SoK: Hybrid Machine Learning and Explainable AI for HTTP/Application-Layer DDoS Detection

Application-layer HTTP Distributed Denial-of-Service (DDoS) attacks have surged by 118% year-over-year (Cloudflare, Q1 2025), exposing the limitations of signature-based and volumetric-focused detection approaches. In response, machine learning (ML), deep learning (DL), and explainable AI (XAI) techniques have emerged as promising defenses. However, the literature remains highly fragmented by lacking unified terminology, comparable evaluation settings, and systematic insights for their practical deployment. This paper presents a comprehensive Systematization of Knowledge (SoK) by reviewing 210 studies (2015-2025) on HTTP DDoS detection using ML/DL and XAI approaches. We organize the field along four dimensions: ML/DL paradigms, explainability methods, evaluation contexts, and deployment domains. Our analysis reveals critical gaps such as only 15% of XAI-enabled studies consider computational overhead, fewer than 5% evaluate adversarial robustness, and 60% rely on limited benchmark datasets. We then synthesize the literature into TRUST-AD, a conceptual, literature-derived four-tier roadmap that highlights the roles of lightweight screening, deep temporal analysis, multi-granular explainability, and adaptive mechanisms for trustworthy detection. This systematization provides actionable directions for researchers and practitioners in building trustworthy HTTP DDoS detection systems.

CCS Concepts: • **Security and privacy** → **Network Security**; **Denial-of-service attacks**; • **Computing methodologies** → **Machine learning**; *XArtificial intelligence*; • **Networks** → *Network protocols.*

Additional Key Words and Phrases: security, privacy, anomaly detection, machine learning, explainability, systematization of knowledge, Anomaly detection, DDoS detection, HTTP attacks

ACM Reference Format:

. 2026. SoK: Hybrid Machine Learning and Explainable AI for HTTP/Application-Layer DDoS Detection. 1, 1 (February 2026), 35 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Distributed Denial-of-Service (DDoS) attacks remain among the most persistent cybersecurity threats [172]. Recent threat-intelligence reports (e.g., Cloudflare Q1 2025) highlight a sharp rise in application-layer incidents, with HTTP-based DDoS attacks increasing by 118% year-over-year [51]. On May 16, 2025, The Hacker News reported a new 200+ HTTP botnet campaign targeting gaming and technology sectors [106]. Unlike volumetric network-layer attacks, these HTTP DDoS attacks are often low-rate and fully protocol-compliant, rendering traditional threshold- and signature-based detection inadequate [36][46][206]. These trends underscore the need for intelligent, adaptive, and transparent detection methods capable of operating in real time [207].

In response, the research community has turned to machine learning (ML) and deep learning (DL) based anomaly detection (AD) methods, spanning supervised, unsupervised, semi-supervised, and reinforcement learning. Despite their high detection accuracy, Ml-based models have a key barrier to operational deployment due to their "black-box" nature. Since security analysts require actionable explanations to support incident response, model debugging, and

Author's Contact Information:

Manuscript submitted to ACM

regulatory compliance, the research community has been exploring the integration of explainable artificial intelligence (XAI) methods for trust and transparency in DDoS detection [110].

Despite substantial research on the related topics, the field lacks a coherent, unified understanding of how ML/DL detection models, XAI techniques, datasets, and deployment environments interact—particularly in the context of HTTP-layer DDoS attacks. Prior surveys (see Section 6) often generalize across attack types, emphasize IoT or network-layer settings, focus narrowly on specific model families, or omit explainability and deployment considerations. As a result, there is a need for a new Systematization of Knowledge (SoK) that unifies the following disparate threads into a coherent structure:

- **Fragmented research landscape:** ML/DL and XAI approaches for HTTP DDoS are scattered across security, networking, and AI venues with inconsistent terminology and incompatible evaluation setups.
- **Missing integration patterns:** No existing survey explains how XAI methods align with specific ML/DL architectures or why certain combinations dominate.
- **Lack of deployment grounding:** Prior surveys rarely address real-time constraints, scalability, interpretability requirements, or operational feasibility.
- **Dataset and methodology inconsistencies:** Studies vary widely in data sources, feature engineering, labeling assumptions, and evaluation conditions, preventing reproducible comparison.

In response to developing such a new SoK and addressing the challenges above, we conduct a PRISMA-guided literature review and systematization of 210 studies on HTTP/application-layer DDoS detection published between 2015 and 2025. This SoK is guided by the following research questions:

RQ1: What are the current state-of-the-art approaches for anomaly detection in HTTP DDoS attacks, and how are they organized across learning paradigms, architectures, and explainability methods?

RQ2: How is explainable AI (XAI) integrated into HTTP DDoS detection systems, and what systematic patterns emerge between model architectures, explainability methods, and operational constraints?

RQ3: What research gaps hinder the development of trustworthy, real-time, and deployable HTTP DDoS detection systems, and how can these gaps be addressed through a unified design perspective?

Respectively, this SoK makes the following contributions:

- **Multi-Dimensional Systematization (answers RQ1):** We systematically analyze 210 studies and organize HTTP DDoS detection research along four interconnected dimensions: (i) ML/DL paradigms, (ii) explainability methods, (iii) evaluation contexts, and (iv) deployment and operational constraints. This systematization reveals recurring structural patterns—such as the coupling between dataset characteristics and learning paradigms, and the influence of deployment constraints on model and XAI choices—that are not captured in prior surveys.
- **Model–XAI Mapping in HTTP DDoS detection (answers RQ2):** We provide a structured analysis of how explainability methods are selected and applied across different ML/DL architectures in HTTP DDoS detection. Our analysis identifies systematic dependencies between model families, XAI techniques, computational overhead, and deployment environments, demonstrating that XAI integration follows consistent architectural and operational patterns rather than ad hoc choices.
- **Research Gaps and Recommendations (answers RQ3):** We identify critical and quantifiable gaps in current research, including limited real-world validation, insufficient reporting of XAI overhead, lack of adversarial robustness evaluation, and overreliance on a small set of benchmark datasets. For each gap, we provide concrete recommendations to support reproducible, trustworthy, and deployable HTTP DDoS detection research.

- **TRUST-AD Conceptual Roadmap:** Based on the identified research gaps and cross-dimensional dependencies, we synthesize the literature into TRUST-AD (Trustworthy Real-time Unified Security for Transparent Anomaly Detection), a four-tier conceptual roadmap for future HTTP DDoS detection systems. TRUST-AD integrates lightweight screening, deep temporal analysis, deployment-aware explainability, and adaptive mechanisms, and provides literature-grounded performance targets and validation considerations. We emphasize that TRUST-AD is a conceptual synthesis—not an implemented or empirically validated system—intended to guide future research and system design.

The remainder of this paper is organized as follows: Section 2 provides background on HTTP DDoS attacks, anomaly detection, and explainable AI. Section 3 describes our RISMA-guided methodology. Section 4 presents our systematization. Section 5 identifies research gaps and implications. Section 6 compares this SoK to prior surveys. Section 7 introduces the TRUST-AD a conceptual roadmap synthesized from literature. Section 8 concludes and outlines future work.

2 BACKGROUND

We introduce core concepts relevant to HTTP-layer DDoS attacks, anomaly detection (AD), and explainable AI (XAI).

2.1 HTTP/Application-layer DDoS Attacks

The broader DDoS landscape is often organized along three dimensions: emerging target systems (e.g., blockchain, IoT, SDN, cellular, routing), emerging protocols across the network stack (IP, DNS, TCP/QUIC, SIP, HTTP), and defense techniques (protocol enforcement, commercial services, and learning-based detection). This SoK focuses on specifically HTTP DDoS attacks, which typically fall into four families:

- **GET/POST floods:** high-volume but syntactically valid requests targeting expensive endpoints.
- **Slowloris:** many partially-open connections maintained indefinitely to exhaust server threads.
- **Slow HTTP attacks:** slow body or header transmission to tie up server resources.
- **HEAD/OPTIONS floods:** semi-lightweight requests used for rapid resource exhaustion.

This SoK concentrates on HTTP-layer DDoS attacks for two main primary reasons:

- **Prevalence and impact:** Although the application layer includes multiple protocols (e.g., DNS, SMTP, FTP), recent threat-intelligence reports (e.g., Cloudflare [51], Akamai [12], Imperva [84]) show that the majority of application-layer DDoS incidents exploit the HTTP protocol. This dominance reflects the central role of HTTP in modern web services, APIs, and microservice architectures, making HTTP-based attacks the most operationally significant Application-layer threat.
- **Detection complexity and methodological focus:** From a detection perspective, HTTP DDoS attacks are low-rate, protocol-compliant, and session-aware, closely mimicking legitimate traffic. Encrypted HTTPS further limits payload inspection, forcing reliance on behavioral, temporal, and statistical features rather than signatures. This motivates anomaly detection and explainable AI-based approaches, while focusing on HTTP enables deep behavioral analysis without protocol heterogeneity and remains extensible to other application-layer protocols. Figure 7 provides a consolidated view of HTTP DDoS attack types and their behavioral challenges; the model and explainability mappings illustrated in the lower portion of the figure are discussed in detail in Section IV.

In the following two sub-sections, we establish the technical landscape of AD and XAI methods. But it is also critical to link specific HTTP DDoS attack types to their behavioral signatures, underlying challenges (low-rate, protocol-compliant, encrypted, and zero-day traffic), and potential AD solutions. This **mapping** between HTTP attack

behaviors, detection challenges, ML-based solutions, and explainability mechanisms is synthesized in Fig. 7 as part of our systematization in Section 4.

2.2 Anomaly Detection for DDoS

Anomaly detection (AD) aims to identify anomalous patterns deviating from expected behavior using statistical modeling, clustering, classification, isolation, or deep learning (DL) [56, 195][79]. Such deviations often signify rare occurrences, cheating activity, network intrusions, or system failures. AD systems are being used across many domains to enhanced the AD accuracy, scalability, and adaptability. Fig. 9 (with supporting details in Table **??**) in Appendix summarizes the broader landscape of AD methods and their core principles across many domains.

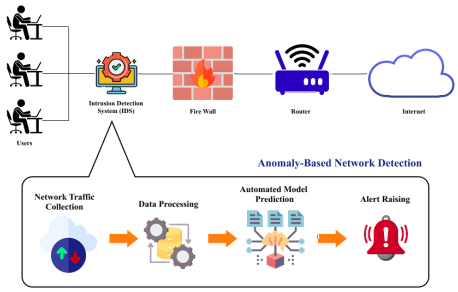

Fig. 1. Overview of network intrusion detection and workflow of anomaly-based network intrusion detection.

In cybersecurity specifically, AD is used to detect intrusions [130], malware, insider threats, and DDoS attacks by analyzing network traffic, system logs, or application-layer behavior [113]. Fig. 8 illustrates how an AD-based intrusion detection system is integrated into a typical network infrastructure. In fig. 2, we present an illustration of a high-level view of the web application ecosystem and the typical architecture of an AD-based HTTP-layer DDoS system.

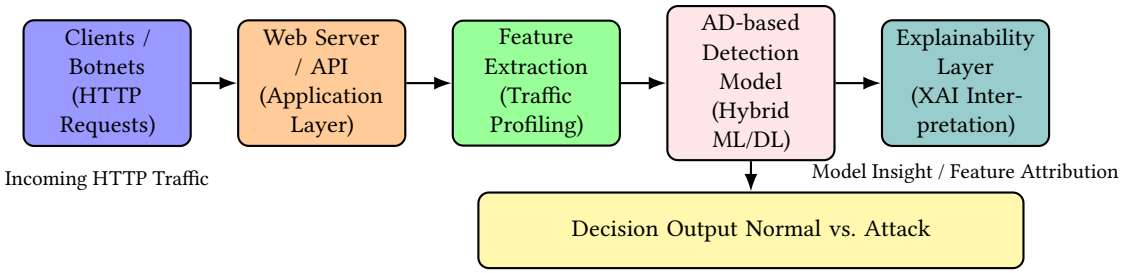

Fig. 2. Architecture of an HTTP-layer AD-based detection system. Incoming requests are processed by the web server, transformed into traffic features, evaluated by a hybrid ML/DL anomaly detector, and enriched with XAI-driven explanations.

2.3 Explainable AI

Explainable AI (XAI) provides transparency by clarifying how and why machine-learning models produce decisions, which is an widely regarded as important for trust, debugging, and regulatory compliance in cybersecurity [95, 127]. As ML-based DDoS detection becomes more complex, XAI provides the interpretability needed for analysts to understand alerts, diagnose misclassifications, and refine defenses.

Table 1. Comparison of XAI techniques.

Method	Scope	Cost	Real Time Suit.
SHAP	Global/Local	High	Moderate
LIME	Local	Medium	High
Counterfactual	Local	Med–High	Low
Permutation	Global	Low	High
Attention Viz.	Model-spec.	Medium	High
Rule/NAM	Ante-hoc	Low	Very High

Table 1 summarizes commonly used XAI approaches and compares their scope, computational cost, and suitability for real-time detection. These existing XAI techniques can broadly be categorized as:

- **Post-hoc vs. ante-hoc:** Post-hoc methods generate explanations after model training, while ante-hoc methods build interpretability into model architecture [119].
- **Global vs. local:** Global explanations describe overall model behavior, while local explanations focus on individual predictions [90][186].
- **Model-specific vs. model-agnostic:** Model-specific methods exploit internal structure of a model, while model-agnostic approaches treat the model as a black boxe [101][90, 95, 119, 153, 155, 186].

Below are the existing XAI frameworks that are most relevant to HTTP-layer DDoS detection:

- **SHAP (Shapley Additive Explanations)** uses game-theoretic feature attributions to quantify how each feature contributes to a model's prediction [90]. SHAP explanations use the additive model:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j, \tag{1}$$

  where g is the explanation model, z' is a binary vector indicating whether each simplified feature is present, M is the number of features, and ϕ_j is the Shapley value representing the contribution of feature j. SHAP provides consistent global and local explanations but at significant computational cost.
- **LIME (Local Interpretable Model-Agnostic Explanations)** generates perturbations around an instance and fits a simple surrogate model (e.g., a sparse linear model or shallow tree) that approximates the black-box model locally [186]. Formally:

$$\xi(x) = \arg\min_{g \in G} \left(L(f, g, w_x) + \Omega(g) \right), \tag{2}$$

  where:
  - x is the instance being explained,
  - f is the original black-box model,
  - $g \in G$ is an interpretable surrogate model,
  - $L(f, g, w_x)$ measures how well g locally approximates f, weighted by w_x so that samples closer to x matter more,
  - $\Omega(g)$ penalizes complexity, encouraging simple explanations [66].

  LIME excels at explaining individual local decisions and is comparatively lightweight, making it suitable for near real-time inspection.

- **Attention-based explanations** highlight the input components a deep model focuses on (e.g., header fields or temporal segments), making them particularly useful for transformer and sequence models [147].
- **Rule-extraction approaches** translate complex models into human-readable rule sets, improving interpretability without retraining the underlying model [101].

3 METHODOLOGY

This Systematization of Knowledge (SoK) follows a PRISMA-guided systematic literature review protocol to ensure transparency, reproducibility, and methodological rigor. Fig. 3 summarizes the four-stage workflow: search, screening, eligibility assessment, and final classification.

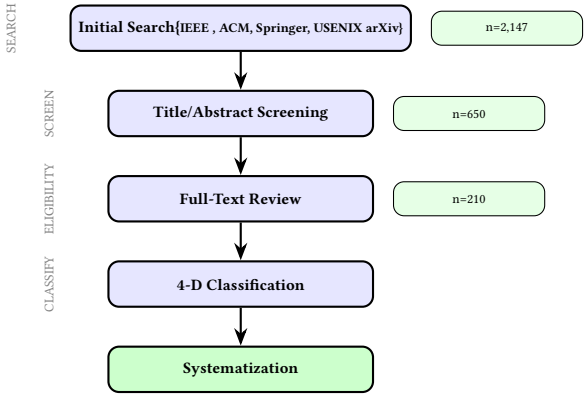

Fig. 3. Systematic literature review process following PRISMA.

