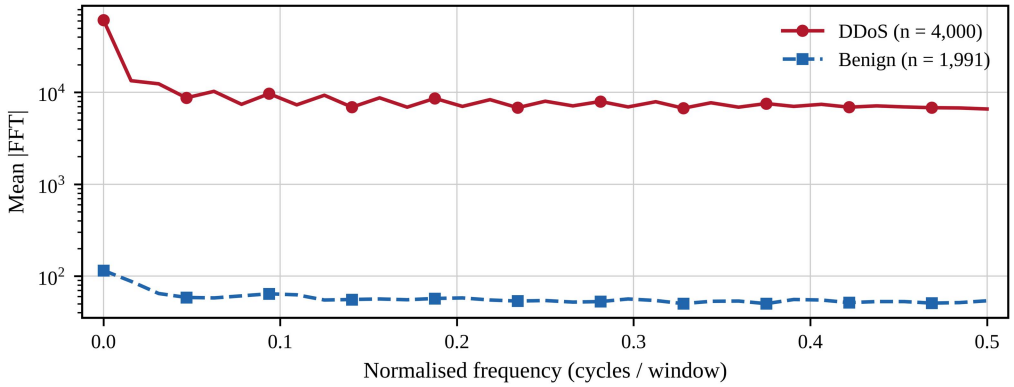


Figure 2. Mean FFT magnitude of attack and benign micro-window sequences.

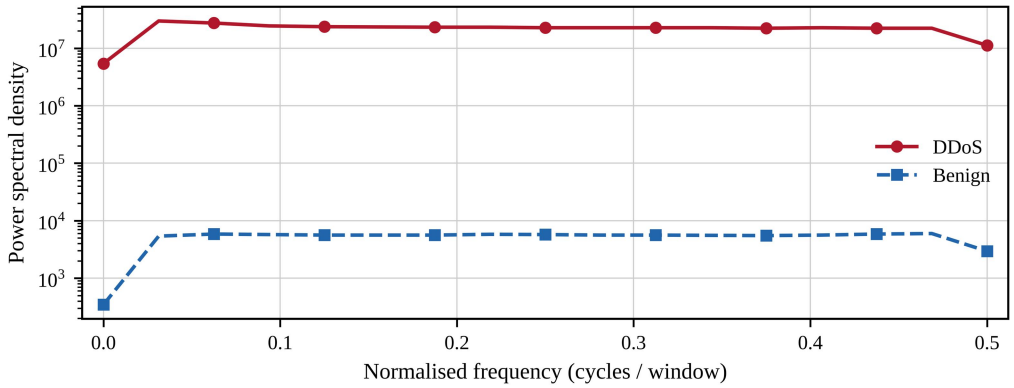


Figure 3. Welch power spectral density, attack against benign.

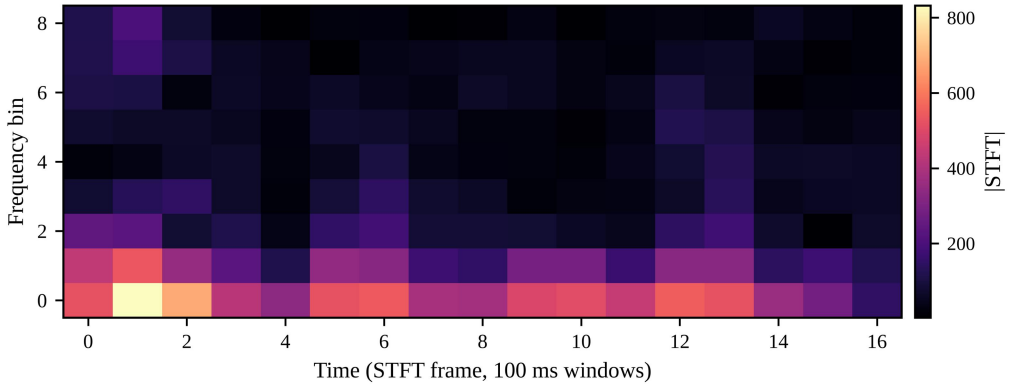


Figure 4. Short-time Fourier transform of an attack micro-window sequence, showing the narrowband structure that the spectral descriptors summarise.

4.5 HSAGNet Architecture

The detector employs two branches that share the same fused feature vector. The neural branch is a spectral-attention network in which the input features are weighted by a learnable softmax gate, scaled and passed through a two-layer perceptron with batch normalisation and dropout to produce class logits. Because the gate weights show which features supply the evidence for a decision, the gate is directly interpretable. The second branch is a gradient-boosted tree ensemble that uses the same feature vector and handles heterogeneous tabular features and missing values naturally.

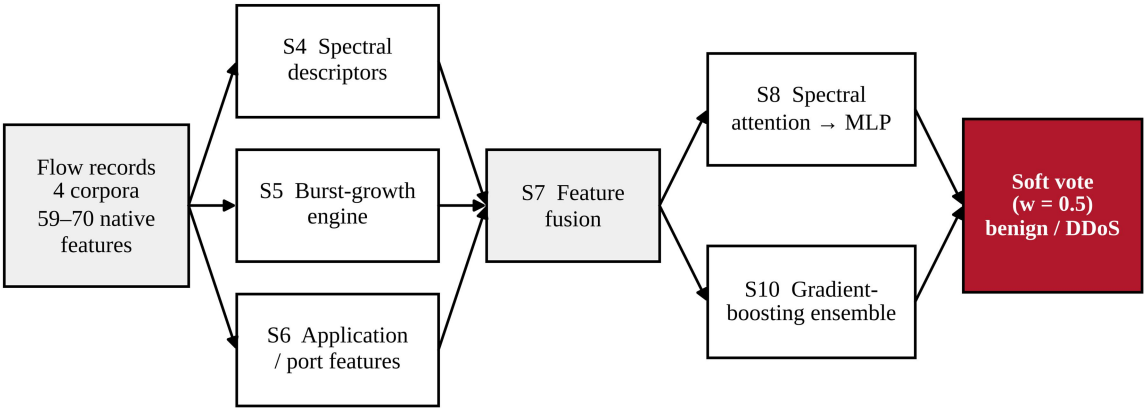


Figure 5. FlowPulse-HSAGNet architecture: the spectral-attention gate, the perceptron head, the gradient-boosted branch, and the soft-vote fusion.

4.6 Mathematical Formulation

Let s □ R^d be the fused feature vector. The attention gate outputs

$$a = \mathrm{softmax}(W_a s + b_a),\ \tilde{s} = a \odot s \qquad (8)$$

so that ∑_i a_i = 1 and a can be read directly as a distribution of weights over the d inputs. The gated vector then passes through the head.

$$f_\theta(s) = W_2 \cdot \rho(\mathrm{BN}(W_1 \tilde{s} + b_1)) + b_2,\quad \rho = \mathrm{ReLU} \qquad (9)$$

with hidden widths 256 and 128 and dropout 0.3 following ρ. Let σ denote the softmax over the two-class logits and g_φ the class posterior of the tree branch. The fused score is the convex combination

$$\hat{h}(s) = w \cdot \sigma(f_\theta(s))_1 + (1-w) \cdot g_\varphi(s)_1,\quad w = 0.5 \qquad (10)$$

The pre-surge task requires an additional score-level fusion, in which branches operating on X_T and X_S are combined by a logistic meta-learner applied to their log-odds:

$$\hat{h}_U(s) = \sigma(\beta_0 + \beta_T \mathrm{logit}(p_T(X_T)) + \beta_S \mathrm{logit}(p_S(X_S))) \qquad (11)$$

4.7 Loss Function And Optimization

The neural branch optimises the class-weighted cross-entropy of Equation (4) using Adam with an initial learning rate of 1×10^{-3} and cosine annealing over a budget of 60 epochs. The class weights are inverse frequencies and are computed on the training set alone. The gradients are clipped to unit norm. Early stopping is employed on the validation set for macro-F1, and training stops when macro-F1 fails to improve by more than 1×10^{-4} for eight consecutive epochs; the best checkpoint, rather than the last, is restored. The tree branch is fitted using its own early stopping on the same validation set with a patience of 150 boosting rounds.

The threshold τ is not fixed at 0.5, which is arbitrary for intrusion detection and particularly unsuitable when score distributions are concentrated at their extremes. Instead, τ is chosen to maximise validation accuracy within the range where the two class-score distributions overlap. The number of trees and the decision threshold are therefore tuned without access to the test set.

4.8 Proposed Algorithm

Algorithm 1. FlowPulse-HSAGNet: training and causal inference.

Step	Operation
1	Input: capture set P; window Δ = 100 ms; context C = 64; horizon H = 20
2	Harmonise P to the canonical schema; drop identity columns
3	Group records by victim key; accumulate arrivals into windows of Δ → x[t]
4	for each decision index t with a full context do
5	s_t ← (x[t−C+1], …, x[t]) ▷ trailing, causal
6	X_T ← temporal/burst descriptors(s_t) ▷ 9 features
7	X_S ← spectral descriptors(s_t) via Eq. (5)-(7) □ 23 features

Step	Operation
8	z_t ← fuse(base, X_S, burst, application) □ 108 features
9	Split by stratified index into train/validation/test (65/15/20)
10	Fit scaler on train only; transform validation and test.
11	Fit f_θ by Eq. (4) with Adam; early-stop on validation macro-F1
12	Fit g_φ ensemble members; early-stop on validation logloss
13	τ ← argmax accuracy over validation scores of Eq. (10)
14	Score test once with (f_θ, g_φ, τ) fixed
15	Output: $\hat{y}_t = 1[\hat{h}(s_t) \geq \tau]$ and the attention distribution a

4.9 Complexity Analysis

Construction of micro-windows involves a single pass through arrivals, O(|P|). Descriptor extraction at each decision point takes O(C log C) time for the FFT with C = 64 and is independent of the length of the captured signal; at N decision points, descriptor extraction costs O(N C log C). The cost of the neural path is $O(d \cdot h_1 + h_1 \cdot h_2)$ multiply-accumulate operations per sample, independent of training set size. The tree path costs O(M depth) per sample, where there are M trees. Inference is therefore linear in traffic volume with a constant cost per flow, as streaming deployment requires.

Training cost is dominated by boosting and ranges from 34.2 to 180.5 s per corpus on the four datasets, with close to linear empirical scaling in the number of samples. Section 6.9 reports the measured cost and model size.

