electronics

MDPI

Article

A Tri-Axis Systematic Literature Review of AI-Powered Cyber Defense: ATT&CK-Aligned Analysis of Cyberattacks, Machine Learning Methods, and Datasets

Mohammad Chizari *, Abu Alam, Qublai Khan Ali Mirza and Hassan Chizari

School of Business, Computing and Social Sciences, University of Gloucestershire, Park Campus, Cheltenham GL50 2RH, UK; aalam@glos.ac.uk (A.A.); qalimirza@glos.ac.uk (Q.K.A.M.); hchizari@glos.ac.uk (H.C.)

* Correspondence: mchizari2@glos.ac.uk

Abstract

The increasing complexity and sophistication of cyberattacks have made machine learning (ML) and artificial intelligence (AI) central to modern cyber defense. However, existing surveys typically examine attacks, ML methods, or datasets separately, limiting understanding of how methodological choices align with adversarial behaviours and benchmark availability. This paper presents a systematic literature review (SLR) of AI- and ML-based cyber defense studies published between 2019 and 2025, framed as an ATT&CK-aligned tri-axis synthesis of cyberattacks, machine learning methods, and datasets. Across 99 primary studies, the review maps 312 attack labels to MITRE ATT&CK tactics and techniques, categorises the ML methods applied, and organizes 96 datasets into a refined taxonomy spanning NIDD, IoT-NIDD, malware, Spam and Phishing, ICS, Insider Threat, custom-collected, and other datasets. Rather than treating attacks, ML methods, and datasets as separate descriptive dimensions, the review analyses them jointly through a tri-axis cross-reference framework, enabling the identification of benchmark dependence, methodological concentration, and underexplored attack–method–dataset intersections that are not visible in single-axis or model-centred surveys. The synthesis shows that the literature is strongly concentrated on externally visible attacks associated with Impact, Initial Access, and Execution, that ensemble and deep learning models dominate high-frequency detection settings, and that dataset usage remains heavily skewed toward a small set of public benchmarks, particularly CSE-CIC-IDS2017, UNSW-NB15, and NSL-KDD. This review further identifies persistent blind spots, including limited coverage of post-compromise ATT&CK behaviours, sparse use of ICS and insider-threat datasets, and weak support for multi-stage or multi-dataset evaluation. These findings provide a more focused and actionable evidence base for future ML-based cyber defense research.

Keywords: cybersecurity; systematic literature review (SLR); MITRE ATT&CK; machine learning; deep learning; cyberattack analysis; intrusion detection; dataset taxonomy; threat detection

Academic Editor: Krzysztof Szczypiorski

Received: 13 March 2026
Revised: 8 May 2026
Accepted: 12 May 2026
Published: 25 June 2026

1. Introduction

Cybersecurity has become one of the most pressing challenges in the digital era, primarily due to the exponential growth of interconnected systems, cloud infrastructures, and Internet of Things (IoT) devices [1]. This rapid expansion of the attack surface has been accompanied by a marked increase in the complexity and sophistication of cyber threats [2].

Traditional attacks such as brute-force password guessing or denial-of-service (DoS) have evolved into highly coordinated, multi-vector campaigns that exploit advanced malware, supply-chain vulnerabilities, and even adversarial artificial intelligence (AI) [3,4]. As a result, conventional signature or rule-based detection mechanisms are increasingly insufficient in addressing modern adversarial tactics, techniques, and procedures (TTPs) [5,6].

The evolution of cyberattacks reflects a growing asymmetry between adversarial innovation and defensive readiness [7]. Recent years have witnessed a proliferation of sophisticated attacks targeting diverse domains, ranging from ransomware and distributed denial-of-service (DDoS) campaigns to stealthy insider threats and industrial control system (ICS) compromises [8,9]. While some attack categories such as intrusion detection [10,11] have attracted significant research attention, others remain comparatively underexplored. This imbalance suggests the presence of critical research gaps where adversaries may enjoy a strategic advantage due to limited academic or operational focus. Understanding not only which attack types dominate the research landscape, but also which ones are neglected, is essential for shaping future defensive strategies [12].

In response to these challenges, the field has increasingly turned to machine learning (ML) methods for adaptive and data-driven cyber defense [13]. The machine learning methods used in cybersecurity span a wide spectrum of techniques, ranging from foundational optimization and classical methods to advanced ensemble and hybrid deep learning architectures [14]. These methods are deployed to address critical challenges, including intrusion detection, malware classification, phishing detection, and insider threat monitoring [3,15]. Deep learning approaches such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) have further demonstrated strong capabilities in capturing temporal, spatial, and multi-modal attack patterns, thereby enhancing detection accuracy [16–18]. However, this diversity of techniques raises important questions: Which ML paradigms are most effective for particular adversarial behaviours or attack stages? Why do certain methods dominate in specific contexts while others remain underutilized?

A systematic way to address these questions is by mapping ML methods onto structured attack frameworks. The MITRE ATT&CK framework provides a comprehensive taxonomy of adversarial TTPs across stages such as Initial Access, Lateral Movement, and Exfiltration [19]. By aligning ML approaches with these tactics and techniques, researchers can gain deeper insight into the suitability of specific algorithms for particular adversarial behaviours. For example, ensemble learning methods have frequently demonstrated utility in classifying attack behaviours, while sequential models such as LSTM and GRU are often applied to detect exfiltration or credential access attacks [17,18]. Such cross-referencing not only illuminates prevailing trends but also highlights underutilized approaches with potential for novel defensive applications.

Equally critical to AI-driven cybersecurity research is the question of datasets. The reliability and generalisability of ML models are fundamentally dependent on the representativeness of their training data [18]. Despite the frequent use of benchmark datasets such as NSL-KDD [20], UNSW-NB15 [21], and CIC-IDS2017 [22], these resources have well-documented limitations, including outdated traffic patterns and limited coverage of modern adversarial techniques [1,8]. While more specialized datasets such as ToN-IoT [23] and BoT-IoT [9] have emerged for IoT security, other critical domains—including insider threats, multi-vector attack simulations, and ICS environments—remain comparatively under-represented. The lack of systematic integration across dataset categories further constrains the robustness of existing ML models and leaves significant blind spots in defensive research.

Several systematic literature reviews (SLRs) have examined the intersection of AI and cybersecurity from different perspectives. Foundational surveys such as those by

Buczak and Guven [6] and Sommer and Paxson [1] established early discussions around machine learning for intrusion detection, while Ring et al. [17] and Yang et al. [18] examined the growing use of deep learning for cybersecurity tasks. More recent reviews have expanded the field in different directions. For example, Sowmya et al. [24] focused on AI-based intrusion detection, Mvula et al. [8] reviewed cybersecurity data repositories and performance metrics for semi-supervised learning, Salem et al. [25] surveyed AI-driven cyber-threat detection techniques, Ofusori et al. [26] provided a broad review of AI applications in cybersecurity, and Rehman et al. [27], Hozouri et al. [28], and Dobler et al. [29] examined IDS-oriented or dataset-oriented aspects of the field.

Despite these valuable contributions, most existing reviews remain limited in three respects:

1. They emphasise ML models without systematically connecting them to adversarial behaviours or structured taxonomies such as MITRE ATT&CK;
2. They treat datasets in isolation, without a comprehensive categorisation across domains such as NIDD, IoT-NIDD, ICS, and Insider Threat datasets;
3. They rarely provide cross-reference analyses that jointly consider attacks, ML paradigms, and datasets.

To address these limitations, this paper is positioned as a systematic literature review with tri-axis synthesis, rather than as a general survey of AI in cybersecurity. This review is organized around three linked evidence dimensions: (i) cyberattacks, represented through MITRE ATT&CK tactics and techniques; (ii) machine learning methods, spanning classical, ensemble, hybrid, and deep learning approaches; and (iii) datasets, categorised into a refined taxonomy of cybersecurity data sources. The novelty of this review lies not only in aligning attacks to MITRE ATT&CK or in organising datasets into a refined taxonomy, but in cross-referencing cyberattacks, ML methods, and datasets jointly as three linked evidence axes. This tri-axis perspective makes it possible to identify concentration patterns, benchmark dependence, and missing attack–method–dataset intersections that are not visible in single-axis or model-centred surveys.

The scope of this review is intentionally bounded. It includes peer-reviewed studies published between 2019 and 2025 that apply ML- or AI-based methods to cyberattack detection, classification, or mitigation and that report identifiable attack types and datasets. It does not aim to review all applications of AI in cybersecurity, such as cryptography, privacy-preserving AI, secure software engineering, purely conceptual discussions, or studies that do not expose a clear attack–method–dataset relationship. This boundary is necessary to support a consistent and reproducible cross-reference synthesis across the three evidence axes.

Accordingly, this paper should be read not as a general review of AI-powered cybersecurity, but as an ATT&CK-aligned SLR that synthesizes the literature through the linked lenses of cyberattacks, machine learning methods, and datasets.

Positioning Against Existing Reviews

Recent review papers provide valuable but partial views of the literature. Sowmya et al. [24] review AI-based intrusion detection and organize work around ML, DL, and ensemble methods, but remain IDS-centred rather than ATT&CK-centred. Mvula et al. [8] provide an SLR of cybersecurity datasets and performance metrics for semi-supervised learning, but their emphasis is dataset repositories and evaluation metrics rather than cross-referencing attacks, methods, and datasets. Salem et al. [25] survey AI-driven detection techniques across more than sixty recent studies and cover broad cyber threats such as malware, network intrusion, and spam, but they do not organize the literature through an ATT&CK-aligned tri-axis synthesis. Ofusori et al. [26] review AI applications

in cybersecurity at a broader level, but without a structured ATT&CK-based threat synthesis. More recent IDS-focused reviews, such as Rehman et al. [27] and Hozouri et al. [28], concentrate on detection models, benchmark datasets, evaluation metrics, and IDS architectures, again without an explicit ATT&CK mapping or a structured attack–method–dataset cross-reference. Dataset-focused work also remains more specialized: Dobler et al. [29], for instance, systematically characterize malicious industrial network traffic datasets for ML evaluation, but their contribution is dataset-centric rather than a broader synthesis across attacks, ML methods, and dataset usage. Table 1 summarises how this review differs from representative recent surveys.

Table 1. Positioning of this review against representative recent surveys.

Review	Time Window/Focus	Main Analytical Emphasis	Attack Taxonomy	Dataset Taxonomy Depth	ATT&CK Mapping	Attack–Method–Dataset Cross-Reference	Main Limitation Relative to This Study
Sowmya et al. (2023) [24]	72-paper review of AI-based IDS	ML, DL, and ensemble methods for intrusion detection	No explicit ATT&CK-style taxonomy	Moderate	No	Limited	IDS-centered rather than ATT&CK-aligned or tri-axis.
Mvula et al. (2023) [8]	SSL-focused SLR on cybersecurity datasets and metrics	Dataset repositories and performance metrics	No explicit attack taxonomy	Strong	No	Limited	Dataset- and metric-centred, not a tri-axis synthesis.
Salem et al. (2024) [25]	Review of more than sixty AI-driven cyber-threat studies	Broad comparison of ML, DL, and metaheuristics	Broad attack coverage, but no ATT&CK-based taxonomy	Moderate	No	Limited	Technique-centred rather than ATT&CK-organized.
Ofusori et al. (2024) [26]	Broad review of AI in cybersecurity	Applications, trends, and future directions	No structured threat taxonomy	Limited–moderate	No	No	Too high-level to expose specific attack–method–dataset gaps.
Rehman et al. (2025) [27]	Systematic review of ML-based intrusion detection	Models, datasets, metrics, and challenges	IDS/domain framing rather than ATT&CK tactics/techniques	Moderate–strong	No	Limited	Close in topic, but still IDS-centric and not ATT&CK-aligned.
Hozouri et al. (2025) [28]	Survey of IDS with ML/DL advances	IDS architectures, benchmark datasets, and emerging challenges	No explicit ATT&CK-based taxonomy	Moderate	No	Limited	Strong IDS synthesis, but not a behavioural cross-reference review.
Dobler et al. (2025) [29]	Systematic review of malicious industrial traffic datasets	Dataset characterization and ML-oriented selection	Industrial attack types, but not ATT&CK as the main frame	Strong	No	Limited	Domain-specific dataset review rather than a broader tri-axis synthesis.
This review	2019–2025; 99 studies	Tri-axis synthesis across attacks, ML methods, and datasets	MITRE ATT&CK-aligned	Strong	Yes	Yes	Designed to expose underexplored intersections, benchmark dependence, and gaps across attack behaviours, model families, and dataset categories.

Dataset taxonomy depth is reported qualitatively. Limited indicates illustrative mention only, moderate indicates structured but bounded discussion, and strong indicates a dedicated taxonomy or systematic dataset characterization. Attack–method–dataset cross-reference indicates whether a review explicitly synthesizes these three dimensions together rather than discussing them separately.

Taken together, prior review papers provide valuable but partial views of the literature. Broad AI–cybersecurity surveys tend to emphasise methods and applications at a high level, IDS-focused reviews concentrate on detection architectures, datasets, and performance measures, and dataset-oriented studies examine repositories or benchmark suitability within narrower methodological settings. In contrast, the present study is designed as an ATT&CK-aligned systematic literature review with tri-axis synthesis. Its contribution is not merely to summarize attacks, models, or datasets separately, but to examine how these three evidence dimensions interact across the literature. This framing makes it possible to identify underexplored ATT&CK behaviours, methodological concentration around particular ML families, benchmark dependence, and missing intersections across attack types, model classes, and dataset categories.

Accordingly, the present review is differentiated not simply by broader coverage, but by its tri-axis analytical design, which cross-references attacks, methods, and datasets jointly rather than discussing each dimension in isolation.

The contributions of this study are threefold:

- **ATT&CK-aligned attack synthesis:** This review identifies and maps 312 attack labels from 99 studies to MITRE ATT&CK tactics and techniques, providing a structured threat-oriented view of the literature.
- **Refined cross-domain dataset taxonomy:** This review organizes 96 datasets into a refined taxonomy spanning NIDD, IoT-NIDD, malware, Spam and Phishing, ICS, Insider Threat, custom-collected, and other datasets, thereby clarifying the benchmark landscape used in AI-driven cyber defense research.
- **Tri-axis cross-reference analysis:** The central contribution of this study is a joint cross-reference of cyberattacks, ML methods, and datasets as three linked evidence axes. This tri-axis analysis reveals methodological concentration, benchmark dependence, and underexplored attack–method–dataset intersections that are not visible when attacks, models, or datasets are examined separately.

Through this systematic approach, the review aims to bridge fragmented research streams in AI-driven cybersecurity and provide actionable insights for both academic and operational communities. The findings highlight not only current practices but also critical gaps—such as the underutilization of insider-threat and ICS datasets, and the persistent over-reliance on single-dataset experiments—that must be addressed to develop more resilient and generalizable ML-based defense systems.

The remainder of this paper is organized as follows. Section 2 outlines the methodology used to conduct the SLR. Section 3 presents a taxonomy and analysis of cyberattacks in line with MITRE ATT&CK tactics and techniques. Section 4 reviews and analyses the ML techniques used in the literature. Section 5 explores the datasets employed, their frequency, and the research gaps associated with them. Section 6 provides a cross-reference analysis of attack tactics, techniques, ML models, and datasets. Section 7 discusses the main gaps and limitations of the SLR, and Section 8 outlines implications for future research. Finally, Section 9 summarizes the main findings.

2. Methodology

Systematic Literature Reviews (SLRs) are widely used to identify research gaps, synthesize evidence, and structure knowledge within a specific domain. According to Alnabhan and Branco (2024) [30], an SLR provides a structured and reproducible approach for synthesising evidence and identifying trends in emerging research areas. This study follows the guidelines for conducting a Systematic Literature Review proposed by Keele (2007) [31], which have been widely applied in software engineering and cybersecurity research. The review protocol was designed to ensure reproducibility, minimise bias, and systematically address the research questions. The process consists of seven linked phases:

1. **Review Scope and Analytical Frame**;
2. **Research Questions**;
3. **Search Strategy and Literature Identification**;
4. **Study Selection and Eligibility Screening**;
5. **Data Extraction and Coding Procedure**;
6. **Attack Mapping, Method Grouping, and Dataset Categorisation Rules**;
7. **Ambiguity Handling, Consistency Checking, and Quality Appraisal**.

2.1. Review Scope and Analytical Frame

This review was designed as a systematic literature review with tri-axis evidence synthesis. The analytical unit of the review is not the ML model alone, nor the attack category or dataset in isolation, but the reported relationship among attack, method, and dataset within each included study. Accordingly, studies were retained only when they provided sufficient evidence to identify: (i) the cyberattack type, family, or ATT&CK-relevant

behaviour under study; (ii) the machine learning or deep learning method applied; and (iii) the dataset or data source used for evaluation. This design choice narrows the scope of the review to the literature that can support consistent cross-reference analysis across the three axes.

To preserve analytical coherence, this review does not seek to cover the full landscape of AI in cybersecurity. Studies were considered outside scope when they addressed cybersecurity in a broad conceptual sense without a clear attack focus, lacked explicit dataset grounding, or did not report a machine learning pipeline suitable for structured comparison. This bounded scope is aligned with the research questions and supports reproducible synthesis of trends, dominant practices, and gaps across attacks, methods, and datasets.

2.2. Review Research Questions

To guide the review and define the research aims, five research questions (RQs) were formulated, as shown in Table 2. These questions address the landscape of cyberattacks, machine learning methods, datasets, and gaps in the literature.

Table 2. Research Questions.

RQ	Research question
RQ1	What types of cyberattacks are most frequently studied, and how have they evolved?
RQ2	Which machine learning and deep learning techniques are applied to mitigate attacks?
RQ3	What datasets are commonly used in AI-powered cybersecurity research?
RQ4	Which ML techniques are associated with specific categories of cyberattacks?
RQ5	What are the key gaps and limitations in applying AI-powered methods for attack mitigation?

2.3. Search Strategy and Literature Identification

An unbiased and rigorous search strategy was adopted to retrieve the literature relevant to AI- and ML-based cybersecurity methods for cyberattack detection, classification, or mitigation. To maximize coverage, four major bibliographic databases were searched: IEEE Xplore, ACM Digital Library, Scopus, and Google Scholar. The search string was derived from the research questions and combined terms related to cyberattacks, machine learning, mitigation, and datasets:

(“cyberattack” OR “cyber security”) AND (“ML” OR “DL” OR “machine learning” OR “deep learning” OR “AI” OR “artificial intelligence”) AND (“mitigate” OR “mitigation”) AND (“dataset” OR “benchmarking”).

The search covered the period from January 2019 to March 2025 and produced an initial retrieval of 1960 papers.

2.4. Study Selection

Figure 1 illustrates the study selection process. An initial total of 1960 articles was retrieved from four databases: Scopus (1075), ACM (661), Google Scholar (82), and IEEE Xplore (142). After the removal of duplicates and non-relevant studies—such as non-English papers, posters, reviews, surveys, non-scientific publications, books, book chapters, newer versions of studies, guideline documents, and editorials—1245 studies remained, resulting in the exclusion of 715 records.

These 1245 studies were then subjected to title and abstract screening, during which the research methodology, results, and contributions were examined. Based on this step, 787 articles were excluded, leaving 458 studies. If the title and abstract did not provide

sufficient clarity about the application domain or contribution of the study, the paper was retained for the next stage.

Subsequently, full-text screening was carried out to assess the originality of the research and its relevance to cyberattacks, machine learning methods, and datasets. This process led to the exclusion of a further 359 studies. Ultimately, 99 primary studies were included in this systematic literature review and formed the basis for the findings and analysis presented in the following sections.

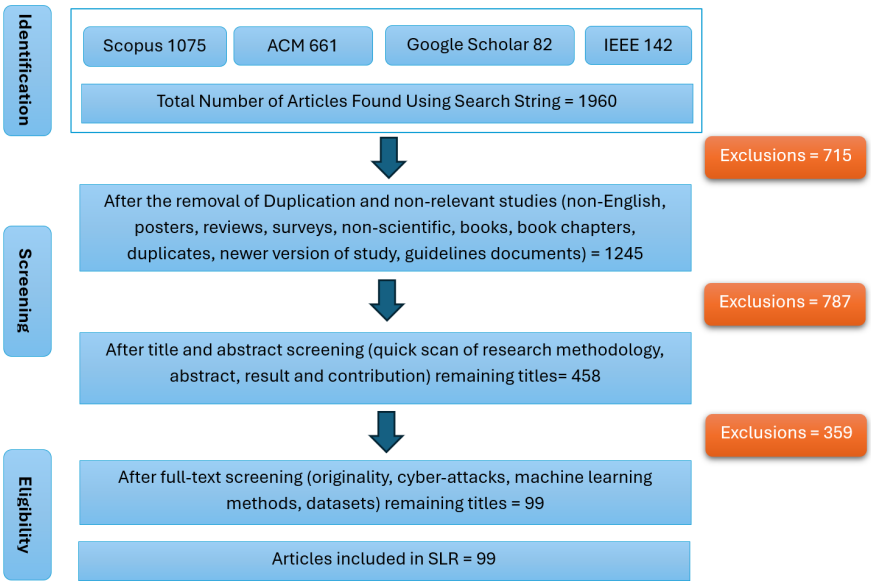

Figure 1. Study selection process: filtering and inclusion workflow.

Inclusion and Exclusion Criteria

Explicit inclusion and exclusion criteria were applied to ensure quality and relevance (Table 3).