3.1 Systematic Literature Review Protocol

- **Search Strategy:** We queried major digital libraries including IEEE Xplore, ACM Digital Library, SpringerLink, USENIX, Elsevier, Science Direct, Scopus, and arXiv.org.
- **Search String (boolean):** ("DDoS" OR "distributed denial of service" OR "distributed denial of service")
  AND ("anomaly detection" OR "intrusion detection" OR "threat detection"))
  AND ("machine learning" OR "deep learning" OR "neural network")
  AND ("application layer" OR "HTTP" OR "Layer 7" OR "L7" OR "web application" OR "application-level)
  AND ("explainable" OR "interpretable" OR "XAI" OR "SHAP" OR "LIME" OR "transparency" OR "interpretability" OR "transparency")
- **Time Window:** January 2015 – October 2025
- **Search Execution Date:** July 15, 2025
- **Inclusion Criteria:** Papers were included if they met all of the following:
  (1) **Scope:** proposed, implemented, or empirically evaluated ML/DL-based or anomaly detection for DDoS attacks
  (2) **Layer Focus:** addressed HTTP/Application-layer attacks OR hybrid approaches combining multiple layers with explicit HTTP component,
  (3) **Empirical Content:** reported quantitative evaluation results (accuracy, F1-score, precision, recall, ROC-AUC, or computational metrics), or XAI integration component,

(4) **Publication Type:** was published in peer-reviewed conferences/journals, or recognized preprints (arXiv with ≥ 10 citations),
(5) **Language:** was written in English,
(6) **Quality assessment:** received scoring ≥ 1.5/3 on our initial tri-criterion assessment: Clarity of methodology (0-1), Rigor of evaluation (0-1), Contribution significance (0-1), resulting in 210 high-quality studies.

- **Exclusion Criteria:** We excluded papers focusing exclusively on network-layer attacks without the application-layer component, presenting only mitigation/response mechanisms without detection components, being purely conceptual without empirical validation, being workshop abstracts, posters, or short papers (< 4 pages), having duplicate content from same authors (kept most recent/comprehensive version), lacking clear evaluation methodology or results.

3.2 Screening Process

- **Stage 1: Database Query (n=2,147)**: After searching across major databases and de-duplicating the same titles, 2,147 unique papers remained.
- **Stage 2: Title/Abstract Screening (n=650)**: After reviving titles and abstracts based on relevance to HTTP DDoS detection and ML/XAI method, 210 papers left for the full-text review in then next stage.
- **Stage 3: Full-Text Assessment and Classification (n = 210):** We applied a full-text review for methodology rigor, contribution clarity, and alignment with our research questions. Studies assessed during this stage include [2–22, 24–39, 41–78, 80–87, 89–130, 132–138, 140–153, 155–213] We identified the four dimensions for our systematization framework presented in the next section:
  - **Learning paradigm:** Supervised, semi-supervised, unsupervised, reinforcement learning, Tree-based (RF, XGBoost), deep learning (CNN, LSTM, Transformer), hybrid (CNN-LSTM, RF-AED), autoencoder, graph neural networks, etc.
  - **Explainability method:** SHAP, LIME, counterfactuals, permutation, attention, rule-based, or none
  - **Evaluation attributes: Datasets** used (CIC-IDS2017, CIC-DDoS2019, NSL-KDD, UNSW-NB15, custom, etc.); **Metrics** reported (accuracy, precision, recall, F1, ROC-AUC, inference time, memory). **Feature set:** Flow-based, packet-based, hybrid
  - **Deployment attributes:** Target context (cloud, edge/IoT, enterprise, SDN or cross domain); Real-time feasibility (yes/no/unclear); Adversarial robustness evaluated (yes/no).

3.3 Synthesis and Systematization

We employed four complementary synthesis methods:

- Thematic coding to identify conceptual patterns across models, XAI methods, and datasets.
- Cross-tabulation (e.g., XAI method × model architecture) to reveal systematic relationships.
- Gap analysis using coverage matrices to highlight underexplored areas.
- Temporal trend analysis comparing early (2015–2020) vs. recent (2021–2025) developments.

4 SYSTEMATIZATION OF THE FIELD

Based on our systematic review (Section 3), we now present a unified systematization of HTTP/application-layer DDoS anomaly detection. Our systematization organizes prior work along four interconnected dimensions: ML/DL paradigms, Explainability (XAI) methods, Evaluation contexts, and Deployment and Operation constraints.

Fig. 4 illustrates this systemization as a multi-dimensional conceptual framework by highlighting not only the four dimensions shaping design choices but also the systematic dependencies (dashed arrows) across these dimensions, such as how computational constraints influence XAI choices or how dataset imbalance affects model paradigms.

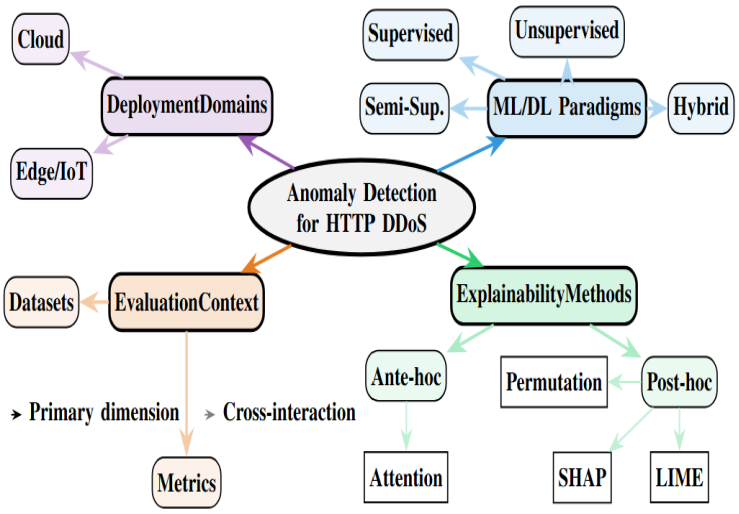

Fig. 4. Systematization framework organizing AD/ML/XAI-based HTTP DDoS research along 4 interconnected dimensions.

The remainder of this section provides a structured view of current state-of-the-art approaches along the four dimensions, and then synthesizes cross-dimensional patterns through empirical model–XAI mappings to demonstrate how XAI methods systematically aligns with different ML models.

4.1 Dimension 1: ML/DL Paradigms

To contextualize the learning paradigms used in HTTP DDoS detection within the broader DDoS landscape, we begin with a high-level taxonomy of DDoS detection approaches (Fig. 5).

This taxonomy organizes prior work by attack characteristics, feature types, and common detection models studied across the literature.

Because this SoK focuses specifically on HTTP-based DDoS detection using ML/AI-driven anomaly detection (AD) systems, we now analyze the major learning paradigms employed in such systems.

Supervised Learning remains the most prevalent paradigm when labeled datasets are available (about 45% of reviewed studies). Common models include tree-based classifiers (RF, XGBoost), instance-based methods (KNN), linear models (SVM), and deep learning architectures (CNN, LSTM, and CNN-LSTM hybrids). These models often report high accuracy (often F1 > 0.95) but require large labeled training sets and often struggle to generalize to unseen or evolving attack variants [2, 30, 40, 67].

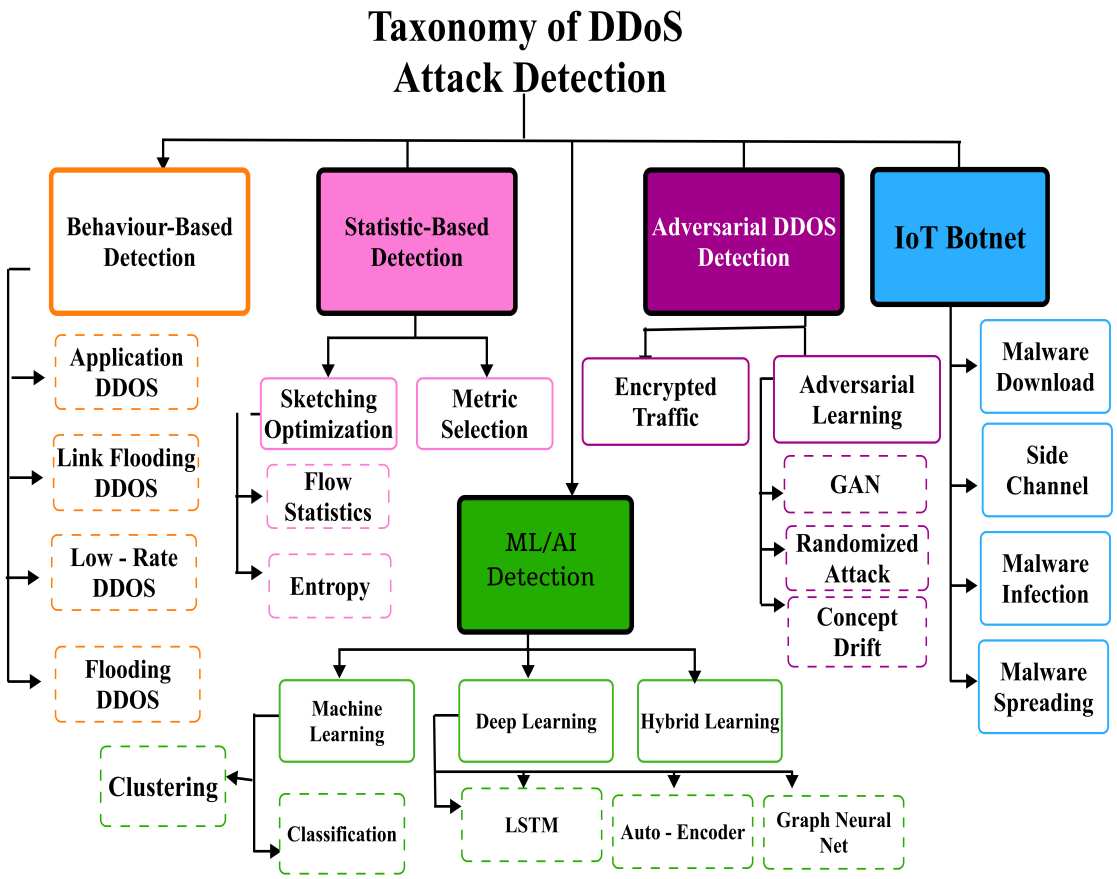

Fig. 5. Taxonomy of DDoS Attack Detection

Unsupervised Learning approaches such as autoencoders [159], clustering algorithms (K-means, DBSCAN), and isolation forests are used when labeled traffic is scarce or non-existent. They are particularly effective for detecting zero-day or mimicry-based attacks but frequently exhibit higher false positive rates. Reported performance varies, with F1 scores typically ranging from 0.85 to 0.92 [184].

Semi-Supervised Learning leverages small amounts of labeled traffic alongside large unlabeled datasets. Techniques include semi-supervised GANs, self-training, and co-training [1, 65, 72, 74, 82, 123]. These approaches reduce labeling costs and often achieve strong performance (F1 $\approx$ 0.90–0.95), making them particularly suitable for HTTP traffic where labeled anomalies are rare.

Reinforcement Learning (RL) treats detection as a sequential decision-making problem, enabling adaptive responses to evolving attack behaviors. Examples include Q-learning, policy gradient, and actor-critic models for dynamic defense [47]. Although RL offers promising adaptability, it remains relatively less common due to high training complexity and the need for stable reward signals in security contexts.

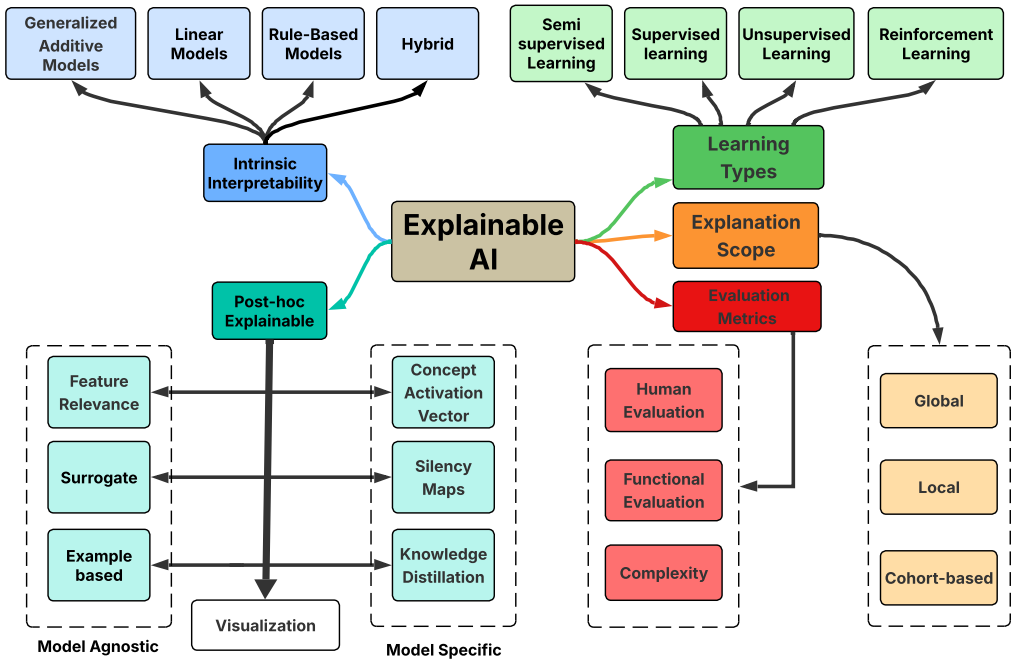

Fig. 6. XAI Taxonomy for HTTP DDoS Attack Detection.

Hybrid Approaches combine multiple paradigms to leverage complementary strengths. Examples include RF+Autoencoder [159], combining supervised classification with unsupervised anomaly detection, and CNN-LSTM models enhanced with attention mechanisms [30]. Hybrid models frequently report higher performance (F1 > 0.97) [40, 159] by exploiting complementary strengths to balance generalization, robustness, and detection precision.

Our systematization reveals a strong structural pattern: unsupervised and semi-supervised methods dominate when datasets exhibit heavy class imbalance or insufficient labeling, whereas supervised methods appear mainly in studies relying on CIC-based datasets with artificially balanced labels. This structural pattern rooted in dataset characteristics rather than algorithmic preference is one of the several gaps across prior surveys. We will further analyze such gaps and contrast this SoK with prior surveys in Section 6.

4.2 Dimension 2: Explainability Methods

Explainability (XAI) is increasingly treated as an essential component for operational ML-based DDoS detection. Our systematization reveals structured patterns in how explainability is adopted, the techniques preferred, and the gaps that remain. Fig. 6 show detailed taxonomy of XAI for HTTP DDoS Attack Detection.

High-level patterns. Among XAI-enabled studies, the majority (≈65%) apply **post-hoc, model-agnostic** methods, reflecting a tendency to attach interpretability after model training rather than integrating it into the model's design. SHAP and LIME are the most frequently used post-hoc tools because they can be applied to a wide range of model families (tree ensembles, neural networks, etc.) [19, 19, 24, 90, 95, 108, 119, 153, 155, 186]. A useful nuance is that *TreeSHAP* (an optimized SHAP variant) computes exact Shapley values efficiently for tree ensembles, which explains

SHAP's prevalence with RF/XGBoost models; generic Kernel-SHAP remains computationally expensive for large models or high-dimensional inputs [24, 153].

In contrast, only a small set of XAI-enabled studies (≈25%) pursue **ante-hoc** or built-in interpretability, for example by (i) using attention mechanisms in sequence models, (ii) employing interpretable model classes (rule sets, GAMs, NAMs), or (iii) extracting rules from trained networks. These approaches aim to produce explanations as part of the model's internal computation rather than as a separate post-processing step.

Substantial portion of DDoS detection studies, unfortunately, have **no explainability** at all. This creates an important operational gap for real-world defense environments that require transparency for auditing, incident response, and operational trust in SOC environments.

Technical caveats and trade-offs. In the current literature, XAI is largely treated as a post-training add-on rather than a foundational design principle. This trend is driven by the field's heavy reliance on black-box models, the convenience of generic post-hoc tools, and limited evaluation standards for XAI in security settings. Despite their convenience, post-hoc methods incur overhead and come with various caveats:

- **Computational cost:** SHAP (kernel SHAP) can be slow for high-dimensional inputs and complex models; TreeSHAP is efficient for tree ensembles but not applicable to all architectures. LIME requires fitting surrogate models per instance, which can add per-instance latency.
- **Local vs. global fidelity:** LIME and many local methods provide explanations that are valid only in a local neighborhood; global interpretability of the original model is not guaranteed.
- **Attention explanation:** Attention weights are frequently used as an ante-hoc interpretability signal in transformers/LSTMs, but the literature cautions that attention is not always a faithful explanation of model behavior and should be validated against alternative attribution methods[88].
- **Operational cost of XAI:** Few works measure XAI latency, memory, or throughput impact metrics that are essential for real-time deployments.

Benefits reported by XAI-enabled studies. A growing number of studies integrate XAI into DDoS detection frameworks [9, 19, 31, 187] because XAI brings several benefits: **Feature importance analysis** identifies critical traffic attributes contributing to classification decisions. **Model transparency** clarifies why traffic is flagged as malicious, facilitates verification of automated decisions by security analysts. **Adaptive defense mechanisms** enable security teams to fine-tune models with deeper understanding of decision-making. **Bias detection** helps avoid flagging legitimate traffic as malicious, reducing false positives and boosting trust in automated defense systems.

Representative SOTA. Table 2 summarizes the state-of-the-art (SOTA) studies on DDoS detection with some integrating XAI. The table also notes strengths, limitations, and gaps that future work should address.