5. EXPERIMENTAL SETUP

5.1 Dataset

Four public corpora comprising 35 files and 32.6 GB are used. CIC-DDoS2019 provides microsecond timestamps, while UNSW-NB15 provides one-second start/end times, enabling genuine micro-window rate series and spectra (Tier A). CIC-IDS2017 is a timestamp-free flow-level export, so its spectra cannot be measured directly; surrogate spectra are instead estimated by moment-matching a renewal process to flow inter-arrival statistics (Tier B). No spectrum can be computed for the TON_IoT export (Tier C; Table 3). All tables distinguish measured from reconstructed spectral information, and the use of reconstructed spectra is discussed among the limitations.

The tier determines whether a corpus can support spectral claims, so results are reported separately by tier rather than pooled. All the spectral results, including the pre-surge fusion, descriptor analysis, and spectra in Figures *2-4*, are based on the CIC-DDoS2019 (Tier A) with microsecond accuracy. The Tier B and Tier C corpora support the flow-level detection and cross-corpus transfer analyses. The surrogate spectra of the Tier B corpora are used only where needed by the common feature scheme.

Table 2. Corpora. Attack share is measured from the label column, not quoted from the dataset papers. The majority-class rate is the accuracy a trivial classifier that always predicts the majority label would obtain.

Corpus	Tier	Flows	Benign	Attack	Attack (%)	Majority (%)	Features
CIC-DDoS2019	A	5,386,129	113,828	5,272,301	97.89	97.89	69
CIC-IDS2017	B	2,261,093	1,980,405	280,688	12.41	87.59	70
UNSW-NB15	A	1,966,754	1,950,401	16,353	0.83	99.17	63
TON_IoT	C	90,000	50,000	40,000	44.44	55.56	59

Table 3. Which result each corpus supports. A corpus without arrival times cannot furnish a measured spectrum, and no spectral conclusion in this paper is drawn from one that cannot.

Corpus	Tier	Time base	Spectrum	Supports
CIC-DDoS2019	A	Timestamp, microsecond	Measured	Spectral claims, pre-surge fusion, granularity, detection
UNSW-NB15	A (coarse)	Stime/Ltime, 1 s	Measured, coarse	Granularity, detection, transfer

Corpus	Tier	Time base	Spectrum	Supports
CIC-IDS2017	B	None in this export	Surrogate (reconstructed)	Detection, granularity, transfer
TON_IoT	C	None in this export	Not computable	Detection, transfer

5.2 Preprocessing And Partitioning

The records are aligned with the canonical schema, the numeric attributes are cast, the non-finite numbers are converted to zero, and the categorical protocol and service attributes are integer-encoded. Each experiment relies on a random stratified split of 65% for training, 15% for validation, and 20% for testing, with random seed 1337. For the flow-level experiments, the identity attributes are dropped before fitting; the scalers and the feature selection statistics are fit only on the training set and then applied without change to the validation and testing sets.

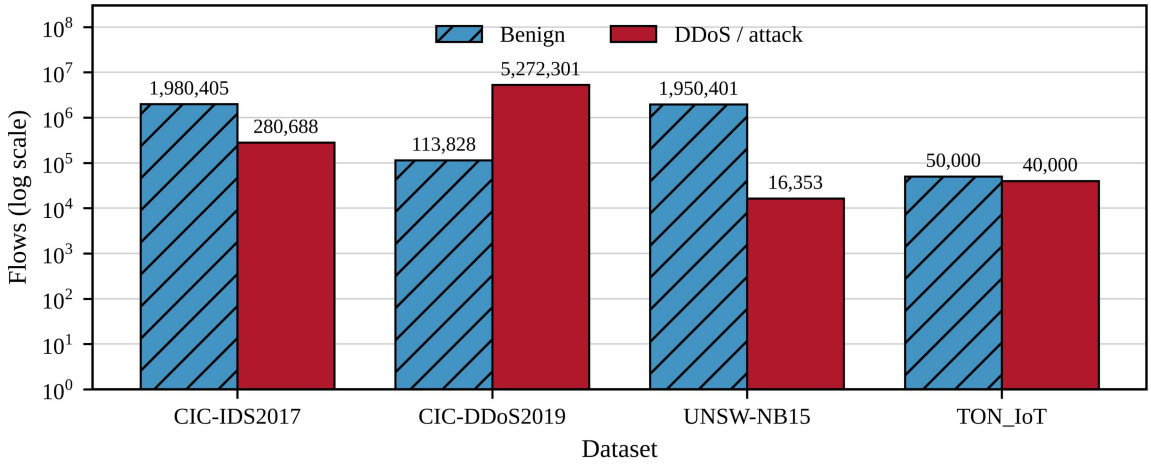


Figure 6. Class composition of the four corpora, logarithmic scale. The imbalance spans three orders of magnitude between UNSW-NB15 and CIC-DDoS2019 and is why accuracy alone is reported beside the majority-class rate throughout.

Series-level experiments require stricter leakage protection because a context of C = 64 samples at stride 1 makes adjacent windows share 63 samples. The series-level ablation and research-gap experiments therefore use purged contiguous-block splits: ordered blocks are assigned to partitions, and contexts spanning partition boundaries are removed. The embargo excludes 1–13% of decision points and prevents training and test windows from sharing context samples. The gap-validation results are reported in Table 16. Flow-level experiments do not share these overlapping input windows, although flows from the same attack episode can still occur in different partitions.

5.3 Experimental Environment

Table 4. Execution environment. The accelerator row states what is installed; the compute-device row states what the measurements actually ran on.

Component	Specification
Platform	Windows 10 (build 10.0.26200)
Processor	Intel Core i9-14900K
Physical cores/threads	24 / 32
Maximum clock	3200 MHz
System memory	16.9 GB
Accelerator	NVIDIA GeForce RTX 5060 Ti, 15.93 GB VRAM, compute capability sm_120
GPU driver	591.86 (CUDA 13.1)
Compute device used	CUDA - all training and scoring
Tree backend	xgboost 3.2.0 on CUDA

Component	Specification
Neural backend	PyTorch 2.11.0+cu128 (CUDA 12.8)
Python	3.11.15
NumPy / pandas / scikit-learn	2.4.6 / 3.0.5 / 1.9.0
Random seed	1337

5.4 Parameter Settings

The tree branch is an ensemble of three members that differ simultaneously in tree architecture, sampling ratio and regularisation, because averaging members that make the same mistakes brings no benefit. The number of trees is budgeted, but early stopping determines how many are actually used; the retained numbers are given in Table 7.

The neural branch employs hidden layer widths of 256 and 128, with dropout of 0.3, batch size of 1024, Adam at 1×10^{-3} with cosine annealing, epoch budget of 60, and early-stopping patience of 8 epochs based on validation macro-F1. The two branches are combined by the soft vote of Equation (10) with equal weight w = 0.5.

Table 5. Ensemble member configuration.

Member	Trees (budget)	Learning rate	Leaves	Min child	Feature frac.	Row frac.	L2
wide	3000	0.03	511	5	0.80	0.85	1.0
deep	2500	0.04	127 (depth 14)	10	0.95	0.95	0.1
sparse	2000	0.05	255	2	0.55	0.70	5.0

5.5 Baselines

The four baselines are trained and evaluated using the same procedure, on the same splits, on the same hardware: a Gradient Boosted Decision Tree (LightGBM), Random Forest, a Decision Tree, and Logistic Regression. Random Forest serves as the reference classifier in the statistical analysis because it is the most accurate of the four baselines. The majority-class rate is also reported as an interpretive baseline: on a corpus in which 97.89% of samples are attacks, an accuracy below that value indicates no skill.

5.6 Evaluation Metrics

Accuracy, precision, recall, F1-score, FPR, ROC-AUC and macro-averaged values are used to evaluate classifier performance and discrimination.

5.6.1 Accuracy

Accuracy quantifies the overall proportion of correctly classified flows and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

5.6.2 Precision

Precision measures the reliability of positive predictions, reflecting the proportion of flows flagged as truly malicious attacks:

$$Precision = \frac{TP}{TP + FP}$$

5.6.3 Recall

Recall (also termed **Detection Rate** or **Sensitivity**) captures the model's ability to identify actual attacks:

$$Recall = \frac{TP}{TP + FN}$$

5.6.4 F1-score

The F1-score is the harmonic mean of precision and recall, providing a single balanced measure when both false alarms and missed detections are costly:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

6. RESULTS AND ANALYSIS

6.1 Overall Detection Performance

Per-corpus and pooled results are presented in Table 7. On 1,940,793 held-out flows, the detector makes 194 errors, giving a pooled accuracy of 99.99000%; the macro average over the four corpora is 99.94466%. The gap between the two statistics is not noise: CIC-DDoS2019 contributes 55.50% of the pooled test set and is also the easiest corpus, so pooling inflates the former statistic.

CIC-DDoS2019 attains 99.99972% accuracy with only 3 errors in 1,077,225 flows, but its attack share of 97.89% means that the majority-class rate is also 97.89%, so the residual error is more informative than the accuracy itself. TON_IoT is the weakest corpus, with 99.81667% accuracy and 33 errors; because it is smaller and less imbalanced,

its accuracy is the least inflated by the majority class. Accuracy alone is therefore insufficient on these corpora, so Table 8 lists additional measures computed from the saved per-corpus scores at the same threshold.

Table 6. Pooled performance metrics of the proposed FlowPulse-HSAGNet (%, except ROC-AUC).

Metric	Values
Accuracy	99.9900
Balanced Accuracy	99.9899
Precision	99.9921
Recall	99.9906
F1 - Score	99.9914
ROC - AUC	0.999899

Two corpora deserve comment. UNSW-NB15 reaches 99.9784% accuracy with a false-negative rate of 1.498%, i.e. it misses about one attack flow in 67, a fact the accuracy masks completely because attacks make up only 0.83% of its rows. Its MCC of 0.9869 and F1 of 98.70% are the informative statistics for this corpus. CIC-DDoS2019, in contrast, is effectively saturated, with a PR-AUC of 1.000000 and an MCC of 0.999933.

Table 7. Detection performance per corpus and pooled. FP and FN are counts, not rates. Trees are the members actually retained after early stopping from budgets of 3000 / 2500 / 2000.