Table 3. Inclusion and exclusion criteria.

Criterion	Description
Inclusion	Peer-reviewed articles published between 2019 and 2025, written in English, addressing AI- or ML-based methods for cyberattack detection, classification, or mitigation, and reporting identifiable model/classifier and dataset information.
Exclusion	Non-peer-reviewed works, duplicate records, papers outside computer science, cybersecurity, or closely related AI-for-security domains, short or insufficiently detailed papers (fewer than five pages), and studies lacking the methodological detail required for structured comparison.

2.5. Data Extraction and Coding Procedure

For each included study, a structured extraction form was used to record bibliographic and analytical metadata. The extracted fields included publication year, application domain, reported attack type(s), machine learning (ML) or deep learning method(s), dataset(s) used, performance metrics, and any study-specific contextual notes required for subsequent coding. The extraction process was designed to support the tri-axis analytical frame introduced in Section 2, in which each study was represented through the linked dimensions of attack, method, and dataset.

The coding process proceeded in three stages. First, raw study-level descriptors were extracted as reported by the original authors wherever possible. Second, these raw descriptors were normalized into comparable analytical units. For example, reported attack names were standardized into consistent attack labels, model names were grouped into method families and subfamilies, and datasets were assigned to a unified taxonomy. Third, the normalized entries were linked across the three axes to enable cross-reference analysis of attack–method, attack–dataset, method–dataset, and attack–method–dataset relationships.

To preserve traceability, coding was based on a principle of closest explicit evidence. When a study directly stated the attack type, ML method, or dataset used, that information was recorded without reinterpretation. Additional interpretation was introduced only when the study used broad or incomplete terminology and a more specific label could be derived from the dataset description, malware family, or evaluation context provided in the paper. This approach reduced unnecessary inference while allowing consistent aggregation across studies.

2.6. Attack Mapping, Method Grouping, and Dataset Categorisation Rules

Attack extraction and normalization. Attack evidence was extracted at the most specific level reported in each study. When authors explicitly named attack types (e.g., DoS, SQL injection, phishing, ransomware, or brute force), these labels were retained and later normalized for consistency across spelling variants, abbreviations, and closely related naming conventions. When a study referred only to a broad task such as "network intrusion detection" without enumerating attack classes, the attack labels were derived from the documented attack classes in the dataset used by that study. When a paper focused on malware families rather than named attack types, the malware family name was first recorded as the raw label and then mapped to the most appropriate ATT&CK-relevant adversarial behaviour based on the primary function or operational objective described in the study.

ATT&CK mapping rules. After normalization, each attack label was mapped to one primary MITRE ATT&CK tactic and one primary technique to maintain a consistent unit of comparison across this review. The mapping followed a hierarchical decision rule: the adversary's operational objective determined the tactic, and the most representative observable behaviour determined the technique. When a label could plausibly correspond to multiple ATT&CK entries, the mapping favoured the behaviour that best matched the study's detection target rather than the full attack chain. For example, a denial-of-service label was mapped to an Impact tactic and the corresponding denial-of-service technique, whereas a malware family used primarily for credential theft was mapped to Credential Access rather than to malware execution in a generic sense. To maintain comparability across domains, MITRE ATT&CK for Enterprise was used as the primary reference framework throughout this review, including for studies involving ICS-related datasets.

Machine learning method grouping. ML methods were first recorded using the terminology provided in the original study. They were then normalized into method families and subcategories to support meaningful aggregation across heterogeneous naming practices. Closely related variants were grouped under shared analytical labels; for example, decision-tree derivatives were grouped under Tree-Based Models, autoencoder variants under Autoencoders, and recurrent sequence models under LSTM/GRU families where appropriate. After this normalization step, the methods were organized into higher-level categories used throughout the review, namely Classical Machine Learning Models, Deep Learning Models, Hybrid/Ensemble/Explainable Methods, and Learning Paradigms and Optimization.

Dataset categorisation rules. Datasets were recorded using the names reported by the study authors and then assigned to a refined taxonomy developed for this review. The categorisation was based on the dataset's operational context, data source, and intended cybersecurity use case rather than on name alone. Accordingly, datasets were grouped into categories such as NIDD, IoT-NIDD, malware, Spam and Phishing, ICS, Insider Threat, custom-collected, and other datasets. When prior survey papers classified a dataset differently, reassignment in this review followed the way the dataset was actually used in the reviewed study. For example, datasets used in industrial monitoring or cyber-physical intrusion contexts were categorised under ICS even if they had been placed elsewhere in earlier surveys. This rule was adopted to prioritise analytical consistency with study usage over inheritance from prior taxonomies.

2.7. Ambiguity Handling, Consistency Checking, and Quality Appraisal

Ambiguities were handled through an explicit decision hierarchy. First, the study's own wording was used whenever it was sufficiently specific. Second, if the wording was broad or underspecified, the dataset documentation and evaluation context reported in the study were used to recover the likely attack class, method category, or dataset role. Third, if multiple interpretations remained plausible, the more conservative and more general label was retained to avoid over-specification. This rule was particularly important for malware-family studies, multi-stage attacks, and papers that used broad labels such as "intrusion" or "malicious traffic".

Initial screening, extraction, and coding were conducted by the first author. To improve coding consistency, ambiguous cases were flagged during extraction and revisited in iterative consistency passes using the shared coding rules described above. This process allowed earlier and later cases to be harmonized under a shared codebook, refine normalization rules, remove duplicate or near-duplicate labels, and ensure that similar study designs were treated under the same analytical criteria across the final review dataset.

A lightweight quality appraisal was also applied at the included-study level. This appraisal did not exclude studies after screening, but it was used to judge the interpretive strength of the evidence base. The appraisal considered whether a study clearly specified the following: (i) the attack target, (ii) the ML method or pipeline, (iii) the dataset or data source, (iv) the evaluation setting or metrics, and (v) sufficient methodological detail to support comparative interpretation.

Representative difficult cases included studies that used broad labels such as "intrusion" without enumerating attack classes, malware-family studies without explicit behavioural labels, and datasets whose operational use in the reviewed study differed from earlier survey classifications. These cases were resolved using the decision hierarchy described above, with preference given to the study's explicit evaluation target and the dataset's documented usage context.

Table 4 summarises the lightweight quality-appraisal criteria used to assess evidence clarity.

The quality-appraisal score was used descriptively to assess evidence clarity and interpretive confidence, rather than as an exclusion threshold after screening.

Together, these procedures established a traceable pipeline from search and screening to coding, normalization, and tri-axis synthesis, thereby supporting reproducible comparison across cyberattacks, machine learning methods, and datasets.

Table 4. Lightweight quality-appraisal criteria applied to included studies.

Criterion	Description	Scoring
Q1	The study clearly specifies the attack type, family, or adversarial behaviour under analysis.	0/1
Q2	The study clearly identifies the ML/DL method, model family, or detection pipeline used.	0/1
Q3	The dataset or data source is clearly reported and sufficiently identifiable.	0/1
Q4	The evaluation setting, metrics, or experimental design is sufficiently described for interpretation.	0/1
Q5	The study provides enough methodological detail to support comparative synthesis.	0/1

3. Cyberattacks Across Reviewed Studies

From the 99 included studies, 312 unique raw attack labels were extracted. Attack evidence was recorded at the most specific level reported in each paper and then normalized according to the coding rules defined in Section 2.6. When studies explicitly named attack classes, those labels were retained and standardized for consistency. When studies used broad task descriptions such as "network intrusion detection" without enumerating attack classes, the corresponding attack labels were derived from the attack classes documented in the dataset used by the study. For studies centred on malware families, the malware family name was first retained as the raw label and then mapped to the most appropriate ATT&CK-relevant adversarial behaviour on the basis of the malware's primary operational objective. This process provided a traceable bridge between heterogeneous study descriptions and the unified analytical framework used in the review.

3.1. Attack Taxonomy Selection and ATT&CK Mapping

Numerous taxonomies have been developed to classify cyberattacks based on various criteria, including attack vectors, targets, impacts, and methodologies [32]. While these taxonomies offer valuable perspectives, not all are well-suited to the scope and objectives of this systematic literature review (SLR), which analyses over 300 distinct cyberattack labels extracted from 99 academic studies. The decision to adopt a single, unified taxonomy necessitated a careful evaluation of several popular existing frameworks, their strengths, and their limitations in relation to the reviewed studies and datasets [33].

Several taxonomies focus on attack methods and locations, particularly within manufacturing and industrial contexts. For instance, research by Wu and Moon [34,35], Pan et al. [36], and Tuptuk and Hailes [37] emphasise physical domain attack vectors and the impact of cyberattacks on manufacturing systems. These taxonomies often target Industry 4.0 applications, additive manufacturing, and quality inspection processes [38–40]. While valuable for sector-specific threat modelling, these approaches are often too domain-specific, lacking generalisability to the broader range of network, malware, and intrusion-based attacks covered in this review.

The Common Attack Pattern Enumeration and Classification (CAPEC) schema [41] provides a comprehensive catalogue of attack patterns. While highly detailed, CAPEC focuses primarily on describing how attacks are executed (i.e., attack patterns), making it more useful for software developers or threat modelling than for high-level threat categorisation or mapping to real-world datasets. Its granularity, while useful for certain applications, presents challenges for mapping general attack types or dataset labels, which are often abstract or ambiguous.

Another taxonomy proposed by Hansman and Hunt [42] classifies attacks based on four dimensions: vector, target, vulnerability, and payload. This multi-dimensional "axe structure" offers a detailed framework but leans heavily toward malware-centric attacks, potentially under-representing human-focused or access-based threats. Similarly,

the taxonomy by Meyers et al. [43] emphasises subtypes and granular descriptions but retains a malware-heavy orientation.

Chapman et al. [44] introduced a novel approach based on access requirements, which is insightful in evaluating preconditions for different attacks. However, this method lacks detail in key aspects such as privilege escalation, and its practical application across diverse attack types is limited. Zhu et al. [32] proposed a taxonomy tailored to Operational Technology (OT) and ICS environments, categorising attacks across hardware, software, and communication stacks. While effective in the ICS domain, this model does not generalize well to IT systems or network-level attacks.

Simmons et al. [45] proposed the AVOIDIT taxonomy (Attack vector, Operational impact, Victim, Objective, Defense, Information impact, and Target), using a tree-based structure. While AVOIDIT provides a well-rounded view of the attack lifecycle, it suffers from uneven detail across categories—some attack vectors are richly described, while others remain vague.

MITRE ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge) [19] is a globally recognized, continuously updated knowledge base that categorises adversarial behaviours observed in real-world cyber incidents. The MITRE ATT&CK framework provides broad coverage, including a wide spectrum of tactics and techniques. This makes it compatible with the diverse nature of the reviewed studies, as it encompasses both malware behaviours and network-based intrusions. The framework's separation of tactics (the adversary's goal) and techniques (how the goal is achieved) enables hierarchical classification and flexible mapping of ambiguous or compound attack labels. MITRE ATT&CK is derived from empirical evidence and adversary behaviours, providing a practical foundation for aligning academic research with operational threat intelligence. While separate frameworks exist for Enterprise and ICS environments, ATT&CK allows for cross-domain integration. In this review, the MITRE ATT&CK for Enterprise was used as a unifying framework to maintain consistency, even when analysing ICS-related datasets. MITRE ATT&CK is supported by a wide ecosystem of tools and is widely adopted in both academia and industry, enhancing the reproducibility and applicability of this review. Given the diversity and often loosely defined nature of attack labels in the reviewed literature and datasets, the MITRE ATT&CK framework was selected as the most suitable unifying taxonomy for this review.

Despite its advantages, ATT&CK is not without limitations. Notably, it assumes the existence of network activity or observable system behaviour, which made it challenging to categorise studies that focused purely on malware binaries without behavioural context. In such cases, attack labels were mapped according to the malware sample's primary function or operational objective (e.g., credential access or lateral movement). Nevertheless, some datasets pertain to Industrial Control Systems (ICS), for which the MITRE ATT&CK for ICS provides a separate taxonomy. However, to maintain consistency across the review, the MITRE ATT&CK Enterprise framework was used as the unified categorisation scheme wherever feasible.

For analytical consistency, each normalized attack label was assigned one primary ATT&CK tactic and one primary technique. This one-to-one assignment was used as a comparative device rather than as a claim that real attacks are single-stage phenomena. In cases where an attack could plausibly map to multiple ATT&CK entries, the selected mapping prioritised the study's explicit detection focus and the most representative observable behaviour. This rule improved comparability across the full corpus while preserving interpretive transparency in ambiguous cases.

3.2. Cyberattack Frequency Analysis

Across the 99 reviewed papers, a total of 312 unique cyberattack labels were identified, each mapped to a single MITRE ATT&CK tactic and technique (see Appendix A). The distribution reveals a strong concentration around a limited subset of highly visible attack behaviours, a pattern examined further in the subsequent trend, co-occurrence, and breadth analyses. The top-ten most frequent labels (counted once per paper) are the following: DoS (30 papers), SQL Injection (26), XSS (26), PortScan (23), Backdoor (23), DoS Slowloris (20), SYN Flood (18), Reconnaissance (18), Worms (18), and Exploits (17). This distribution highlights the dominance of Denial-of-Service (DoS) attacks (including specific variants such as Slowloris and SYN Flood) and common web-based attacks (SQL Injection; XSS). These categories remain popular due to their high visibility, ease of deployment, and immediate operational consequences.

Table 5 summarises the distribution of attack labels across MITRE ATT&CK tactics. The most frequent tactic is Impact (72, 14.55%), reflecting the prevalence of attacks aimed at disrupting availability and causing direct operational harm—most notably Denial-of-Service (DoS) and its sub-variants. Following closely are Initial Access (59, 11.92%), Execution (58, 11.72%), and Command and Control (55, 11.11%), all of which focus on the adversary's ability to establish and sustain an initial foothold. Reconnaissance-related behaviour (Reconnaissance, 54, 10.91%) also features prominently, underscoring its role in both attack preparation and intrusion attempts.

Table 5. MITRE ATT&CK tactic counts.

Tactic	Count	Percentage
Impact	72	14.55%
Initial Access	59	11.92%
Execution	58	11.72%
Command and Control	55	11.11%
Reconnaissance	54	10.91%
Credential Access	42	8.48%
Defense Evasion	31	6.26%
Discovery	28	5.66%
Lateral Movement	23	4.65%
Exfiltration	21	4.24%
Persistence	17	3.43%
Collection	17	3.43%
Privilege Escalation	13	2.63%
Resource Development	5	1.01%

By contrast, tactics that represent deeper compromise stages—such as Persistence (3.43%), Collection (3.43%), Privilege Escalation (2.63%), and Resource Development (1.01%)—appear far less frequently. This suggests that current datasets and experimental studies focus disproportionately on early-stage, externally observable threats rather than long-term adversarial presence within systems.

Importantly, the dominance of Impact, Initial Access, and Execution in the literature should not be interpreted as evidence that these are the only or most consequential phases in real intrusions. Rather, these phases are more visible, easier to simulate, and better supported by public datasets, which makes them more attractive for benchmark-driven research.

The ten most common MITRE ATT&CK techniques are shown in Table 6. A clear skew exists towards availability-related behaviours: Network Denial of Service (61, 7.71%) and Endpoint Denial of Service (44, 5.56%) together account for over 13% of all labels. Other frequently observed techniques include Exploit Public-Facing Application (5.44%), Active

Scanning (5.06%), and Gather Victim Host Information (4.55%), reflecting the emphasis on perimeter-based exploitation and preparatory reconnaissance. Credential-related attacks (Brute Force, 4.30%) and command-driven execution (Command-Line Interface, 4.05%) also feature prominently.

Notably, techniques associated with stealthier, post-compromise activity—such as privilege escalation, persistence, or covert exfiltration—do not appear in the top 10, again illustrating the field's bias towards high-visibility attacks that are easier to simulate and label in controlled datasets.

Table 6. Top 10 most frequent MITRE ATT&CK techniques.

Technique	Count	Percentage
Network Denial of Service	61	7.71%
Endpoint Denial of Service	44	5.56%
Exploit Public-Facing Application	43	5.44%
Active Scanning	40	5.06%
Gather Victim Host Information	36	4.55%
Brute Force	34	4.30%
Application Layer Protocol	32	4.05%
Command-Line Interface	32	4.05%
Phishing	25	3.16%
Input Capture	23	2.91%

Taken together, these findings reveal a research landscape that strongly favours high-impact, externally observable attacks (e.g., DoS, scanning, and web exploitation). This emphasis is likely driven by the relative ease of reproducing such attacks in experimental testbeds, as well as the availability of public datasets containing these behaviours. However, the relative scarcity of advanced, low-and-slow tactics such as privilege escalation, persistence, lateral movement, and exfiltration highlights a notable gap between experimental research and real-world adversary tradecraft. Addressing this imbalance would require richer datasets and experimental designs capable of capturing stealthier post-compromise activities, which are often more critical in determining the true severity of a cyber intrusion.

3.3. Cyberattack Trends (2019–2024)

Since the data for 2025 are incomplete, including them would not provide an accurate reflection of the overall trend. Therefore, the trend analysis presented here is limited to the period 2019–2024. The longitudinal view of tactic usage (Figure 2) shows a clear maturation pattern in the literature between 2019 and 2024. Three tactics—Impact, Initial Access, and Execution—dominate the growth trajectory, rising steadily after 2021 and reaching sharp peaks in 2024 (Impact = 30, Initial Access = 23, and Execution = 24). This reflects an increased focus on high-visibility behaviours such as denial-of-service and exploitation of external-facing assets, which are easier to reproduce in experimental datasets.

Command and Control (23 in 2024) and Reconnaissance (23 in 2024) also display consistent upward trends, suggesting stronger coverage of adversarial activities both before and after gaining access. Credential Access grows more moderately but still peaks at 18 in 2024, highlighting rising attention to authentication-based threats.

By contrast, stealthier or deeper-compromise tactics such as Privilege Escalation (5 in 2024), Persistence (10 in 2024), and Exfiltration (8 in 2024) remain under-represented throughout the period. While they do show modest increases, they never surpass 10 annual instances, indicating that datasets and studies still place less emphasis on long-term intrusion behaviours compared to disruptive, high-impact attacks. Similarly, Resource Development appears sporadically and with negligible representation.

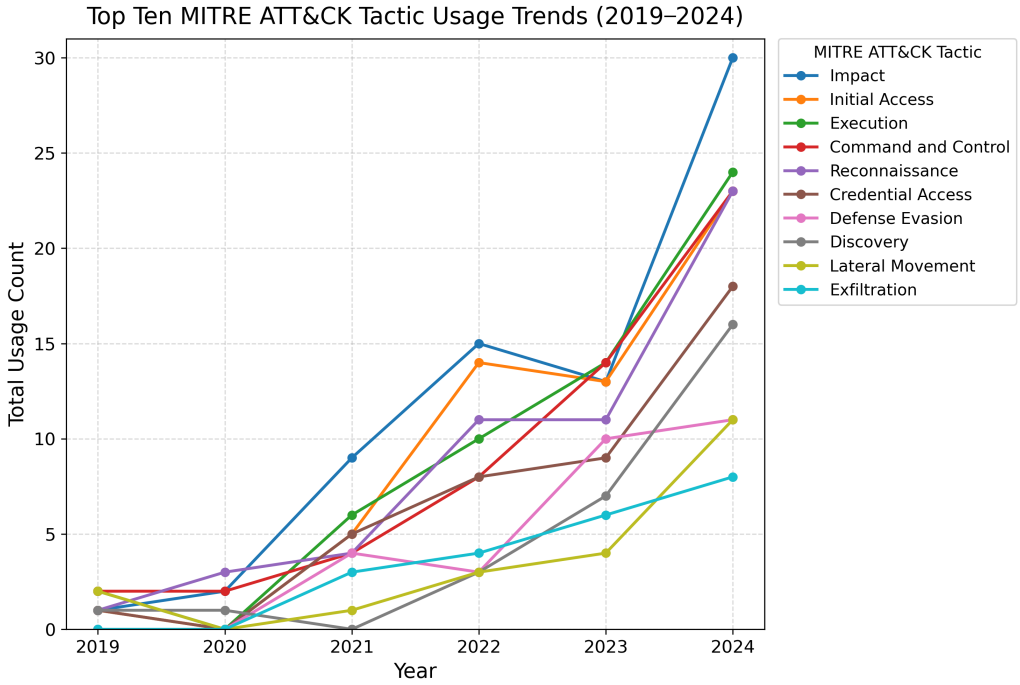

Figure 2. Trend of the ten most frequently represented MITRE ATT&CK tactics across the reviewed studies from 2019 to 2024. Each line shows the yearly frequency with which a given tactic appeared in the reviewed literature.

At the technique level (Figure 3), the dominance of denial-of-service behaviours is unmistakable. Network Denial of Service rises from just 1 instance in 2019 to 27 in 2024, while Endpoint Denial of Service follows a similar trajectory, reaching 20 in 2024. These two techniques together form the bulk of observed attacks, underlining the field's reliance on availability-disruption scenarios.

Other techniques focused on exploiting the external perimeter—Exploit Public-Facing Application (18 in 2024) and Active Scanning (19 in 2024)—also show strong upward trends. Meanwhile, Gather Victim Host Information peaks at 16 in 2024, reflecting a growing but secondary emphasis on reconnaissance and victim profiling. Because 2025 records are incomplete, they are excluded from this longitudinal trend analysis to avoid distorting the observed pattern.