4.3 Dimension 3: Evaluation Contexts

Evaluation practices across HTTP DDoS anomaly detection studies show substantial inconsistency. Two elements dominate the evaluation landscape: (i) the datasets used for experimentation, and (ii) the metrics used to assess both predictive performance and computational feasibility.

Datasets. Our review reveals a strong dependence on a small set of publicly available benchmark datasets. Specifically:

Table 2. Current SOTA AD Systems for DDoS and XAI Approaches

Author and Year	Problems Addressed	Proposed Solution	Results Obtained	Advantages	Disadvantages	Research Gaps
Hekmati et al., 2024[78]	Stealthy DDoS attacks in IoT networks often evade detection due to low-volume distributed traffic and lack of cross-node correlation analysis	Correlation-aware neural network architectures (MLP, CNN, LSTM, Transformer, Autoencoder) using inter-node traffic data and selective feature sharing via SHAP and correlation metrics	MM-WC on LSTM achieved 81% F1 score, significantly outperforming non-correlation models (35%) on a 4,060-node urban IoT dataset	Improved detection of stealthy attacks, scalable architecture, efficient data sharing, strong real-world evaluation	Increased system complexity, dependence on accurate correlation metrics, possible false positives	Need for real-time adaptation, integration with existing IoT security frameworks, and testing on more diverse datasets
Akgun et al., 2022[13]	Difficulty in accurately detecting DDoS attacks using traditional methods.	Developed a DL-based intrusion detection system for DDoS attacks.	Achieved high accuracy and low false-positive rates.	Enhanced detection capabilities using DL.	Potential overfitting on specific datasets.	Need for evaluation on diverse real-world datasets.
Arreche et al., 2024[23]	Lack of interpretability in AI-based intrusion detection systems.	Proposed XAI-IDS, an explainable AI framework enhancing network IDS.	Improved model transparency and decision understanding.	Facilitates trust in AI decisions through explanations.	May introduce computational overhead.	Integration with real-time systems.
Okporokpo et al., 2025[145]	Vulnerabilities in Smart Home IoT networks to DDoS attacks.	Proposed a hybrid trust-based DDoS detection technique.	Enhanced detection and mitigation in SH-IoT networks.	Adaptive mechanisms and trust evaluation models.	May require significant computational resources.	Scalability to larger IoT ecosystems.
Roshan & Zafar, 2022[156]	Optimization of network AD models.	Employed Kernel SHAP for explainability and model optimization.	Improved accuracy, recall, precision, and F-score.	Provides insights into feature importance.	May not generalize across all network types.	Validation on diverse network conditions.
Gaspar & Silva, 2024 [66]	Lack of transparency in IoT IDS using black-box ML models	LIME + SHAP integrated with MLP for interpretable detection	Detailed comparison of explanation effectiveness	Provides deep interpretability insights	Limited to a single model architecture; computational overhead	Need for testing on deeper DL architectures and real IoT traffic
Alghazzawi, 2021[18]	Efficient detection of DDoS attacks with feature selection.	Developed a hybrid DL model combining CNN and BiLSTM.	Achieved high detection accuracy with improved feature selection.	Captures spatial and temporal features effectively.	Potential overfitting on specific datasets.	Testing on real-time traffic data.
Bamber et al., 2025[30]	Intelligent cyber intrusion detection using DL.	Proposed a hybrid CNN-LSTM approach for IDS.	Demonstrated improved intrusion detection performance.	Leverages strengths of both CNN and LSTM models.	Computationally intensive training process.	Real-time deployment considerations.
Wei et al., 2023[195]	Need for explainability in DDoS attack detection.	Combined ML with SHAP for classification and explanation.	Achieved over 99% accuracy with interpretable results.	Enhances trust through model transparency.	Potential computational overhead.	Application to other types of cyber attacks.
Ahmed Bensaoud and Jugal Kalita, 2025[33]	Optimized detection of IoT cyber-attacks	SOM + DBN + Autoencoder optimized via PSO	High performance on test datasets	Effective hybrid optimization	May require hyperparameter tuning	Real-time and resource-limited deployment
Gniewkowski et al., 2022[70]	Detection of varied DDoS attack types	Hybrid CNN-LSTM AD model	Demonstrated high accuracy	Optimized for multiple attack types	Lacks real-time deployment focus	Explicit real-time implementation missing
Alzu et al., 2024[19]	Classification and explanation of DDoS detection	Multilayer Perceptron with SHAP explanations	SHAP enhances interpretability, good classification results	Global and local explanation support	SHAP is computationally expensive	Real-time system efficiency
Abdulatif, 2024[14]	Real-time IoT DoS/DDoS detection	ML-based AD with adaptive thresholds	Real-time detection achieved	Behavior-based adaptation	Lacks comparison to existing benchmarks	Evaluation on different IoT scenarios
Doriguzzi-Corin & Siracusa, 2025[58]	Distributed DDoS detection with privacy constraints	FLAD: Federated Learning for DDoS AD	Effective distributed detection	Dynamic model updates, privacy-preserving	Requires sufficient network bandwidth	Scalability in low-resource environments
Wang, 2022[188]	Real-time malicious traffic detection	AI-powered IDS with adaptive learning	Real-time threat prediction achieved	Continuous retraining adapts to changes	Model drift risk	Maintaining model performance over time

- **CIC-IDS2017** and **CIC-DDoS2019** appear in more than 60% of studies, making them the de facto evaluation standards for DDoS detection.
- **NSL-KDD** remains surprisingly common (≈35%), despite known limitations such as outdated traffic profiles and simplified feature engineering.
- **UNSW-NB15** and **CICIDS2018** are used less frequently (≈15% each), typically as secondary validation datasets.
- **Real-world production traffic** is extremely rare (<5%), limiting external validity and raising concerns of dataset overfitting.

This overdependence on synthetic and partially balanced datasets inflates reported performance and obscures generalization challenges, particularly problematic for HTTP DDoS, where real-world attack characteristics evolve rapidly.

Metrics. Nearly all studies report standard classification metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. However, metrics needed to evaluate realistic operational feasibility remain largely absent:

- Only **30%** report inference latency or computational cost.
- **15%** provide memory footprint or model-size constraints.
- About **10%** quantify additional overhead introduced by XAI methods.
- Fewer than **5%** evaluate *adversarial robustness* or sensitivity to evasive behaviors.

This imbalance heavy emphasis on academic predictive metrics but limited reporting of practical performance indicators creates a gap between published results and deployable systems.

Evaluation Gaps and Reporting Requirements. A recurring issue across the surveyed literature is the absence of key operational measurements that affect real-world viability. To enable meaningful comparison and reproducible evaluation, studies should report, at minimum:

- **per-instance explanation latency (ms)**,
- **memory footprint** introduced by XAI computation,
- whether explanations are computed **synchronously** (inline) or **asynchronously** (offline),
- **fidelity and stability** of explanations (e.g., consistency under small perturbations),
- the **impact of XAI on detection throughput** and latency.

These metrics are essential for determining whether an ML/XAI-enabled detection system can meet the demands of real-time HTTP DDoS defense. Their omission complicates comparison across studies and obscures the practical implications of proposed methods.

4.4 Dimension 4: Deployment and Operation Constraints

Deployment environments play a decisive role in shaping the design of HTTP DDoS detection systems. Different domains impose distinct computational, latency, and integration constraints, which in turn influence both model selection and the choice of XAI techniques.

Cloud/Data Center (≈50%). Cloud-based deployments assume abundant compute, memory, and storage. As a result, deep learning architectures (CNNs, LSTMs, Transformers, and hybrid models) are common. These studies typically prioritize throughput and detection accuracy over resource efficiency. Post-hoc explainability methods such as SHAP or integrated gradients are feasible here due to relaxed latency constraints and batch-processing pipelines.

Edge/IoT (≈25%). Edge environments exhibit strict resource limitations: low CPU frequency, limited RAM, intermittent connectivity, and battery constraints, making detection latency and energy consumption the key design factors. Accordingly, the lightweight models (e.g., tree-based classifiers, shallow neural networks, statistical anomaly detectors) are commonly used. Explainability choices also reflect these constraints and select methods with low computational overhead (permutation importance, lightweight rule extraction, simplified attention mechanisms) over expensive methods such as SHAP[120].

Enterprise Networks (≈15%). Enterprise deployments balance accuracy and cost by often employing tiered architectures that use a lightweight screening followed by deep analysis. Tiered detection architectures include a fast front-end filter (e.g., statistical detector or tree-based model) to screen flows before forwarding suspicious traffic to a deeper ML/DL model. Likewise, explainability methods vary depending on integration requirements with SIEM/SOC tools. Localized explanations (e.g., LIME or attention heatmaps) are valued for analyst workflows.

Software-Defined Networking Environments (≈10%). SDN–based solutions leverage centralized controllers and programmable switches to coordinate detection and mitigation. Because SDN control-plane operations must meet tight latency budgets, many SDN studies restrict analysis to flow-level features and employ lightweight or limited explainability. XAI is often applied offline rather than inline to avoid interfering with forwarding-path performance.

Operational considerations. Regardless of deployment setting, practitioners emphasize priorities that remain underrepresented in academic work [24, 119, 130]:

- strict false-positive control to reduce alert fatigue,
- real-time detection and explanation latency,
- interoperability with SIEM, WAF, and SDN,
- robustness against adversarial and mimicry-based attacks,
- adaptation to concept drift in evolving HTTP traffic,
- transparency and auditability for regulatory compliance.

These priorities highlight that model selection cannot be evaluated independently of deployment context; operational constraints must guide both architecture and explainability design. Moreover, systems protecting critical infrastructure (e.g., financial services, healthcare) face stricter requirements for model transparency and regulatory compliance compared to general web services.

4.5 Model–XAI Mapping in HTTP DDoS detection

To understand how explainability methods align with different ML/DL architectures, we analyzed cross-paper patterns across 210 studies. Our findings reveal four recurring structural relationships:

Post-hoc dominance for tree-based models. SHAP (TreeSHAP) and LIME are used in roughly 80% of XAI-enabled studies involving Random Forests, XGBoost, and other ensemble methods. This stems from their model-agnostic nature and ability to handle complex feature interactions. Specifically, TreeSHAP became the dominant choice for feature attribution because of its polynomial-time computation for tree ensembles. This aligns with deployment trends in cloud or enterprise settings, where structured-feature attribution is efficient and scalable.

Attention as a dual-purpose mechanism. In Transformer- and LSTM-based detectors (≈30% of deep learning studies), attention mechanisms serve a dual role: (i) improving temporal feature modeling, and (ii) providing ante-hoc interpretability through attention weight visualization. However, attention weights must be interpreted with caution; they provide useful but sometimes not a faithful explanations. Post-hoc attributions (e.g., SHAP, integrated gradients) are used when finer-grained explanations are required, reflecting the practical pattern observed in sequence models.

Explainability–computational tradeoffs. Deployment constraints strongly influence XAI method choice:

- Edge and IoT systems favor lightweight techniques such as permutation importance or rule extraction to maintain low-latency operation.
- Cloud and enterprise environments can accommodate heavier post-hoc methods such as SHAP and integrated gradients.

These tradeoffs appear in nearly every deployment-aware study (≈90%) and reinforce that XAI choices are tightly coupled with computational and latency constraints.

Correlation-aware hybrid systems. A recent cluster of works (≈15%, 2022–2025) integrates explicit correlation analysis, combining SHAP attributions with domain-specific relational features (e.g., inter-node relationships in IoT). These systems aim to improve explainability and reduce false positives by incorporating structural traffic relationships.

While the previous sections analyze ML paradigms and XAI methods independently, operational HTTP DDoS detection requires aligning attack behaviors with detection challenges, learning models, and explainability mechanisms. To capture this end-to-end dependency, Fig. 7 synthesizes the recurring patterns observed across the literature into a pipeline integrating representative HTTP-layer attack types to their associated detection challenges, commonly adopted ML-based solutions, and the explainability mechanisms. In contrast to the individual studies addressing specific portions of this pipeline, the integrated view highlights the need for end-to-end alignment across attack modeling, detection, explanation, and operational constraints. Moreover, this reinforces the need to evaluate ML-based HTTP DDoS detectors as holistic model–XAI–deployment units rather than isolated components, and motivates the research gaps identified in the next section.

5 RESEARCH GAPS AND IMPLICATIONS

Our systematization reveals that, despite rapid progress in HTTP DDoS detection, existing work exhibits fundamental gaps that limit real-world deployability and long-term effectiveness. These gaps cluster around three interrelated dimensions: scalability and architectural efficiency, robustness and generalization, and operational validity.

5.1 Scalability and Architectural Efficiency

Explainability Overhead at Scale: While explainable AI (XAI) techniques are increasingly adopted, current methods such as SHAP and LIME introduce substantial latency (20-100ms), making them impractical for high-throughput or edge deployment [120][153]. The literature lacks lightweight, deployment-aware interpretability mechanisms that can operate under strict latency budgets, highlighting a need for explainability approaches explicitly designed for real-time environments.

Lack of System-Level Architectural Evaluation: Most studies evaluate detection models in isolation, without considering how multiple models might be composed within a single detection pipeline. As a result, there is limited understanding of how to balance screening, deep analysis, and explanation under resource constraints. Future work

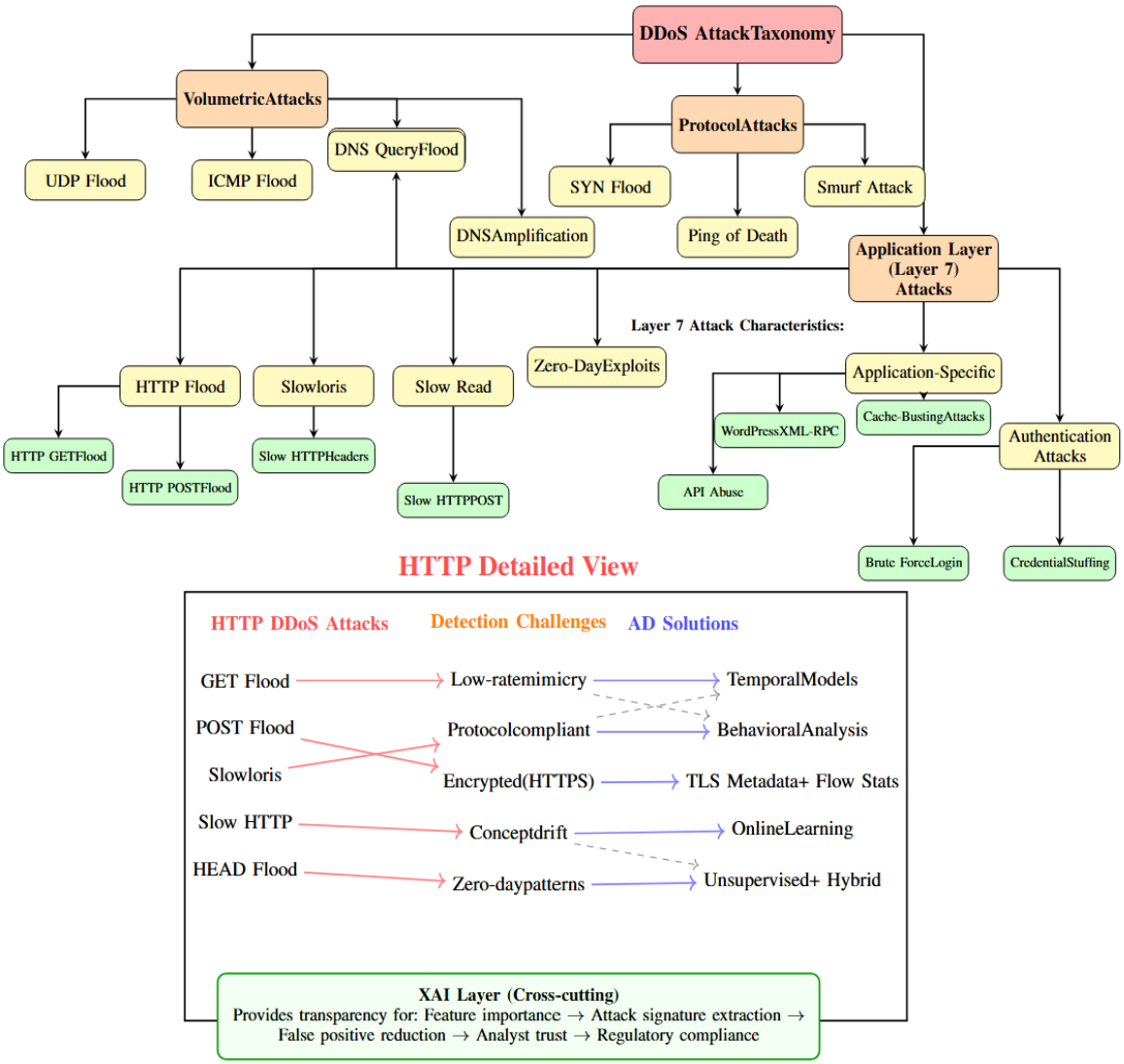

Fig. 7. Taxonomy of HTTP DDoS attacks, their mappings to detection challenges, and AD solutions with an XAI layer.

must move beyond single-model evaluations toward system-level designs that explicitly account for routing decisions, computational budgets, and end-to-end latency.