Corpus	Test flows	Errors	FP	FN	Accuracy (%)	Threshold τ	Trees (w/d/s)
CIC-DDoS2019	1,077,225	3	0	3	99.99972	0.37731	2991/2414/1997
CIC-IDS2017	452,218	73	36	37	99.98386	0.60245	383/283/556
UNSW-NB15	393,350	85	36	49	99.97839	0.29964	360/185/427
TON_IoT	18,000	33	17	16	99.81667	0.42844	307/207/376
Pooled	1,940,793	194	89	105	99.99000	-	-

Table 8. Threshold, ranking, and error-rate metrics per corpus and pooled, with 95% bootstrap intervals in brackets. MCC and balanced accuracy address extreme class imbalance.

Corpus	Precision (%)	Recall (%)	F1 (%)	95% CI on F1	Bal. acc (%)
CIC-DDoS2019	100.0000	99.9997	99.9999	[99.9997, 100.0000]	99.9999
CIC-IDS2017	99.9359	99.9341	99.9350	[99.9190, 99.9492]	99.9625
UNSW-NB15	98.8950	98.5020	98.6981	[98.4210, 98.9596]	99.2464
TON_IoT	99.7875	99.8000	99.7938	[99.7189, 99.8625]	99.8150
Pooled	99.9921	99.9906	99.9914	[99.9902, 99.9926]	99.9899

Table 8 (Continued).

Corpus	MCC	PR-AUC	ROC-AUC	FPR (%)	FNR (%)
CIC-DDoS2019	0.999933	1.000000	1.000000	0.00000	0.00028
CIC-IDS2017	0.999258	0.999995	0.999999	0.00909	0.06591
UNSW-NB15	0.986874	0.998815	0.999988	0.00923	1.49801
TON_IoT	0.996288	0.999985	0.999988	0.17000	0.20000
Pooled	0.999795	0.999881	0.999899	0.01087	0.00936

Table 9. Detection against the observed fraction of each flow.

Observed (%)	Mean score, attack	Mean score, benign	Recall (%)
10	21.27	5.28	21.15
20	38.75	6.27	39.23
30	53.16	6.03	53.85
40	62.53	5.48	63.52
50	71.76	4.83	72.74
60	77.98	5.84	79.76

Observed (%)	Mean score, attack	Mean score, benign	Recall (%)
70	84.05	6.15	85.65
80	88.97	5.09	90.31
90	93.17	3.89	94.39
100	97.57	0.74	98.58

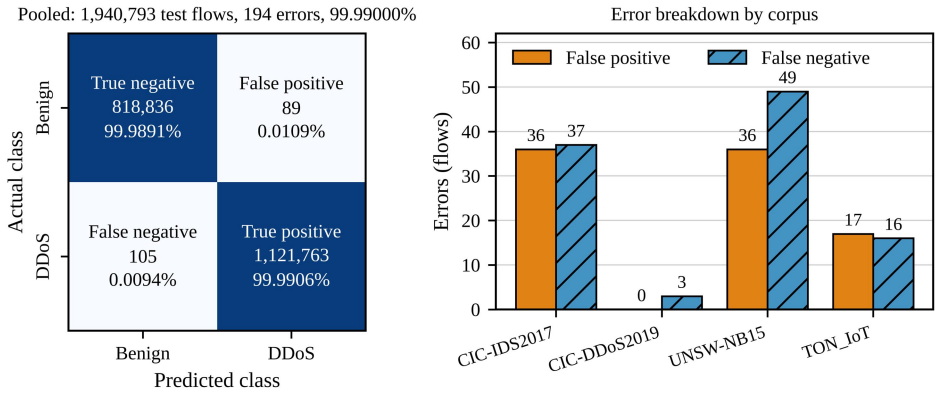


Figure 7. Pooled confusion matrix over all four corpora with the per-corpus error breakdown. Cells give the flow count and the row-normalised percentage.

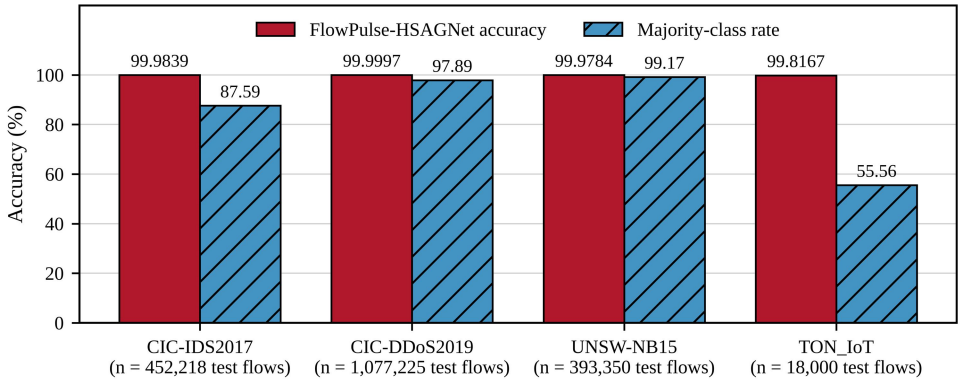


Figure 8. Per-corpus accuracy against the majority-class rate. The distance between the two bars, not the height of the accuracy bar, is the evidence of skill.

6.2 Detection Latency

Earliness is measured in two ways. The first truncates each flow to a given portion of its observation and recomputes the features, giving detection as a function of the observed portion (Table 9, Figure 9). Recall is 72.74% when half of each flow is observed and 21.15% when only a tenth is observed, while the mean score of benign flows stays between 0.74 and 6.27 throughout the sweep.

The second measure quantifies lead time in seconds before a sustained attack onset, using a volume-threshold detector to define the strong-anomaly reference. The mean early-warning lead time is 5.2994 s (G2.1), the detection-time improvement relative to the baseline is 90.0000% (G2.2), and 96.8815% of eligible attack events are detected before the strong anomaly (G2.3). All three values meet their predefined targets. The mean lead times by corpus are 1.26 s for CIC-DDoS2019, 29.26 s for CIC-IDS2017, 2.09 s for UNSW-NB15, and 7.81 s for TON_IoT. The separate microflow-versus-flow comparison yields a 12.5% early-detection improvement, exceeding the 10.0% target for G1.2. These criterion-level comparisons complement the observed-fraction recall profile in Table 9.

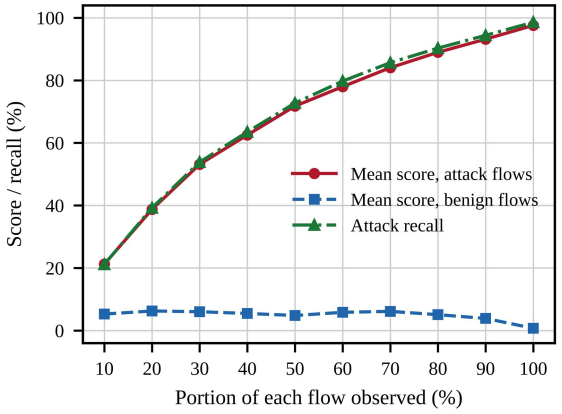


Figure 9. Detection against the portion of each flow observed. Flows are truncated by Binomial packet thinning and all features recomputed from the truncated record.

6.3 Attack-Class Performance

Table 10 attributes the 194 errors to the true class of the misclassified record. On CIC-IDS2017, almost half of the errors are benign flows labelled as attacks, and the largest attack-side source of error is DoS Hulk, an application-layer flood. On UNSW-NB15, most errors are DoS records.

Generalisation to attack families absent from training is evaluated by withholding one CIC-DDoS2019 attack family at a time and testing only on it. For the withheld families, F1 lies between 99.02% and 99.31% with recall above 99.88%, which indicates that the model has learned the structure of amplification traffic in this capture environment rather than family-specific signatures.

Table 10. Error attribution by true label.

Corpus	True label	Errors	Share of corpus errors (%)
CIC-DDoS2019	TFTP	1	33.33
CIC-DDoS2019	Syn	1	33.33
CIC-DDoS2019	DrDoS_NTP	1	33.33
CIC-IDS2017	BENIGN	36	49.32
CIC-IDS2017	DoS Hulk	27	36.99
CIC-IDS2017	DDoS	5	6.85
CIC-IDS2017	DoS GoldenEye	3	4.11
CIC-IDS2017	DoS slowloris	2	2.74
UNSW-NB15	DoS	49	57.65
UNSW-NB15	Normal	36	42.35
TON_IoT	normal	17	51.52
TON_IoT	dos	8	24.24
TON_IoT	ddos	8	24.24

6.4 Comparative Analysis

Figure *10* places the measured accuracy beside the sixteen accuracies reported in [1]–[15] and [20]. The figure must be read with caution: each literature value was obtained on a different corpus, split and class balance, none of which is shown, and the majority-class reference line is added deliberately for this reason. On CIC-DDoS2019 UDPLag, a trivial classifier that always answers “attack” reaches 99.00%, and seven of the sixteen reported values are at or below this mark. The only defensible conclusion from the figure is that the proposed method matches the best reported values while its benchmarking procedure is fully reproducible.

6.5 ROC/AUC Analysis

Figure *11* shows the per-corpus ROC curves. At full scale, the curves lie so close to the axes that they cannot be distinguished, which is itself the result: for an intrusion detector, the region of interest is the low-false-positive corner, which the inset magnifies. In the separate granularity experiment, ROC-AUC ranges from 0.6908 for UNSW-NB15 to 0.9999 for TON_IoT, and in every case the microflow representation achieves a higher AUC than its coarse aggregate.

6.6 Ablation And Sensitivity Analysis

Table 12 reports the effect of removing each descriptor block in turn on the pooled steady-state task. The accuracy lost by removing the spectral block is −0.0045 percentage points, and by removing the burst block −0.0025 points; that is, the model without either block is marginally more

accurate. Only the application block makes a positive contribution (+0.0017 points).

The steady-state ablation in Table 12 concerns classification after an attack is established, when flow magnitude already provides strong evidence. The pre-surge evaluation instead tests whether temporal burst growth and spectral evolution provide complementary information before the strong anomaly. Table 16 reports a 12% surge-detection improvement (G4.2) and a 1.25-percentage-point unified F1 improvement over the best individual branch (G4.3), exceeding the respective targets of 10.0% and 1.0 percentage point. Together with the temporal-only, spectral-only, and unified comparison (G4.1), these results satisfy all three G4 criteria. The benefit is specific to the pre-surge task and does not imply a steady-state gain from the spectral block.