Taken together, these patterns reveal that between 2019 and 2024, cyberattack research and datasets have increasingly centred on high-impact, high-detection-rate behaviours—particularly denial-of-service, scanning, and web-facing exploitation. This emphasis likely reflects the relative ease of simulating and labelling such behaviours in testbeds. However, stealth-oriented tactics such as persistence, lateral movement, and exfiltration remain consistently under-represented, pointing to a persistent gap between experimental focus and the realities of advanced adversarial operations. Addressing this gap requires richer datasets that capture the "low-and-slow" phases of intrusions, which are often decisive in real-world incidents but challenging to model experimentally.

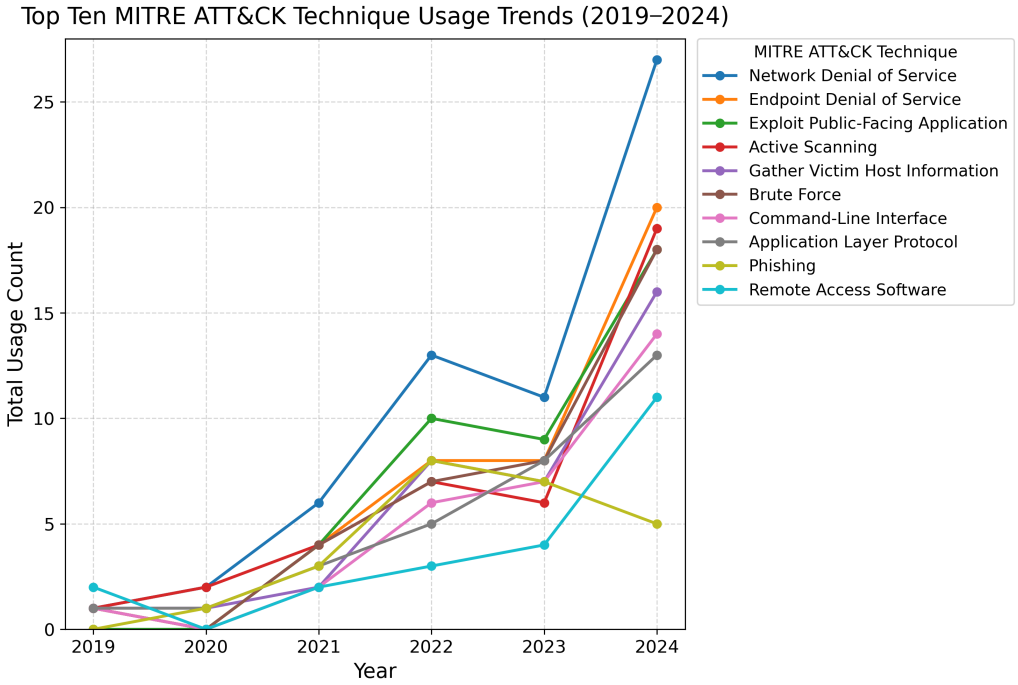

Figure 3. Trend of the ten most frequently represented MITRE ATT&CK techniques across the reviewed studies from 2019 to 2024. Each line shows the yearly frequency with which a given technique appeared in the reviewed literature.

3.4. Co-Occurrence of ATT&CK Tactics and Techniques

Examining the co-occurrence of MITRE ATT&CK tactics and techniques across the 99 reviewed papers reveals which adversarial behaviours most frequently appear together within the same study (Figures 4 and 5). Such analysis highlights patterns of behavioural clustering, where specific adversarial actions are operationally dependent or strategically complementary.

These co-occurrences should be interpreted as patterns of joint coverage within the reviewed studies rather than as direct evidence of temporal sequencing within individual real-world incidents.

At the tactic level, the strongest association is between Execution and Command and Control (50 co-occurrences), illustrating that once a foothold is established, adversaries often execute commands or payloads directly through their C2 infrastructure. This reflects the operational reality that C2 channels are not passive but serve as the primary conduit for directing malicious activity.

Impact co-occurs heavily with Reconnaissance (47) and Initial Access (46), suggesting a frequent attack progression: adversaries probe and map the target environment, establish access, and then deliver disruptive or destructive payloads. Similarly, Execution pairs strongly with Impact (42) and Reconnaissance (42), reinforcing the tight coupling between operational actions and observable consequences.

Reconnaissance itself is present in seven of the top-ten tactic pairings, underscoring its role as both a preparatory and supporting behaviour across different attack stages. Meanwhile, deeper-compromise tactics such as Persistence, Privilege Escalation, and Exfiltration appear less frequently in high-ranking pairs, reflecting their secondary treatment in experimental and dataset-driven research.

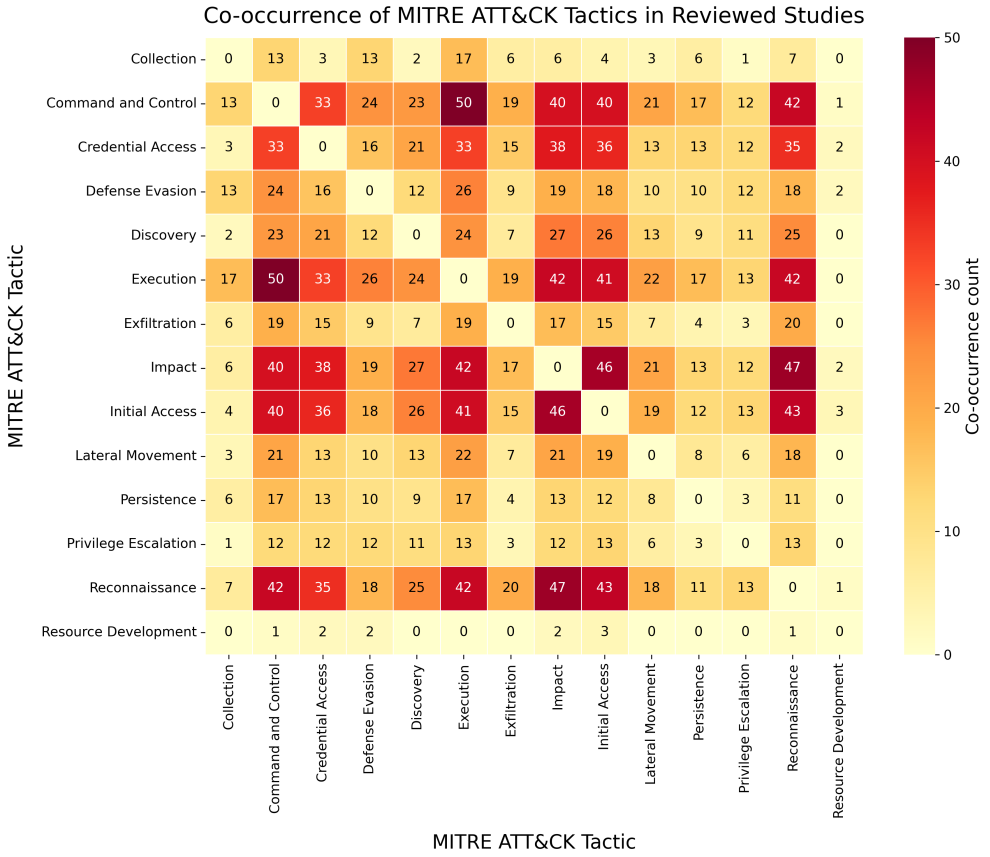

Figure 4. Co-occurrence heatmap of MITRE ATT&CK tactics across the 99 reviewed studies. Each cell reports the number of papers in which the corresponding pair of tactics co-appeared within the same study. Darker cells indicate higher co-occurrence frequencies, and diagonal values are set to zero because self-co-occurrence is suppressed.

At the technique level, the dominance of denial-of-service behaviours is clear. The strongest pairing is Network Denial of Service with Endpoint Denial of Service (43), highlighting how many DoS scenarios combine both network-level saturation and endpoint disruption for maximum effect. Closely following are Network Denial of Service with Exploit Public-Facing Application (41) and Network Denial of Service with Active Scanning (39), reflecting a pattern where service exploitation and reconnaissance are embedded within disruptive campaigns.

Techniques linked to credential compromise and direct execution—such as Brute Force with Command-Line Interface (30) and Brute Force with Active Scanning (32)—appear frequently in mid-ranked pairings. This points to an integrated attack workflow: reconnaissance to identify weak points, brute-forcing to obtain credentials, and CLI-based execution to solidify control.

By contrast, stealthier or persistence-related techniques, such as Phishing and Remote Access Software or Remote Access Services, show weaker associations overall, appearing less prominently in the top co-occurrence patterns. This reflects the broader trend across the literature, where high-impact, high-visibility behaviours are emphasised, while subtle post-compromise techniques remain underexplored.

Overall, co-occurrence analysis reinforces the conclusion that cyberattack research between 2019 and 2025 has centred on externally visible, easily simulated behaviours (DoS, scanning, public-facing exploitation). The lack of strong co-occurrence signals for stealth-oriented tactics and techniques highlights a persistent gap between academic research and real-world adversary operations, where multi-stage, low-and-slow intrusions are often decisive.

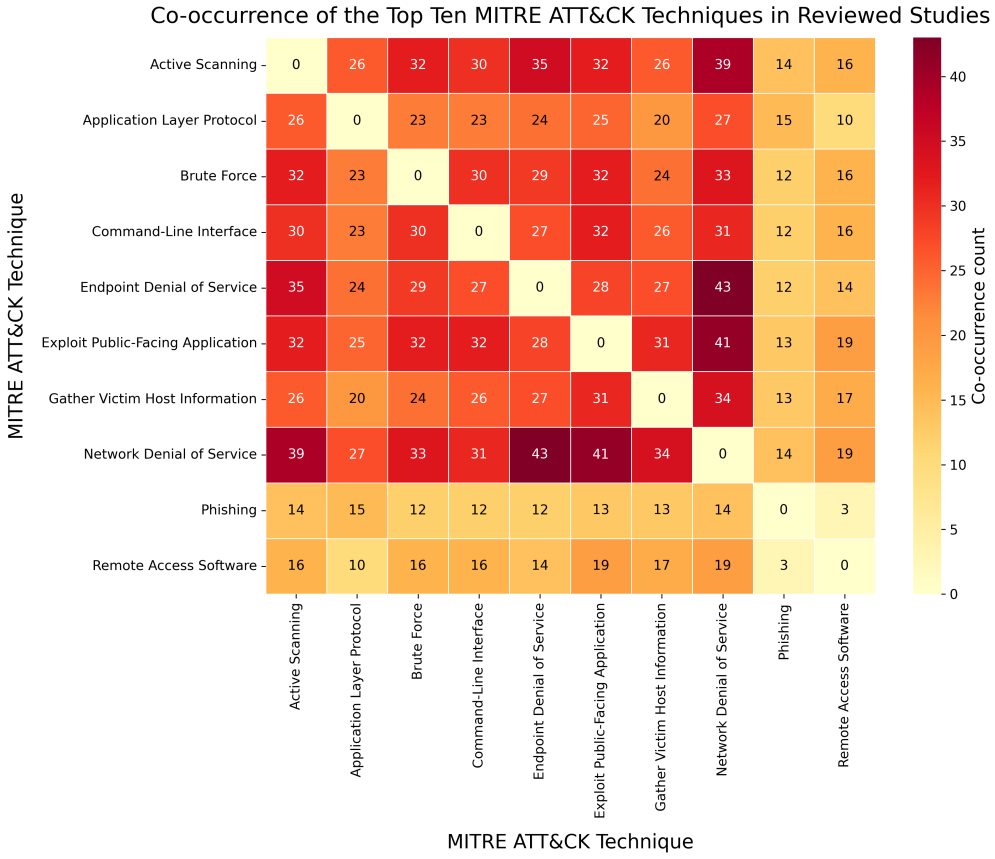

Figure 5. Co-occurrence heatmap of the ten most frequently represented MITRE ATT&CK techniques across the 99 reviewed studies. Each cell reports the number of papers in which the corresponding pair of techniques co-appeared within the same study. Darker cells indicate higher co-occurrence frequencies, and diagonal values are set to zero because self-co-occurrence is suppressed.

3.5. Breadth of ATT&CK Coverage per Paper

To better understand the breadth of adversarial coverage in the literature, both the number of distinct MITRE ATT&CK tactics and techniques addressed in each of the 99 papers were analysed. The results are summarised in Tables 7 and 8.

Table 7. Distribution of tactics per paper (total papers: 99).

Number of Tactics	Number of Papers	Percentage
1	19	19.19%
2	8	8.08%
3	11	11.11%
4	11	11.11%
5	7	7.07%
6	8	8.08%
7	10	10.10%
8	9	9.09%
9	7	7.07%
10	6	6.06%
11	1	1.01%
12	2	2.02%

Of the 99 papers, 19 (19.2%) concentrated exclusively on a single tactic. These were often highly specialised studies targeting niche attack behaviours (e.g., denial-of-service, phishing, or command-and-control detection). While such targeted research yields deep insights into specific adversary techniques, it inherently limits coverage of the wider attack

lifecycle. For instance, a DoS study examining only Impact overlooks how attackers gain initial access—a critical gap for defensive planning.

At the other end, only 16 papers (16%) addressed nine tactics or more, highlighting that comprehensive coverage of the ATT&CK framework remains uncommon. Notably, around half of the reviewed studies (49.5%) engaged with four tactics or fewer, reflecting either a deliberate focus on niche adversarial behaviour, dataset limitations, or methodological simplicity. This suggests that while breadth of coverage is possible, many works prioritise depth in specific phases of the attack lifecycle.

Table 8. Distribution of techniques per paper (total papers: 99).

Number of Techniques	Number of Papers	Percentage
1	14	14.14%
2	11	11.11%
3	6	6.06%
4	12	12.12%
5	5	5.05%
6	4	4.04%
7	4	4.04%
8	11	11.11%
9	1	1.01%
10	1	1.01%
11	7	7.07%
12	5	5.05%
13	2	2.02%
16	1	1.01%
17	2	2.02%
18	3	3.03%
19	1	1.01%
22	2	2.02%
23	2	2.02%
24	1	1.01%
26	3	3.03%
27	1	1.01%

The analysis of techniques reveals a similar skew. Fourteen papers (14.1%) employed only a single ATT&CK technique, while a comparable proportion (12.1%) examined four techniques. A small subset of highly comprehensive works mapped over 20 techniques, with the most extensive study covering 27. These broad-spectrum studies are rare (9%), but they demonstrate the potential of multi-technique modelling to capture the complexity of real-world adversarial behaviour.

Overall, the distributions for both tactics and techniques confirm a trend towards narrower research scopes: roughly half of the studies explored only a limited set of adversarial behaviours, while only a minority attempted wide coverage across the attack lifecycle. This reflects a balance in the field between depth (specialised studies) and breadth (holistic approaches).

3.6. Cyberattack-Side Key Findings and Research Gaps

The cyberattack analysis reveals a structurally imbalanced evidence base. The literature is concentrated on ATT&CK tactics such as Impact, Initial Access, Execution, Reconnaissance, and Command and Control, while post-compromise phases such as Persistence, Privilege Escalation, Lateral Movement, Collection, and Exfiltration remain comparatively under-represented. This pattern should not be interpreted as evidence that the latter tactics are less important in practice. Rather, it reflects the fact that externally visible, short-horizon,

and high-signal attacks are easier to reproduce in laboratory settings, easier to label in public datasets, and therefore easier to benchmark.

This concentration has two important implications for the field. First, model development is being guided disproportionately by attacks that are operationally loud and experimentally convenient, such as DoS/DDoS, scanning, and web-facing exploitation. As a result, the literature gives stronger coverage to disruption-centric threats than to the stealthier phases that often determine the success, persistence, and severity of real intrusions. Second, the relative scarcity of studies covering broader ATT&CK trajectories suggests that many detection pipelines are being evaluated on isolated attack moments rather than on full adversarial workflows. This limits their value for understanding how AI-based cyber defense performs against coordinated, multi-stage campaigns.

The main research gap exposed by this section is therefore not simply the absence of certain attack labels, but the lack of sustained empirical attention to low-and-slow post-compromise behaviour. Future work should prioritise datasets and experimental designs that better capture persistence, lateral movement, privilege escalation, exfiltration, and other ATT&CK phases that remain difficult to observe but are central to real-world adversary operations.

This attack-side imbalance becomes even more significant when interpreted alongside the model and dataset distributions examined in the following sections.

4. Machine Learning Methods Across Reviewed Studies

In this section, the machine learning (ML) approaches employed in the reviewed studies are examined through the normalization and grouping rules defined in Section 2.6. Across the 99 included studies, 143 unique raw ML method labels were identified. Because the literature uses heterogeneous naming conventions, closely related algorithms and architectural variants were first standardized into comparable analytical units before higher-level aggregation. For example, decision-tree derivatives were grouped under Tree-Based Models, autoencoder variants under Autoencoders, and recurrent sequence models under their corresponding sequence-learning families. This two-stage coding process allowed this review to compare methodological patterns without losing the connection to the original study terminology.

Following this initial classification, this study adopted the high-level taxonomy proposed by Emmert-Streib et al. [46], which categorises machine learning approaches into Classical Machine Learning Models, Deep Learning Models, and Learning Paradigms. This framework encompassed the majority of the identified method groups. However, in line with von Rueden et al. [47], an additional category—Hybrid and Ensemble Learning—was introduced, reflecting the importance of integrating multiple models or knowledge sources to enhance performance and robustness.

Two remaining categories, Optimization Algorithms (e.g., Stochastic Gradient Descent, Adam, and RMSProp) and Explainable AI (XAI) methods (e.g., SHAP; LIME), did not directly align with these existing groups. For the purposes of this review, explainable AI techniques were incorporated into the Hybrid, Ensemble, and Explainable Methods category, acknowledging that interpretability methods are often developed and applied alongside hybrid or ensemble approaches [48]. Optimization algorithms were grouped under the Learning Paradigms and Optimization category because they function primarily as training and performance-enhancement mechanisms rather than as standalone detection models [46].

4.1. ML Method Frequency Analysis

The most frequently reported individual method across the 99 reviewed papers is the Random Forest (RF) algorithm, appearing in 22 studies. RF's popularity in cybersecurity research is linked to its robustness and strong performance on tabular, high-dimensional datasets [49]. Long Short-Term Memory (LSTM) networks appear in 18 studies, while Convolutional Neural Networks (CNNs) appear in 15. The frequent use of LSTM and CNN reflects the growing application of deep learning models to sequential (e.g., network traffic) and spatial/temporal data representations in intrusion detection. Support Vector Machines (SVMs) (15 studies) and Decision Trees (DTs) (12 studies) also remain widely used.

Table 9 presents the frequency of machine learning models across the main categories and their subcategories. Note that one paper may employ multiple models, leading to a total frequency larger than the number of reviewed studies (99).

Table 9. Distribution of machine learning models across main categories and subcategories.

Main Category	Count	Subcategories	Count
Deep Learning Models	72	LSTM & Variants	27
		Feedforward Networks & Variants	24
		Core CNN Architectures	22
		Transformer-Based Models	9
		Autoencoders	8
		Specialized/Advanced CNNs	8
		GAN & Variants	5
		GRU & Variants	4
		Graph Neural Networks (GNN)	3
Hybrid, Ensemble & Explainable	46	Ensemble Learning Methods	29
		Boosting	16
		Hybrid Architectures	13
		Interpretability	4
Classical Machine Learning Models	34	Statistical Models	17
		SVM & Variants	17
		Tree-Based Models	16
		Bayesian Models	9
		Clustering	6
		Hidden Markov Models	1
Learning Paradigms and Optimization	18	Optimization Algorithms	11
		Learning Paradigms & Feature Selection	7

Accordingly, the prevalence of deep learning and ensemble methods should be interpreted in light of the data regimes most commonly studied. Their dominance reflects not only methodological strength, but also the structure of available benchmark tasks.

At the main category level, Deep Learning Models dominate the literature, appearing in 72 studies, followed by Hybrid, Ensemble, and Explainable Methods in 46 studies and Classical Machine Learning Models in 34 studies. Learning Paradigms and Optimization represent the smallest share, appearing in 18 studies, typically appearing as supporting techniques rather than standalone detection frameworks. The prominence of deep learning is consistent with its ability to automatically extract complex features from raw data, making it highly effective for processing the high-volume and high-velocity data streams typical of cybersecurity contexts [50,51].

At the subcategory level, Ensemble Learning Methods were the most common (29 papers), reflecting the strong emphasis on improving accuracy and reducing variance through model combination—particularly important in imbalanced intrusion detection datasets [52]. Within deep learning, LSTM and Variants (27 papers), Feedforward Networks

(24 papers), and Core CNN Architectures (22 papers) were the leading approaches, demonstrating the relevance of sequence modelling and convolution-based feature extraction for time-series and packet-level cybersecurity data.

Classical methods still retain a significant role, with Statistical Models and SVM variants each appearing in 17 studies, showing that interpretable and well-established algorithms remain valuable, especially in settings requiring transparency or where computational resources are limited.

The results also show the importance of hybridization: Boosting (16 papers) and Hybrid Architectures (13 papers) frequently appeared, often combining deep learning with traditional models to exploit complementary strengths. Meanwhile, optimization techniques (11 papers) and explainability-focused methods (4 studies) were less frequent, indicating that while they are recognized as critical areas, they remain underexplored in the current cybersecurity ML literature.