5.2 Robustness and Generalization

Limited Adversarial Evaluation: Fewer than 5% of reviewed studies assess robustness against adversarial behaviors such as evasion or poisoning attacks. This omission is particularly concerning given the adaptive nature of DDoS

adversaries. The field lacks standardized adversarial threat models and evaluation protocols tailored to HTTP DDoS detection.

Dataset Dependence and Generalization Risk: Approximately 60% of existing studies rely on variants of CIC-IDS2017/2019, raising concerns about dataset-specific bias and overfitting[154]. Cross-dataset evaluation and domain adaptation techniques remain underexplored, limiting confidence in model performance across heterogeneous network environments [131, 167].

5.3 Operational Validity and Human Utility

Absence of Real-World Validation: Only a small fraction ($\approx$ 5%) of studies report validation using production or longitudinal traffic traces. Consequently, little is known about how detection performance degrades over time, how models respond to traffic evolution, or whether reported gains translate to operational settings.

Underdeveloped Human-Centered Evaluation: Explainability is typically evaluated using proxy metrics (e.g., feature attribution consistency) rather than through studies involving security analysts. The literature provides limited evidence that current XAI techniques improve analyst decision-making, response time, or trust.

5.4 Evidence-Based Gaps in Current Literature

While the above challenges describe broad systemic limitations, we now ground them in concrete, recurring gaps observed across the literature. These evidence-based gaps provide a more granular view of where current detection pipelines fall short and collectively motivates the architectural and design choices introduced later in the next section.

Gap 1: Real-time detection constraints. Although approximately 75% of reviewed studies report high detection accuracy, only about 30% provide inference-time measurements, and fewer than 10% demonstrate feasibility under real-time, high-throughput conditions. Tree-based models are consistently reported to achieve low-latency inference, motivating their use for initial traffic screening.

Gap 2: Temporal pattern modeling. Roughly 60% of HTTP DDoS attacks exhibit temporal characteristics such as burstiness and connection timing. Studies employing CNN-LSTM and related temporal architectures consistently report F1 scores exceeding 0.95, motivating the consideration of a dedicated deep analysis stage.

Gap 3: Explainability integration. Only about 40% of studies incorporate any form of explainability, and nearly 90% of those rely primarily on computationally expensive post-hoc methods without accounting for runtime overhead. Reported SHAP latencies range from 20–100ms per instance, motivating selective rather than universal explainability.

Gap 4: Concept drift adaptation. Fewer than 5% of reviewed studies address concept drift or long-term adaptation. Existing work on online learning and drift detection remains largely disconnected from complete detection pipelines, motivating an explicit adaptability component.

5.5 Implications for Future Research

These gaps suggest several concrete shifts in how HTTP DDoS detection research should be conducted and evaluated.

System-Level Evaluation: Future work should move beyond single-model benchmarks toward end-to-end pipelines that explicitly report latency, memory footprint, and routing behavior under realistic traffic loads.

Robustness and Adaptation: Adversarial evaluation and concept drift handling should be treated as first-class requirements rather than optional extensions, with standardized threat models and longitudinal evaluations.

Reproducibility and Transparency: To enable meaningful comparison and deployment readiness, researchers should provide reproducibility artifacts, including code, data preprocessing pipelines, and deployment configurations.

5.6 Implications for Practitioners

Beyond research considerations, the identified gaps have direct consequences for practitioners deploying HTTP DDoS detection systems in operational environments.

Model Selection Guidance: Based on our systematization:

- **For high accuracy:** Hybrid approaches (ensemble + deep learning) consistently achieve F1 > 0.97
- **For low latency:** Tree-based methods (RF, XGBoost) provide less than 2s inference with F1 > 0.95
- **For interpretability:** Models with SHAP/LIME integration offer good transparency performance balance, but evidence suggests that human-centered XAI designs may achieve superior analyst utility at lower operational cost.
- **For edge deployment:** Lightweight RF with 15-20 features optimizes resource constraints

SOC Integration Requirements: Successful deployment requires: Explanation interfaces integrated with SIEM systems, Configurable confidence thresholds for different threat levels, Audit trails linking detections to model explanations, Analyst feedback loops for continuous model improvement.

Deployment Risk Mitigation: Address critical operational concerns: Plan for concept drift through monitoring and retraining schedules, Implement fallback mechanisms when models encounter distribution shifts, Test adversarial robustness before production deployment, Validate performance on representative production traffic before full deployment

5.7 Implications for Publication Venues

Finally, several of the identified gaps stem not only from technical challenges but also from prevailing publication and review practices in the field.

Artifact Evaluation: Security venues should:

- Require reproducibility artifacts (code, data, models) for ML security papers
- Establish artifact evaluation committees for verification
- Recognize reproducible research through badges and awards

Standardized Reporting: Establish reporting requirements:

- Mandatory computational cost reporting (training time, inference latency, memory)
- XAI overhead quantification for interpretable systems
- Cross-dataset evaluation to demonstrate generalization
- Adversarial robustness testing for security-critical systems

Review Criteria Updates: Update review guidelines to emphasize: Practical deployability alongside theoretical contributions, Real-world validation beyond benchmark dataset performance, Comprehensive threat model analysis including adversarial scenarios, Explainability evaluation including computational overhead.

Taken together, these gaps highlight a fundamental disconnect between model-centric advances and system-level requirements for HTTP-layer DDoS detection. While individual studies report strong performance along isolated dimensions, the literature lacks a coherent perspective that jointly considers learning paradigms, explainability integration, evaluation practices, and deployment constraints. Importantly, this fragmentation is not limited to primary studies but also manifests at the level of existing survey literature. To contextualize these findings and position our contribution within the broader survey landscape, we next compare our systematization with prior surveys and SoK efforts.

6 COMPARISON WITH EXISTING SURVEYS

Having identified evidence-based research gaps in the preceding section, we now examine how existing survey and SoK papers address these dimensions. The goal of this section is not to re-identify gaps, but to contextualize them at the level of survey literature and to clarify how our SoK complements and extends prior surveys by focusing explicitly on HTTP-layer DDoS detection and by jointly analyzing learning paradigms, explainability mechanisms, evaluation practices, and deployment environments within a single systematization.

In general, the fragmentation observed in individual detection studies persists even across comprehensive surveys such as [128], [101], [203], [87],[183],[150], [93], [185] and [55]. Existing surveys offer valuable but largely orthogonal perspectives on DDoS detection, explainability, and deployment. In addition to revealing such methodological distinctions, our review identifies several gaps across earlier survey papers. Specifically, prior surveys typically (i) lack a unified taxonomy linking HTTP attack behaviors to AD/XAI techniques, (ii) overlook deployment-driven design constraints, (iii) treat XAI as an isolated topic without mapping it to model families, and (iv) provide limited or no cross-dataset analysis. These omissions leave an incomplete picture of how ML, datasets, XAI, and deployment environments interact in practice, motivating the need for a unified, system-level systematization focused on HTTP-layer DDoS detection, as targeted in this SoK.

Several closely related surveys illustrate the diversity of prior perspectives. **Mittal et al.** [128] provide a comprehensive taxonomy of deep learning approaches for DDoS detection, but focus exclusively on DL models without examining hybrid ML/DL designs or explainability integration. **Kohli et al**. [101] offer a broad survey of anomaly detection techniques across domains, but do not specialize in application-layer DDoS attacks or HTTP-specific traffic behaviors. **Zamanzadeh et al**. [203] review deep learning methods for time-series anomaly detection with limited discussion of explainability and deployment considerations, and without DDoS-specific focus. **Najafimehr et al.** [134] focus on ML taxonomies for DDoS detection, but do not analyze explainability integration or operational deployment contexts. **Neupane et al**. [139] survey explainable intrusion detection systems, but do not systematically relate XAI methods to model families or evaluate their suitability for real-time DDoS detection.

To ensure consistency with the research gaps previously identified and to facilitate a structured comparison, we compare representative surveys in Table 3, where the final column ("How Our SoK Advances") highlights how our work extends or complements each survey, along the following dimensions showing what each survey covers:

- **Attack-layer focus**, distinguishing network-layer from application-layer (HTTP) DDoS attacks;
- **Learning paradigms**, including classical machine learning, deep learning, and hybrid approaches;
- **Explainable AI (XAI)**, indicating whether interpretability is explicitly analyzed and how it is integrated;
- **Deployment considerations**, such as edge, enterprise, cloud, or high-throughput environments;
- **Systematic methodology**, distinguishing narrative surveys from structured or PRISMA-style systematizations.

Several consistent patterns emerge from this comparison. Most existing surveys provide deep coverage along one or two dimensions—such as deep learning architectures, explainability techniques, IoT environments, or SDN-based defenses—but relatively few jointly analyze learning paradigms, XAI integration, datasets, and deployment contexts. Moreover, surveys that address application-layer attacks often do so without systematically linking model choices to explainability or operational constraints.

These observations reinforce the findings of the Research Gaps in Section 5 by showing that the lack of cross-dimensional integration persists even at the survey level. Motivated by these observations and the research gaps,

Table 3. Comparison with previous surveys

Papers	Scope	Year	Taxonomy	HTTP Focus	XAI	Deployment	Systematic	Main Limitation	How Our SoK Advances
Mittal et al. [128]	Deep learning for DDoS	2022	✓					DL only; no hybrid models or XAI emphasis	Hybrid ML/DL models + XAI taxonomy
Kohli et al. [101]	General anomaly detection	2025	✓					General AD; not specialized in DDoS or application-layer	HTTP DDoS-specific focus + dataset analysis
Zamanzadeh et al. [203]	Time-series anomaly detection	2023	✓		✓			Time-series focus; no real-time DDoS deployment	Real-time deployment considerations
Jadhav et al. [87]	Edge computing for IDS	2023	✓			✓		Edge computing focus; lacks application-layer DDoS	Application-layer HTTP DDoS specificity
Thudumu et al. [183]	High-dimensional anomaly detection	2020	✓					General high-dim AD; outdated, no DL/DDoS specifics	2015–2025 coverage; DL + Hybrid models
Ping et al. [150]	AI in cybersecurity	2024						Broad AI overview; lacks depth on application-layer	Systematic taxonomy + dataset mapping
Najafimehr et al. [134]	ML for DDoS detection	2023	✓					ML taxonomy only; no XAI integration or deployment	XAI integration + deployment framework
Neupane et al. [139]	XAI methods in security	2022			✓			XAI methods only; no technique comparison or deployment	DDoS-specific XAI + model comparison
Singh & Behal [173]	SDN-based DDoS detection	2024	✓					SDN plane-wise analysis; lacks HTTP DDoS specificity	HTTP application-layer focus
Arreche et al. [23]	XAI benchmarking	2024			✓			XAI benchmarks only; no model comparison or deployment	Model comparison + deployment guidelines
Kohli et al.[101]	Comprehensive AD survey	2025	✓	✓	✓	✓		General AD focus; not specialized in DDoS or application-layer	DDoS-specialized taxonomy + XAI
Zamanzadeh et al.[203]	Time-series + XAI	2023	✓	✓	✓	✓	✓	Time-series focused; no coverage of real-time DDoS deployment	Real-time HTTP DDoS deployment
Jadhav et al.[87]	Edge ML/DL systems	2023	✓	✓	✓			Limited to edge computing; lacks application-layer DDoS focus	Broader deployment + HTTP focus
Thudumu et al.[183]	AD with XAI	2020	✓	✓	✓			General high-dim AD; outdated, no DL/DDoS specifics	Updated DL/DDoS taxonomy
Ping et al.[150]	AI overview + XAI	2024		✓	✓	✓		Broad AI overview; lacks depth on application-layer or XAI	Systematic DDoS-focused analysis
Kumari & Jain[104]	IoT DDoS countermeasures	2023	✓	✓	✓			IoT DDoS countermeasures; limited to 2016-2022 literature	2015–2025 comprehensive coverage
Pakmehr et al.[147]	IoT network DDoS	2024	✓	✓	✓			IoT network survey; no application-layer or XAI coverage	Application-layer HTTP DDoS + XAI
Choraś et al.[50]	XAI challenges	2024			✓		✓	XAI challenges only; lacks technique survey or solutions	XAI solutions + technique survey
Ables et al.[3]	XAI integration review	2025	✓				✓	XAI integration systematic review; no empirical evaluation	Empirical HTTP DDoS evaluation
Adel & Fuad[5]	IoT DL+XAI	2025		✓	✓	✓		IoT-specific; DL+XAI but no survey or taxonomy	Comprehensive taxonomy
Islam et al.[85]	XAI-based IDS	2024	✓	✓		✓	✓	XAI-IDS framework; single dataset evaluation	Multi-dataset + systematic comparison
Kavitha et al.[92]	SDN-IoT DDoS	2023	✓	✓	✓	✓		Focus on SDN-IoT; limited XAI emphasis	XAI taxonomy integration
Dilip et al.[55]	General DDoS survey	2023	✓	✓	✓	✓		Broad survey; overlooks application-layer and XAI	Systematic HTTP DDoS + XAI framework
Najafimehr et al.[134]	DDoS ML models	2023	✓	✓	✓			ML taxonomy only; no DL architectures or XAI integration	DL/Hybrid architectures + XAI
Singh & Behal[173]	SDN DDoS mitigation	2024	✓	✓	✓	✓		SDN plane-wise analysis; lacks application-layer HTTP DDoS	HTTP application-layer specificity
Haseeb-Ur-Rehman et al.[77]	High-speed network detection	2023	✓	✓	✓			High-speed networks focus; limited to detection, not mitigation	Detection + mitigation framework
Neupane et al.[139]	XAI + DDoS	2022	✓		✓		✓	XAI methods only; lacks technique comparison or empirical study	Comprehensive model + XAI comparison
Gelgi et al.[69]	IoT botnet DDoS	2024	✓	✓	✓			IoT botnet focus; taxonomy only, no XAI or model reproduction	HTTP DDoS + reproducible XAI
This paper	HTTP DDoS Detection (ML/XAI)	**2025**	✓	✓	✓	✓	✓	The roadmap requires empirical validation	First unified ML/XAI/dataset for HTTP DDoS

we introduce TRUST-AD in the next section as a conceptual roadmap to synthesize these dimensions into a unified framework.

7 TRUST-AD: A CONCEPTUAL ROADMAP

The preceding analysis reveals that both primary studies and existing surveys address HTTP DDoS detection through largely fragmented lenses, with limited integration across learning paradigms, explainability mechanisms, evaluation practices, and deployment contexts. To synthesize the recurring design patterns, deployment constraints, and explainability trade-offs across the literature into a unified, deployment-aware architectural framework, we introduce TRUST-AD as a conceptual roadmap. TRUST-AD focuses on how to integrate existing empirical insights (e.g., lightweight screening, deep temporal analysis, selective explainability, and long-term adaptability) into a coherent structure rather than proposing a new detection algorithm. Accordingly, TRUST-AD serves both as a research agenda and as a reference architecture for analyzing and designing trustworthy HTTP DDoS detection systems.

Specifically, TRUST-AD makes the following contributions in direct response to the limitations identified above:

- It introduces a multi-tier detection architecture that separates lightweight screening from deep temporal analysis, enabling scalability under real-time constraints.
- It formalizes *selective explainability* as a first-class design principle, aligning XAI methods with operational and deployment requirements.
- It identifies adaptability and concept drift handling as essential system-level components and incorporates them as a cross-layer capability rather than an afterthought.
- It reframes HTTP DDoS detection systems as *holistic model–XAI–deployment units*, providing a structured lens for evaluating existing and future approaches.

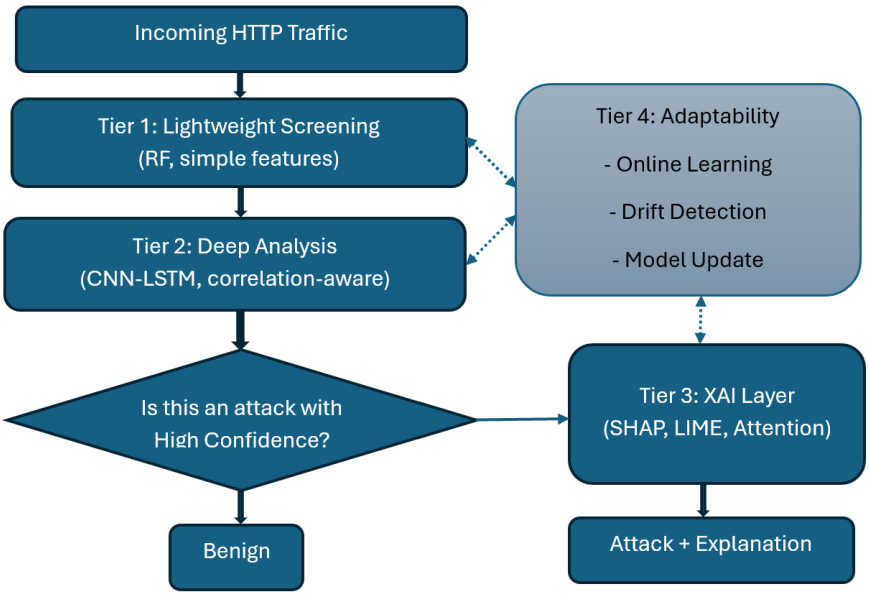

Fig. 8. TRUST-AD architecture with four tiers.