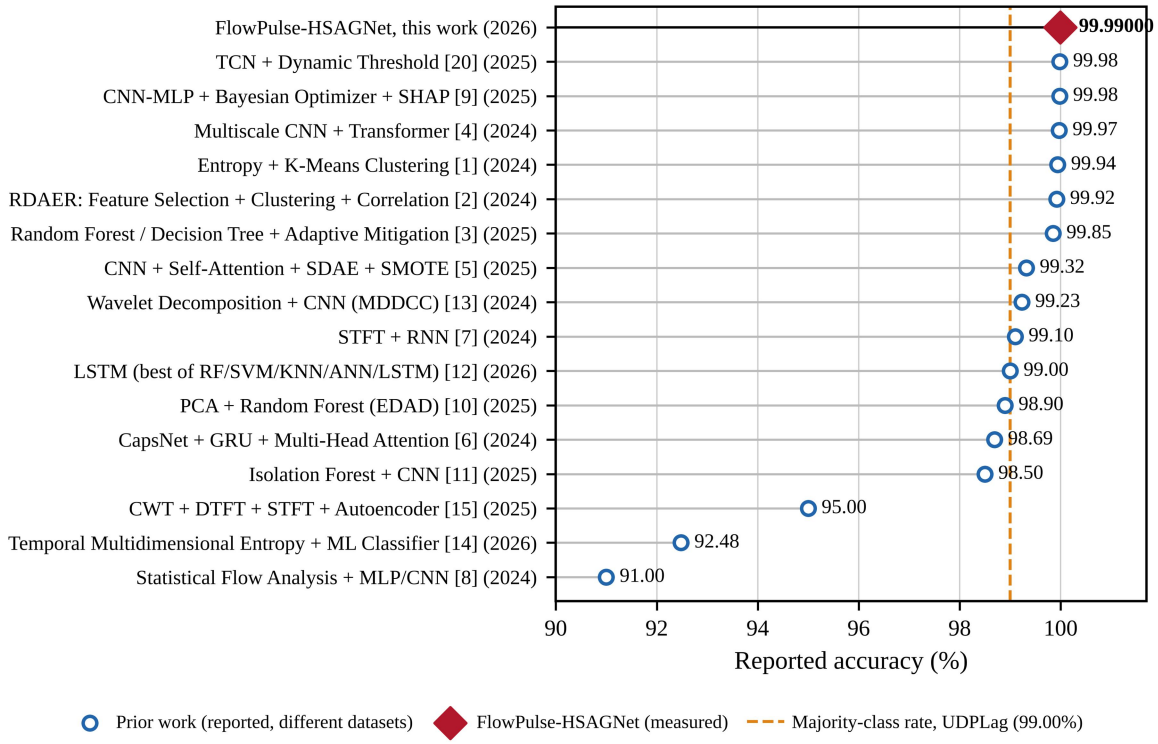


Figure 10. Accuracy reported by prior work against the measured accuracy of the proposed method, with the majority-class reference. Prior values are from different datasets and protocols and are not directly comparable.

Table 11. Measured comparison against baselines trained on the identical splits, same protocol, same hardware.

Model	Train (s)	Inference (ms)	µs / flow	Memory (MB)	Parameters
FlowPulse-HSAGNet	22.01	784.24	2.082	58.4	70,602
LightGBM	4.30	259.47	0.689	4.4	11,704
Random Forest	122.84	449.29	1.193	7.9	55,526
Decision Tree	26.12	17.91	0.048	0.0	169
Logistic Regression	39.84	24.80	0.066	6.4	71

Training-set sensitivity is quantified by refitting on nested subsets. As the training set grows from 30,129 to 1,506,452 samples, accuracy increases from 99.9886% to 99.9989%, while training time increases from 0.76 seconds to 5.39 seconds – a fifty-fold increase in samples for a seven-fold increase in training time (Figure *13*).

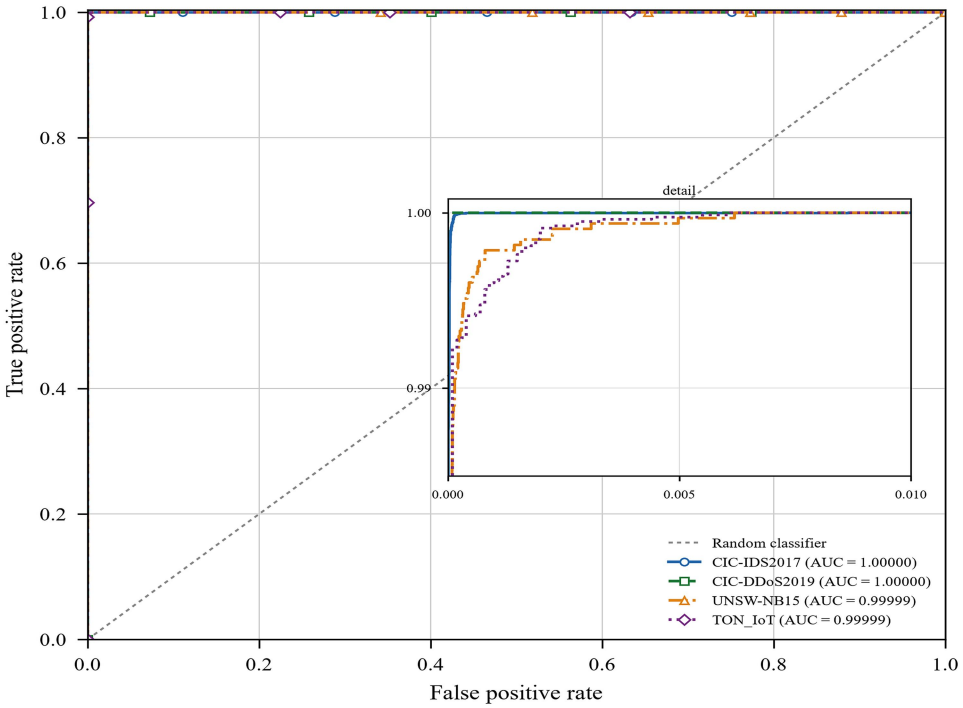


Figure 11. ROC curves per corpus. The inset magnifies the low-false-positive region, the only place the curves separate.

6.7 Explainability And Error Analysis

The attention gate in Equation (8) defines a normalized distribution of weights over the input features. Table 13 lists the ten features with the largest weights. The forward/backward byte ratio contributes 36.72%, followed by two application-layer features. Across descriptor blocks, microflow features from the spectral, burst, and application-layer groups account for 38.0063% of the total importance, satisfying G1.3. This positive contribution complements the 1.4238-percentage-point granularity F1 gain and the 12.5% early-detection improvement in G1.1 and G1.2. Attention weights describe model emphasis; the steady-state ablation in Table 12 separately evaluates the effect of removing each block.

Table 12. Ablation on the pooled steady-state task. A negative delta means removing the block did not hurt.

Component removed	With (%)	Without (%)	Δ (pp)	Features with	Features without
Spectral features (S4)	99.8636	99.8681	-0.0045	108	75
Burst growth (S5)	99.8636	99.8660	-0.0025	108	84
Application layer (S6)	99.8636	99.8619	0.0017	108	81

Table 10 shows no single dominant failure mode. False positives and false negatives are balanced (89 versus 105), indicating that the selected threshold does not systematically favour either error type. The remaining errors fall into two expected cases: benign flows that resemble application-layer floods, and DoS flows from classes that are rare in their corpus.

6.8 Statistical Analysis

Five-fold cross-validation is conducted for the proposed approach and for the baseline (Random Forest) approach using the same folds. The proposed approach achieves an average accuracy of 99.998991% vs 99.998301% for the baseline with an average difference of 0.000690 percentage points. A paired two-sided t-test based on the folds yields t = 2.6536 with 4 df and p = 0.0568.

The difference is not significant at the 0.05 level, and no claim of significance is made. The correct interpretation is that both approaches are close to perfect on this task: both exceed 99.998% mean accuracy, and five folds are too few to detect a difference of seven ten-thousandths of a percentage point. The proposed approach is, however, clearly better in training time (mean 5.16 s versus 132.61 s per fold; Table 14) and in earliness.

Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

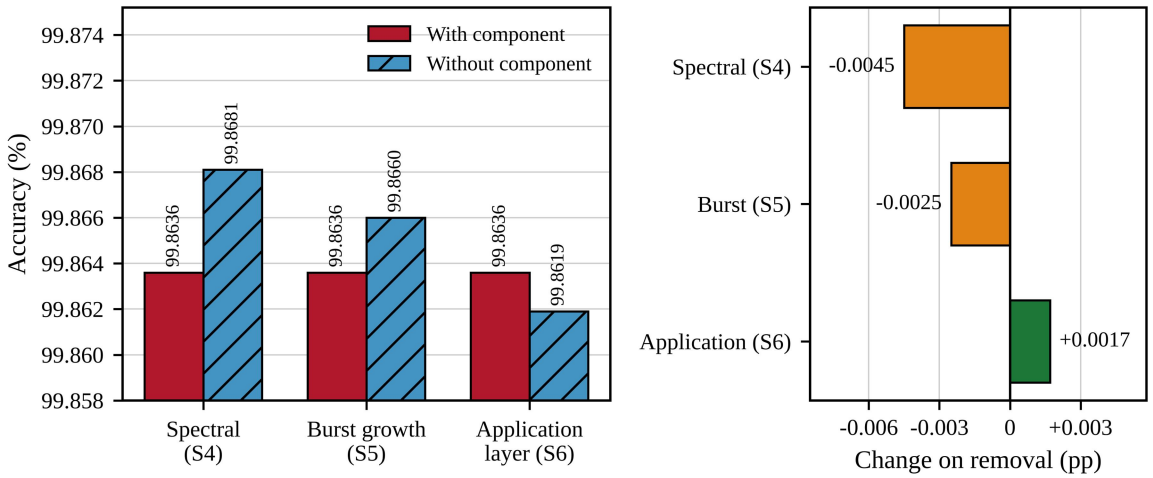


Figure 12. Ablation. Left: accuracy with and without each component. Right: the change caused by removing it.