4.2. ML Method Trends (2019–2024)

As the data for 2025 are incomplete, including them would not provide an accurate reflection of the overall trend. Consequently, the trend analysis presented here is limited to the period from 2019 to 2024. As illustrated in Figure 6, the adoption of machine learning (ML) methods in cybersecurity shows dynamic growth patterns between 2019 and 2024. Among the main categories, Deep Learning Models show the most consistent rise, starting from a single study in 2019 and reaching a peak of 26 in 2024. This trend highlights both the rapid adoption of deep architectures in cybersecurity research and the possible transition toward more integrated frameworks in recent years.

Hybrid, Ensemble, and Explainable Methods also demonstrate a steady increase, with no recorded studies in 2019 but 17 studies in 2024. Their emergence reflects increasing recognition of the benefits of combining multiple models and prioritising interpretability, which is particularly relevant in security-sensitive contexts.

Classical Machine Learning Models maintain a relatively stable presence across the review period, peaking at nine studies in 2022 and sustaining eight studies during 2023–2024. This suggests that while traditional algorithms such as SVMs and decision trees remain relevant, they are increasingly supplemented by more complex architectures. Learning Paradigms and Optimization methods appear later in the timeline, from 2022 onward, gradually increasing to eight studies in 2024 and indicating a growing interest in adaptive learning strategies and optimization-based improvements.

At the subcategory level, as shown in Figure 7, the top ten approaches reveal more nuanced patterns. Ensemble Learning Methods steadily gain traction from 2020, reaching their highest point in 2024 with 11 studies, reinforcing their reputation for robustness across heterogeneous cyber datasets. LSTM and Variants show sharp growth from 2021 onward, peaking at 10 studies in 2024, consistent with their strength in modelling sequential attack patterns.

Feedforward Networks and Variants demonstrate continuous adoption, rising from one study in 2020 to eight studies in 2023 before stabilising, confirming their value as general-purpose deep models. Similarly, Core CNN Architectures surge in 2024, with nine studies, reflecting renewed interest in image-like cybersecurity representations. Classical approaches such as SVM and Variants, Tree-Based Models, and Statistical Models experience moderate but significant usage during 2022–2024, often in hybrid or explainable frameworks.

A notable observation is the trajectory of Boosting Methods, which rise steadily to five studies in 2024, suggesting sustained relevance as a performance-enhancement strategy in evolving attack settings. Hybrid Architectures maintain a low but stable presence across

the years, while Optimization Algorithms display a late surge in 2024, with six studies, indicating growing experimentation with parameter tuning and adaptive optimization for cybersecurity tasks.

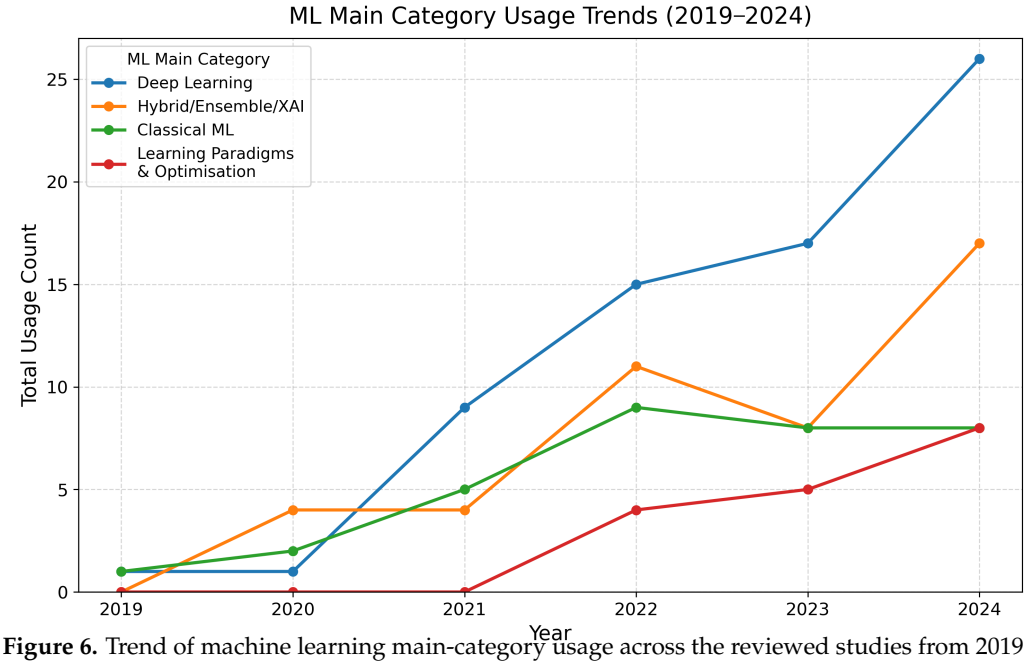

Figure 6. Trend of machine learning main-category usage across the reviewed studies from 2019 to 2024. Each line shows the yearly frequency with which a given ML main category appeared in the reviewed literature.

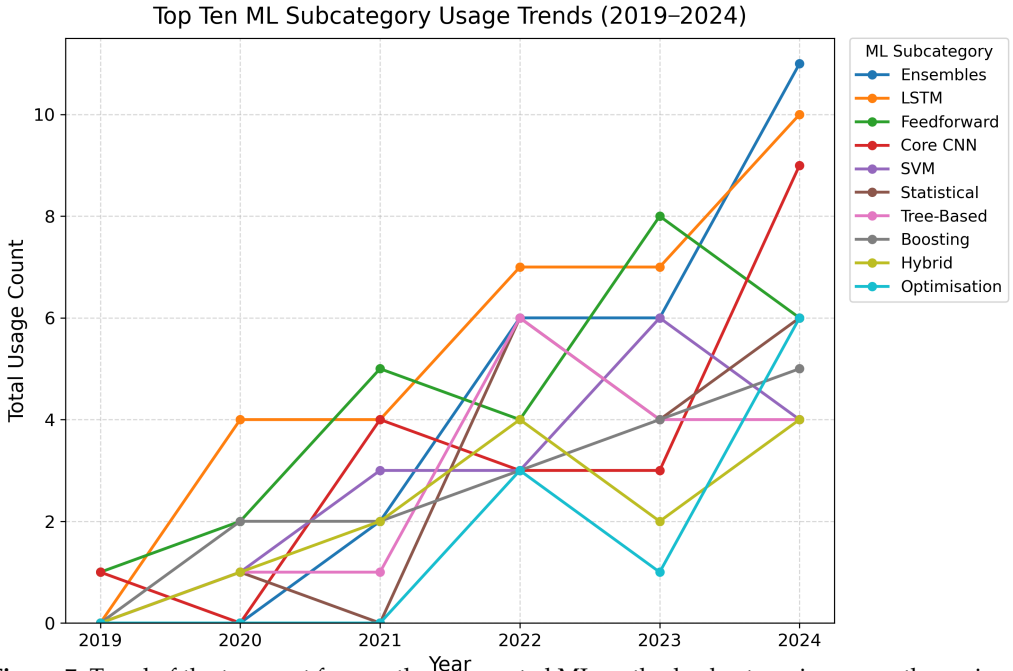

Figure 7. Trend of the ten most frequently represented ML method subcategories across the reviewed studies from 2019 to 2024. Each line shows the yearly frequency with which a given subcategory appeared in the reviewed literature.

4.3. Co-Occurrence of Main Categories and Subcategories

At the main category level, as shown in Figure 8, the strongest co-occurrence is observed between Deep Learning Models and Hybrid, Ensemble, and Explainable Methods (26 instances, 26.26%). This reflects a consistent trend in the literature of augmenting deep neural architectures with ensemble strategies to improve accuracy, robustness, and

generalization—especially in adversarial or imbalanced cybersecurity contexts. The second most frequent pairing is between Classical Machine Learning Models and Hybrid, Ensemble, and Explainable Methods (21 instances, 21.21%), suggesting that ensembles remain a vital strategy to leverage the diversity of traditional models such as decision trees, SVMs, and statistical techniques. Classical Machine Learning Models and Deep Learning Models co-occurred 18 times (18.18%), often in comparative evaluations or hybrid frameworks designed to combine interpretability and efficiency with the representation learning power of deep networks. Finally, Learning Paradigms and Optimization methods appeared as secondary but non-negligible companions to Deep Learning Models (14 instances), indicating growing interest in reinforcement learning, optimization algorithms, and training strategies to fine-tune model performance.

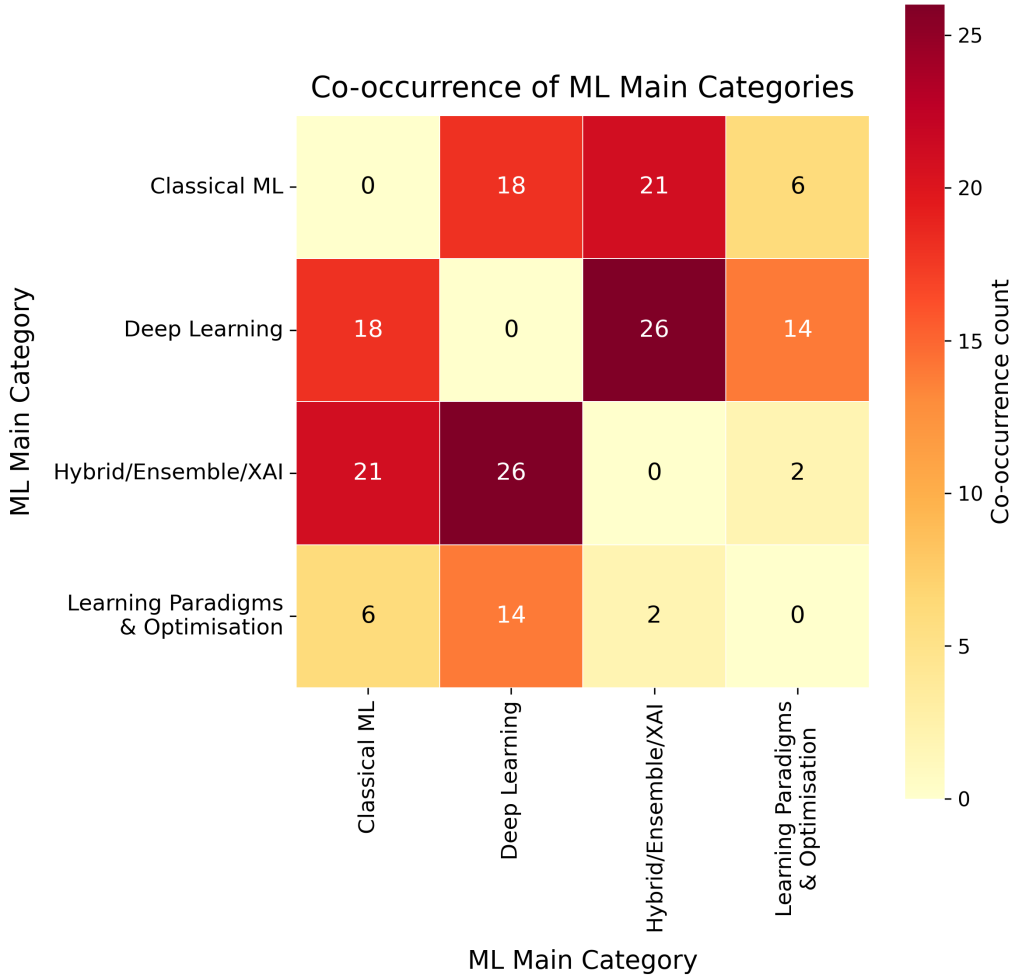

Figure 8. Co-occurrence heatmap of ML main categories across the reviewed studies. Each cell reports the number of papers in which the corresponding pair of ML main categories co-appeared within the same study. Darker cells indicate higher co-occurrence frequencies, and diagonal values are set to zero because self-co-occurrence is suppressed.

These co-occurrences should be interpreted as patterns of joint methodological use within the reviewed studies rather than as direct evidence that the paired methods were always integrated into a single deployable detection pipeline.

At the subcategory level, the co-occurrence heatmap in Figure 9 highlights several important patterns. The strongest link is observed between Tree-Based Models and Ensemble Learning Methods (12 instances, 12.12%), reaffirming the widespread adoption of ensemble strategies such as bagging and boosting applied to decision-tree families (e.g., Random Forest; XGBoost). Ensemble Learning Methods also co-occur frequently with SVM

and Variants (nine instances) and with Statistical Models (nine instances), underlining a broader trend of combining interpretable, mathematically grounded models with ensemble techniques to balance transparency and predictive power.

On the deep learning side, a notable co-occurrence is seen between LSTM and Variants and Core CNN Architectures (10 instances, 10.10%). This indicates the prevalence of hybrid temporal–spatial deep models, where CNNs extract structural or spatial patterns (e.g., from network flows or packet payloads) while LSTMs capture sequential dependencies, making them well suited, at least conceptually, to attacks with both spatial and temporal signatures. Feedforward Networks and Variants also frequently pair with LSTM and Variants (seven instances) and Ensemble Learning Methods (six instances), highlighting their role as baseline or complementary components in more complex architectures.

Boosting techniques are another recurring theme, with frequent co-occurrence alongside Tree-Based Models (seven instances), SVM and Variants (six instances), and Statistical Models (five instances). This suggests that boosting remains a key strategy to elevate the performance of both classical and statistical learners in cybersecurity tasks. Finally, Optimization Algorithms, though less frequently paired overall, appear most often in combination with LSTM and Variants (six instances), reflecting ongoing research into improving deep sequence models through enhanced optimization strategies.

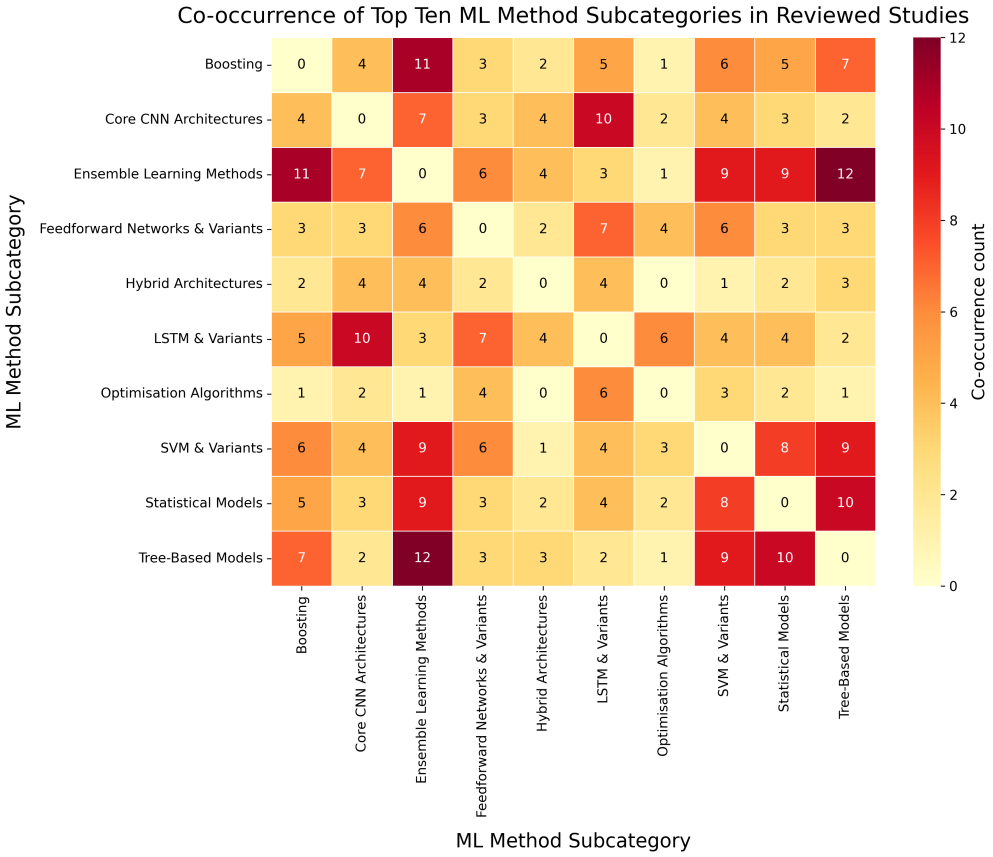

Figure 9. Co-occurrence heatmap of the ten most frequently represented ML method subcategories across the reviewed studies. Each cell reports the number of papers in which the corresponding pair of ML subcategories co-appeared within the same study. Darker cells indicate higher co-occurrence frequencies, and diagonal values are set to zero because self-co-occurrence is suppressed.

Overall, these patterns suggest a field that is moving increasingly towards integrative strategies: ensembles to improve robustness, hybrid architectures that blend classical and deep models, and optimization-driven refinements for specialized deep learning subcategories. This co-occurrence analysis highlights not only the dominance of ensemble

thinking across both classical and deep paradigms but also a gradual convergence between spatial and temporal deep networks, which appears especially promising for detecting multi-faceted cyber threats.

4.4. Method Breadth per Paper

Table 10 presents the number of different methods employed within individual papers. A notable 22.2% of studies applied only a single method, indicating a focused, single-technique approach—often for benchmarking or proof-of-concept purposes. The most common configuration, however, involved two methods (25.3%), reflecting a tendency to either pair complementary algorithms for improved robustness and performance, or to compare novel techniques against established baselines. Only a small number of studies adopted extensive methodological portfolios, with as many as seven (3.0%), eight (1.0%), eleven (1.0%), or even thirteen (1.0%) methods tested within a single paper. Such high counts are typically associated with large-scale comparative studies or advanced ensemble frameworks that aim for comprehensive performance evaluations.

Table 10. Distribution of machine learning methods in papers by count.

Number of ML Methods Used	Number of Papers	Percentage
1	22	22.2%
2	25	25.3%
3	17	17.2%
4	13	13.1%
5	7	7.1%
6	9	9.1%
7	3	3.0%
8	1	1.0%
11	1	1.0%
13	1	1.0%

At the main category level, the analysis (Table 11) shows that the majority of papers restricted themselves to a limited number of categories. Specifically, 43 papers (43.4%) relied on only one main category, while another 42 papers (42.4%) combined two categories. This suggests that most research either focused narrowly on one paradigm (e.g., tree-based models, neural networks, or probabilistic methods) or paired two distinct categories to balance interpretability and predictive strength. Only 13 papers (13.1%) employed three categories, and just 1 study (1.0%) explored four, highlighting the rarity of highly diversified category-level integrations. Overall, more than 85% of the studies used one or two main categories, underscoring the preference for simplicity or targeted methodological design.

Table 11. Main categories used per paper.

Number of ML Main Categories Used	Number of Papers	Percentage
1	43	43.4%
2	42	42.4%
3	13	13.1%
4	1	1.0%

At the subcategory level, shown in Table 12, a similar but more nuanced trend emerges. Around one-quarter of the papers (25.3%) limited themselves to a single subcategory, while the largest share (37.4%) combined two subcategories. Taken together, more than 60% of studies relied on only one or two subcategories, indicating a preference for simpler designs or narrowly targeted approaches to cyberattack detection. More complex integrations also

appear, with 12.1% of papers using three subcategories, another 12.1% using four, and smaller proportions using five (7.1%), six (4.0%), or seven (2.0%). These higher counts are characteristic of ensemble and hybrid designs, where multiple algorithm families contribute to a unified detection system.

Table 12. Subcategory methods used per paper.

Number of ML Subcategories Used	Number of Papers	Percentage
1	25	25.3%
2	37	37.4%
3	12	12.1%
4	12	12.1%
5	7	7.1%
6	4	4.0%
7	2	2.0%

4.5. ML Method-Side Key Findings and Research Gaps

The machine learning results indicate that current AI-driven cyber defense research is shaped as much by benchmark structure as by algorithmic suitability. The strong presence of deep learning and ensemble methods suggests that the field is optimizing for settings in which large volumes of structured or semi-structured traffic data are available and where classification performance can be improved through feature learning, model combination, and robustness to class imbalance. This helps explain why Random Forest, LSTM, CNN, and hybrid ensemble architectures appear so frequently across the reviewed studies.

At the same time, the dominance of these families should not be read as proof that they are universally the best choices for cyber defense. Their concentration is partly a function of the tasks and datasets most commonly studied. High-volume intrusion datasets and overt attack classes naturally favour methods that perform well on tabular traffic features, sequential traces, or benchmark-style supervised learning problems. By contrast, method families that may be valuable for relational reasoning, adaptive defense, or sparse high-context environments—such as graph neural networks, reinforcement learning, and more systematic explainability-oriented approaches—remain comparatively rare. This indicates a mismatch between the richness of the available ML toolbox and the narrowness of its current application.

A second gap concerns operational realism. Classical models remain present not only because they are familiar, but because they are computationally efficient, interpretable, and easier to deploy in constrained environments. This suggests that real-world deployment constraints continue to matter, even when the literature favours high-capacity deep models. Future work should therefore move beyond benchmark-driven performance comparisons and examine which method families are most appropriate for specific ATT&CK stages, data regimes, interpretability requirements, and deployment contexts. In particular, more work is needed on methods suited to stealthy, long-horizon, and low-resource detection settings rather than only high-signal benchmark tasks.

These methodological patterns become more interpretable when considered alongside the benchmark concentration and category imbalances in the dataset landscape examined in the next section.