7.1 Multi-tier Detection Architecture of TRUST-AD

To systematically address the identified gaps, TRUST-AD employs a multi-tier architecture illustrated in Figure **??**. Tiers 1-3 form a sequential decision pipeline that progressively increases analytical complexity, while Tier 4 operates as a cross-layer adaptability plane that continuously monitors and updates the system. Each tier targets a specific gap while ensuring seamless integration with other layers

Tier 1: Lightweight Screening Layer
(Addresses Gap 1: Real-time detection constraints)

Literature Justification: Across multiple studies, tree-based models such as Random Forests and XGBoost are reported to achieve high accuracy (95–99%) with low inference latency, making them well suited for high-throughput screening [23, 67]. These models are commonly used to filter the majority of traffic at minimal computational cost.

Proposed Design: TRUST-AD places resource-efficient Random Forest models at the network edge or gateway, using compact feature sets (e.g., flow duration, packet rate statistics, request timing). Prior work indicates that 15–20 features can achieve performance comparable to larger feature sets while significantly reducing latency.

Deployment Context: Edge devices, network gateways, and IoT environments with strict latency constraints.

Tier 2: Deep Analysis Layer
(Addresses Gap 2: Temporal pattern modeling)

Literature Justification: CNN-LSTM hybrid architectures consistently outperform shallow models for detecting low-rate and stealthy HTTP DDoS attacks, with reported F1 scores exceeding 0.95 [30]. Correlation-aware extensions further improve detection of coordinated multi-source attacks by explicitly modeling relationships between traffic entities [78].

Proposed Design: Suspicious traffic flagged by Tier 1 is forwarded to a deep analysis layer that processes temporal sequences using CNN-LSTM models augmented with correlation-aware mechanisms. Reported window sizes range from 10 to 60 seconds, balancing detection fidelity and computational cost.

Deployment Context: Cloud or data-center environments where moderate inference latency (10–100ms per flow) is acceptable.

Tier 3: Explainability Layer
(Addresses Gap 3: Explainability integration)

Literature Justification: Post-hoc explainability methods such as SHAP and LIME are shown to provide valuable insights for analysts but incur non-trivial runtime overhead [23, 40, 159]. Prior studies emphasize that explainability is most valuable for investigation and auditing rather than for every prediction. Attention mechanisms embedded in deep models offer low-cost, ante-hoc interpretability.

Proposed Design: TRUST-AD adopts a selective, multi-granular explainability strategy:

- Global SHAP for understanding overall model behavior,
- Local SHAP or LIME for individual incident investigation,
- Attention visualization from Tier 2 models,
- Feature contribution analysis for root cause identification.

Comprehensive XAI is applied only to confirmed attacks, typically representing 5–10% of traffic, minimizing overhead while preserving transparency for security analysts.

Deployment Context: Post-detection analysis, SOC workflows, and regulatory reporting.

Tier 4: Adaptability Layer (Cross-Layer)
(Addresses Gap 4: Concept drift adaptation)

Literature Justification: Although rarely addressed in existing work, studies on online learning and drift detection show that continuous adaptation can substantially reduce performance degradation over time [98]. However, these mechanisms are not integrated into end-to-end detection architectures.

Proposed Design: Rather than acting as a sequential detection stage, Tier 4 operates as a cross-layer adaptability plane that supports all other tiers through:

- Statistical drift detection via feature and confidence monitoring,
- Incremental online learning using recent traffic,
- Model switching during detected drift periods,
- Dynamic ensemble weighting based on recent performance.

Deployment Context: Continuous background processes updating Tier 1 and Tier 2 models.

7.2 Deployment Guidelines from Literature

Building on the multi-tier architecture, we outline the following deployment strategies based on the observations in prior work to (i) respond to practitioner challenges discussed in Section 5, and (ii) show how TRUST-AD can be instantiated under different operational constraints.

High-Throughput Environments (> 10 Gbps):

- Deploy Tier 1 only with highly optimized Random Forest
- Literature suggests this handles 90% of traffic with <5ms latency
- Add Tier 2 selectively for flagged flows (<10% of traffic)

Security-Critical Environments (Financial, Healthcare):

- Full three-tier deployment (screening + deep analysis + XAI)
- Comprehensive explainability for regulatory compliance
- Literature indicates this achieves F1 > 0.97 with full audit trails

Resource-Constrained Environments (Edge/IoT):

- Tier 1 only at edge with minimal features (10-15)
- Offload Tier 2 analysis to cloud for suspicious traffic
- Literature shows this balances edge constraints with detection quality

Enterprise Networks (Mixed Requirements):

- Tiered deployment with Tier 1 at perimeter
- Tier 2 in data center for internal analysis
- Tier 3 integrated with SIEM systems
- Literature demonstrates this optimizes cost-performance tradeoffs

7.3 Expected Benefits Inferred from Prior Work

By integrating lightweight screening, deep analysis, selective explainability, and adaptability, TRUST-AD offers the following benefits that directly addresses the operational, robustness, and transparency gaps identified earlier.

Accuracy: Hybrid approaches in literature consistently achieve F1 > 0.97, with resource-efficient models reaching 0.99+ on benchmark datasets. Tiered architecture should maintain this performance while improving efficiency.

Efficiency: As introduced in TRUST-AD, lightweight screening (Tier 1) processes 80-90% of traffic with <2s inference per 1000 flows. Deep analysis (Tier 2) handles remaining 10-20% with acceptable latency (<100ms per flow).

Interpretability: Selective XAI deployment (Tier 3) provides transparency where needed without impacting real-time detection. Literature shows SHAP explanations achieve >80% consistency in feature attribution.

Adaptability: Continuous learning mechanisms (Tier 4) should maintain performance against evolving attacks. Limited existing work shows online learning reduces accuracy degradation from 15-20% to <5% over 6-month deployment periods.

Overall, TRUST-AD consolidates fragmented findings from the HTTP DDoS detection literature into a unified, deployment-aware architectural blueprint. By explicitly aligning model selection, explainability, and adaptability with operational constraints, the framework provides a structured lens for analyzing existing systems and guiding the design of future ones. Empirical instantiations of this roadmap offer a natural next step for evaluating trade-offs between accuracy, latency, interpretability, and long-term robustness, as we plan to do in future work.

8 CONCLUSION AND FUTURE WORK

This paper presented a comprehensive systematization of knowledge on hybrid machine learning and explainable AI approaches for HTTP DDoS detection. Through a PRISMA-style review spanning a decade of research (2015–2025), we organized a fragmented literature into a unified analytical framework encompassing learning paradigms, explainability mechanisms, evaluation practices, and deployment contexts.

8.1 Summary of Contributions

Our analysis provides several key insights into the current state and limitations of the field.

First, we showed that while supervised learning dominates existing work and delivers strong performance under controlled conditions, **the literature increasingly relies on hybrid architectures** that combine lightweight classical models with deep temporal learners. These hybrid designs emerge as a practical response to real-time constraints and evolving attack behaviors, rather than purely accuracy-driven choices.

Second, our systematization revealed that explainability remains inconsistently integrated. Although post-hoc XAI methods dominate existing implementations, **explainability is often decoupled from deployment realities**, with limited attention to computational overhead, analyst workflows, or selective usage. As a result, explainability is frequently treated as an auxiliary feature rather than a system-level design decision.

Third, we identified and quantified **several persistent research gaps**, including limited adversarial evaluation, heavy dataset reuse, scarce real-world validation, and minimal consideration of concept drift. These gaps collectively indicate that progress in model accuracy has not translated into deployable, long-lived detection systems.

To synthesize these findings, we introduced TRUST-AD as a conceptual, literature-driven roadmap for trustworthy HTTP DDoS detection. Rather than proposing a new system, **TRUST-AD organizes recurring design principles observed across prior work into a coherent, deployment-aware architecture**. The framework highlights how

lightweight screening, deep temporal analysis, selective explainability, and continuous adaptation can be integrated within a single detection pipeline, providing a unifying lens for future empirical research.

8.2 Implications for the Field

For Researchers: This systematization highlights the need to **move beyond isolated model improvements toward system-level evaluation**. Priority directions include lightweight and selective explainability, standardized adversarial testing, longitudinal deployment studies, domain adaptation across datasets, and empirical validation of multi-tier detection pipelines.

For Practitioners: Our findings provide guidance for aligning model choice with operational constraints, illustrating when hybrid architectures, selective XAI, or cloud-assisted analysis are most appropriate. The identified deployment patterns **offer practical strategies** for balancing accuracy, latency, and interpretability in real-world environments.

For Publication Venues: We recommend stronger emphasis on reproducibility, explicit reporting of computational and deployment costs, and evaluation protocols that reflect operational realities rather than benchmark-only performance.

Overall, this SoK advances the understanding of HTTP-layer DDoS detection by reframing the problem as an integrated learning–explainability–deployment challenge. By consolidating fragmented findings into a unified systematization and conceptual roadmap, we aim to support the development of detection systems that are not only accurate, but also interpretable, adaptable, and deployable. Our companion empirical study, currently in preparation, will build upon this foundation by implementing and evaluating key TRUST-AD components in controlled and production settings.

Acknowledgments

The authors acknowledge the use of generative AI tools, specifically OpenAI ChatGPT, in an iterative and assistive manner to obtain reviewer-style feedback, refine organization, and improve clarity and language in selected sections of the paper. All technical content, interpretations, analyses, and final editorial decisions remain the sole responsibility of the authors.

Artifact Availability

This paper provides conceptual artifacts derived from our systematization, including taxonomy figures, classification tables, and a structured bibliographic dataset of reviewed studies. These artifacts support transparency and reproducibility of the survey process.
GitHub Repository: https://github.com/TagBiz-lang/Anomaly-Detection-for-HTTP-Layer-DDoS-Comparative-Evaluation-of-Hybrid-ML-and-XAI-Approaches.

References

[1] Muhammad Aamir and Syed Mustafa Ali Zaidi. 2021. Clustering based semi-supervised machine learning for DDoS attack classification. *Journal of King Saud University-Computer and Information Sciences* 33, 4 (2021), 436–446.

[2] S Abiramasundari and V Ramaswamy. 2025. Distributed denial-of-service (DDOS) attack detection using supervised machine learning algorithms. *Scientific Reports* 15, 1 (2025), 13098.

[3] Jesse Ables et al. 2025. A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity. *Frontiers in Artificial Intelligence* (2025). doi:10.3389/frai.2025.1526221

[4] Giuseppe Aceto, Fabio Giampaolo, Ciro Guida, Stefano Izzo, Antonio Pescape, Francesco Piccialli, and Edoardo Prezioso. 2024. Synthetic and privacy-preserving traffic trace generation using generative AI models for training Network Intrusion Detection Systems. *Journal of Network and Computer Applications* 229 (2024), 103926. doi:10.1016/j.jnca.2024.103926 Dataset available at https://codeberg.org/CiroGuida/GenAI-network-traffic.

[5] Ali Adel and Muhammad Fuad. 2025. An intrusion detection system over the IoT data streams using eXplainable artificial intelligence (XAI). *Sensors* 25, 3 (2025), 847. doi:10.3390/s25030847

[6] Charan Ahmad, Sridhar Adepu, and Aditya Mathur. 2017. A Testbed and Dataset for Cyber-Physical Intrusion Detection. *Proceedings of the International Conference on Prognostics and Health Management (ICPHM)* (2017).

[7] Zahra Ahmad, A. Shahid Khan, Chee Wai Shiang, Jamaludin Abdullah, and Fadi Ahmad. 2021. Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches. *Transactions on Emerging Telecommunications Technologies* 32, 1 (2021), e4150.

[8] Khatereh Ahmadi and Reza Javidan. 2024. A Trust Based Anomaly Detection Scheme Using a Hybrid Deep Learning Model for IoT Routing Attacks Mitigation. *IET Information Security* (2024).

[9] Shakil Ibne Ahsan, Phil Legg, and SM Iftekharul Alam. 2025. An explainable ensemble-based intrusion detection system for software-defined vehicle ad-hoc networks. *Cyber Security and Applications* (2025).

[10] AI-RAN Alliance. 2024. *AI-RAN Alliance Vision and Mission White Paper.* Technical Report. AI-RAN Alliance.

[11] Noor Ul Ain, Muhammad Sardaraz, Muhammad Tahir, Mohamed W Abo Elsoud, and Abdullah Alourani. 2025. Securing IoT Networks Against DDoS Attacks: A Hybrid Deep Learning Approach. *Sensors* (2025).

[12] Akamai. 2024. Why Modern Layer 7 DDoS Protections Are Crucial for Web Security in 2024. https://www.akamai.com/blog/security/why-modern-layer-7-ddos-protections-crucial-web-security-2024. Accessed: 2025-01-15.

[13] Devrim Akgun, Selman Hizal, and Unal Cavusoglu. 2022. A new DDoS attacks intrusion detection model based on deep learning for cybersecurity. *Computers & Security* (2022).

[14] Abdullah Alabdulatif, Navod Neranjan Thilakarathne, and Mohamed Aashiq. 2024. Machine Learning Enabled Novel Real-Time IoT Targeted DoS/DDoS Cyber Attack Detection System. *Computers, Materials & Continua* (2024).

[15] Abdussalam Ahmed Alashhab, Mohd Soperi Zahid, Babangida Isyaku, Asma Abbas Elnour, Wamda Nagmeldin, Abdelzahir Abdelmaboud, Talal Ali Ahmed Abdullah, and Umar Danjuma Maiwada. 2024. Enhancing DDoS attack detection and mitigation in SDN using an ensemble online machine learning model. *IEEE access* 12 (2024), 51630–51649.

[16] Pol Alemany, Raul Munoz, Juan Castaneda Cisneros, Mehmet Karaca, Pawani Porambage, Hien Quoc Tran, Pierluigi G. Giardina, Ioannis Tzanettis, X. R. Sousa, Jose Maria Jorquera Valero, Behnam Ojaghi, Ricard Vilalta, Per Rugeland, Mathieu Boussard, Giacomo Landi, Alexandros Zafeiropoulos, Sergio Rodriguez, Manuel Gil Perez, Stylianos Barmpounakis, Mikko A. Uusitalo, Diego Lopez, and Stephane Kerboeuf. 2025. Defining Intent-Based Service Management Automation for 6G Multi-Stakeholders Scenarios. *IEEE Open Journal of the Communications Society* 6 (2025), 2373–2396.

[17] Ali Alfatemi, Mohamed Rahouti, Ruhul Amin, Sarah ALJamal, Kaiqi Xiong, and Yufeng Xin. 2024. Advancing ddos attack detection: A synergistic approach using deep residual neural networks and synthetic oversampling. *arXiv preprint* (2024).

[18] Daniyal Alghazzawi, Omaimah Bamasag, Hayat Ullah, and Muhammad Zubair Asghar. 2021. Efficient detection of DDoS attacks using a hybrid deep learning model with improved feature selection. *Applied Sciences* (2021).

[19] Ahmad Alzu'bi, Amjad Albashayreh, Abdelrahman Abuarqoub, and Mai AM Alfawair. 2024. Explainable AI-Based DDoS Attacks Classification Using Deep Transfer Learning. *Computers, Materials & Continua* 80, 3 (2024).

[20] Mulualem Bitew Anley, Angelo Genovese, Davide Agostinello, and Vincenzo Piuri. 2024. Robust DDoS attack detection with adaptive transfer learning. *Computers & Security* (2024).

[21] Fahim Arif, Nauman Ali Khan, Javed Iqbal, Faten Khalid Karim, Nisreen Innab, Samih M Mostafa, et al. 2024. DQQS: Deep Reinforcement Learning based Technique for Enhancing Security and Performance in SDN-IoT Environments. *IEEE Access* (2024).

[22] Osvaldo Arreche, Tanish R Guntur, Jack W Roberts, and Mustafa Abdallah. 2024. E-xai: Evaluating black-box explainable ai frameworks for network intrusion detection. *IEEE Access* (2024).

[23] Omar Arreche, Tejaswi R. Guntur, Jacob W. Roberts, and Mustafa Abdallah. 2024. E-XAI: Evaluating black-box explainable AI frameworks for network intrusion detection. *IEEE Access* 12 (2024), 23954–23988. doi:10.1109/ACCESS.2024.3365140

[24] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* (2020).