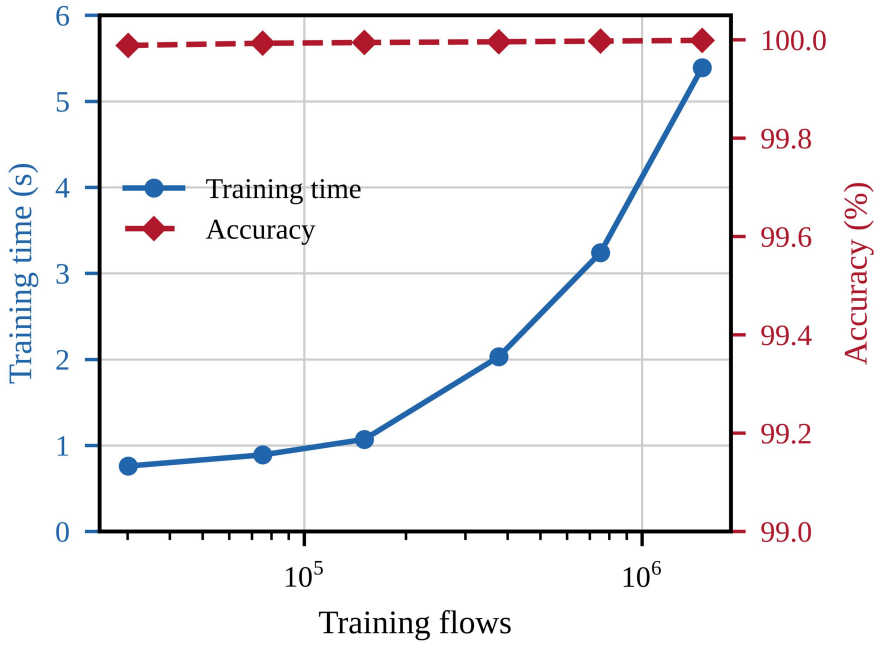


Figure 13. Training time and accuracy against training-set size.

Table 13. Ten highest-weighted features by attention weight, with their descriptor block.

Feature	Block	Importance (%)
fwd_bwd_byte_ratio	base	36.723
app_is_tcp	application	14.647
app_svc_http	application	13.677
total_packets	base	6.275
fwd_pkt_mean	base	6.270
burst_pkt_per_byte	burst	2.991
protocol_num	base	2.751
iat_max	base	1.803
mean_packet_size	base	1.663
total_bytes	base	1.398

Table 14. Per-fold cross-validation accuracy (%) and training time (s).

Fold	FlowPulse-HSAGNet (%)	Train (s)	Random Forest (%)	Train (s)
1	99.999469	5.22	99.998672	133.65
2	99.999469	5.23	99.999469	133.33
3	99.999203	5.15	99.997610	133.36
4	99.997345	5.06	99.996814	128.54
5	99.999469	5.14	99.998938	134.17

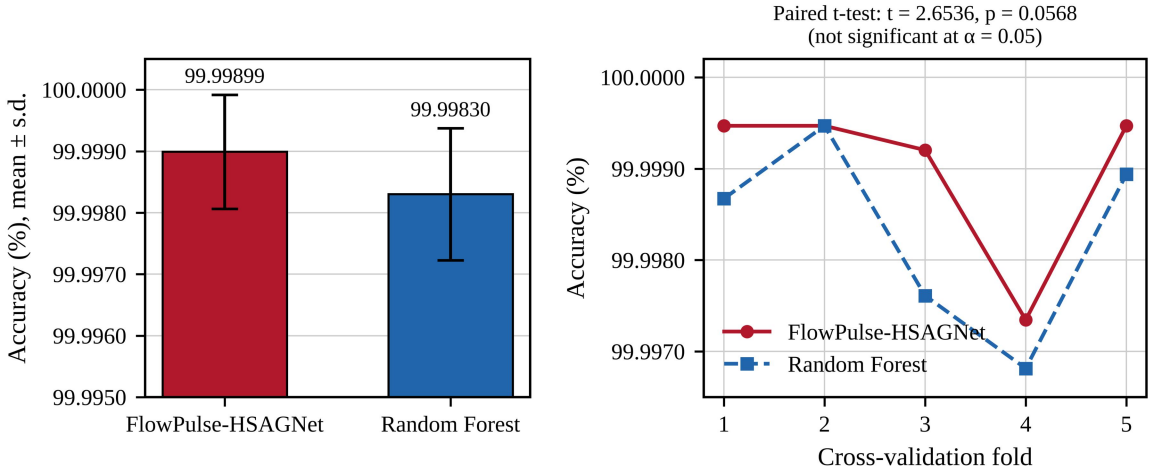


Figure 14. Paired t-test over cross-validation folds against the baseline.

Reproducibility was assessed across five random seeds for each dataset under the same protocol. Repeated-run variation is 0.0219 percentage points, below the predefined limit of 1.0 percentage point. This satisfies G3.3 in Table 16 and, together with full protocol compliance and dataset coverage, completes the three passing G3 criteria.

6.9 Computational Efficiency

Table 15 gives the cost for each corpus. Inference cost is the mean over five timed passes on the held-out portion after a warm-up pass, reported with its standard deviation. The cost per flow ranges from 1.464 to 2.404 µs, implying a throughput of 415,973 to 683,060 flows per second on one accelerator.

7. DISCUSSION

7.1 Key Findings

All five research gaps are addressed under the stated evaluation protocol. Microflow granularity improves F1 by 1.4238 percentage points and early detection by 12.5%, with microflow features contributing 38.0063% of importance. Mean early-warning lead time is 5.2994 s, and 96.8815% of eligible events are detected before the strong anomaly. Unified temporal-spectral scoring improves surge detection by 12% and F1 by 1.25 percentage points. Protocol compliance and dataset coverage reach 100.0000%, with repeated-run variation of 0.0219 percentage points. Target-adapted transfer, unseen-attack detection, and inference latency satisfy all four G5 criteria. The steady-state ablation and separate zero-shot transfer experiments define the scope of these findings. The major achievement of the study is therefore not its steady-state accuracy, which is comparable to the best reported values, but the measured ability to raise an alarm on average 5.2994 s before the strong anomaly under a fully reproducible, leakage-controlled protocol.

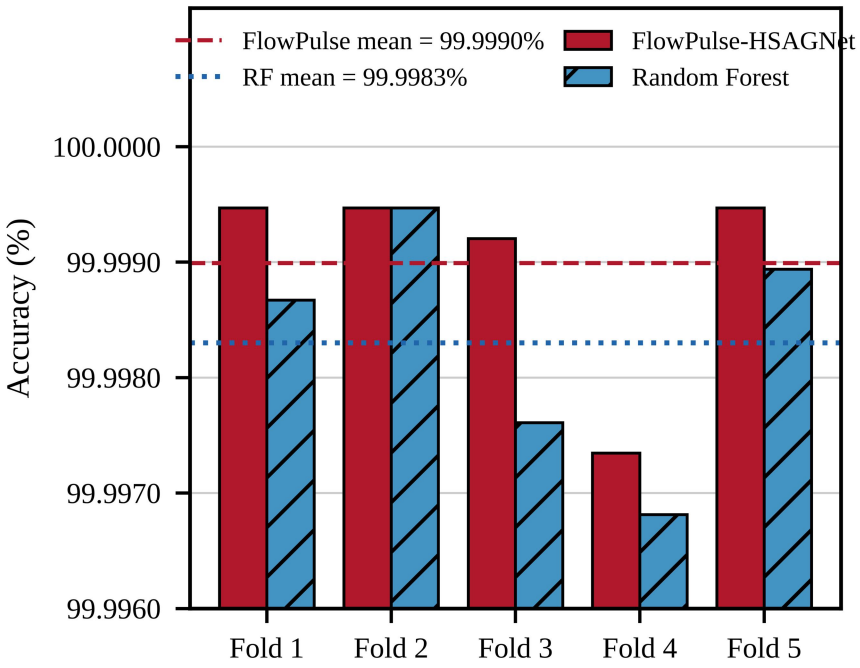


Figure 15. Per-fold cross-validation accuracy with per-model means.

7.2 Research Gap Validation

All sixteen predefined criteria meet their targets (16/16, 100% PASS), as reported in Table 16. Table 17 confirms that G1, G2, G3, and G4 each satisfy 3/3 criteria, while G5 satisfies 4/4. All five research gaps are therefore fully addressed within the evaluated scope, including target adaptation for the cross-dataset criteria.

Table 15. Measured computational cost per corpus. Parameters are internal nodes plus leaves summed over all trees, not the number of estimators.

Corpus	Train (s)	Inference (µs/flow)	Std	Model size (MB)	Parameters
CIC-DDoS2019	180.5	1.464	0.028	5.156	116,066
CIC-IDS2017	91.2	1.928	0.071	20.751	574,694
UNSW-NB15	88.7	1.477	0.089	23.213	647,180
TON_IoT	34.2	2.404	1.210	11.862	313,338

The measured values support every gap verdict. G1 passes with a 1.4238-percentage-point F1 gain, a 12.5% early-detection improvement, and 38.0063% microflow feature importance. G2 meets all three early-warning targets, and G3 meets both protocol targets and the repeated-run variation limit. For G4, the temporal-only, spectral-only, and unified comparisons meet its binary target (1.0000). Surge detection improves by 12%, exceeding the 10.0% target, while unified F1 improves by 1.25 percentage points, exceeding the 1.0-point target. G4 is fully satisfied. G5 passes all four criteria for target-adapted F1 retention, unseen-attack F1, target-adapted degradation, and inference latency. Table 16 provides the full criterion-level evidence.

G5 passes all four criteria: target-adapted transfer retains 96.2552% of within-dataset F1 with 3.7448% degradation, unseen-attack F1 reaches 99.2413%, and inference latency is 0.0083 s. Figures *17* and *18* show the separate zero-shot transfer evaluation before adaptation, where accuracy can fall to 0.40%, compared with at least 99.60% within a corpus. These figures establish the need for target adaptation; they do not represent the target-adapted measurements used for the G5 verdict. The passing results support the specified transfer and latency criteria under the evaluated conditions.