5. Datasets Across Reviewed Studies

The datasets used across the reviewed literature through the dataset-categorisation rules are defined in Section 2.6. In total, 96 unique datasets were identified across the 99 included studies. Each dataset was recorded using the name reported by the study authors

and then assigned to a refined taxonomy based on operational context, data source, and intended cybersecurity use case. This approach allowed datasets with similar names but different analytical roles to be categorised consistently, while also permitting reassignment when prior survey classifications did not match how the dataset was actually used in the reviewed study.

Selecting an appropriate dataset is central to AI-driven cybersecurity research, because the reliability and generalisability of machine learning models depend heavily on how well the underlying data represent operational threats and environments [8,53]. To support comparative analysis, datasets in this review were organised using prior survey frameworks [17,18] and then refined where necessary to better reflect their observed use in the reviewed studies.

On this basis, the datasets were divided into seven primary categories:

- NIDD (Network-based Intrusion Detection Datasets);
- IoT-NIDD (IoT-specific Network-based Intrusion Detection Datasets);
- S&P (Spam and Phishing Datasets);
- ICS (Industrial Control System Datasets);
- Insider Threat;
- Custom-Collected Datasets;
- Other (e.g., computer vision, NLP, or behavioural datasets).

Two refinements are particularly important. First, an additional IoT-NIDD category was introduced to distinguish IoT-specific intrusion datasets from broader NIDD resources, reflecting the distinct traffic characteristics and communication constraints of IoT environments. Second, some datasets were reassigned when their actual use in the reviewed studies differed from earlier survey classifications. For example, the IEEE Bus Distribution and Gas Pipeline datasets [54] were categorised under ICS because they were used in industrial or cyber–physical contexts in the reviewed literature. Likewise, computer-vision datasets such as Udacity, GTSRB, and UAVid were placed in the "Other" category because of their relevance to autonomous systems, smart mobility, and broader cyber–physical settings. The full dataset-level information used to support this taxonomy is provided in Appendix B.

5.1. Dataset Frequency Analysis

Table 13 summarizes the frequency of dataset usage across the refined dataset categories. Network-based Intrusion Detection Datasets (NIDDs) are by far the most frequently used category, appearing in 65 studies. They are followed by IoT-NIDD datasets, which appear in 31 studies, less than half the usage frequency of standard NIDD resources. Malware datasets are used in 20 studies, while Spam and Phishing (S&P) datasets appear in 17. At the lower end, ICS datasets appear in 12 studies, Other datasets in 7, and Insider Threat datasets in only 4.

This distribution indicates a strong research concentration around NIDD-, IoT-NIDD-, and malware-related benchmarks, while insider-threat and ICS-oriented data remain comparatively under-represented. In practical terms, this means that the literature is richest in domains where public benchmark datasets are already mature and easily reusable, and much thinner in settings where data collection is operationally sensitive, difficult to label, or harder to share.

Table 13. Dataset usage frequency by category.

Category	Frequencies	Most Used Datasets within Category
NIDD	65	CSE-CIC-IDS2017 (14), UNSW-NB15 (11), NSL-KDD (10), CSE-CIC-IDS2018 (4)
IoT-NIDD	31	ToN-IoT (5), EdgeIIoT 2023 (5), BoT-IoT (4), N-BaIoT (3)
Malware	20	Malimg (4), BIG 2015 (3)
S&P	17	Phishing Email Collection (4), PhishTank (3)
Custom-Collected Datasets	16	
ICS	12	SWaT dataset (3), Gas Pipeline (2)
Other	7	
Insider Threat	4	CERT Insider Threat (4)

The benchmark concentration is also visible within categories. In NIDDs, a small set of widely reused public datasets dominates, especially CSE-CIC-IDS2017, UNSW-NB15, and NSL-KDD. Similar concentration patterns appear within IoT-NIDD and S&P categories. This repeated reliance on a limited family of public datasets has methodological consequences: it may inflate apparent model maturity while masking sensitivity to domain shift, class imbalance, and attack distributions outside the benchmark setting.

5.2. Dataset Trends (2019–2024)

Because the 2025 data are incomplete, the trend analysis is limited to the period 2019–2024. Figure 10 shows the yearly usage patterns of the major dataset categories across the reviewed studies.

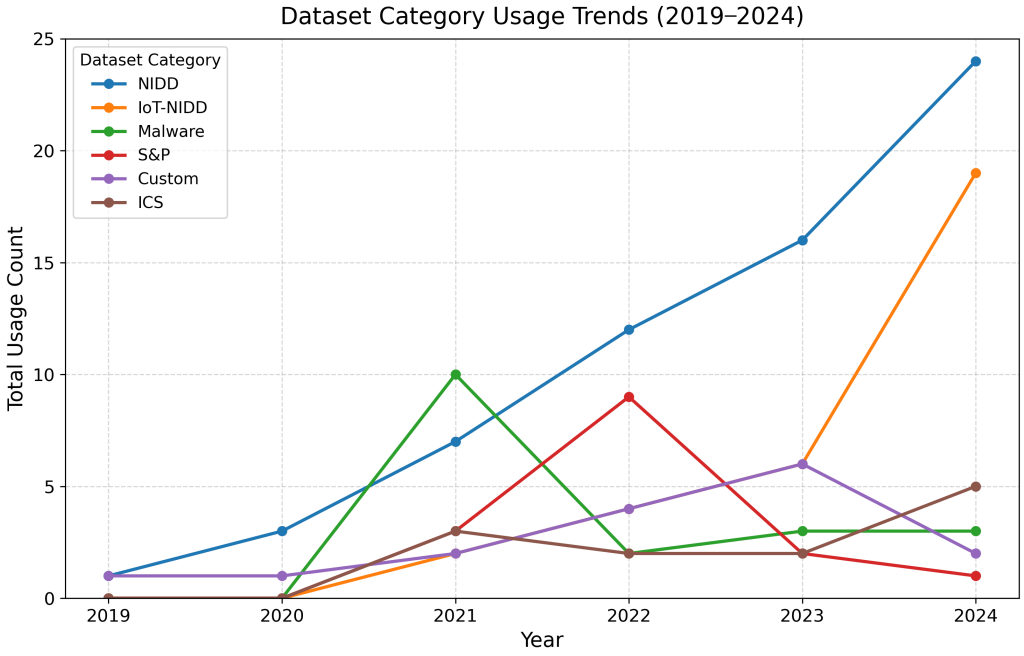

Figure 10. Trend of dataset-category usage across the reviewed studies from 2019 to 2024. Each line shows the yearly frequency with which datasets from a given category appeared in the reviewed literature.

NIDD exhibits the clearest and most sustained growth, increasing steadily from 2019 to a peak in 2024. This confirms its continued centrality in AI-based cybersecurity research and reflects the long-standing availability of reusable public intrusion-detection benchmarks. IoT-NIDD also grows gradually through 2023 before rising more sharply in 2024, suggesting

increasing academic and industrial attention to the security of connected and resource-constrained devices.

Malware datasets peak earlier, particularly in 2021, and then decline before stabilising. A similar pattern appears for S&P datasets, which rise to a peak in 2022 and then level off. These trajectories may indicate bursts of concentrated research attention followed by partial saturation, rather than the steady benchmark reuse observed in NIDD. By contrast, ICS datasets remain relatively sparse throughout the period, although they show modest growth toward 2024, indicating gradually increasing interest in critical infrastructure and cyber–physical security.

Custom-collected datasets increase through 2023 and then decline in 2024. This pattern may reflect the growing availability of standardized public datasets, which reduce the need for researchers to construct new datasets for every study. Taken together, the trend analysis suggests that dataset usage is not only shaped by threat priorities, but also by the relative availability, maturity, and reusability of public benchmark resources.

5.3. Co-Occurrence of Dataset Categories

Only 42 of the 99 reviewed studies used more than one dataset. For this subset, co-occurrence analysis was used to examine which dataset categories most often appeared together within the same paper (Figure 11). These co-occurrences should be interpreted as patterns of joint dataset use within the reviewed studies rather than as evidence that the paired categories are naturally equivalent or directly interoperable.

The strongest co-occurrence is between NIDD and IoT-NIDD datasets (eight studies). This suggests that researchers increasingly view conventional network intrusion benchmarks and IoT-specific intrusion datasets as complementary resources for evaluating detection methods across heterogeneous network environments. However, the relatively small number of such studies also indicates that broad comparative validation across conventional and IoT traffic settings remains limited.

ICS datasets co-occur less frequently, appearing three times with NIDD and two times with IoT-NIDD. These pairings suggest growing efforts to adapt or compare intrusion-detection methods across conventional, IoT, and industrial cyber–physical environments. At the same time, the low co-occurrence counts indicate that cross-domain evaluation remains uncommon, especially in settings that bridge ICS and IIoT security. This is an important gap, as industrial systems increasingly incorporate IoT-like devices, communication layers, and distributed sensing infrastructures.

Overall, the co-occurrence results suggest that multi-domain dataset evaluation is still relatively rare. Most studies continue to remain within a single benchmark family or a narrow pair of related categories, limiting evidence of how well published methods transfer across substantially different operational settings.

5.4. Dataset Breadth per Paper

Figure 12 shows the distribution of dataset usage per paper across the 99 reviewed studies. A substantial majority of studies rely on a small number of datasets: 58 papers use only one dataset, 21 use two, and 12 use three. Only nine studies incorporate four or more datasets.

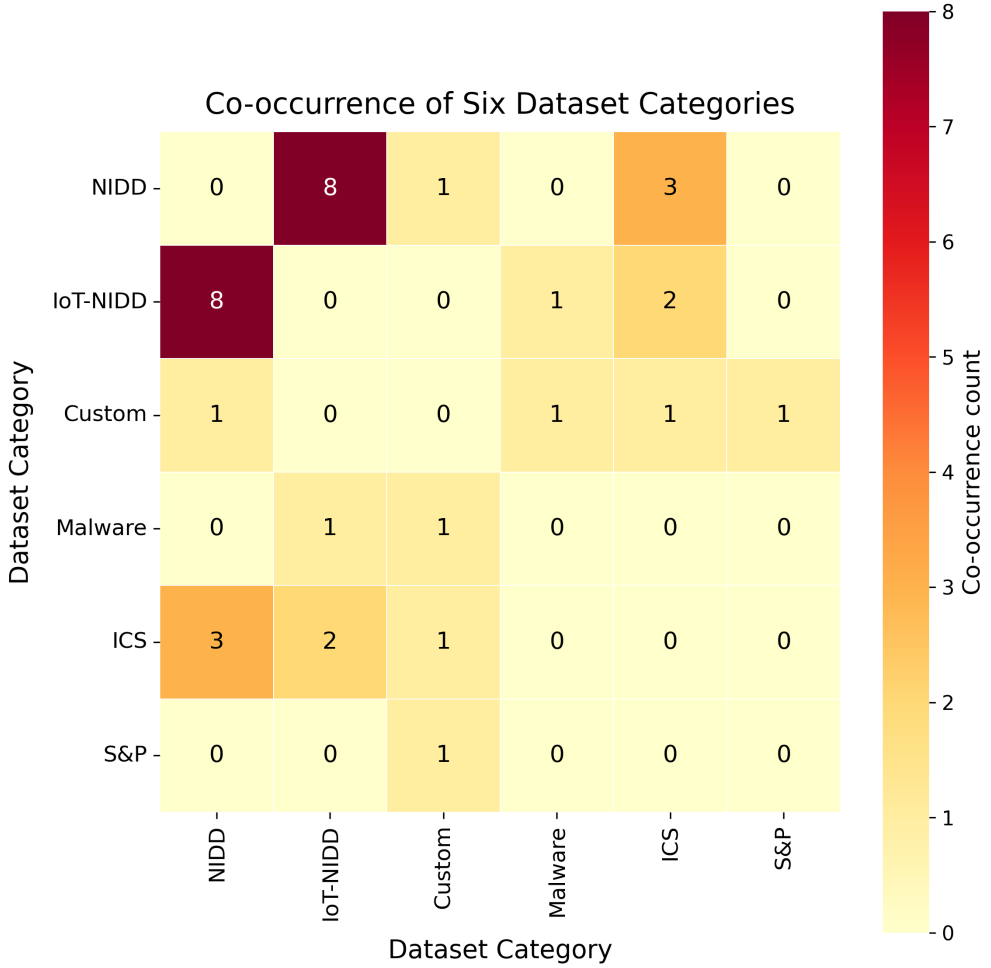

Figure 11. Co-occurrence heatmap of the six most frequently represented dataset categories across the reviewed studies. Each cell reports the number of papers in which the corresponding pair of dataset categories co-appeared within the same study. Darker cells indicate higher co-occurrence frequencies, and diagonal values are set to zero because self-co-occurrence is suppressed.

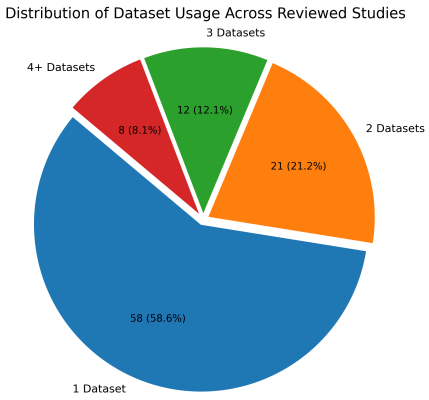

Figure 12. Distribution of dataset usage across the reviewed studies. Most studies relied on a single dataset, whereas a smaller proportion evaluated models using multiple datasets. Percentages are reported relative to the total number of reviewed papers represented in the figure.

This pattern indicates that narrow evaluation settings still dominate the literature. In some cases, such focused designs are appropriate because the study targets a specific attack type, domain, or detection scenario. However, the prevalence of single-dataset experiments also raises concerns about limited generalisability, as methods validated in one

benchmark environment may not transfer well to other organisations, traffic conditions, or operational domains.

The small number of studies using four or more datasets suggests that broad validation remains the exception rather than the norm. This is important because multi-dataset evaluation provides stronger evidence of robustness, transferability, and resistance to overfitting benchmark-specific properties. Overall, the distribution reinforces the concern that much of the current literature remains benchmark-convenient but only weakly validated across diverse real-world conditions.

5.5. Dataset-Side Key Findings and Research Gaps

The dataset analysis shows that current AI-based cyber defense research is strongly conditioned by benchmark availability. Network intrusion datasets dominate the literature, with repeated reliance on a relatively small set of public benchmarks such as NSL-KDD, UNSW-NB15, and CICIDS2017. This concentration creates an availability bias: researchers are more likely to study attack classes that are already well represented in public data and are correspondingly less likely to investigate attack types, environments, and operational contexts for which benchmark datasets are scarce.

This has important consequences for how published performance should be interpreted. Highly reported results on widely reused datasets do not necessarily indicate broad real-world robustness; in many cases, they may instead reflect repeated optimization against familiar data distributions, familiar feature structures, and familiar attack classes. The prevalence of single-dataset studies further reinforces this concern, as it limits evidence of transferability across domains, organisations, or operational settings. As a result, benchmark progress may be overstating methodological maturity while understating generalization risk.

The weakest parts of the current evidence base are also the most strategically important. ICS, insider-threat, and other under-represented categories remain sparse, while multi-dataset evaluations are still uncommon. This means that the literature is richest where public data are easiest to obtain, not necessarily where defensive need is greatest. Future work should therefore prioritise broader dataset diversity, clearer reporting of dataset provenance and limitations, and more systematic evaluation across multiple datasets and domains. Without this shift, progress in AI-driven cyber defense risks remaining benchmark-strong but deployment-fragile.

These dataset-side constraints become even more revealing when interpreted alongside the attack and model distributions in the cross-reference analysis presented in the next section.

6. Cross-Reference Analysis

6.1. Cyberattacks × ML Overview

This section examines how machine learning (ML) approaches are distributed across adversarial behaviours by cross-referencing MITRE ATT&CK tactics and techniques with ML model families. Heatmaps are used to make relative concentration patterns visible across attack phases, model categories, and dataset-linked problem settings. Unless otherwise noted, counts are aggregated at the paper level across the 2019–2025 review window and should be interpreted as patterns of research attention rather than direct evidence of model superiority or causal attack sequencing.

The heatmaps should be interpreted as showing relative concentration of research activity and paper-level co-occurrence, not comparative model performance.

6.2. Tactics × ML Main Categories

Figure 13 shows a pronounced concentration of studies on Impact, Execution, Initial Access, Command and Control (C2), and Reconnaissance when paired with Deep Learning Models. In particular, Impact (55), Execution (42), and Initial Access (40) form the hottest region, followed by C2 (38) and Reconnaissance (35). Hybrid, Ensemble, and Explainable Methods also show notable intensity for Impact (36), Execution (26), and Initial Access (28), indicating that ensembles are frequently adopted alongside deep architectures in critical, high-visibility phases.

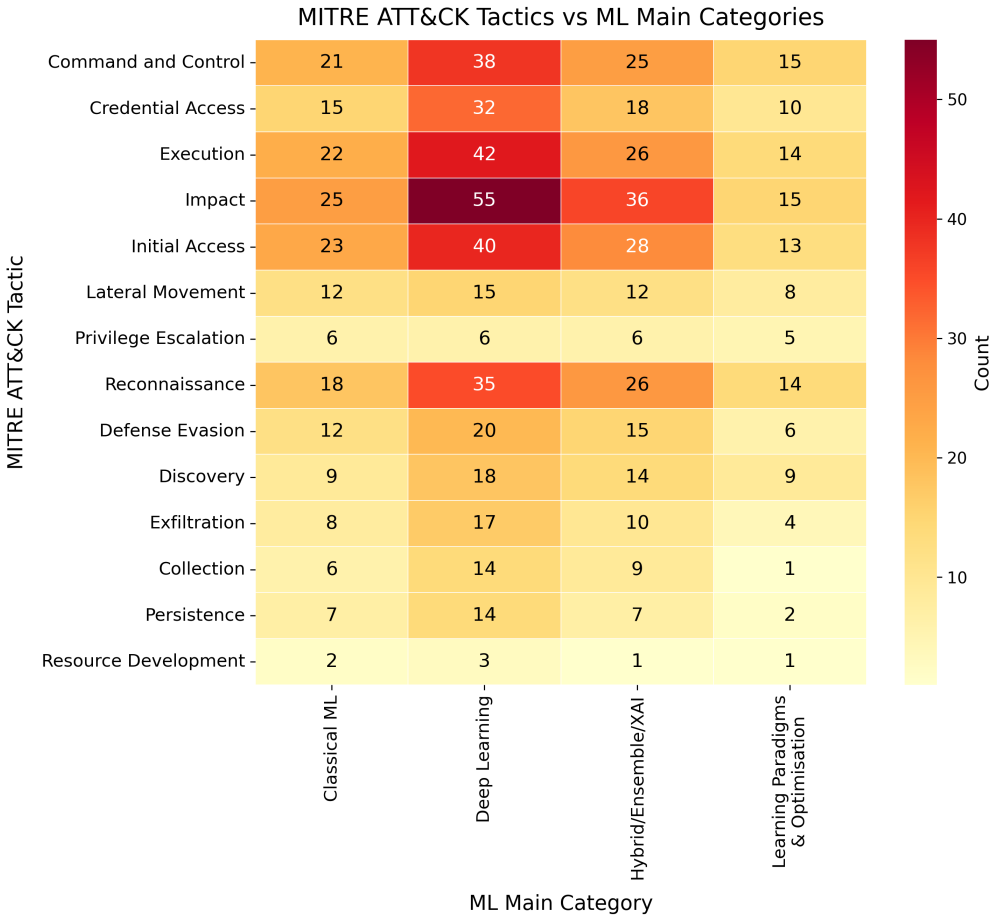

Figure 13. Heatmap of MITRE ATT&CK tactics versus ML main categories across the reviewed studies. Each cell reports the number of papers in which a given tactic was associated with a given ML main category. Darker cells indicate higher frequencies of co-occurrence in the literature.

Viewed jointly, these patterns suggest that benchmark availability is shaping both which attacks are studied and which ML paradigms appear most successful, thereby narrowing the practical meaning of published performance gains.

By contrast, Collection, Persistence, and Resource Development remain consistently cool across all main categories, with maximum counts of 14, 14, and 3, respectively. Privilege Escalation exhibits a relatively flat profile across families: Classical = 6; Deep = 6; Hybrid = 6; Learning Paradigms and Optimization = 5.

Interpretation of Tactics × ML Main Categories Patterns

(1) The strong emphasis on Impact and Execution aligns with datasets that are easier to generate and label, such as volumetric DoS/DDoS, overt service disruption, and malware execution traces. These settings favour deep models that learn complex, non-linear signatures from high-dimensional traffic or telemetry.

(2) The elevated presence of ensembles in Impact, Initial Access, and Execution suggests a stabilizing role for bagging/boosting and model stacking when class imbalance and dataset heterogeneity are pronounced.

(3) The consistently low-intensity regions, including Persistence, Collection, and Resource Development, likely reflect data scarcity and measurement challenges associated with stealthy, low-and-slow behaviours and therefore represent a substantive research gap.

6.3. Tactics × ML Subcategories

Figure 14 resolves the main-category picture into specific architectures. Three patterns stand out:

- **Ensemble Learning as a cross-tactic workhorse.** The hottest subcategory cells appear under Impact (23), Execution (19), and Initial Access (19), with consistently high values for C2 (17) and Reconnaissance (15). This indicates that ensembles are the default stabilizer across diverse data modalities, including netflows, logs, and host events, and across different class distributions.
- **Temporal models where sequences matter.** LSTM and Variants are prominent for Impact (21), Initial Access (16), Execution (16), C2 (15), and Reconnaissance (15), which is consistent with the appeal of sequence modelling in settings where temporal ordering, staged activity, or periodic communication patterns are expected.
- **CNNs for structured/spatialized representations.** Core CNN Architectures show meaningful intensity for Impact (16), Execution (15), Initial Access (13), Credential Access (10), and C2 (14). This is consistent with representations that transform packets, flows, binaries, or logs into grids, images, or embeddings exhibiting local spatial patterns.