[25] G Arulselvan and A Rajaram. 2023. RETRACTED: Hybrid trust-based secure routing protocol for detection of routing attacks in environment monitoring over MANETs. *Journal of Intelligent & Fuzzy Systems* (2023).

[26] Julien Audibert, Pietro Michiardi, Frédéric Guyard, Sébastien Marti, and Maria A Zuluaga. 2020. Usad: Unsupervised anomaly detection on multivariate time series. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*.

[27] Aimira Baitieva, David Hurych, Victor Besnier, and Olivier Bernard. 2024. Supervised Anomaly Detection for Complex Industrial Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[28] Balakrishnan Balasubramanian, Edward S. Daniels, Markku Hiltunen, Rittwik Jana, Kaushik Joshi, Raghunathan Sivaraj, Tuan X. Tran, and Chih-Lin Wang. 2021. RIC: A RAN Intelligent Controller Platform for AI-Enabled Cellular Networks. *IEEE Internet Computing* 25, 2 (2021), 7–17.

[29] Omaimah Bamasag, Alaa Alsaeedi, Asmaa Munshi, Daniyal Alghazzawi, Suhair Alshehri, and Arwa Jamjoom. 2022. Real-time DDoS flood attack monitoring and detection (RT-AMD) model for cloud computing. *PeerJ Computer Science* (2022).

[30] Sukhvinder Singh Bamber, Aditya Vardhan Reddy Katkuri, Shubham Sharma, and Mohit Angurala. 2025. A hybrid CNN-LSTM approach for intelligent cyber intrusion detection system. *Computers & Security* 148 (2025), 104146.

[31] Sudipto Baral, Sajal Saha, and Anwar Haque. 2024. An Adaptive End-to-End IoT Security Framework Using Explainable AI and LLMs. In *2024 IEEE 10th World Forum on Internet of Things (WF-IoT)*.

[32] Raj Kumar Batchu, Thulasi Bikku, Srinivasarao Thota, Hari Seetha, and Abayomi Ayotunde Ayoade. 2024. A novel optimization-driven deep learning framework for the detection of DDoS attacks. *Scientific Reports* (2024).

[33] Ahmed Bensaoud and Jugal Kalita. 2025. Optimized detection of cyber-attacks on IoT networks via hybrid deep learning models. *Ad Hoc Networks* (2025).

[34] Abdelilah Benziker, R. Arunagiri, and G. Maheswari. 2023. Improved IDS for Vehicular Ad-Hoc Network using Deep Learning Approaches. In *IEEE 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*. 341–346.

[35] Daniel S Berman, Anna L Buczak, Jeffrey S Chavis, and Cherita L Corbett. 2019. A survey of deep learning methods for cyber security. *Information* (2019).

[36] Ketan Bhardwaj, Joaquin Chung Miranda, and Ada Gavrilovska. 2018. Towards {IoT-DDoS} prevention using edge computing. In *USENIX workshop on hot topics in edge computing (HotEdge 18)*.

[37] Ashwin Bhat, Adou Sangbone Assoa, and Arijit Raychowdhury. 2022. Gradient backpropagation based feature attribution to enable explainable-ai on the edge. In 2022 IFIP/IEEE 30th International Conference on Very Large Scale Integration (VLSISoC).

[38] Jalal Bhayo, Riaz Jafaq, Awais Ahmed, Sufian Hameed, and Syed Attique Shah. 2021. A time-efficient approach toward DDoS attack detection in IoT network using SDN. *IEEE Internet of Things Journal* 9, 5 (2021), 3612–3630.

[39] Ivo Afonso Bispo. 2025. *Explainable AI (XAI) for Cybersecurity: Intrusion Detection System (IDS).* Technical Report. Computer Science and Communication Research Center.

[40] Ivo Afonso Bispo. 2025. Explainable AI (XAI) for Cybersecurity: Intrusion Detection System (IDS). Computer Science and Communication Research Centr.

[41] Jonas Bushart and Christian Rossow. 2023. Anomaly-based filtering of application-layer DDoS against DNS authoritatives. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*. IEEE, 558–575.

[42] Antonio Carvajal and VR Garcia-Colon. 2003. High capacity motors on-line diagnosis based on ultra wide band partial discharge detection. In *4th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, 2003. SDEMPED 2003.*

[43] Jasmeen Kaur Chahal, Puninder Kaur, and Avinash Sharma. 2021. Distributed Denial of Service (DDoS) Attacks in Software-defined Networks (SDN). In *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)*.

[44] Ayan Chatterjee and Bestoun S Ahmed. 2022. IoT anomaly detection methods and applications: A survey. *Internet of Things* (2022).

[45] C Chen and Hsin-Chiao Chen. 2016. A Hybrid Approach Combining Rule-Based And Anomaly-Based Detection Against DDoS Attacks. *International Journal Of Network Security Its Applications* (2016).

[46] Chin-Ling Chen. 2009. A New Detection Method for Distributed Denial-of-Service Attack Traffic based on Statistical Test. *J. Univers. Comput. Sci.* (2009).

[47] Shao-Rui Chen, Shiang-Jiun Chen, and Wen-Bin Hsieh. 2025. Enhancing Machine Learning-Based DDoS Detection Through Hyperparameter Optimization. *Electronics* 14, 16 (2025), 3319.

[48] Yifei Chen, Rui Li, Zhiwei Zhao, Chao Peng, Jun Wu, Ekram Hossain, and Honggang Zhang. 2024. NetGPT: An AI-Native Network Architecture for Provisioning Beyond Personalized Generative Services. *IEEE Network* 38, 6 (2024), 404–413.

[49] Zhi Chen, Jiang Duan, Li Kang, and Guoping Qiu. 2022. Supervised anomaly detection via conditional generative adversarial network and ensemble active learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[50] Michał Choraś et al. 2024. The survey on the dual nature of xAI challenges in intrusion detection and their potential for AI innovation. *Artificial Intelligence Review* (2024). doi:10.1007/s10462-024-10972-3

[51] Cloudflare. 2025. Hyper-volumetric DDoS attacks skyrocket: Cloudflare's 2025 Q2 DDoS threat report. https://blog.cloudflare.com/ddos-threat-report-for-2025-q2/. Accessed: 2025-01-15.

[52] Jian Cui, Jun Xiao, Hao Zhong, Jian Zhang, Liang Wei, Irina Bolodurina, and Debiao He. 2024. LH-IDS: Lightweight Hybrid Intrusion Detection System Based on Differential Privacy in VANETs. *IEEE Transactions on Mobile Computing* 23, 12 (2024), 12195–12210.

[53] Henning Cüppers, Simon Schoen, Gregory Blanc, and Pierre-François Gimenez. 2024. FlowChronicle: Synthetic Network Flow Generation through Pattern Set Mining. In *Proceedings of the 20th International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3651205.3651217

[54] Zhihao Dai, Ligang He, Shuanghua Yang, and Matthew Leeke. 2024. SARAD: Spatial association-aware anomaly detection and diagnosis for multivariate time series. *Advances in Neural Information Processing Systems* (2024).

[55] Kumar Dilip and Kumar Yashwant. 2025. Detecting and Mitigating DDoS Attacks: The Role of AI and Machine Learning. *International Journal of Trend in Scientific Research and Development* (2025).

[56] Marinos Dimolianis, Adam Pavlidis, and Vasilis Maglaris. 2021. SYN flood attack detection and mitigation using machine learning traffic classification and programmable data plane filtering. In *2021 24th conference on innovation in clouds, internet and networks and workshops (ICIN)*.

[57] Roberto Doriguzzi-Corin, Stuart Millar, Sandra Scott-Hayward, Jesus Martinez-del Rincon, and Domenico Siracusa. 2020. LUCID: A practical, lightweight deep learning solution for DDoS attack detection. *IEEE Transactions on Network and Service Management* (2020).

[58] R Doriguzzi-Corin and D Siracusa. 2022. FLAD: Adaptive federated learning for DDoS attack detection. *arXiv preprint* (2022).

[59] Keval Doshi, Yasin Yilmaz, and Suleyman Uludag. 2021. Timely detection and mitigation of stealthy DDoS attacks via IoT networks. *IEEE Transactions on Dependable and Secure Computing* 18, 5 (2021), 2164–2176.

[60] Hani Elubeyd and Derya Yiltas-Kaplan. 2023. Hybrid deep learning approach for automatic DoS/DDoS attacks detection in software-defined networks. *Applied Sciences* (2023).

[61] Cheng Fan, Jian Cui, Hui Jin, Hao Zhong, Irina Bolodurina, and Debiao He. 2024. Auto-Updating Intrusion Detection System for Vehicular Network: A Deep Learning Approach Based on Cloud-Edge-Vehicle Collaboration. *IEEE Transactions on Vehicular Technology* 73, 10 (2024), 15372–15384.

[62] Yebo Feng, Jun Li, and Thanh Nguyen. 2020. Application-layer DDoS defense with reinforcement learning. In *2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS)*.

[63] Yebo Feng, Jun Li, Devkishen Sisodia, and Peter Reiher. 2023. On explainable and adaptable detection of distributed denial-of-service traffic. *IEEE Transactions on Dependable and Secure Computing* (2023).

[64] Mohamed Amine Ferrag, Othmane Friha, Djallel Hamouda, Leandros Maglaras, and Helge Janicke. 2022. Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. *IEEE Access* 10 (2022), 40281–40306.

[65] Umang Garg, Maninder Kaur, Malvika Kaushik, and Neha Gupta. 2021. Detection of DDoS attacks using semi-supervised based machine learning approaches. In *2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST)*. IEEE, 112–117.

[66] Diogo Gaspar, Paulo Silva, and Catarina Silva. 2024. Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron. *IEEE Access* 12 (2024), 30164–30175.

[67] Nikola Gavric, Guru Prasad Bhandari, and Andrii Shalaginov. 2024. Towards resource-efficient DDoS detection in IoT: leveraging feature engineering of system and network usage metrics. *Journal of Network and Systems Management* 32, 4 (2024), 69.

[68] Alexander Geiger, Dongyu Liu, Sarah Alnegheimish, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. 2020. Tadgan: Time series anomaly detection using generative adversarial networks. In *2020 ieee international conference on big data (big data)*.

[69] Metehan Gelgi, Yueting Guan, Sanjay Arunachala, Maddi Samba Siva Rao, and Nicola Dragoni. 2024. Systematic literature review of IoT botnet DDOS attacks and evaluation of detection techniques. *Sensors* 24, 11 (2024), 3571. doi:10.3390/s24113571

[70] Mateusz Gniewkowski, Henryk Maciejewski, and Tomasz Surmacz. 2022. Anomaly detection techniques for different ddos attack types. In *International Conference on Dependability and Complex Systems*.

[71] Jonathan Goh, Sridhar Adepu, Mian Tan, and Zhi Wei Lee. 2017. A Dataset to Support Research in the Design of Secure Water Treatment Systems. *Proceedings of the International Conference on Prognostics and Health Management (ICPHM)* (2017).

[72] Yonghao Gu, Kaiyue Li, Zhenyang Guo, and Yongfei Wang. 2019. Semi-supervised K-means DDoS detection method using hybrid feature selection algorithm. *IEEE Access* 7 (2019), 64351–64365.

[73] Ahmed Gueriani, Hichem Kheddar, and Abdelkader C. Mazari. 2024. Enhancing IoT Security with CNN and LSTM-based Intrusion Detection Systems. In *IEEE 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. 1–7.

[74] Ehsan Hallaji, Vaishnavi Shanmugam, Roozbeh Razavi-Far, and Mehrdad Saif. 2025. A Study on Semi-Supervised Detection of DDoS Attacks under Class Imbalance. *arXiv preprint arXiv:2506.22949* (2025).

[75] Amal Hamada, Salwa Mohamed Hassan, Salma Samy, Mohamed Azab, and Efat Fathalla. 2024. A review: State-of-the-art of integrating AI models with moving-target defense for enhancing IoT networks security. In *2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 108–114.

[76] Han and Others. 2012. Massive DDoS Attack on Cryptocurrency Exchange: A Case Study. *IEEE Access* 10 (2012), 1435–1448.

[77] R. M. A. Haseeb-ur Rehman et al. 2023. High-speed network DDoS attack detection: A survey. *Sensors* 23, 15 (2023), 6850. doi:10.3390/s23156850

[78] Arvin Hekmati, Jiahe Zhang, Tamoghna Sarkar, Nishant Jethwa, Eugenio Grippo, and Bhaskar Krishnamachari. 2024. Correlation-aware neural networks for DDOS attack detection in IoT systems. *IEEE/ACM Transactions on Networking* (2024).

[79] Mohammad Hiari, Yousef Alraba'nah, and Iyas Qaddara. 2025. A Deep Learning-Based Intrusion Detection System using Refined LSTM for DoS Attack Detection. *Engineering, Technology & Applied Science Research* 15, 4 (2025), 25627–25633.

[80] Carol Hildebrand. 2021. The Beat Goes On. https://www.netscout.com/blog/asert/beat-goes. NETSCOUT ASERT Blog, Accessed: May 21, 2025.

[81] Subhadeep Hore, Quang H. Nguyen, Yuchen Xu, Aayush Shah, Nathan D. Bastian, and Tien Le. 2023. Empirical Evaluation of Autoencoder Models for Anomaly Detection in Packet-based NIDS. In *IEEE Conference on Dependable and Secure Computing (DSC)*. 1–8.

[82] Mohamed Idhammad, Karim Afdel, and Mustapha Belouch. 2018. Semi-supervised machine learning approach for DDoS detection. *Applied Intelligence* 48, 10 (2018), 3193–3208.

[83] Fatemeh Imani, Masoud Kargar, Alireza Assadzadeh, and Ali Bayani. 2024. Integrating CNN-LSTM Networks with Statistical Filtering Techniques for Intelligent IoT Intrusion Detection. In *2024 8th International Conference on Smart Cities, Internet of Things and Applications (SCIoT)*. IEEE, 189–195.

[84] Imperva. 2023. Imperva Global DDoS Threat Landscape Report 2023. https://www.imperva.com/resources/reports/the-imperva-global-ddos-threat-landscape-report-2023.pdf. Accessed: 2025-02-15.

[85] Md Saiful Islam et al. 2024. XAI-IDS: Toward proposing an explainable artificial intelligence framework for enhancing network intrusion detection systems. *Applied Sciences* 14, 10 (2024), 4170. doi:10.3390/app14104170

[86] K. I. Iyer. 2021. From Signatures to Behavior: Evolving Strategies for Next-Generation Intrusion Detection. *European Journal of Advances in Engineering and Technology* 8, 6 (2021), 165–171.

[87] Sonali Jadhav and Arun Kulkarni. [n. d.]. Comprehensive Survey on Detection of Anomalies in Edge Computing Network and Deep Learning Solutions. ([n. d.]).

[88] Sarthak Jain and Byron C. Wallace. 2019. Attention Is Not Explanation. *NAACL* (2019).

[89] Donghyun Je, Jihyun Jung, and Sunghyun Choi. 2021. Toward 6G Security: Technology Trends, Threats, and Solutions. *IEEE Communications Standards Magazine* 5, 3 (2021), 64–71.

[90] Md Sarwar Kamal, Nilanjan Dey, Linkon Chowdhury, Syed Irtija Hasan, and KC Santosh. 2022. Explainable AI for glaucoma prediction analysis to understand risk factors in treatment planning. *IEEE Transactions on Instrumentation and Measurement* (2022).

[91] Parneet Kaur, Manish Kumar, and Abhinav Bhandari. 2017. A review of detection approaches for distributed denial of service attacks. *Systems Science & Control Engineering* 5, 1 (2017), 301–320.

[92] D Kavitha and R Ramalakshmi. 2024. Machine learning-based DDOS Attack Detection and Mitigation in SDNs for IoT environments. *Journal of the Franklin Institute* (2024).

[93] M Kavitha, M Suganthy, Aniket Biswas, R Srinivsan, R Kavitha, and A Rathesh. 2022. Machine learning techniques for detecting ddos attacks in sdn. In *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*. IEEE, 634–638.

[94] Thura Jabbar Khaleel and Nadia Adnan Shiltagh. 2023. DDoS Cyber-Attacks Detection-Based Hybrid CNN-LSTM. In *International Conference on Computing and Communication Networks*.

[95] Izhar Ahmed Khan, Nour Moustafa, Imran Razzak, Muhammad Tanveer, Dechang Pi, Yue Pan, and Bakht Sher Ali. 2022. XSRU-IoMT: Explainable simple recurrent units for threat detection in Internet of Medical Things networks. *Future generation computer systems* (2022).

[96] Maryam Mahsal Khan and Mohammed Alkhathami. 2024. Anomaly detection in IoT-based healthcare: machine learning for enhanced security. *Scientific reports* (2024).