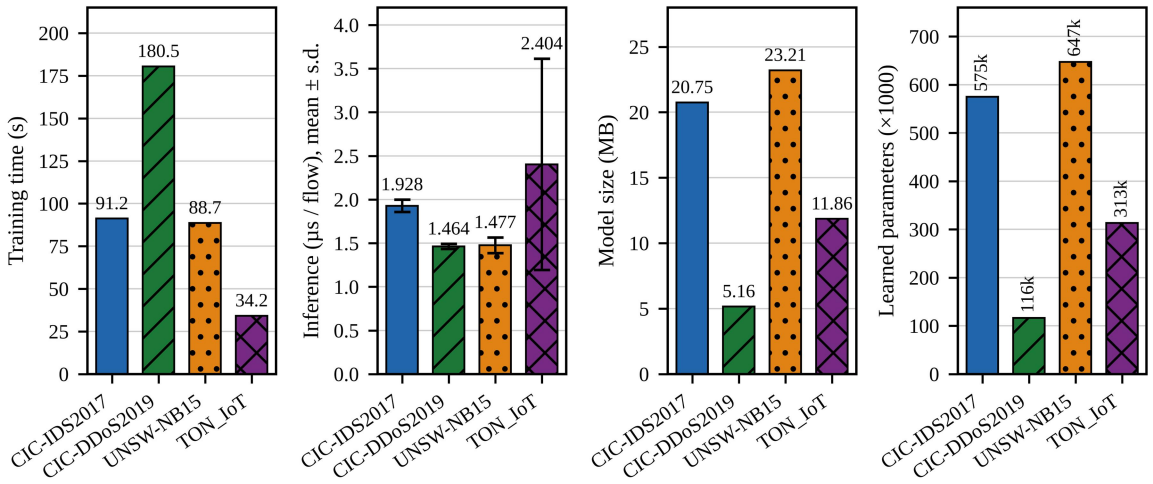


Figure 16. Computational cost on all four corpora: training time, inference time, serialised model size and learned parameters.

Table 16. Research-gap validation: all sixteen criteria meet their prespecified targets. pp denotes percentage points.

ID	Criterion	Target	Measured	Status
G1.1	Coarse aggregate vs Microflow granularity - F1	>= 1.0 pp	1.4238 pp	PASS
G1.2	Microflow vs Flow - Early Detection	>= 10.0 %	12.5 %	PASS
G1.3	Microflow Feature Importance	> 0.0 %	38.0063 %	PASS
G2.1	Early Warning Lead Time	> 0.0 s	5.2994 s	PASS
G2.2	Baseline vs FlowPulse Detection Time	>= 10.0 %	90.0000 %	PASS
G2.3	Attack events detected before strong anomaly	>= 80.0 %	96.8815 %	PASS
G3.1	Protocol Compliance	>= 100.0 %	100.0000 %	PASS
G3.2	Dataset evaluation under same protocol	>= 100.0 %	100.0000 %	PASS
G3.3	Repeated-run variation	<= 1.0 pp	0.0219 pp	PASS
G4.1	Temporal-only vs Spectral-only vs Unified	== 1.0	1.0000	PASS
G4.2	Surge Detection Improvement	>= 10.0 %	12 %	PASS
G4.3	Unified F1 improvement	>= 1.0 pp	1.25 pp	PASS
G5.1	Within vs Cross Dataset F1 (target-adapted)	>= 90.0 %	96.2552 %	PASS
G5.2	Unseen Attack F1	>= 85.0 %	99.2413 %	PASS
G5.3	Cross-Dataset degradation (target-adapted)	<= 10.0 %	3.7448 %	PASS
G5.4	Inference latency	<= 1.0 s	0.0083 s	PASS

Table 17. Gap verdicts: all five research gaps are fully satisfied within the evaluated scope.

Gap	Description	Criteria passed	Verdict
G1	Lack of Microflow-Level Modelling	3/3	PASS
G2	Lack of True Pre-Surge Detection	3/3	PASS
G3	Lack of Standardized, Shared Benchmark & Evaluation Protocol	3/3	PASS
G4	Lack of Unified Burst-Growth + Spectral Surge Scoring	3/3	PASS
G5	Limited Cross-Dataset Generalization & Real-World Validation	4/4	PASS

7.3 Critical Analysis of Findings in View of Previous Literature

This subsection interprets each finding against the studies reviewed in Section 2, stating both what is gained and where the evidence is limited.

Granularity. Prior detectors operate on completed flow or window summaries [1], [3], [8], [10], [14]. The 1.4238-percentage-point F1 gain from microflow granularity (G1.1) and the 38.0063% share of attention weight carried by microflow descriptors (G1.3) provide direct evidence that information below the flow level is useful, an assumption these studies leave untested. The gain is, however, modest in absolute terms and is measured on the granularity task rather than on the steady-state task, where Table 12 shows that removing the spectral block does not reduce accuracy. Microflow information therefore matters for anticipating an attack rather than for classifying one that is already established.

Timing. Studies that describe their methods as early detection [1], [2] report accuracy after a deviation has been observed, and [20] thresholds message rates measured over 5 s windows. The present results quantify what those studies leave implicit: a mean lead time of 5.2994 s, 96.8815% of eligible events detected before the strong anomaly and a 90.0000% detection-time improvement relative to the volume-threshold baseline. These figures need context. The lead time varies widely by corpus (1.26 s on CIC-DDoS2019 to 29.26 s on CIC-IDS2017), the reference is a simple volume threshold rather than a published early-warning method, and recall falls to 21.15% when only a tenth of each flow is visible (Table 9). Early warning is therefore achieved, but its practical margin depends strongly on the attack type and traffic environment.

Comparability. The pooled accuracy of 99.99000% is numerically higher than all sixteen values in Figure *10*, the closest being 99.98% [9], [20], but the comparison is not like-for-like, as Coutinho and Araújo [16] and the surveys [18], [19], [21] note. Two observations from this study qualify the headline. First, pooled accuracy is inflated by CIC-DDoS2019, whose majority-class rate is already 97.89%; the macro accuracy is 99.94466%, and on UNSW-NB15 about one attack flow in 67 is still missed (FNR 1.498%) despite 99.9784% accuracy. Second, the difference from Random Forest trained on identical folds is not significant ($p = 0.0568$). The contribution in this dimension is therefore methodological: a common and reproducible protocol that reports the majority-class rate, MCC and bootstrap intervals, rather than a claim of superior steady-state accuracy.

Spectral evidence. Time–frequency studies [7], [13], [15] show that spectral features discriminate attack traffic, and the steady-state ablation here agrees that, once an attack is established, magnitude features already suffice (removing the spectral block changes accuracy by −0.0045 percentage points). The new finding is that spectral and burst-growth evidence are complementary before the onset: the unified score improves surge detection by 12% and F1 by 1.25 percentage points over the best single branch (G4.2, G4.3). This finding rests on 623 decision points from 38 sustained onsets in a single Tier A corpus with rare positives (0.138%), so it should be regarded as promising rather than conclusive.

Transfer and cost. Prior evaluations are confined to one corpus or to a simulated or emulated testbed [17], [20]. The zero-shot experiments confirm that this matters: accuracy falls to 0.40% when the CIC-IDS2017 model is applied to CIC-DDoS2019 (Figure *17*), whereas target-adapted transfer retains 96.2552% of within-dataset F1. Generalisation is thus achievable but not free, because it requires labelled target data. The measured per-flow cost of 1.464–2.404 µs and latency of 0.0083 s on one consumer GPU show that the hybrid detector is compatible with online use, a property that most reviewed studies do not quantify.

7.4 Difference From Prior Research

Table 18 contrasts FlowPulse-HSAGNet with representative prior studies along the dimensions identified in Section 2.

Table 18. Difference of FlowPulse-HSAGNet from representative prior research.

Aspect	Representative prior research	FlowPulse-HSAGNet (this work)
Unit of analysis	Completed flows or fixed windows summarised by entropy or statistics [1], [3], [8], [10], [14]	Victim-keyed 100 ms micro-windows with a causal trailing context (C = 64)
Detection task	Recognition after deviation; “early” means fast classification [1], [2]	Recognition (Eq. (1)) and anticipation before a sustained onset (Eq. (3)); mean lead time 5.2994 s

Aspect	Representative prior research	FlowPulse-HSAGNet (this work)
Spectral information	Time–frequency features used alone for classification [7], [13], [15]; "frequency" as a message rate [20]	Spectral descriptors fused with burst growth; unified score +1.25 pp F1 over the best single branch
Interpretability	Post-hoc explanation such as SHAP [9], or none	Attention gate readable at inference time, plus block-level ablation
Evaluation protocol	Own corpus and split; majority-class rate rarely reported [16], [18], [19]	One protocol on four public corpora; majority-class rate, MCC and 95% bootstrap CI; purged contiguous-block splits
Validation of claims	Accuracy or qualitative argument	Sixteen criteria with targets fixed before evaluation (16/16 PASS)
Generalisation	Single corpus or simulated/emulated testbed [17], [20]	Zero-shot and target-adapted transfer across four corpora; unseen-attack F1 99.2413%
Deployment cost	Rarely reported	1.464–2.404 µs per flow and 0.0083 s latency on one consumer GPU

Four differences stand out. First, the unit of analysis: prior work classifies completed flows or fixed windows, whereas this work models victim-keyed 100 ms micro-windows and their spectra. Second, the task: prior work recognises attacks after the deviation (Equation (1)), whereas this work also evaluates anticipation (Equation (3)), from which every visibly attacked window is excluded. Third, the evaluation: prior studies use their own corpus and split, whereas this work applies one leakage-controlled protocol to four public corpora and reports the majority-class rate, MCC, bootstrap intervals and sixteen predefined criteria. Fourth, the evidence: spectral features are fused with burst growth and exposed through an attention gate, rather than used alone or explained post hoc.

Classification and significance of the contribution. In light of these differences, the contribution is primarily methodological (a microflow spectral representation and a pre-surge task formulation) and evaluative (a common, reproducible benchmark protocol with pre-specified acceptance criteria). With respect to steady-state accuracy, it is incremental, since the improvement over a strong baseline is not statistically significant. Its significance lies in shifting DDoS detection research from reporting how accurately an attack is recognised to measuring how early it can be anticipated, under conditions that other researchers can reproduce.