Supporting families appear in narrower roles: SVM and Variants and Tree-Based Models maintain moderate presence across Execution, Initial Access, and Impact, while Optimization Algorithms and Statistical Models are frequently used as feature selection, calibration, or baseline components rather than as final detectors.

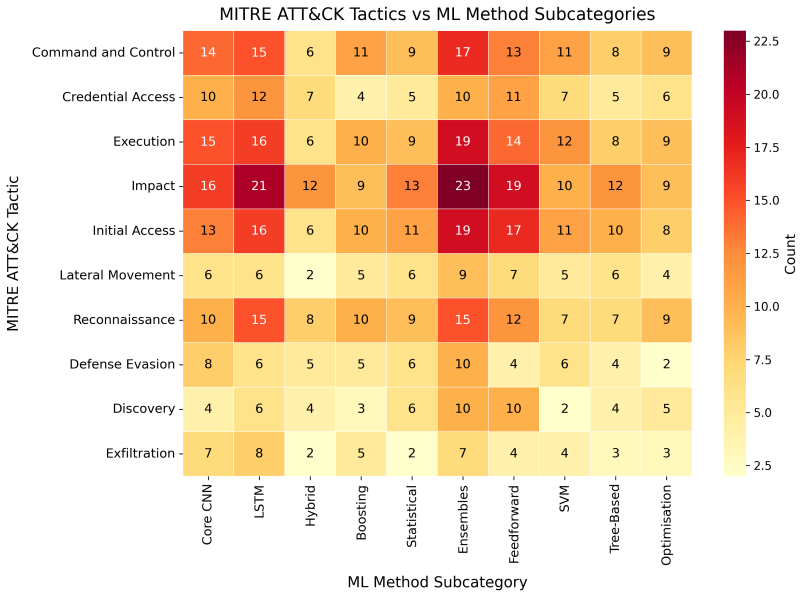

Figure 14. Heatmap of MITRE ATT&CK tactics versus the ten most frequently represented ML method subcategories across the reviewed studies. Each cell reports the number of papers in which a given tactic was associated with a given ML subcategory. Darker cells indicate higher frequencies of co-occurrence in the literature.

Interpretation of Tactics × ML Subcategories

(1) Researchers combine deep sequence learners, such as LSTM models, and spatial learners, such as CNN models, to capture complementary temporal and structural cues; ensembles are then layered to improve generalization under distribution shift.

(2) The persistence of classical learners within high-frequency tactics implies that interpretability and efficiency remain important in operational contexts, such as on-appliance detection and explainable triage.

(3) Cooler subcategories under Defense Evasion, Discovery, and Exfiltration likely result from sparser ground truth and harder labelling, providing further evidence of underexplored, high-value detection targets.

6.4. Techniques × ML Subcategories

At the technique level, Figure 15 shows that the dataset-driven bias becomes most visible:

- **The DoS/DDoS family is hottest.** Network Denial of Service shows high co-occurrence with Ensembles (18), LSTM (17), Feedforward (17), and CNN (12), with Statistical Models (11) also notable. Endpoint Denial of Service exhibits a similar spread, with Ensembles/LSTM = 13 and Feedforward = 12.
- **Perimeter techniques are broadly covered.** Exploit Public-Facing Application is strongly represented across Ensembles (14), LSTM (12), CNN (10), and Feedforward (11). Active Scanning shows a balanced profile, with Ensembles 12, LSTM 11, CNN 10, and Hybrid 8, indicating that reconnaissance is modelled with both temporal and structural features.
- **Protocol and host-information signals skew simpler.** Application Layer Protocol (Ensembles 13; CNN 9; Feedforward 8) and Gather Victim Host Information (Feedforward 11; LSTM 10; Optimization 8) suggest that tabular/engineered features remain competitive where semantics are well captured by aggregate statistics or handcrafted indicators.

Less intense regions, such as Remote Access Software or Remote Access Services, likely reflect a gap in the literature rather than diminished operational importance.

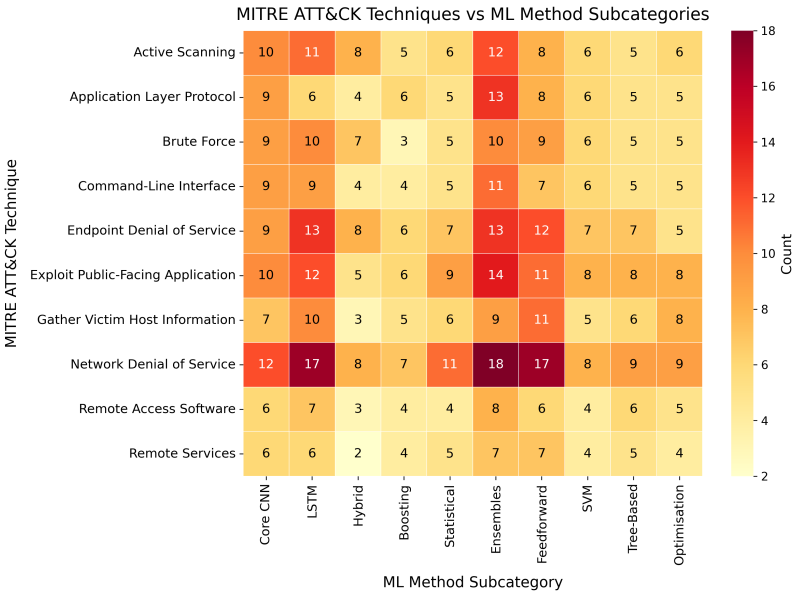

Figure 15. Heatmap of high-frequency MITRE ATT&CK techniques versus the ten most frequently represented ML method subcategories across the reviewed studies. Each cell reports the number of papers in which a given technique was associated with a given ML subcategory. Darker cells indicate higher frequencies of co-occurrence in the literature.

Interpretation of Techniques × ML Subcategories

(1) Technique-level intensities mirror data availability: where labelled, shareable corpora exist, such as DoS, scanning, and exploitation datasets, the literature shows deep coverage across multiple families.

(2) Diverse co-occurrence for the same technique, such as Network Denial of Service across LSTM/CNN/FFN/Ensembles, indicates algorithmic complementarity: temporal cues, localized structure, and robust aggregation each capture distinct signal facets.

(3) For textual/metadata-centric problems, such as phishing, feedforward models remain frequently used, consistent with structured features and a lower need for temporal context.

6.5. Heatmaps Meaning for Research and Practice

Bias toward visible, high-signal phases. Across all views, the field concentrates on Impact, Execution, and Initial Access—phases that are easier to emulate and label and that produce strong signatures. This improves benchmark progress but risks blind spots in stealthy, post-compromise behaviours, including Persistence, Exfiltration, Privilege Escalation, and Resource Development.

Architectural division of labour. LSTM models are especially prominent where order and timing are likely to matter, such as beaconing, staged execution, or flow-based exfiltration patterns. Ensembles appear to serve a stabilizing role under class imbalance, heterogeneous datasets, and potential domain shift, which may help explain their ubiquity across the most intensively studied tactics and techniques.

Operational and data constraints shape model choice. Classical learners, including SVMs, tree-based models, and statistical models, persist in mid-to-hot zones because they are computationally efficient, interpretable, and deployable on constrained sensors. Optimization and feature selection appear as enablers that make higher-capacity detectors viable in production.

6.6. Cyberattacks × ML × Datasets

This subsection examines the combined relationship among cyberattacks, ML paradigms, and datasets. When the attack, model, and dataset dimensions are examined together, a clearer picture of current research practice emerges. Methodological choices are not made independently; they are strongly shaped by the types of attacks represented in public benchmarks and by the availability of reusable datasets. This helps explain why denial-of-service, brute force, scanning, and other highly visible network-centric behaviours are repeatedly associated with deep learning, tree-based ensembles, and hybrid methods.

The tri-axis view also exposes important absence patterns. Post-compromise ATT&CK phases such as Persistence, Lateral Movement, Privilege Escalation, Collection, and Exfiltration are not only less studied overall, but are also supported by a much narrower dataset base and a thinner range of ML methods. Likewise, ICS and insider-threat settings remain methodologically sparse, with limited evidence of broad model exploration or robust cross-dataset validation.

Taken together, these patterns suggest that current progress in AI-driven cyber defense is constrained less by a shortage of algorithms than by the narrowness of representative evaluation settings. This motivates the integrated synthesis presented in the following subsection.

6.7. Key Findings and Research Gaps

The cross-reference analysis provides the clearest evidence that methodological choices in AI-driven cyber defense are not independent of the attacks and datasets being studied. Instead, they are tightly coupled. High-frequency benchmark datasets are strongly asso-

ciated with a narrow band of ATT&CK tactics and techniques, especially those linked to visible network intrusion, scanning, brute force, and denial-of-service behaviour. These settings, in turn, favour model families such as deep learning architectures, tree-based ensembles, and hybrid methods that perform well on high-volume supervised traffic analysis tasks. As a result, the apparent dominance of particular ML paradigms is partly a reflection of benchmark structure rather than a neutral indicator of universal suitability.

This combined perspective also makes absence patterns more visible. Some of the most important gaps in the literature do not appear when attacks, models, or datasets are viewed separately, but emerge only at their intersections. For example, post-compromise ATT&CK phases such as Persistence, Lateral Movement, Privilege Escalation, Collection, and Exfiltration are not only less studied overall; they are also associated with a much narrower range of datasets and a thinner range of ML methods. Likewise, ICS and insider-threat settings remain methodologically sparse, with limited evidence of diverse model exploration, robust cross-dataset validation, or sustained attention to long-horizon adversarial behaviour.

The main implication is that current progress in AI-driven cyber defense is constrained less by a shortage of algorithms than by a shortage of representative evaluation settings. The field has developed a rich catalogue of ML methods, but these methods are repeatedly validated on a relatively narrow subset of benchmark tasks. Consequently, there is a risk that methodological sophistication is being mistaken for broad defensive readiness. Future work should therefore target missing attack–model–dataset intersections, especially those involving stealthy post-compromise tactics, under-represented operational domains, and multi-dataset or cross-domain validation settings.

These tri-axis gaps directly motivate the broader limitations and structural bottlenecks discussed in the following section.

7. Gaps and Limitations

This section synthesises findings across attacks, ML methods, and datasets to identify where current AI-driven cyber defense research is well covered, where it is thin, and why. We also acknowledge threats to validity in our own review.

7.1. Coverage Gaps Across ATT&CK Tactics and Techniques

The cross-reference heatmaps reveal a systematic concentration on Initial Access, Execution,Command and Control, Reconnaissance, and Impact. By contrast, Persistence, Collection, and Resource Development remain consistently underexplored. Post-compromise behaviours such as Privilege Escalation, Lateral Movement, and Exfiltration also exhibit flatter, fragmented coverage. These phases are inherently stealthier, unfold over longer horizons, and are harder to label with reliable ground truth, which depresses dataset availability and skews model development toward high-signal, easily benchmarked problems (e.g., DoS/DDoS, scanning, and overt malware execution). The result is an ecosystem of models that perform strongly on visible, short-horizon events but provide limited assurance against low-and-slow, multi-stage adversaries.

7.2. Methodological Limitations in Model Development and Evaluation

Over-reliance on single-dataset studies. A substantial proportion of studies relied on only one dataset for model development and evaluation. While this can support focused benchmarking or proof-of-concept analysis, it also reflects simplified experimental settings that do not adequately capture the variability, complexity, and unpredictability of real-world cyber threats. As a result, many published models are optimized for narrow benchmark environments rather than for transfer across organisations, infrastructures,

or attack settings. This contributes to limited generalisability and weak evidence for robustness against multi-stage or multi-vector adversaries.

Shallow alignment to ATT&CK semantics. Although many papers reference ATT&CK, the alignment between model inputs, labels, and outputs and specific tactics or techniques is often indirect. Models trained on coarse labels such as "attack" versus "benign" are sometimes discussed as though they detect fine-grained TTPs. Without semantically aligned labels and evaluation procedures, it is difficult to claim genuine coverage of stealthier adversarial behaviours such as credential abuse, persistence, or living-off-the-land activity.

Limited treatment of drift and adversarial pressure. Few studies explicitly address concept drift, evolving baselines, software changes, or adversarial ML threats such as poisoning, evasion, or backdoor manipulation. Consequently, many proposed detectors may be brittle in long-running deployments and vulnerable to targeted manipulation or degradation over time.

Underuse of complementary learning paradigms. While ensembles, CNNs, and LSTMs dominate the literature, several potentially valuable paradigms remain under-explored. These include graph-based learning for host–process–identity relationships, long-context sequence models for low-and-slow behaviours, self-/semi-supervised pre-training on large unlabelled telemetry, and reinforcement learning for adaptive defense or sensing. Where such methods do appear, they are typically validated on the same narrow set of benchmark datasets, which limits the strength of the conclusions that can be drawn about their broader practical value.

7.3. Dataset-Specific Gaps

Despite the growing diversity of publicly available cybersecurity datasets, this review reveals several persistent gaps in how these resources are used. These gaps directly affect the generalisability, applicability, and resilience of AI-powered cyber defense solutions.

Under-representation of critical domains. Some of the most strategically important domains remain sparsely represented in the literature. Insider-threat datasets appear in only a very small number of studies, despite the operational importance of insider risk in enterprise environments. ICS datasets also remain limited, and only a small number of studies explore their overlap with IoT-NIDD settings. Likewise, phishing and social-engineering datasets, although present, are rarely integrated with network or malware datasets, which restricts the study of multi-vector attacks that cross technical and human attack surfaces.

Limited heterogeneity and co-occurrence. Multi-dataset studies do exist, but they typically combine closely related sources, such as NIDD with IoT-NIDD, rather than truly heterogeneous pairings such as NIDD with ICS, insider-threat, or phishing data. As a result, models are rarely evaluated on composite, multi-stage scenarios such as phishing leading to credential theft, lateral movement, and exfiltration. This limits the realism of current evaluation practice.

Scarce use of dataset expansion and adaptation. Only a limited number of studies use custom-collected datasets, and even fewer combine them with public datasets to broaden coverage. In addition, techniques such as data augmentation, synthetic generation, simulation-based expansion, and domain adaptation remain comparatively rare. This constrains the ability of models to generalise to unseen environments, novel attack vectors, and changing threat conditions.

7.4. Limitations

This systematic literature review is subject to several limitations. First, the search strategy was bounded by the selected databases, search string, and time window

(2019–2025), which means that some relevant studies may not have been captured. Second, the review included only English-language, peer-reviewed studies, so potentially relevant preprints, technical reports, and practitioner-oriented system documents were excluded. Third, the mapping of studies to ATT&CK tactics and techniques, ML categories, and dataset categories involved structured judgement; although coding rules and consistency checks were applied, some residual misclassification remains possible. Fourth, the review necessarily relies on author-reported metrics and experimental settings, which differ across preprocessing choices, sampling procedures, feature engineering, and evaluation protocols. This limits the comparability of reported performance across studies. Finally, the counts and heatmap intensities presented in this paper reflect patterns of emphasis within the literature, not the true prevalence of attacks in operational environments, and should therefore be interpreted as indicators of research attention rather than incident frequency.

Taken together, these gaps and limitations suggest that the main challenge facing AI-driven cyber defense is not simply model selection, but the interaction among semantic labelling quality, benchmark diversity, and realistic cross-domain evaluation.

8. Future Research Directions

Building on the gaps identified across attacks, methods, and datasets, this section outlines several actionable directions for advancing AI-driven cyber defense research. These directions are intended not as generic future-work statements, but as priorities derived directly from the tri-axis synthesis presented in this review.

8.1. Develop Multi-Domain and Multi-Stage Benchmarking Datasets

Future work should move beyond single-domain benchmarks and embrace multi-dataset integration. Composite datasets—linking, for example, NIDD traffic with phishing payloads, insider threat activity, and ICS telemetry—would enable the study of adversaries who operate across vectors and stages. Curating or simulating multi-stage attack chains (e.g., phishing $\rightarrow$ credential theft $\rightarrow$ lateral movement $\rightarrow$ data exfiltration) will be critical to closing the current fragmentation and testing model resilience in realistic, multi-hop kill chains.

8.2. Generate Fine-Grained Datasets Aligned with ATT&CK TTPs

Most current benchmarks offer only coarse-grained attack labels (e.g., “DoS” or “malware”). Future datasets should explicitly encode MITRE ATT&CK tactics and techniques, enabling models to learn at the semantic level of adversary behaviour. This includes collecting detailed telemetry for stealthy tactics (Persistence, Credential Access, and Lateral Movement) and ensuring balanced representation of underexplored techniques. Synthetic augmentation (e.g., GAN-based traffic generation, attack simulations in controlled labs) could help fill blind spots where real-world data are scarce.

8.3. Expand Methodological Horizons Beyond Standard Models

While ensembles, CNNs, and LSTMs dominate, more diverse ML paradigms should be tested against cyber defense challenges. Promising directions include the following:

- Graph-based learning for modelling host-process-identity relationships, naturally capturing lateral movement or privilege escalation.
- Long-context sequence models (e.g., Transformers) for detecting low-and-slow behaviours.
- Self-/semi-supervised pretraining on large unlabelled telemetry for better generalisation.
- Reinforcement learning for adaptive detection and proactive defense.

Crucially, such methods should not be validated only on narrow benchmarks but on heterogeneous, realistic scenarios.

8.4. Address Under-Represented Critical Domains

Insider threats, ICS/SCADA, IoT convergence, and phishing/social engineering remain substantially underexplored. Researchers should prioritise these domains, not in isolation, but by integrating them with traditional NIDD or malware datasets to reflect enterprise and cyber-physical system realities. Realistic cross-domain datasets would allow defense models to detect hybrid threats (e.g., phishing-initiated attacks against IoT-enabled critical infrastructure).

8.5. Promote Heterogeneity and Co-Occurrence in Evaluation

Future evaluation protocols should require models to demonstrate robustness across heterogeneous datasets and attack contexts, not just within one dataset family. Cross-domain benchmarking (e.g., training on NIDD and testing on IoT or ICS) would stress-test generalisation and reduce overfitting to dataset artefacts. Co-occurrence evaluation—where models face blended attack types within the same session—would better reflect real-world complexity.

8.6. Advance Data Expansion, Domain Adaptation, and Drift-Resilience

There is a clear need for systematic methods and evaluation frameworks that can cope with evolving adversarial environments. Techniques such as domain adaptation, transfer learning, and drift detection should be embedded into experimental pipelines. Data expansion through adversarial generation, simulation environments, and continual learning frameworks can provide models with resilience against concept drift and novel attack vectors. This direction is particularly critical for long-horizon tactics like persistence and stealthy credential abuse.

8.7. Key Research Priorities

Taken together, the future directions outlined in this section point toward a common priority: AI-driven cyber defense research must move beyond narrow benchmark optimisation and toward richer, ATT&CK-aligned, and cross-domain evaluation ecosystems. The strongest pattern across the identified priorities is the need for multi-stage datasets, heterogeneous validation settings, and broader testing of methods capable of handling long-horizon and structurally complex adversarial behaviour. The central gap is therefore not the absence of promising algorithms, but the absence of sufficiently realistic data and evaluation conditions in which such methods can be meaningfully assessed. This matters because without these shifts, the field will continue to produce strong benchmark results without corresponding confidence in real-world defensive readiness.

Without this shift, future AI-driven cyber defense research is likely to remain strong in benchmark performance but limited in operational generalisability.

9. Conclusions

This review mapped the contemporary landscape of AI-powered cyber defense across three linked evidence axes: cyberattacks, machine learning methods, and datasets. By analysing 99 peer-reviewed studies published between 2019 and 2025, aligning 312 attack labels to the MITRE ATT&CK framework, and organising 96 datasets into a refined taxonomy, this study provides an integrated perspective that goes beyond model-centric or dataset-centric surveys. In particular, it contributes the following: (i) an ATT&CK-aligned view of the attack landscape, (ii) a structured synthesis of ML method usage across attack contexts, and (iii) a tri-axis cross-reference analysis showing how attacks, models,

and datasets interact to shape current research practice. Among these, the tri-axis cross-reference analysis is the distinctive contribution: by examining attacks, ML methods, and datasets jointly, it surfaces benchmark dependencies and missing intersections that are obscured in model-centred or dataset-centred reviews.

Three findings stand out. First, research attention remains strongly concentrated on high-visibility phases such as Initial Access, Execution, Command and Control, Reconnaissance, and Impact, while stealthier and longer-horizon phases such as Persistence, Privilege Escalation, Lateral Movement, and Exfiltration remain comparatively underexplored. Second, deep learning and ensemble-based approaches dominate the most intensively studied settings, especially those supported by large public intrusion-detection benchmarks, while classical models remain important because of their efficiency, interpretability, and deployability in constrained environments. Third, dataset availability exerts a strong shaping effect on the field: repeated reliance on a narrow group of public NIDD benchmarks supports progress on visible threats such as DoS/DDoS, scanning, and web exploitation, but provides far weaker evidence of robustness against multi-stage, cross-domain, and low-and-slow adversaries.