[97] Bum-Sok Kim, Hye-Won Suk, Yong-Hoon Choi, Dae-Sung Moon, and Min-Suk Kim. 2024. Optimal Cyber Attack Strategy Using Reinforcement Learning Based on Common Vulnerability Scoring System. *CMES-Computer Modeling in Engineering & Sciences* (2024).

[98] Dongmin Kim, Sunghyun Park, and Jaegul Choo. 2024. When model meets new normals: test-time adaptation for unsupervised time-series anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*.

[99] Song-Kyoo Kim and Hou Cheng Vong. 2025. Secured Network Architectures Based on Blockchain Technologies: A Systematic Review. *Comput. Surveys* (2025).

[100] Richard Kimanzi, Peter Kimanga, Dedan Cherori, and Patrick K Gikunda. 2024. Deep Learning Algorithms Used in Intrusion Detection Systems–A Review. *arXiv preprint* (2024).

[101] Mudita Kohli and Indu Chhabra. 2025. A comprehensive survey on techniques, challenges, evaluation metrics and applications of deep learning models for anomaly detection. *Discover Applied Sciences* (2025).

[102] Heena Kousar, Mohammed Moin Mulla, Pooja Shettar, and DG Narayan. 2021. Detection of DDoS attacks in software defined network using decision tree. In *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*. IEEE, 783–788.

[103] Prashant Kumar, Chitra Kushwaha, Dimple Sethi, Debjani Ghosh, Punit Gupta, and Ankit Vidyarthi. 2025. Investigating the performance of multivariate LSTM models to predict the occurrence of Distributed Denial of Service (DDoS) attack. *PloS one* (2025).

[104] Pooja Kumari and Ashish Kumar Jain. 2023. A comprehensive study of DDoS attacks over IoT network and their countermeasures. *Computer Security* 127 (2023), 103096. doi:10.1016/j.cose.2023.103096

[105] Lagnajit Kundu, Xiaoming Lin, Raghav Gadiyar, Jean-Francois Lacasse, and Suman Chowdhury. 2025. AIRAN: Transforming RAN with AI-driven Computing Infrastructure. *arXiv preprint arXiv:2501.09007* (2025).

[106] Ravie Lakshmanan. 2025. New HTTPBot Botnet Launches 200+ Precision DDoS Attacks on Gaming and Tech Sectors. https://thehackernews.com/2025/05/new-httpbot-botnet-launches-200.html. Accessed: May 21, 2025.

[107] Arash Habibi Lashkari, Gerard Draper-Gil, Md. S. I. Mamun, and Ali A. Ghorbani. 2020. Characterization of DDoS Attacks Using CIC-DDoS2019 Dataset. *Proceedings of the International Conference on Information Systems Security and Privacy (ICISSP)* (2020).

[108] Huafeng Li, Chen Zhang, Zhanxuan Hu, Yafei Zhang, and Zhengtao Yu. 2024. Interactive attack-defense for generalized person re-identification. *Neural Networks* (2024).

[109] Haibin Li, Yi Zhao, Wenbing Yao, Ke Xu, and Qi Li. 2023. Towards real-time ML-based DDoS detection via cost-efficient window-based feature extraction. *Science China Information Sciences* (2023).

[110] Qing Li, He Huang, Ruoyu Li, Jianhui Lv, Zhenhui Yuan, Lianbo Ma, Yi Han, and Yong Jiang. 2023. A comprehensive survey on DDoS defense systems: New trends and challenges. *Computer Networks* (2023).

[111] Sifan Li, Yue Cao, Shuohan Liu, Yuping Lai, Yongdong Zhu, and Naveed Ahmad. 2024. Hda-ids: A hybrid dos attacks intrusion detection system for iot by using semi-supervised cl-gan. *Expert Systems with Applications* (2024).

[112] Zezhong Li, Wei Guo, Jianpeng An, Qi Wang, Yingchun Mei, Rongshun Juan, Tianshu Wang, Yang Li, and Zhongke Gao. 2025. AD2T: Multivariate Time Series Anomaly Detection with Association Discrepancy Dual-Decoder Transformer. *IEEE Sensors Journal* (2025).

[113] Zhuo Li, Yuhao Yan, Xiangheng Wang, Yifei Ge, and Lin Meng. 2025. A survey of deep learning for industrial visual anomaly detection. *Artificial Intelligence Review* 58, 9 (2025), 279.

[114] Cheng Liang, Hao Du, Yiming Sun, Dusit Niyato, Jiawen Kang, Dong Zhao, and Muhammad Ali Imran. 2025. Generative AI-driven Semantic Communication Networks: Architecture, Technologies, and Applications. *IEEE Transactions on Cognitive Communications and Networking* 11, 1 (2025), 27–47.

[115] Tiago Linhares, Ahmed Patel, Ana Luiza Barros, and Marcial Fernandez. 2023. SDNTruth: innovative DDoS detection scheme for software-defined networks (SDN). *Journal of Network and Systems Management* (2023).

[116] Haitao Liu and Haifeng Wang. 2023. Real-time anomaly detection of network traffic based on CNN. *Symmetry* (2023).

[117] Zhenpeng Liu, Yihang Wang, Fan Feng, Yifan Liu, Zelin Li, and Yawei Shan. 2023. A DDoS detection method based on feature engineering and machine learning in software-defined networks. *Sensors* (2023).

[118] LiveMint News Staff. 2020. Google Services Including YouTube, Gmail, Google Drive Face Outage Due to DDoS Attack. https://www.livemint.com/technology/apps/google-services-youtube-gmail-google-drive-face-outage-11607947475759.html. Accessed: May 21, 2025.

[119] Gaur Loveleen, Bhandari Mohan, Bhadwal Singh Shikhar, Jhanjhi Nz, Mohammad Shorfuzzaman, and Mehedi Masud. 2023. Explanation-driven HCI model to examine the mini-mental state for Alzheimer's disease. *ACM Transactions on Multimedia Computing, Communications and Applications* (2023).

[120] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*.

[121] Ruikui Ma, Qiuqian Wang, Xiangxi Bu, and Xuebin Chen. 2023. Real-time detection of DDoS attacks based on random forest in SDN. *Applied Sciences* (2023).

[122] Manisha Malik, Maitreyee Dutta, et al. 2023. Feature engineering and machine learning framework for DDoS attack detection in the standardized internet of things. *IEEE Internet of Things Journal* (2023).

[123] Innocent Mbona and Jan HP Eloff. 2022. Detecting zero-day intrusion attacks using semi-supervised machine learning approaches. *IEEE Access* 10 (2022), 69822–69838.

[124] Yair Meidan, Michael Bohadana, Asaf Shabtai, Juan Guarnizo, Martin Ochoa, Nils Ole Tippenhauer, and Yuval Elovici. 2018. N-BaIoT: Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders. *IEEE Pervasive Computing* 17, 3 (2018), 12–22.

[125] Jelena Mirkovic and Peter Reiher. 2005. D-WARD: a source-end defense against flooding denial-of-service attacks. *IEEE Transactions on Dependable and Secure Computing* (2005).

[126] Chris Misa, Ramakrishnan Durairajan, Arpit Gupta, Reza Rejaie, and Walter Willinger. 2024. Leveraging prefix structure to detect volumetric ddos attack signatures with programmable switches. In *2024 IEEE Symposium on Security and Privacy (SP)*.

[127] Harsh Mittal, Hitesh Saluja, and Shubham Behal. 2023. DDoS-AT-2022: a distributed denial of service attack dataset for evaluating DDoS defense system. *Proceedings of the Indian National Science Academy* 89, 2 (2023), 565–573. doi:10.1007/s43538-023-00114-3

[128] Meenakshi Mittal, Krishan Kumar, and Sunny Behal. 2023. Deep learning approaches for detecting DDoS attacks: A systematic review. *Soft computing* (2023).

[129] Mohammadreza Mohammadi, Rakesh Shrestha, Sima Sinaei, Alberto Salcines, David Pampliega, Raul Clemente, and Ana Lourdes Sanz. 2023. Anomaly detection using lstm-autoencoder in smart grid: A federated learning approach. In *Proceedings of the 2023 7th International Conference on Cloud and Big Data Computing*.

[130] Reza Mohammadi, Reza Javidan, and Mauro Conti. 2017. Slicots: An sdn-based lightweight countermeasure for tcp syn flooding attacks. *IEEE Transactions on Network and Service Management* (2017).

[131] Nour Moustafa and Jill Slay. 2015. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 military communications and information systems conference (MilCIS)*. IEEE, 1–6.

[132] Mohsin Munir, Shoaib Ahmed Siddiqui, Andreas Dengel, and Sheraz Ahmed. 2018. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE Access* 7 (2018), 1991–2005.

[133] Nura Shifa Musa, Nada Masood Mirza, Saida Hafsa Rafique, Amira Mahamat Abdallah, and Thangavel Murugan. 2024. Machine learning and deep learning techniques for distributed denial of service anomaly detection in software defined networks current research solutions. *IEEE Access* (2024).

[134] Mohammad Najafimehr, Sajjad Zarifzadeh, and Seyedakbar Mostafavi. 2023. DDoS attacks and machine-learning-based detection methods: A survey and taxonomy. *Engineering Reports* (May 2023). doi:10.1002/eng2.12697

[135] Alessandro Nascita, Giuseppe Aceto, Domenico Ciuonzo, Antonio Montieri, Valerio Persico, and Antonio Pescape. 2024. A Survey on Explainable Artificial Intelligence for Internet Traffic Classification and Prediction, and Intrusion Detection. *IEEE Communications Surveys & Tutorials* (2024).

[136] Jamal Abdul Nasir, Osama Subhani Khan, and Iraklis Varlamis. 2021. Fake news detection: A hybrid CNN-RNN based deep learning approach. *International journal of information management data insights* (2021).

[137] Euclides Carlos Pinto Neto, Sajjad Dadkhah, Raphael Ferreira, Alireza Zohourian, Rongxing Lu, and Ali A Ghorbani. 2023. CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors* (2023).

[138] Euclides Carlos Pinto Neto, Sajjad Dadkhah, Raphael Ferreira, Alireza Zohourian, Rongxing Lu, and Ali A Ghorbani. 2023. CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors* 23, 13 (2023), 5941.

[139] Subash Neupane et al. 2022. Explainable intrusion detection systems (X-IDS): A survey of current methods, challenges, and opportunities. *arXiv preprint arXiv:2207.06236* (2022). https://arxiv.org/abs/2207.06236

[140] Suman Neupane, Jared Ables, William Anderson, Sparsh Mittal, Saeed Rahimi, Ioan Banicescu, and Michael Seale. 2022. Explainable Intrusion Detection Systems (X-IDS): A Survey of Current Methods, Challenges, and Opportunities. *IEEE Access* 10 (2022), 112392–112415.

[141] O-RAN Alliance. 2018. *O-RAN White Paper: Towards an Open and Smart RAN.* Technical Report. O-RAN Alliance.

[142] O-RAN Alliance. 2025. *O-RAN WG3 E2 General Aspects and Principles (E2GAP).* Technical Report v07.00. O-RAN Alliance.

[143] O-RAN Alliance. 2025. *O-RAN WG3 E2 Service Model (E2SM) KPM Specification.* Technical Report v06.00. O-RAN Alliance.

[144] Behnam Ojaghi, Ferran Adelantado, Angelos Antonopoulos, and Christos Verikoukis. 2022. SlicedRAN: Service-Aware Network Slicing Framework for 5G Radio Access Networks. *IEEE Systems Journal* 16, 2 (2022), 2556–2567.

[145] Oghenetejiri Okporokpo, Funminiyi Olajide, Nemitari Ajienka, et al. 2025. Detection of DDoS Cyberattack Using a Hybrid Trust-Based Technique for Smart Home Networks. *International Journal of Advanced Computer Science & Applications* (2025).

[146] Mohamed Ouhssini, Karim Afdel, Elhafed Agherrabi, Mohamed Akouhar, and Abdallah Abarda. 2024. DeepDefend: A comprehensive framework for DDoS attack detection and prevention in cloud computing. *Journal of King Saud University-Computer and Information Sciences* (2024).

[147] Ali Pakmehr, Arno Aßmuth, Neda Taheri, and Ali Ghaffari. 2024. DDoS attack detection techniques in IoT networks: a survey. *Cluster Computing* 27, 10 (2024), 14637–14668. doi:10.1007/s10586-024-04662-6

[148] Bo Peng, Eric Alcaide, Quentin Anthony, Alon Albalak, Samuel Arcadinho, Stella Biderman, Huanqi Cao, Xin Cheng, Michael Chung, Matteo Grella, et al. 2023. Rwkv: Reinventing rnns for the transformer era. *arXiv preprint* (2023).

[149] Trung V Phan, Tri Gia Nguyen, Nhu-Ngoc Dao, Truong Thu Huong, Nguyen Huu Thanh, and Thomas Bauschert. 2020. DeepGuard: Efficient anomaly detection in SDN with fine-grained traffic flow monitoring. *IEEE Transactions on Network and Service Management* (2020).

[150] Zhe Ping and Dingyang Jiao. 2024. A Study on the Application of Artificial Intelligence in DDoS Attack Defense: A Literature Review. In *Proceedings of the 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management.*

[151] Michele Polese, Leonardo Bonati, Salvatore D'Oro, Stefano Basagni, and Tommaso Melodia. 2023. Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges. *IEEE Communications Surveys & Tutorials* 25, 2 (2023), 1376–1411.

[152] Marcos Aurélio Ribeiro, Mauro Sergio Pereira Fonseca, and Juliana de Santi. 2023. Detecting and mitigating DDoS attacks with moving target defense approach based on automated flow classification in SDN networks. *Computers & Security* 134 (2023), 103462.

[153] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. In *KDD*.

[154] Markus Ring et al. 2019. Flow-based Benchmark Data Sets for Intrusion Detection. *European Conference on Cyber Warfare and Security* (2019).

[155] Gaith Rjoub, Jamal Bentahar, Omar Abdel Wahab, Rabeb Mizouni, Alyssa Song, Robin Cohen, Hadi Otrok, and Azzam Mourad. 2023. A survey on explainable artificial intelligence for cybersecurity. *IEEE Transactions on Network and Service Management* (2023).

[156] Khushnaseeb Roshan and Aasim Zafar. 2022. Using kernel shap xai method to optimize the network anomaly detection model. In *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom).*

[157] Ahoora Rostamian. 2024. Applications of Deep Learning Models in Financial Forecasting. University of Essex.

[158] Mohammad A Salahuddin, Vahid Pourahmadi, Hyame Assem Alameddine, Md Faizul Bari, and Raouf Boutaba. 2021. Chronos: Ddos attack detection using time-based autoencoder. *IEEE Transactions on Network and Service Management* (2021).

[159] AL Samed and Seref Sagiroglu. 2025. Explainable artificial intelligence models in intrusion detection systems. *Engineering Applications of Artificial Intelligence* (2025).

[160] Yogesh B Sanap and Pushpalata Aher. 2023. A Comprehensive Survey On Detection And Mitigation Of DDoS Attacks Enabled With Deep Learning Techniques In Cloud Computing. In *2023 6th International Conference on Advances in Science and Technology (ICAST).*

[161] Reneilson Santos, Danilo Souza, Walter Santo, Admilson Ribeiro, and Edward Moreno. 2020. Machine learning algorithms to detect DDoS attacks in SDN. *Concurrency and Computation: Practice and Experience* 32, 16 (2020), e5402.

[162] Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. 2022. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *European Conference on Computer Vision.*

[163] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. 2022. Anomaly detection in time series: a comprehensive evaluation. *Proceedings of the VLDB Endowment* (2022).

[164] R Raja Sekar, Ardhala Mounika Jenny, Dubba Sreshta, M Vikas, Dasari Badri Nageshwar Ajay, and Mankena Ganesh. 2023. Prediction of distributed denial of service attacks in SDN using machine learning techniques. In *2023 3rd International Conference on Intelligent Technologies (CONIT).* IEEE, 1–5.

[165] Jamshed Ali Shaikh, Chengliang Wang, Muhammad Wajeeh Us Sima, Muhammad Arshad, Muhammad Owais, Dina SM Hassan, Reem Alkanhel, and Mohammed Saleh Ali Muthanna. 2025. A deep Reinforcement learning-based robust Intrusion Detection System for securing IoMT Healthcare Networks. *Frontiers in Medicine* (2025).

[166] N. S. Shaji, R. Muthalagu, and P. M. Pawar. 2024. SD-IIDS: Intelligent Intrusion Detection System for Software-Defined Networks. *Multimedia Tools and Applications* 83, 4 (2024), 11077–11109.

[167] Iman Sharafaldin, Arash Habibi Lashkari, Ali A Ghorbani, et al. 2018. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp* 1, 2018 (2018), 108–116.

[168] Bhavya Sharma, Lalit Sharma, Chhagan Lal, and Sajal Roy. 2023. Anomaly Based Network Intrusion Detection for IoT Attacks Using Deep Learning Technique. *Computers and Electrical Engineering* 107 (2023), 108626.