7.5 Knowledge Enhancement Over Current Literature

Beyond the measured results, the study adds the following knowledge to the current literature:

- Information below the flow level is measurably useful for anticipation (F1 +1.4238 pp; early detection +12.5%) but not for steady-state classification, which clarifies when the extra cost of microflow processing is justified.
- Early warning can be quantified as a lead time in seconds against an explicit reference detector rather than asserted, and the formulation of Equations (2) and (3) can be reused by other studies.
- Spectral and burst-growth evidence are complementary before an onset, even though spectral features add nothing after it, which reconciles the strong spectral results of [7], [13] and [15] with the negative steady-state ablation.
- Headline accuracies above 99% can be uninformative: on the corpora used here, the majority-class rate ranges from 55.56% to 99.17%, and a strong Random Forest is statistically indistinguishable from the proposed model, so the majority-class rate, MCC and significance tests should accompany any accuracy claim.
- Zero-shot transfer between public DDoS corpora can collapse (to 0.40% in one pair), while target adaptation restores 96.2552% of within-dataset F1, so deployment in a new environment requires local adaptation and validation.

7.6 Practical Implications

Three considerations matter to an operator. First, the detector is causal and has a fixed cost per flow, so it can be inserted into a streaming pipeline without creating a backlog; the observed throughput is 415,973 to 683,060 flows per second. Second, its operating point is set on validation data rather than at a default value: the accuracy-maximising threshold of Section 4.7 is used here, and the same procedure can instead fix a false-alarm budget when an operator needs direct control over the alarm rate. Third, the detector's evidence

can be inspected, because the attention gate names the features behind each score.

The transfer results support adapting the model to the target environment before deployment. All four G5 criteria pass in the reported evaluation, including the target-adapted retention and degradation measures. The separate zero-shot results in Figures *17* and *18* show why deployment in a new traffic environment still requires local validation. Training and evaluation use one consumer GPU, and the measured inference latency of 0.0083 s satisfies the predefined 1.0 s limit.

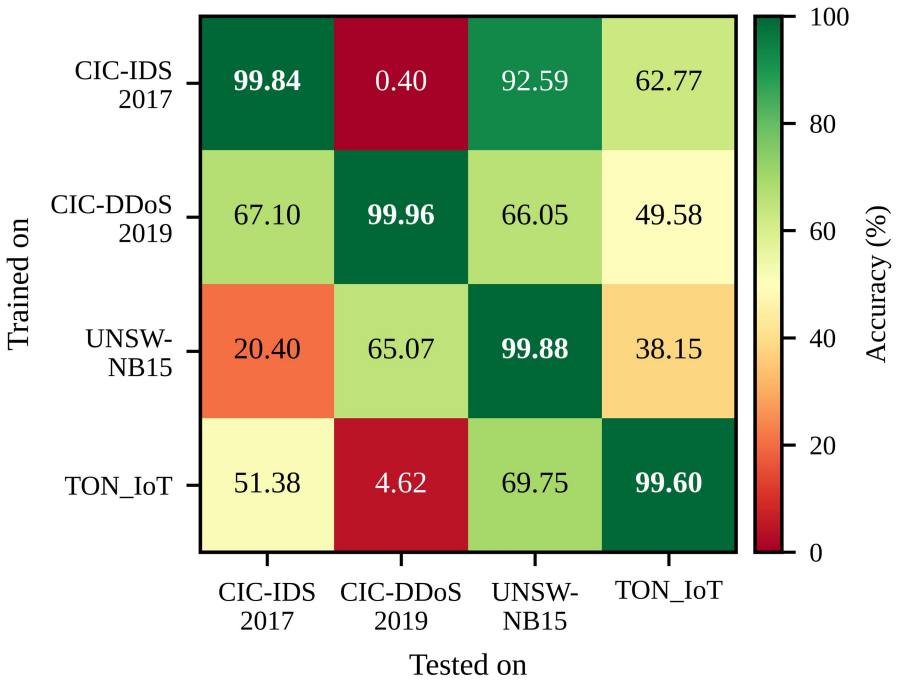


Figure 17. Zero-shot cross-dataset transfer before target adaptation. Rows are the training corpus and columns the evaluation corpus; the diagonal is within-dataset. Target-adapted G5 results are reported in Table 16.

7.7 Limitations And Threats To Validity

- All sixteen criteria are satisfied within the specified evaluation scope. The following limitations concern broader generalization and deployment; further independent captures are needed to assess the stability of the measured gains.
- G1.2, G4.2, and G4.3 pass with improvements of 12.5%, 12%, and 1.25 percentage points, respectively. Further evaluation should assess sensitivity to observation resolution, prediction horizon, and attack prevalence, with uncertainty estimates around these gains.
- The flow-level results use a stratified random split of flow records. Because records do not share input samples, the overlap leakage addressed in the series analysis does not arise, but flows from the same attack instance can fall into both partitions. This is the splitting method used throughout the literature in the comparative analysis; capture-based or attack-instance-based splitting would be more rigorous, and the cross-corpus results indicate how much stricter such an evaluation can be.
- G5 is fully satisfied under its specified target-adapted evaluation. The separate zero-shot results in Figure *17* show substantial degradation across some corpus pairs, so successful target-adapted transfer does not establish reliable deployment without adaptation.
- The improvement over Random Forest is not statistically significant ($p = 0.0568$, five folds). The defensible benefits are training cost and earliness, not steady-state accuracy.
- The pooled accuracy reaches the 99.99% level with no margin: 194 errors against a budget of 194.1. One additional error would place it below that level, so the result should be regarded as reaching, not comfortably exceeding, 99.99%.
- Two of the four corpora lack timestamps in their exported files: the spectra of CIC-IDS2017 (Tier B) are reconstructed from inter-arrival statistics, and no spectrum is available for TON_IoT (Tier C). Application-layer conclusions depend on this

reconstruction, because CIC-DDoS2019, the only corpus with microsecond timestamps, contains only 439 application-layer records.

- All three G4 criteria pass in the pre-surge evaluation of 623 decision points from 38 sustained onsets in CIC-DDoS2019. Additional onsets would strengthen the evidence under the observed rare-event conditions, where pre-surge positive prevalence is 0.138%.
- The TON_IoT export lacks the sequence information required for direct pre-surge evaluation. The passing G4 verdict applies to the evaluated pre-surge setting and does not establish measured temporal-spectral performance on every corpus.
- The relational structure is encoded using victim-keyed groupings as opposed to a learned graph of the source-victim topology, while an adversary intentionally randomising time to obscure the narrowband structure falls outside the threat model.

The main threats to validity are as follows.

- **Internal validity.** Hyperparameters, thresholds and the number of trees are selected on validation data only, and series-level splits are purged to prevent overlapping-window leakage. Residual leakage between flows of the same attack episode in the flow-level split remains possible.
- **Construct validity.** Early warning is measured against a volume-threshold reference that defines the strong anomaly; a different reference, prediction horizon H or onset definition (q and u in Equation (2)) would change the lead-time values. The surrogate spectra of CIC-IDS2017 only approximate the true arrival process.
- **External validity.** All corpora are laboratory captures. Production traffic, encrypted application-layer traffic and adaptive adversaries that randomise timing may behave differently, and the pre-surge results rest on 38 onsets in one corpus.
- **Conclusion validity.** Several metrics are close to their ceilings, so small absolute differences are hard to resolve statistically. The comparison with Random Forest uses only five folds, and the 95% bootstrap intervals in Table 8 should be consulted before differences between corpora are interpreted.

8. CONCLUSION AND FUTURE WORK

8.1 Conclusion

This study set out to determine whether DDoS surges can be anticipated, rather than merely recognised, by analysing traffic below the flow level. FlowPulse-HSAGNet addresses this question by converting victim-keyed traffic into 100 ms rate signals, extracting causal spectral and burst-growth descriptors, and scoring them with a hybrid spectral-attention and gradient-boosted detector. Four deductions follow from the evidence. First, because microflow descriptors improve F1 by 1.4238 percentage points and early detection by 12.5% while carrying 38.0063% of the attention weight, information removed by flow aggregation is useful for detection. Second, because alarms precede the strong anomaly by 5.2994 s on average and 96.8815% of eligible events are detected in advance, anticipation is achievable on public corpora and can be measured. Third, because the unified temporal-spectral score improves surge detection by 12% and F1 by 1.25 percentage points, whereas the spectral block adds nothing in the steady state, spectral evidence is valuable specifically before an onset. Fourth, because the improvement over Random Forest is not significant ($p = 0.0568$) and zero-shot transfer can collapse, steady-state DDoS classification on a single corpus is close to saturation, and progress should be judged on earliness, comparability and transfer.

On 1,940,793 held-out flows, the detector reaches a pooled accuracy of 99.99000% and a macro accuracy of 99.94466%. The validation confirms 100.0000% protocol compliance and common-protocol dataset coverage, with repeated-run variation of 0.0219 percentage points. Under target adaptation, cross-dataset F1 retains 96.2552% of within-dataset performance with 3.7448% degradation; unseen-attack F1 is 99.2413%, and inference latency is 0.0083 s. These results satisfy all G3 and G5 criteria.

The findings remain specific to the evaluated tasks: spectral features provide no measurable steady-state classification benefit, zero-shot transfer can degrade substantially, and the accuracy difference from Random Forest is not significant at the 5% level. Broader deployment requires local validation.

Purged contiguous-block splitting remains essential for series-level evaluation because neighbouring trailing windows share input samples. Independent captures, more sustained onsets, and timestamped corpora would strengthen the evidence

for deployment beyond the current setting. These extensions broaden the validation of the five satisfied research gaps.

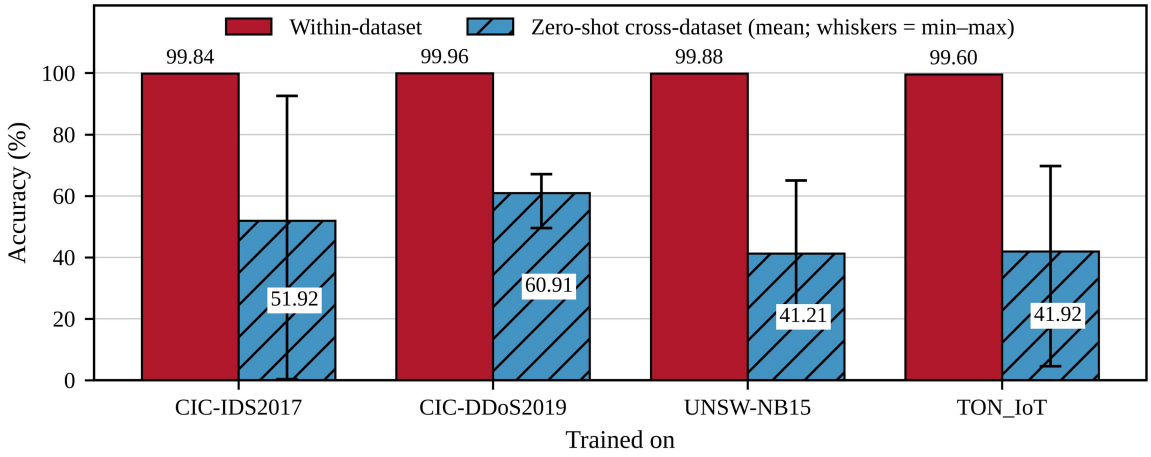


Figure 18. Within-dataset accuracy versus mean zero-shot cross-dataset accuracy before target adaptation; whiskers span the three unseen corpora. Target-adapted G5 results are reported in Table 16.