These findings make the principal gaps in the field more explicit. Current AI-driven cyber defense research remains heavily concentrated on benchmark-friendly, high-visibility attack settings, leaving the post-compromise ATT&CK phases—Persistence, Privilege Escalation, Lateral Movement, and Exfiltration—comparatively underexplored. The literature is also shaped by strong benchmark dependence, with repeated reliance on a small set of public NIDD datasets and only limited evidence from multi-dataset, cross-domain, or long-horizon evaluation. As a result, under-represented but strategically important domains—including ICS/IIoT environments, insider threats, and phishing-led multi-stage attack scenarios—remain much less studied than conventional intrusion-detection settings. Together, these gaps reflect three structural bottlenecks that constrain real-world impact: dataset concentration, which channels evaluation toward a narrow set of reusable benchmarks; semantic shallowness, where coarse labels are sometimes interpreted as though they provide fine-grained ATT&CK coverage; and weak evaluation diversity, reflected in limited cross-dataset, cross-domain, and long-horizon validation.

Future research should therefore prioritise the development of richer ATT&CK-aligned datasets, especially for stealthy and long-horizon adversarial behaviours that remain difficult to observe in existing benchmarks. Greater emphasis is needed on multi-domain and multi-stage evaluation settings that combine network, IoT, phishing, insider-threat, and ICS/IIoT evidence, rather than validating models only within a single benchmark family. In methodological terms, the field would benefit from broader testing of models suited to structurally complex and low-and-slow environments, together with stronger evaluation under dataset shift, concept drift, and cross-organisational variation. From a practical standpoint, security operations and platform teams can translate these insights into near-term gains by combining heterogeneous telemetry sources, balancing interpretable and high-capacity detectors, and evaluating systems using operator-relevant criteria such as transferability, latency, alert burden, and investigation cost. Without these changes, progress in AI-driven cyber defense risks remaining strong at the benchmark level but weak in deployment realism.

Ultimately, the next stage of progress in AI-powered cyber defense will depend less on adding new models and more on building richer datasets, stronger cross-domain evaluations, and more realistic ATT&CK-aligned evidence.

Author Contributions: Conceptualization, M.C., A.A., and Q.K.A.M.; Methodology, M.C., A.A., and Q.K.A.M.; Software, M.C.; Validation, M.C. and H.C.; Formal Analysis, M.C. and H.C.; Investigation, M.C.; Data Curation, M.C.; Writing—Original Draft Preparation, M.C.; Writing—Review and Editing, A.A., Q.K.A.M., and H.C.; Visualization, M.C.; Supervision, A.A. and Q.K.A.M.; Project Administration, M.C. All authors have read and agreed to the published version of this manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive comments and helpful suggestions, which helped improve the quality and clarity of this manuscript.

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

Appendix A. Mapping of Identified Cyberattacks to MITRE ATT&CK Tactics and Techniques

Table A1. Mapping of identified cyberattacks to MITRE ATT&CK tactics, techniques, and IDs.

Attack/Keyword	Tactic	Technique	Technique ID
Reconnaissance			
Vulnerability Scanning	Reconnaissance	Active Scanning: Vulnerability Scanning	T1595.002
mscan	Reconnaissance	Active Scanning: Scanning IP Blocks	T1595.001
saint	Reconnaissance	Active Scanning: Vulnerability Scanning	T1595.002
portsweep	Reconnaissance	Active Scanning: Scanning IP Blocks	T1595.001
satan	Reconnaissance	Active Scanning: Vulnerability Scanning	T1595.002
ipsweep	Reconnaissance	Active Scanning: Scanning IP Blocks	T1595.001
nmap	Reconnaissance	Active Scanning	T1595
PortScan	Reconnaissance	Active Scanning	T1595
PortScan OS	Reconnaissance	Gather Victim Host Information: Client Configurations	T1592.004
Reconnaissance	Reconnaissance	Active Scanning, Gather Victim Host Information	T1595, T1592
OS Fingerprinting	Reconnaissance	Gather Victim Host Information: Client Configurations	T1592.004
Probe	Reconnaissance	Active Scanning	T1595
OS Scanning	Reconnaissance	Gather Victim Host Information: Client Configurations	T1592.004
Scanning	Reconnaissance	Active Scanning	T1595
Browsing job-hunting or competitor websites	Reconnaissance	Gather Victim Org Information	T1591
Information Gathering	Reconnaissance	Gather Victim Host Information	T1592
SYN Scan	Reconnaissance	Active Scanning	T1595
TCP Connect Scan	Reconnaissance	Active Scanning	T1595
UDP Scan	Reconnaissance	Active Scanning	T1595
Network Sweep (IP range scanning)	Reconnaissance	Active Scanning	T1595
Mirai–Junk/Scan	Reconnaissance	Active Scanning	T1595
Bashlite–Junk/Scan	Reconnaissance	Active Scanning	T1595
Service Scan	Reconnaissance	Active Scanning	T1595
pingsweep	Reconnaissance	Active Scanning	T1595
Resource Development			
Domain Abuse	Resource Development	Acquire Infrastructure: Domains	T1583.001
Malicious Domain	Resource Development	Acquire Infrastructure: Domains	T1583.001
Account creation for fake reviews	Resource Development	Establish Accounts: Social Media Accounts	T1585.001
Phishing Kits	Resource Development	Develop Capabilities	T1587
Hosting Infra	Resource Development	Acquire Infrastructure: Web Services	T1583.006
Botnet fake account	Resource Development	Establish Accounts: Social Media Accounts	T1585.001
Sybil Attack	Resource Development	Compromise Accounts	T1586
Initial Access			
sendmail	Initial Access	Exploit Public-Facing Application	T1190
named	Initial Access	Exploit Public-Facing Application	T1190
ftp_write	Initial Access	Valid Accounts	T1078
Web Attack–SQL Injection	Initial Access	Exploit Public-Facing Application	T1190
LDAP Injection	Initial Access	Exploit Public-Facing Application	T1190
XPath Injection	Initial Access	Exploit Public-Facing Application	T1190
mysql	Initial Access	Valid Accounts	T1078
Unintentional Illegal Requests	Initial Access	Exploit Public-Facing Application	T1190
sqlattack	Initial Access	Exploit Public-Facing Application	T1190
SSI Injection	Initial Access	Exploit Public-Facing Application	T1190
phf	Initial Access	Exploit Public-Facing Application	T1190
Exploits	Initial Access	Exploit Public-Facing Application	T1190
Phishing Email	Initial Access	Phishing	T1566

Table A1. *Cont.*

Attack/Keyword	Tactic	Technique	Technique ID
Infiltration	Initial Access	Phishing	T1566
CVE-2017-5638 (Struts2)	Initial Access	Exploit Public-Facing Application	T1190
proftpd	Initial Access	Exploit Public-Facing Application	T1190
apache-struts	Initial Access	Exploit Public-Facing Application	T1190
Spam Email	Initial Access	Phishing: Spearphishing via Email	T1566.001
Spam	Initial Access	Phishing	T1566
Phishing	Initial Access	Phishing	T1566
Fake Pages	Initial Access	Phishing	T1566
Phishing Site Deployment	Initial Access	Phishing: Spearphishing Link	T1566.002
AI-generated phishing URLs	Initial Access	Phishing: Spearphishing Link	T1566.002
Payload via Email	Initial Access	Phishing: Spearphishing via Email	T1566.001
Telnet exploit	Initial Access	Exploit Public-Facing Application	T1190
Generic (Generic Exploits)	Initial Access	Exploit Public-Facing Application	T1190
Code Red Worm	Initial Access	Exploit Public-Facing Application	T1190
Parameter Tampering	Initial Access	Exploit Public-Facing Application	T1190
Path Traversal	Initial Access	Exploit Public-Facing Application	T1190
Opportunistic Service Attack (OSA)	Initial Access	Exploit Public-Facing Application	T1190
Replay Attacks	Initial Access	Valid Accounts	T1078
Spearphishing	Initial Access	Phishing: Spearphishing Attachment	T1566.001
S7 unauthorized access	Initial Access	Exploit Public-Facing Application	T1190
Unauthorized access to HMI or SCADA	Initial Access	Valid Accounts	T1078
Execution			
Web Attack–XSS	Execution	Command-Line Interface: JavaScript	T1059.007
Server-Side Include	Execution	Server Software Component	T1505.003
loadmodule	Execution	Process Injection: Dynamic-Link Library Injection	T1055.001
Web Attack-Command Injection	Execution	Command-Line Interface	T1059
perl	Execution	Command-Line Interface	T1059
xterm	Execution	Command-Line Interface	T1059
Shellcode	Execution	Exploitation for Client Execution	T1203
OS Command Execution	Execution	Command and Scripting Interpreter	T1059
Downloaders/Droppers	Execution	Ingress Tool Transfer	T1105
S7 command injection	Execution	Manipulation of Control/Command Message	T0851
Command injection to PLC or RTU	Execution	Command and Scripting Interpreter	T1059
warezmaster	Execution	Command-Line Interface	T1059
JavaMeterpreter	Execution	Command-Line Interface: JavaScript	T1059.007
Windows-RCE	Execution	Exploitation for Client Execution	T1203
Trojan	Execution	User Execution	T1204
Installing unauthorized software	Execution	User Execution: Malicious File	T1204.002
Unauthorized command via MODBUS	Execution	Command and Scripting Interpreter	T1059
Viruses	Execution	User Execution	T1204
Fileless Malware	Execution	Command and Scripting Interpreter: PowerShell	T1059.001
Allaple.A/L	Execution	User Execution	T1204
Agent.FYI	Execution	Command and Scripting Interpreter	T1059
Dialplatform.B	Execution	User Execution	T1204
Instantaccess	Execution	User Execution	T1204
VB.AT	Execution	Command and Scripting Interpreter	T1059
Yuner.A	Execution	Command and Scripting Interpreter	T1059
Gatak	Execution	Command and Scripting Interpreter	T1059
Lollipop	Execution	Command and Scripting Interpreter	T1059
Tracur	Execution	User Execution	T1204
Simda	Execution	Command and Scripting Interpreter	T1059
Htbot	Execution	Command and Scripting Interpreter	T1059
Miuref	Execution	Command and Scripting Interpreter	T1059
Neris	Execution	Command and Scripting Interpreter	T1059
Kenjiro	Execution	Command and Scripting Interpreter	T1059
Hide and Seek	Execution	Command and Scripting Interpreter	T1059
Mirai	Execution	Command and Scripting Interpreter	T1059
Persistence			
Wintrim.BX	Persistence	Boot or Logon Autostart Execution: Registry Run Keys/Startup Folder	T1547.001
Autorun.K	Persistence	Boot or Logon Autostart Execution	T1547
Murlo (TCP-based backdoor)	Persistence	Create or Modify System Process	T1569
Browser Hijacking	Persistence	Software Extensions: Browser Extensions	T1176.001
Uploading Attack	Persistence	Server Software Component: Web Shell	T1505.003
Modification of control logic	Persistence	Event Triggered Execution	T1546
Malware with Persistence via Registry	Persistence	Boot or Logon Autostart Execution: Registry Run Keys/Startup Folder	T1547.001
Alueron.gen!J	Persistence	Boot or Logon Autostart Execution	T1547
Dontovo.A	Persistence	Boot or Logon Autostart Execution	T1547
Lolyda.AA1/AA2/AA3/AT	Persistence	Boot or Logon Autostart Execution	T1547
Malex.gen!J	Persistence	Boot or Logon Autostart Execution	T1547
Skintrim.N	Persistence	Boot or Logon Autostart Execution	T1547
Kelihos_ver3	Persistence	Boot or Logon Autostart Execution	T1547
Kelihos_ver1	Persistence	Boot or Logon Autostart Execution	T1547
Vundo	Persistence	Boot or Logon Autostart Execution	T1547
Shifu	Persistence	Boot or Logon Autostart Execution	T1547
Torii	Persistence	Boot or Logon Autostart Execution	T1547

Table A1. *Cont.*

Attack/Keyword	Tactic	Technique	Technique ID
Privilege Escalation			
xlock	Privilege Escalation	Exploitation for Privilege Escalation	T1068
Buffer Overflow	Privilege Escalation	Exploitation for Privilege Escalation	T1068
Adduser (Unauthorized)	Privilege Escalation	Create Account	T1136
Defense Evasion			
CRLF Injection	Defense Evasion	Obfuscated Files or Information	T1027
Rootkit	Defense Evasion	Rootkit	T1014
Logging in outside working hours	Defense Evasion	Valid Accounts	T1078
Obfuscator.AD	Defense Evasion	Obfuscated Files or Information	T1027
Obfuscator.ACY	Defense Evasion	Obfuscated Files or Information	T1027
URL Obfuscation	Defense Evasion	Obfuscated Files or Information	T1027
Avoiding spam detection	Defense Evasion	Email Obfuscation/Content Spoofing	T1566/T1027
Log deletion or obfuscation	Defense Evasion	Indicator Removal on Host	T1070.001
Decreased Rank Attack	Defense Evasion	Impair Defenses	T1562
hash-based malware	Defense Evasion	Obfuscated Files or Information	T1027
Spoofing	Defense Evasion	Modify Authentication Process	T1556
Disabling alarms or event logs	Defense Evasion	Indicator Removal	T1070
AI-generated malware	Defense Evasion	Obfuscated Files or Information	T1027
Obfuscated Malware	Defense Evasion	Obfuscated Files or Information	T1027
Polymorphic malware	Defense Evasion	Obfuscated Files or Information	T1027
Metamorphic malware	Defense Evasion	Obfuscated Files or Information	T1027
Packed malware	Defense Evasion	Obfuscated Files or Information: Software Packing	T1027.002
Fakerean	Defense Evasion	Obfuscated Files or Information	T1027
Swizzor.gen!E/I	Defense Evasion	Obfuscated Files or Information	T1027
Nsis-ay	Defense Evasion	Obfuscated Files or Information	T1027
Credential Access			
imap	Credential Access	Exploitation of Remote Services	T1210
Infiltration-mitm	Credential Access	Adversary-in-the-Middle	T1557
MITM	Credential Access	Adversary-in-the-Middle	T1557
Telnet remote access attempts	Credential Access	Brute Force	T1110
xsnoop	Credential Access	Input Capture	T1056
spy	Credential Access	Input Capture	T1056
Keylogger	Credential Access	Keylogging	T1056.001
Web Brute Force	Credential Access	Brute Force	T1110
guess_passwd	Credential Access	Brute Force	T1110
FTP-Patator	Credential Access	Brute Force	T1110
SSH-Patator	Credential Access	Brute Force	T1110
SSH Brute Force	Credential Access	Brute Force	T1110
SSH Attack	Credential Access	Brute Force	T1110
RDP Brute Force	Credential Access	Brute Force	T1110.003
Password Cracking	Credential Access	Brute Force: Password Spraying	T1110.004
RTSP Brute Force	Credential Access	Brute Force	T1110
Credential Harvesting	Credential Access	Input Capture/Phishing	T1056/T1566
SFTP attack	Credential Access	Brute Force	T1110
Brute Force	Credential Access	Brute Force	T1110
Heartbleed	Credential Access	Exploitation for Credential Access	T1212
Hydra-FTP (FTP brute-force attacks)	Credential Access	Brute Force: Password Guessing	T1110.001
Hydra-SSH (SSH brute-force attacks)	Credential Access	Brute Force: Password Guessing	T1110.001
Bashlite–brute	Credential Access	Brute Force: Password Guessing	T1110.001
DNS Spoofing	Credential Access	Adversary-in-the-Middle	T1557
Dictionary BruteForce	Credential Access	Brute Force: Password Guessing	T1110.001
Wormhole Attack (WHA)	Credential Access	Adversary-in-the-Middle	T1557
Hijacking or spoofing PLC communications	Credential Access	Adversary-in-the-Middle	T1557
Evil Twin Attacks	Credential Access	Adversary-in-the-Middle	T1557
Ramnit	Credential Access	Credentials from Password Stores	T1555
Cridex	Credential Access	Credentials from Password Stores	T1555
Zeus	Credential Access	Credentials from Password Stores	T1555
Tinba	Credential Access	Credentials from Password Stores	T1555
Discovery			
Fuzzers Attack	Discovery	Network Service Scanning	T1046
snmpguess	Discovery	Network Service Scanning	T1046
ps	Discovery	Process Discovery	T1057
snmpgetattack	Discovery	Network Service Scanning	T1046
Analysis	Discovery	Network Sniffing	T1040
Sniffing Attacks	Discovery	Network Sniffing	T1040
ARP Spoofing	Discovery	Network Sniffing	T1040
Host Discovery	Discovery	Remote System Discovery	T1018
Sinkhole Attack (SHA)	Discovery	Network Sniffing	T1040
Lateral Movement			
smb-exploit	Lateral Movement	Exploitation of Remote Services	T1210
Worms	Lateral Movement	Remote Services	T1021
W32.Blaster Worm	Lateral Movement	Exploitation of Remote Services	T1210
Reaper Worm	Lateral Movement	Exploitation of Remote Services	T1210
Virut (Malware propagation)	Lateral Movement	Replication Through Removable Media	T1091
Conficker	Lateral Movement	Exploitation of Remote Services	T1210
Hakai	Lateral Movement	Remote Services	T1021
Muhstik	Lateral Movement	Remote Services	T1021

Table A1. *Cont.*

Attack/Keyword	Tactic	Technique	Technique ID
Collection			
File Disclosure	Collection	Data from Local System	T1005
Searching for sensitive documents	Collection	Data from Local System	T1005
Spyware	Collection	Input Capture	T1056
ARP MitM	Collection	Adversary-in-the-Middle	T1557
Active Wiretap	Collection	Adversary-in-the-Middle	T1557
Accessing sensitive files repeatedly	Collection	Data from Local System	T1005
Printing sensitive information	Collection	Data from Local System	T1005
Infostealers	Collection	Input Capture	T1056
Adialer.C	Collection	Input Capture	T1056
Alueron.gen!J	Collection	Ingress Tool Transfer	T1105
Command and Control			
Backdoor	Command and Control	Remote Access Software	T1219
httptunnel	Command and Control	Application Layer Protocol: Web Protocols	T1071.001
multihop	Command and Control	Proxy: Multi-hop Proxy	T1090.003
warezclient	Command and Control	Application Layer Protocol	T1071
Botnet ARES	Command and Control	Application Layer Protocol: Web Protocols	T1071.001
Spam botnet	Command and Control	Application Layer Protocol	T1071
Meterpreter (Metasploit post-exploitation activity)	Command and Control	Application Layer Protocol	T1071
IRC Botnet	Command and Control	Application Layer Protocol	T1071
HTTP Botnet	Command and Control	Application Layer Protocol	T1071
DNS Tunneling	Command and Control	Application Layer Protocol: DNS	T1071.004
Botnet	Command and Control	Application Layer Protocol	T1071
Neris	Command and Control	Application Layer Protocol	T1071
Rbot	Command and Control	Application Layer Protocol	T1071
Menti (P2P botnet)	Command and Control	Non-Application Layer Protocol	T1095
Sogou (HTTP-based C2)	Command and Control	Application Layer Protocol: Web Protocols	T1071.001
NSIS.ay (Downloader trojan)	Command and Control	Ingress Tool Transfer	T1105
Remote Access Trojan (RAT)	Command and Control	Remote Access Tools	T1219
Use of anonymizing tools	Command and Control	Proxy: Multi-hop Proxy	T1090.003
Command & Control using HTTP	Command and Control	Application Layer Protocol: Web Protocols	T1071.001
C2LOP.P/gen!g	Command and Control	Non-Application Layer Protocol	T1095
Rbot!gen	Command and Control	Ingress Tool Transfer	T1105
Geodo	Command and Control	Ingress Tool Transfer	T1105
Gafgyt	Command and Control	Ingress Tool Transfer	T1105
Linux.Hajime	Command and Control	Ingress Tool Transfer	T1105
Exfiltration			
Sending confidential info to competitors	Exfiltration	Exfiltration Over Web Service	T1041
Data Exfiltration	Exfiltration	Exfiltration Over Command and Control Channel	T1041
Data Exfiltration Tools	Exfiltration	Exfiltration Tools	T1567
Uploading to personal email/cloud	Exfiltration	Exfiltration Over Web Service	T1567.002
Use of removable media	Exfiltration	Exfiltration Over Physical Medium: Exfiltration over USB	T1052.001
Impact			
mailbomb	Impact	Email Bombing	T1667
smurf (ICMP amplification)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
neptune	Impact	Network Denial of Service: Direct Network Flood	T1498.001
back	Impact	Service Stop	T1489
teardrop	Impact	Endpoint Denial of Service: Application or System Exploitation	T1499.004
pod (Ping of Death)	Impact	Network Denial of Service: Direct Network Flood	T1498.001
land DoS	Impact	Endpoint Denial of Service: OS Exhaustion Flood	T1499.001
apache2	Impact	Endpoint Denial of Service: Service Exhaustion Flood	T1499.002
processtable	Impact	Endpoint Denial of Service: OS Exhaustion Flood	T1499.001
udpstorm	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Packts fragmentation attack	Impact	Network Denial of Service: Direct Network Flood	T1498.001
UDP Fragmentation	Impact	Network Denial of Service: Direct Network Flood	T1498.001
ACK Fragmentation	Impact	Network Denial of Service: Direct Network Flood	T1498.001
RSTFIN Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
PSHACK Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
ICMP Fragmentation	Impact	Network Denial of Service: Direct Network Flood	T1498.001
SynonymousIP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
SYN Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
UDP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
ICMP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
HTTP Flood	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
Apache Range Header	Impact	Endpoint Denial of Service: Application or System Exploitation	T1499.004
Slow POST	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
HTTP/2 Rapid Reset	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
GraphQL Overload	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
API Parameter Abuse	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003