[169] Lifeng Shen, Zhuocong Li, and James Kwok. 2020. Timeseries anomaly detection using temporal hierarchical one-class network. *Advances in neural information processing systems* (2020).

[170] Ravid Shwartz-Ziv and Amitai Armon. 2022. Tabular data: Deep learning is not all you need. *Information Fusion* 81 (2022), 84–90.

[171] Ayesha Siddique, Khurram Khalil, and Khaza Anuarul Hoque. 2025. Explainable AI-Guided Efficient Approximate DNN Generation for Multi-Pod Systolic Arrays. In *2025 26th International Symposium on Quality Electronic Design (ISQED).* IEEE, 1–8.

[172] Jagdeep Singh and Sunny Behal. 2020. Detection and mitigation of DDoS attacks in SDN: A comprehensive review, research challenges and future directions. *Computer Science Review* (2020).

[173] Prabhjot Singh et al. 2024. A comprehensive plane-wise review of DDoS attacks in SDN: Leveraging detection and mitigation through machine learning and deep learning. *Computer Networks* (2024). doi:10.1016/j.comnet.2024.110583

[174] Seyed Soltani, Amir Amanloo, Mohammad Shojafar, and Rahim Tafazolli. 2025. Intelligent Control in 6G Open RAN: Security Risk or Opportunity? *IEEE Open Journal of the Communications Society* 6 (2025), 840–880.

[175] Ta Sowmya and EA Mary Anita. 2023. A comprehensive review of AI based intrusion detection system. *Measurement: Sensors* (2023).

[176] Doddi Srilatha and N Thillaiarasu. 2022. DDoSNet: A deep learning model for detecting network attacks in cloud computing. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*.

[177] Ya Su, Yungang Zhao, Chenfeng Niu, Rong Liu, Wei Sun, and Dan Pei. 2019. Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 2828–2837.

[178] K Muthamil Sudar, M Beulah, P Deepalakshmi, P Nagaraj, and P Chinnasamy. 2021. Detection of Distributed Denial of Service Attacks in SDN using Machine learning techniques. In *2021 international conference on Computer Communication and Informatics (ICCCI)*.

[179] S Sumathi, R Rajesh, and N Karthikeyan. 2022. DDoS attack detection using hybrid machine learning based IDS models. *NIScPR-CSIR, India* (2022).

[180] Yiming Sun, Lei Zhang, Liang Guo, Jie Li, Dusit Niyato, and Yuguang Fang. 2025. S-RAN: Semantic-aware Radio Access Networks. *IEEE Communications Magazine* 63, 4 (2025), 207–213.

[181] Qiang Tang, Onur Ermis, Cuong D. Nguyen, Andre D. Oliveira, and Alexandre Hirtzig. 2022. A Systematic Analysis of 5G Networks With a Focus on 5G Core Security. *IEEE Access* 10 (2022), 18298–18319.

[182] Geethapriya Thamilarasu and Shiven Chawla. 2019. Towards deep-learning-driven intrusion detection for the internet of things. *Sensors* (2019).

[183] Srikanth Thudumu, Philip Branch, Jiong Jin, and Jugdutt Singh. 2020. A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of big data* (2020).

[184] Shreshth Tuli, Giuliano Casale, and Nicholas R Jennings. 2022. Tranad: Deep transformer networks for anomaly detection in multivariate time series data. *arXiv preprint arXiv:2201.07284* (2022).

[185] Akpan Itoro Udofot, Omotosho Moses Oluseyi, and Edim Bassey. 2024. Advanced Machine Learning–A Comprehensive Survey and New Research Directions. *International Journal of Advanced Engineering and Management* (2024).

[186] Cláudia M Viana, Maurício Santos, Dulce Freire, Patrícia Abrantes, and Jorge Rocha. 2021. Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach. *Ecological Indicators* (2021).

[187] Syed Wali, Yasir Ali Farrukh, and Irfan Khan. 2025. Explainable AI and random forest based reliable intrusion detection system. *Computers & Security* (2025).

[188] Bo-Xiang Wang, Jiann-Liang Chen, and Chiao-Lin Yu. 2022. An AI-powered network threat detection system. *IEEE Access* (2022).

[189] Jie Wang, Raphael C-W Phan, John N Whitley, and David J Parish. 2010. Augmented attack tree modeling of distributed denial of services and tree based attack detection method. In *2010 10th IEEE International Conference on Computer and Information Technology*. IEEE, 1009–1014.

[190] Jincheng Wang, Le Yu, John Lui, and Xiapu Luo. 2025. Modern DDoS Threats and Countermeasures: Insights into Emerging Attacks and Detection Strategies. *arXiv preprint* (2025).

[191] Longhui Wang, Xu Zhou, Weiping Ding, Lifang Chen, and Qi Dai. 2025. Ensemble Intrusion Detection Based on Heterogeneous Data Augmentation and Knowledge Distillation. *IEEE Transactions on Industrial Informatics* (2025).

[192] Rui Wang and Jiayao Li. 2025. VAHMSE: an efficient anomaly detection model based on variational autoencoder and heterogeneous multi-stacking ensemble learning. *Applied Intelligence* 55, 13 (2025), 1–26.

[193] Xiaofeng Wang, Yonghong Wang, Zahra Javaheri, Laila Almutairi, Navid Moghadamnejad, and Osama S Younes. 2023. Federated deep learning for anomaly detection in the internet of things. *Computers and Electrical Engineering* (2023).

[194] Yalu Wang, Jie Li, Wei Zhao, Zhijie Han, Hang Zhao, Lei Wang, and Xin He. 2023. N-STGAT: Spatio-temporal graph neural network based network intrusion detection for near-earth remote sensing. *Remote Sensing* 15, 14 (2023), 3611.

[195] Yuanyuan Wei, Julian Jang-Jaccard, Fariza Sabrina, Wen Xu, Seyit Camtepe, and Aeryn Dunmore. 2023. Reconstruction-based lstm-autoencoder for anomaly-based ddos attack detection over multivariate time-series data. *arXiv preprint* (2023).

[196] Yuanyuan Wei, Julian Jang-Jaccard, Amardeep Singh, Fariza Sabrina, and Seyit Camtepe. 2023. Classification and explanation of distributed denial-of-service (DDoS) attack detection using machine learning and shapley additive explanation (SHAP) methods. *arXiv preprint* (2023).

[197] Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C Cattin. 2022. Diffusion models for medical anomaly detection. In *International Conference on Medical image computing and computer-assisted intervention*.

[198] Haowen Xu, Wenxiao Chen, Nengwen Zhao, Zeyan Li, Jiahao Bu, Zhihan Li, Ying Liu, Youjian Zhao, Dan Pei, Yang Feng, et al. 2018. Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In *Proceedings of the 2018 world wide web conference*.

[199] Hongzuo Xu, Yijie Wang, Guansong Pang, Songlei Jian, Ning Liu, and Yongjun Wang. 2023. Rosas: Deep semi-supervised anomaly detection with contamination-resilient continuous supervision. *Information Processing & Management* (2023).

[200] Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. 2021. Anomaly transformer: Time series anomaly detection with association discrepancy. *arXiv preprint* (2021).

[201] Zhenyu Yin, Shang Liu, and Guangyuan Xu. 2025. DrLLM: Prompt-Enhanced Distributed Denial-of-Service Resistance Method with Large Language Models. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

[202] Noe M Yungaicela-Naula, Cesar Vargas-Rosales, Jesús Arturo Pérez-Díaz, and Diego Fernando Carrera. 2022. A flexible SDN-based framework for slow-rate DDoS attack mitigation by using deep reinforcement learning. *Journal of network and computer applications* (2022).

[203] Zahra Zamanzadeh Darban, Geoffrey I Webb, Shirui Pan, Charu Aggarwal, and Mahsa Salehi. 2024. Deep learning for time series anomaly detection: A survey. *Comput. Surveys* (2024).

[204] Daochen Zha, Kwei-Herng Lai, Mingyang Wan, and Xia Hu. 2020. Meta-AAD: Active anomaly detection with deep reinforcement learning. In *2020 IEEE International Conference on Data Mining (ICDM)*.

[205] Chuxu Zhang, Dongjin Song, Yuncong Chen, Xinyang Feng, Cristian Lumezanu, Wei Cheng, Jingchao Ni, Bo Zong, Haifeng Chen, and Nitesh V Chawla. 2019. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In *Proceedings of the AAAI conference on artificial intelligence*.

[206] Guoxing Zhang, Shengming Jiang, Gang Wei, and Quansheng Guan. 2009. A prediction-based detection algorithm against distributed denial-of-service attacks. In *Proceedings of the 2009 international conference on wireless communications and mobile computing*.

[207] Lan Zhang and Others. 2020. High-Level Cyber Attacks Detection Using AI. *IEEE Transactions on Information Forensics and Security* 15 (2020), 1234–1245.

[208] Tao Zhang, Fanyu Kong, Dongshang Deng, Xiangyun Tang, Xuangou Wu, Changqiao Xu, Liehuang Zhu, Jiqiang Liu, Bo Ai, Zhu Han, et al. 2025. Moving target defense meets artificial intelligence-driven network: A comprehensive survey. *IEEE Internet of Things Journal* (2025).

[209] Zhiyi Zhang, Guorui Xiao, Sichen Song, R Can Aygun, Angelos Stavrou, Lixia Zhang, and Eric Osterweil. 2024. Revealing Protocol Architecture's Design Patterns in the Volumetric DDoS Defense Design Space. *IEEE Communications Surveys & Tutorials* (2024).

[210] Junjie Zhao, Yongmin Liu, Qianlei Zhang, and Xinying Zheng. 2023. CNN-AttBiLSTM mechanism: a DDoS attack detection method based on attention mechanism and CNN-BiLSTM. *IEEE Access* (2023).

[211] Jia-Xing Zhong, Nannan Li, Weijie Kong, Shan Liu, Thomas H Li, and Ge Li. 2019. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

[212] Yingjie Zhou, Xucheng Song, Yanru Zhang, Fanxing Liu, Ce Zhu, and Lingqiao Liu. 2021. Feature encoding with autoencoders for weakly supervised anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[213] Berkay Özçam, H Hakan Kilinc, and Abdül Halim Zaim. 2021. Detecting tcp flood ddos attack by anomaly detection based on machine learning algorithms. In *2021 6th International Conference on Computer Science and Engineering (UBMK)*.

A Appendix

Table 4. Summary of Anomaly Detection (AD) Systems Across Research Domains

Author and Year	Problems Addressed	Proposed Solution	Results Obtained	Advantages	Disadvantages	Research Gaps	Data Source
Li et al., 2025[112]	Unified AD in multivariate time series via reconstruction, prediction, and association discrepancy learning	AD2T (Association Discrepancy Dual-decoder Transformer)	Superior SOTA performance	Joint reconstruction, prediction, and association learning	High computational cost	Limited real-time evaluation	SWaT, WADI, SMD, SMAP, MSL
Dai et al., 2024[54]	AD and root-cause diagnosis in multivariate time series	SARAD	Outperforms SOTA on detection and diagnosis	Spatial dependency modeling; interpretability	Extra computation overhead	Streaming deployment	SWaT, WADI, SMD, SMAP, MSL
Mohammedza et al., 2023[129]	Federated AD in smart grids	LSTM-NDT	High accuracy in distributed environments	Privacy-preserving; scalable	Communication overhead	More attack types	Smart grid logs
Xu et al., 2022[200]	Explainable AD for time-series	Anomaly Transformer (attention discrepancy)	High performance on benchmarks	Built-in explainability	Resource-intensive	Extend to multimodal data	SMAP, MSL
Audibert et al., 2020[26]	Unsupervised AD for finance	USAD	High detection accuracy	Reconstruction + adversarial learning	Architecture tuning complexity	Noisy data robustness	Fin logs
Munir et al., 2019[132]	Unsupervised AD	DeepAnT	High precision	No labels required	Limited multivariate support	Complex MTS	Sensor streams
Geiger et al., 2020[68]	Time-series AD	TadGAN (GAN-based)	High accuracy	Seasonal pattern modeling	Mode collapse risk	Training stability	TTA
Shen et al., 2020[169]	Temporal AD	THOC	High detection accuracy	Hierarchical temporal modeling	Domain tuning	Cyberattack datasets	Ind sensors
Tuli et al., 2022[184]	Transformer-based AD	TranAD (Transformer + attention)	High accuracy under distribution shift	Handles temporal context well	Large compute requirement	Optimize for edge deployment	SWaT

Table 4. Summary of Anomaly Detection (AD) Systems Across Research Domains (continued)

Zhang et al., 2019[205]	Multivariate AD	MSCRED	Improved accuracy	Spatio-temporal correlation	Labeled data dependency	Streaming AD	Server KPIs
Su et al., 2019[177]	IoT multivariate AD	OmniAnomaly	Robust detection	Uncertainty modeling	Low interpretability	XAI integration	IoT logs
Xu et al., 2018[198]	Seasonal KPI AD	Donut	Effective unsupervised AD	Periodicity handling	Sensitive tuning	Robustness improvement	Web KPIs

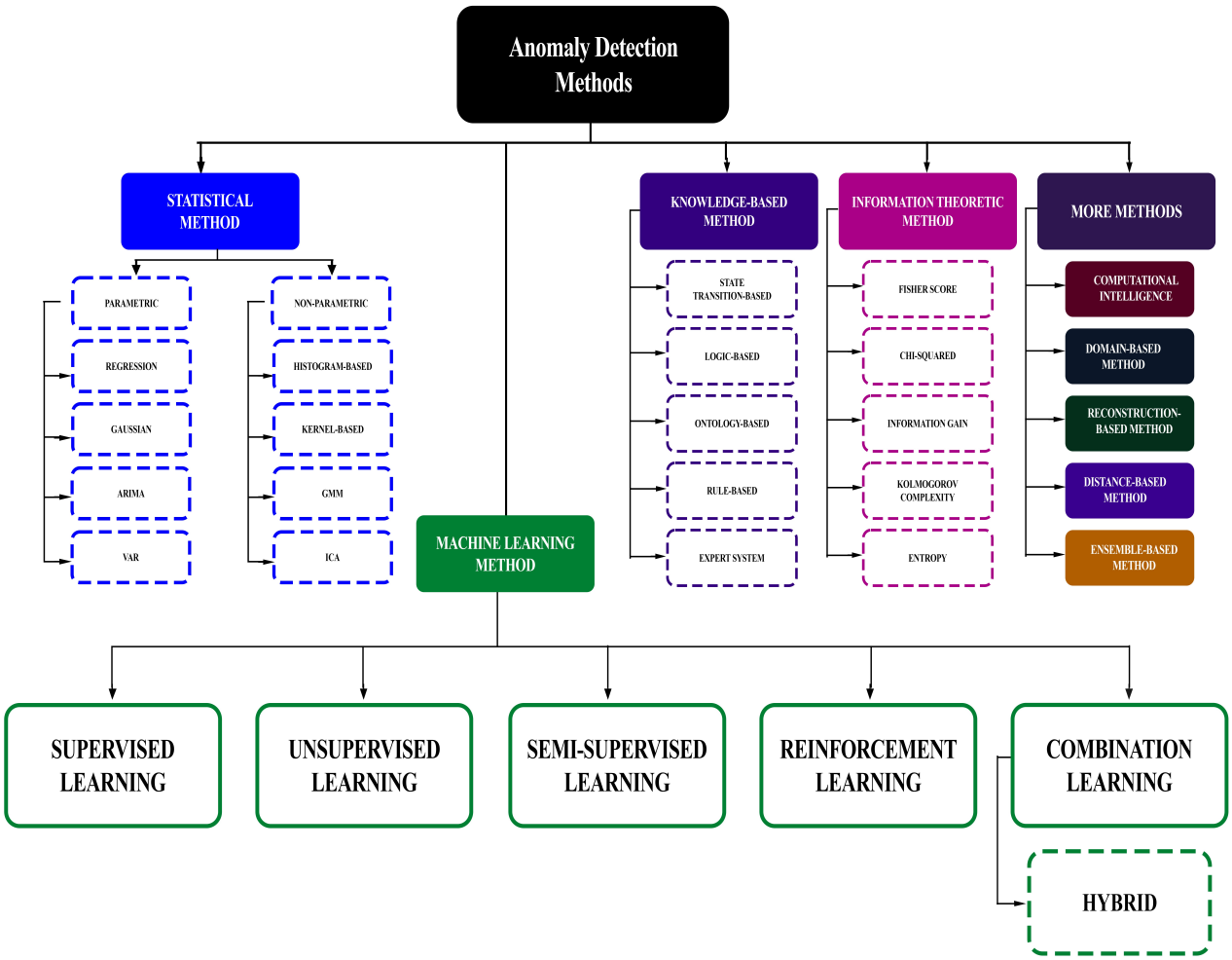

Fig. 9. Taxonomy of AD