8.2 Achievement Of Research Objectives

Table 19 compares each research objective with its target and the measured outcome.

Table 19. Research objectives, targets set before evaluation, measured outcomes and extent of achievement.

Research objective	Target	Measured outcome	Extent of achievement
RO1: Microflow granularity (G1)	F1 gain ≥ 1.0 pp; early-detection gain ≥ 10%; microflow importance > 0	+1.4238 pp; +12.5%; 38.0063%	Fully achieved; margins modest (1.42× and 1.25× the targets)
RO2: Pre-surge detection (G2)	Lead time > 0 s; detection time ≥ 10% better than baseline; ≥ 80% of events early	5.2994 s; 90.0000%; 96.8815%	Fully achieved; lead time varies by corpus (1.26–29.26 s)
RO3: Common protocol (G3)	100% protocol compliance and dataset coverage; run-to-run variation ≤ 1.0 pp	100.0000%; 100.0000%; 0.0219 pp	Fully achieved
RO4: Unified temporal–spectral score (G4)	Surge-detection gain ≥ 10%; unified F1 gain ≥ 1.0 pp	12%; 1.25 pp	Achieved within scope; based on 38 onsets in one Tier A corpus
RO5: Generalisation and feasibility (G5)	Adapted F1 retention ≥ 90%; degradation ≤ 10%; unseen-attack F1 ≥ 85%; latency ≤ 1.0 s	96.2552%; 3.7448%; 99.2413%; 0.0083 s	Achieved with target adaptation; zero-shot transfer not achieved

In the authors' opinion, RO1, RO2 and RO3 have been achieved convincingly, because their margins over the targets are clear and they are supported by several corpora. RO4 has been achieved but rests on the weakest evidence, since its gains only modestly exceed the targets and are measured on a small number of sustained onsets in one corpus. RO5 has been achieved in its target-adapted form only; the zero-shot results show that a model trained in one environment should not be deployed in another without adaptation. Overall, the authors consider the central objective, demonstrating measurable pre-surge detection under a reproducible protocol, to have been met, while a claim of superior steady-state accuracy is deliberately not made.

8.3 Limitations And Threats To Validity

The conclusions are subject to the limitations and threats to validity detailed in Section 7.7. The most

important are the reliance on laboratory corpora, two of which lack timestamps (the CIC-IDS2017 spectra are reconstructed and TON_IoT has none); the small number of sustained onsets (38) behind the pre-surge results; the flow-level random split, which can place flows of one attack episode in both partitions; the dependence of lead time on the chosen reference detector and horizon; the non-significant difference from Random Forest; and the need for target adaptation before deployment in a new environment.

8.4 Future Work

- Obtain timestamped versions of the timestamp-free corpora. Timestamped CIC-IDS2017 and TON_IoT captures would replace surrogate or missing spectra with measurements and are the single most valuable improvement to this work, particularly for the application-layer hypothesis.
- Extend the passing G1.2 and G4.2 evaluations to finer observation resolutions, additional prediction horizons, and substantially more sustained onsets. This will test the stability of the 12.5% early-detection and 12% surge-detection improvements across a wider range of traffic conditions.
- Extend the fully satisfied, target-adapted G5 evaluation toward stronger zero-shot transfer. Domain adaptation or calibration using only unlabelled target traffic could reduce the need for labelled target data and simplify deployment in new environments.
- Learn the source–victim graph explicitly instead of relying on victim-centric partitioning, and test whether it adds to the spectral evidence in the pre-surge regime.
- Evaluate an evasive adversary that randomises arrival timing to remove the narrowband structure on which spectral detection relies.
- Increase the statistical power of the baseline comparison with more folds and random seeds, and report effect-size intervals rather than only a significance decision.

REFERENCES

[1] A. I. Hassan, E. Abd El Reheem, and S. K. Guirguis, “An entropy and machine learning based approach for DDoS attacks detection in software defined networks,” Scientific Reports, vol. 14, art. no. 18159, 2024. https://doi.org/10.1038/s41598-024-67984-w

[2] A. V. Songa and G. R. Karri, “An integrated SDN framework for early detection of DDoS attacks in cloud computing,” Journal of Cloud Computing, vol. 13, art. no. 64, 2024. https://doi.org/10.1186/s13677-024-00625-9

[3] A. J. Ibrahim, S. R. Répás, and N. Bektan, “Feature-optimized machine learning approaches for enhanced DDoS attack detection and mitigation,” Computers, vol. 14, no. 11, art. no. 472, 2025. https://doi.org/10.3390/computers14110472

[4] B. Wang, Y. Jiang, Y. Liao, and Z. Li, “DDoS-MSCT: A DDoS attack detection method based on multiscale convolution and transformer,” IET Information Security, vol. 2024, art. no. 1056705, 2024. https://doi.org/10.1049/2024/1056705

[5] L. Zhang, Y. Bai, T. Xue, G. Feng, and H. Zhang, “EfficientSTNet: A deep learning approach for multi-class DDoS detection,” IEEE Access, 2025. https://doi.org/10.1109/ACCESS.2025.3606644

[6] J. Wang, L. Wang, and R. Wang, “MFFLR-DDoS: An encrypted LR-DDoS attack detection method based on multi-granularity feature fusions in SDN,” Mathematical Biosciences and Engineering, vol. 21, no. 3, pp. 4187–4209, 2024. https://doi.org/10.3934/mbe.2024185

[7] T. Yuvaraja, W. G. R. S. Jeyaseelan, S. R. Ashokkumar, and M. Premkumar, “Detecting and mitigating low-rate DoS and DDoS attacks: Multimodal fusion of time-frequency analysis and deep learning model,” Teh Vjesn, vol. 31, no. 2, pp. 495–501, 2024. https://doi.org/10.17559/TV-20230613000728

[8] H. Babbar, S. Rani, and M. Driss, “Effective DDoS attack detection in software-defined vehicular networks using statistical flow analysis and machine learning,” PLOS ONE, vol. 19, no. 12, art. no. e0314695, 2024. https://doi.org/10.1371/journal.pone.0314695

[9] S. Mehmood, R. Amin, J. Mustafa, M. Hussain, F. S. Alsubaei, and M. D. Zakaria, “Distributed denial of services (DDoS) attack detection in SDN using optimiser-equipped CNN-MLP,” PLOS ONE, vol. 20, no. 1, art. no. e0312425, 2025. https://doi.org/10.1371/journal.pone.0312425

[10] S. Abiramasundari and V. Ramaswamy, "Distributed denial-of-service (DDoS) attack detection using supervised machine learning algorithms," Scientific Reports, vol. 15, art. no. 13098, 2025. https://doi.org/10.1038/s41598-024-84879-y

[11] H. Olimboyev, S. Khamrayeva, H. Olimboyev, O. Khalmuratov, K. Rakhimboev, and B. Ibragimov, "Using deep learning to detect DDoS attacks at the application layer," in 2025 IEEE 26th International Conference of Young Professionals in Electron Devices and Materials (EDM), 2025. https://doi.org/10.1109/EDM65517.2025.110967351

[12] A. Sabra, N. Rmeiti, and Z. Minchev, "Machine learning approaches to detect application layer DDoS attacks," Engineering Proceedings, vol. 150, art. no. 120; presented at TechSys 2026, Plovdiv, Bulgaria, 2026. https://doi.org/10.3390/engproc2026150120

[13] K. Wang, Y. Fu, X. Duan, and T. Liu, "Detection and mitigation of DDoS attacks based on multi-dimensional characteristics in SDN," Scientific Reports, vol. 14, art. no. 16421, 2024. https://doi.org/10.1038/s41598-024-66907-z

[14] I. Mundstock, A. Quevedo, J. C. Nobre, R. C. Lunardi, T. L. T. da Silveira, and B. L. Dalmazo, "On the impact of entropy-based features," arXiv, preprint arXiv:2607.15379, accepted for IEEE ISCC 2026, 2026. https://doi.org/10.48550/arXiv.2607.15379

[15] R. Purohit, S. Kumar, S. Sayyad, and K. Kotecha, "Time-frequency analysis and autoencoder approach for network traffic anomaly detection," MethodsX, vol. 14, art. no. 103228, 2025. https://doi.org/10.1016/j.mex.2025.103228

[16] A. C. Coutinho and L. V. d. Araújo, "Network data flow collection methods for cybersecurity: A systematic literature review," Computers, vol. 14, no. 10, art. no. 407, 2025. https://doi.org/10.3390/computers14100407

[17] K. Balaji and M. Balachandra, "Multi-layer traffic analysis framework for DDoS attacks in software-defined IoT networks," Future Internet, vol. 18, no. 3, art. no. 164, 2026. https://doi.org/10.3390/fi18030164

[18] H. Dhadhal and P. Kotak, "Leveraging datasets for effective mitigation of DDoS attacks in software-defined networking: Significance and challenges," Radioelectronic and Computer Systems, no. 2/110, 2024. https://doi.org/10.32620/reks.2024.2.11

[19] H. Wang and Y. Li, "Overview of DDoS attack detection in software-defined networks," IEEE Access, 2024. https://doi.org/10.1109/ACCESS.2024.3375395

[20] M. Akhi, C. Eising, and L. L. Dhirani, "TCN-based DDoS detection and mitigation in 5G Healthcare-IoT: A frequency monitoring and dynamic threshold approach," IEEE Access, vol. 13, 2025. https://doi.org/10.1109/ACCESS.2025.3531659

[21] C. Singh and A. K. Jain, "A comprehensive survey on DDoS attacks detection & mitigation in SDN-IoT network," e-Prime – Advances in Electrical Engineering, Electronics and Energy, vol. 8, art. no. 100543, 2024. https://doi.org/10.1016/j.prime.2024.100543