Table A1. *Cont.*

Attack/Keyword	Tactic	Technique	Technique ID
WS Amplification	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
DoS GoldenEye	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
DoS Hulk	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
DoS Slowhttptest	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
DoS Slowloris	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
DoS	Impact	Network Denial of Service	T1498
DDoS	Impact	Network Denial of Service	T1498
DDoSsim	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
DDoS LOIC-UDP	Impact	Network Denial of Service: Direct Network Flood	T1498.001
DDoS LOIC-HTTP	Impact	Network Denial of Service: Direct Network Flood	T1498.001
DDoS HOIC	Impact	Network Denial of Service: Direct Network Flood	T1498.001
DDoS Bot	Impact	Network Denial of Service	T1498
DDoS Stomp	Impact	Network Denial of Service: Direct Network Flood	T1498.001
DDoS DYN	Impact	Network Denial of Service: Direct Network Flood	T1498.001
DDoS TCP	Impact	Network Denial of Service: Direct Network Flood	T1498.001
DNS (DNS amplification)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
LDAP (UDP flood)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
MSSQL (UDP flood)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
NetBIOS (UDP flood)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
NTP (NTP amplification)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
Portmap (UDP flood)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
SNMP (SNMP amplification)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
CLDAP Reflection	Impact	Network Denial of Service: Reflection Amplification	T1498.002
SSDP (SSDP amplification)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
UDP (Generic UDP flood)	Impact	Network Denial of Service: Direct Network Flood	T1498.001
UDPLag (UDP with response lag)	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
TFTP (UDP flood)	Impact	Network Denial of Service: Reflection Amplification	T1498.002
WebDDoS (HTTP flood)	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
Memcached	Impact	Network Denial of Service: Reflection Amplification	T1498.002
Mirai–TCP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Mirai–UDP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Mirai–HTTP Flood	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
Mirai-GREIP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Mirai-Greeth Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Mirai-UDPPlain	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Bashlite–TCP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Bashlite–UDP Flood	Impact	Network Denial of Service: Direct Network Flood	T1498.001
Bashlite–HTTP Flood	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
Bashlite–ACK/other floods	Impact	Network Denial of Service: Direct Network Flood	T1499.003
SSL Renegotiation DoS	Impact	Endpoint Denial of Service: Application Exhaustion Flood	T1499.003
Flooding Attack	Impact	Network Denial of Service	T1498
DODAG Version Number Attack	Impact	Endpoint Denial of Service (ICS-specific)	T1499.004
Ransomware	Impact	Data Encrypted for Impact	T1486
Blackhole Attack	Impact	Network Denial of Service	T1498
Video Injection	Impact	Defacement	T1491
Gear Spoofing Attack	Impact	Data Manipulation	T1565
RPM Spoofing Attack	Impact	Data Manipulation	T1565
False Data Injection Attack	Impact	Data Manipulation	T1565
Tolerable FDIA	Impact	Data Manipulation	T1565
Posting fake reviews	Impact	Data Manipulation	T1565
Opinion Spam (Fake Review Attack)	Impact	Data Manipulation	T1565
Disinformation/fake content	Impact	Data Manipulation	T1565
Coordinated campaign (review farms)	Impact	Data Manipulation	T1565
Targeting businesses' reputation	Impact	Data Manipulation	T1565
Review flooding	Impact	Data Manipulation	T1565
Malicious actuator control	Impact	Endpoint Denial of Service	T1499
PTP Attack (time sync manipulation)	Impact	Data Manipulation	T1565
De-authentication DoS	Impact	Endpoint Denial of Service	T1499
Fake Landing (tricking the UAV into landing)	Impact	Data Manipulation	T1565
Adware	Impact	Resource Hijacking	T1496

Appendix B. Reviewed Datasets Information and Refined Taxonomy

Table A2. SLR Dataset Used Properties.

Dataset	Category	Year	Normal	Attack	Metadata	Format	Count	Duration	Kind	Network	Complete	Splits	Balanced	Labeled	Classes
KDD Cup 99 [55]	NIDD	1999	yes	yes	no	other	5 M	-	emulated	small	yes	yes	no	yes	4
NSL-KDD [20]	NIDD	2009	yes	yes	no	other	148,517	-	emulated	small	yes	yes	no	yes	4
UNSW-NB15 [21]	NIDD	2015	yes	yes	yes	packet, other	2,540,044	31 h	emulated	small	yes	yes	no	yes	9
CSE-CIC-IDS2017 [22]	NIDD	2017	yes	yes	yes	packet, bi.flow	3,100,000	5 days	emulated	small	yes	no	no	yes	9
CSE-CIC-IDS2018 [56]	NIDD	2018	yes	yes	yes	packet, bi.flow	16,000,000	10 days	emulated	small	yes	no	no	yes	15
CIC-DDoS2019 [57]	NIDD	2019	yes	yes	yes	packet, bi.flow	50,000,000	5 days	emulated	small	yes	no	no	yes	11
LITNET-2020 [58]	NIDD	2020	yes	yes	yes	flow, packet	50,000,000	months	real	large ISP	yes	no	no	yes	2
NetML-2020 [59]	NIDD	2020	yes	yes	yes	flow	3,000,000	days	emulated	small	yes	no	yes	yes	10
5G-NIDD [60]	NIDD	2021	yes	yes	yes	flow	15,000	hours	emulated	5G wireless	yes	no	yes	yes	2
FLNET2023 [61]	NIDD	2023	yes	yes	yes	flow	6,000,000	24 h	real + emulated	various	yes	no	no	yes	11
NGIDS-DS [62]	NIDD	2022	yes	yes	yes	flow	20,000,000	days	emulated	small	yes	no	no	yes	9
CAIDA 2007 [63]	NIDD	2007	no	yes	yes	packet	huge	minutes	real	various	partial	no	no	no	-
BoNeSi Dataset [64]	NIDD	2010	no	yes	no	packet	100,000	minutes	emulated	lab	no	no	no	no	-
DDoSDB [65]	NIDD	since 2020	varies	yes	yes	packet	-	-	real + synthetic	various	yes	no	no	no	-
App Layer DoS [66]	NIDD	2017	yes	yes	yes	flow	1,000,000	8 h	emulated	small	yes	no	no	yes	4
CSIC HTTP 2010 [67]	NIDD	2010	no	yes	yes	HTTP logs	67,000	sessions	emulated	application-level	yes	yes	yes	yes	2
ECML/PKDD 2007 [68]	NIDD	2007	no	yes	yes	session	600,000	weeks	real	telecom	partial	yes	yes	yes	2
NPS-2009-Casper-Rw [69]	NIDD	2009	no	yes	yes	packet, flow	1,000,000	hours	emulated	small	no	no	no	no	-
NCC Dataset [70]	NIDD	2022	yes	yes	yes	flow	5,000,000	hours	real + emulated	huge	yes	no	no	yes	14
NCC-2 Dataset [71]	NIDD	2023	yes	yes	yes	flow	10,000,000	hours	real + emulated	huge	yes	yes	no	yes	18
InSDN Dataset [72]	NIDD	2022	yes	yes	yes	flow	4,000,000	hours	emulated	SDN	yes	no	yes	yes	10
Benign and Malicious [73]	NIDD	2021	yes	yes	yes	other	90,000	-	real	various	no	-	yes	yes	2
CTU-13 Dataset [74]	NIDD	2011	yes	yes	yes	packet	13 scenarios	days	real + emulated	botnet traffic	yes	no	no	yes	13
USTC-TFC2016 [75]	NIDD	2017	yes	yes	yes	packet	750,000	hours	real	malware dataset	yes	no	no	yes	10
N-BaIoT [76]	IoT-NIDD	2018	yes	yes	yes	flow	100,000	days	emulated	IoT	yes	no	no	yes	2
BoT-IoT [9]	IoT-NIDD	2018	yes	yes	yes	flow	70,000,000	days	emulated	IoT	yes	yes	no	yes	5
IoTPOT [77]	IoT-NIDD	2015	no	yes	yes	packet	-	weeks	real	honeypot	partial	no	no	no	-
ToN-IoT [23]	IoT-NIDD	2020	yes	yes	yes	flow, syslogs	25,000,000	days	emulated + real	IoT	yes	yes	yes	yes	9
IoT-23 [78]	IoT-NIDD	2020	yes	yes	yes	flow	20,000,000	days	emulated+real	IoT	yes	no	no	yes	10
EdgeIIoT 2023 [79]	IoT-NIDD	2023	yes	yes	yes	flow	2,000,000	days	emulated	edge IoT	yes	no	yes	yes	15
CIC-IoT2022 [80]	IoT-NIDD	2022	yes	yes	yes	flow, packet	4,000,000	days	emulated	small IoT lab	yes	no	no	yes	6
CICIoT-2023 [81]	IoT-NIDD	2023	yes	yes	yes	flow, packet	10,000,000	days	emulated	IoT/5G hybrid	yes	no	no	yes	8
UNSW IoT Traffic [82]	IoT-NIDD	2019	yes	yes	yes	flow	1,000,000	hours	emulated	IoT	yes	no	yes	yes	10
Distributed IoT [83]	IoT-NIDD	2021	yes	yes	yes	flow	3,000,000	hours	emulated	IoT	yes	no	yes	yes	10
ROUT-4-2023 [84]	IoT-NIDD	2023	yes	yes	yes	flow	2,000,000	hours	emulated	hybrid SDN & IoT	yes	no	no	yes	9
Kitsune Dataset [85]	IoT-NIDD	2018	yes	yes	yes	packet, flow	100,000,000	days	emulated	IoT & smart home	yes	no	no	yes	21
Wi-Fi Dataset [86]	IoT-NIDD	2022	yes	yes	yes	flow	1,000,000	hours	emulated	Wi-Fi lab	yes	no	yes	yes	7
HCRL CAN [87]	IoT-NIDD	2020	yes	yes	yes	other, packet	4,500,000	hours	real + synthetic	vehicle CAN bus	yes	no	no	yes	5
HCRL Car Hacking [88]	IoT-NIDD	2020	yes	yes	yes	other	4,300,000	40 min	real	vehicle CAN bus	yes	no	no	yes	5
Malimg [89]	Malware	2011	no	yes	yes	images	9339	-	static	malware dataset	yes	no	no	yes	25

Table A2. *Cont.*

Dataset	Category	Year	Normal	Attack	Metadata	Format	Count	Duration	Kind	Network	Complete	Splits	Balanced	Labeled	Classes
BIG 2015 [90]	Malware	2015	no	yes	yes	binaries	10 GB files	-	static	malware dataset	yes	yes	no	yes	9
MaleVis [91]	Malware	2020	no	yes	yes	images	14,226	-	static	malware dataset	yes	no	yes	yes	26
Malicia [92]	Malware	2013	no	yes	yes	binaries	11,668	-	static	malware dataset	yes	no	no	yes	2
Drebin project [93]	Malware	2014	no	yes	yes	APK files	129,013	-	static	mobile malware	yes	yes	yes	yes	2
VX-Heavens [94]	Malware	since 2010	no	yes	limited	binaries	30,000	-	static	malware dataset	no	no	no	no	-
VirusShare [95]	Malware	since 2010	no	yes	limited	binaries	-	-	static	malware dataset	no	no	no	no	-
VirusTotal [96]	Malware	since 2004	no	yes	yes	binaries	-	-	static + dynamic	malware dataset	no	no	no	yes	1
CIC-MalMem-2022 [97]	Malware	2022	no	yes	yes	memory	100,000	hours	dynamic	malware memory	yes	yes	yes	yes	6
MemMal-D2024 [98]	Malware	2024	no	yes	yes	memory	100,000	hours	dynamic	malware memory	yes	yes	yes	yes	2
CIC-CMD-2024 [99]	Malware	2024	yes	yes	yes	flow	10,000,000	days	real + emulated	malware dataset	yes	yes	no	yes	-
SpamEmail [100]	S&P	1999	yes	yes	yes	other	4601	-	static	spam emails	-	no	no	yes	2
SpamClassification [101]	S&P	2021	yes	yes	yes	other	5796	-	emulated	spam messages	-	yes	no	yes	2
Email Spam [102]	S&P	2020	yes	yes	yes	other	5172	-	emulated	spam emails	-	yes	yes	yes	2
SpamAssassin [103]	S&P	since 2021	yes	yes	partial	other	6047	1 year	real	spam emails	-	yes	no	yes	2
Benign Email [104]	S&P	2013	yes	no	yes	other	14,043	-	real	benign emails	-	no	-	yes	1
Phishing Email [105]	S&P	2020	yes	yes	yes	other	-	-	emulated	phishing emails	-	yes	no	yes	2
Bot Account [106]	S&P	2023	yes	yes	yes	other	8574	-	real	social media	no	no	yes	yes	2
STIX & Curated [107]	S&P	2015	no	yes	yes	other	-	-	emulated	Threat Indicators	no	no	-	yes	-
Alexa Phishing [108]	S&P	since 2020	yes	yes	yes	other	1M+	-	real	Phishing URLs	no	no	no	yes	2
PhishTank [109]	S&P	-	no	yes	yes	other	100K+	-	real	Phishing URLs	yes	no	no	yes	2
OpenPhish [110]	S&P	-	no	yes	yes	other	-	-	real	Phishing URLs	no	no	no	yes	2
Anti-Phishing WG [111]	S&P	-	no	yes	yes	other	-	months	real	Phishing incidents	no	no	-	yes	2
YelpChi [112]	S&P	2013	yes	yes	yes	other	45,000+	years	real	reviews	yes	yes	yes	yes	2
YelpNYC [113]	S&P	2015	yes	yes	yes	other	160,000+	years	real	reviews	yes	yes	yes	yes	2
YelpZip [113]	S&P	2015	yes	yes	yes	other	60,000+	years	real	reviews	yes	yes	yes	yes	2
Gas Pipeline [54,114]	ICS	2011	yes	yes	yes	flow, packet	100,000	hours	emulated	ICS network	yes	no	no	yes	2
SWaT dataset [115]	ICS	2015	yes	yes	yes	other	946,722	11 days	real	ICS network	yes	yes	no	yes	2
Necon-IIUM ICS Dataset [116]	ICS	2022	yes	yes	yes	other	1,500,000	7 days	emulated	ICS network	yes	no	no	yes	5
ERENO IEC-61850 [117]	ICS	2020	yes	yes	yes	packet, flow	-	2 h	real	ICS network	yes	no	no	yes	5
IEEE 118-bus dataset [118]	ICS	2001	yes	yes	yes	other	118 buses	-	synthetic	ICS network	yes	no	-	yes	3
IEEE 123-bus dataset [118]	ICS	1991	yes	no	yes	other	123 buses	-	synthetic	ICS network	yes	no	no	yes	3
IEEE 13-bus dataset [118]	ICS	1992	yes	no	yes	other	13 buses	-	synthetic	ICS network	yes	no	no	yes	2
IEEE-14-bus dataset [118]	ICS	2018	yes	yes	yes	other	14 buses	24 h	synthetic	ICS network	yes	no	-	yes	2
CERT Insider Threat [119]	Insider Threat	2016	yes	yes	yes	user logs	1,000,000	months	emulated	enterprise	yes	no	no	yes	2
Udacity dataset [120]	Other	2016	yes	no	yes	images	-	-	real	simulation	yes	no	no	yes	-
GTSRB [121]	Other	2011	-	-	yes	images	51,839	-	real	-	-	yes	no	yes	43
UAVid dataset [122]	Other	2020	-	-	yes	images	3000	-	real	UAV/Aerial	no	yes	no	yes	8
ConsumerComplaint [123]	Other	2018	-	-	yes	other	1,200,000	8 years	real	-	-	no	no	yes	10
SpeechCommands [124]	Other	2017	-	-	yes	wav	105,829	-	real	voice command	-	yes	-	yes	35
IMDB [125]	Other	2011	-	-	yes	other	50,000	-	real	-	-	yes	yes	yes	2
CIFAR-10 [126]	Other	2009	-	-	yes	images	60,000	-	real	-	-	yes	yes	yes	10

References

1. Sommer, R.; Paxson, V. Outside the closed world: On using machine learning for network intrusion detection. In *Proceedings of the 2010 IEEE Symposium on Security and Privacy (SP)*; IEEE: Los Alamitos, CA, USA, 2010; pp. 305–316.
2. Li, Y.; Liu, Q. A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. *Energy Rep.* **2021**, *7*, 8176–8186. [CrossRef]
3. Hizal, S.; Cavusoglu, U.; Akgun, D. A novel deep learning-based intrusion detection system for IoT DDoS security. *Internet Things* **2024**, *28*, 101336. [CrossRef]
4. Jada, I.; Mayayise, T.O. The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data Inf. Manag.* **2024**, *8*, 100063. [CrossRef]
5. Baron Garcia, A. Machine Learning and Artificial Intelligence Methods for Cybersecurity Data Within the Aviation Ecosystem. Ph.D. Thesis, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA, 2022.
6. Buczak, A.L.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2501–2528. [CrossRef]
7. Kaur, R.; Gabrijeic, D.; Klobucar, T. Artificial intelligence for cybersecurity: Literature review and future research directions. *Inf. Fusion* **2023**, *97*, 101804. [CrossRef]
8. Mvula, P.K.; Branco, P.; Jourdan, G.V.; Viktor, H.L. A systematic literature review of cyber-security data repositories and performance assessment metrics for semi-supervised learning. *Discov. Data* **2023**, *1*, 4. [CrossRef]
9. Koroniotis, N.; Moustafa, N.; Turnbull, B.; Choo, K.K.R. Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: The Bot-IoT Dataset. In *Proceedings of the Future Generation Computer Systems*; Elsevier: Amsterdam, The Netherlands, 2019.
10. Guhan, N.K.; Ramachandran, M.; Ravindran, S.; Vijean, V. A Deep and Systematic Review of the Intrusion Detection Systems based on Machine Learning and Deep Learning Techniques. In *Proceedings of the 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*; IEEE: New York, NY, USA, 2024; p. 1564. [CrossRef]
11. Bhavyashree, Y.R.; Kavyashree, M.K.; Amrutha, K.R. Systematic Review on Frameworks for Intrusion Detection using Machine Learning and Deep Learning Algorithms. In *Proceedings of the 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*; IEEE: New York, NY, USA, 2024; pp. 1–12. [CrossRef]
12. Ali, T.; Eleyan, A.; Bejaoui, T. Detecting Conventional and Adversarial Attacks Using Deep Learning Techniques: A Systematic Review. In *Proceedings of the 2023 International Symposium on Networks, Computers and Communications (ISNCC)*; IEEE: New York, NY, USA, 2023. [CrossRef]
13. Gamage, S.; Samarabandu, J. Deep learning methods in network intrusion detection: A survey and an objective comparison. *J. Netw. Comput. Appl.* **2020**, *169*, 102767. [CrossRef]
14. Tsai, C.F.; Hsu, Y.F.; Lin, C.Y.; Lin, W.Y. Intrusion detection by machine learning: A review. *Expert Syst. Appl.* **2009**, *36*, 11994–12000. [CrossRef]
15. Pingala Suthishni, D.N.; Kumar, K.S.S. A Review on Machine Learning based Security Approaches in Intrusion Detection System. In *Proceedings of the 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*; IEEE: Piscataway, NJ, USA, 2022; pp. 101–105. [CrossRef]
16. Azmoodeh, A.; Al-Rawi, W.; Al-Dahhan, M.; Ghita, B. Detecting Cyber Attacks in Industrial Control Systems Using Convolutional Neural Networks. *arXiv* **2018**. [CrossRef]
17. Ring, M.; Wunderlich, S.; Scheuring, D.; Landes, D.; Hotho, A. A survey of network-based intrusion detection data sets. *Comput. Secur.* **2019**, *86*, 147–167. [CrossRef]
18. Yang, Z.; Liu, X.; Li, T.; Wu, D.; Wang, J.; Zhao, Y.; Han, H. A systematic literature review of methods and datasets for anomaly-based network intrusion detection. *Comput. Secur.* **2022**, *116*, 102675. [CrossRef]
19. Strom, B.; Applebaum, A.; Miller, D.; Nickels, K.; Pennington, A.; Thomas, C. *MITRE ATT&CK™: Design and Philosophy*; MITRE Corporation: Bedford, MA, USA, 2018.
20. Tavallaee, M.; Bagheri, E.; Lu, W.; Ghorbani, A.A. A detailed analysis of the KDD CUP 99 data set (NSL-KDD). In *Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), Ottawa, ON, Canada, 8–10 July 2009*; IEEE: Piscataway, NJ, USA, 2009; pp. 1–6. [CrossRef]
21. Moustafa, N.; Slay, J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS)*; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6. [CrossRef]
22. Sharafaldin, I.; Lashkari, A.H.; Ghorbani, A.A. *CSE-CIC-IDS2017 Dataset: Intrusion Detection Evaluation Dataset*; Canadian Institute for Cybersecurity (CIC), University of New Brunswick: Fredericton, NB, Canada, 2017.
23. Gad, A.R.; Nashat, A.A.; Barkat, T.M. Intrusion Detection System Using Machine Learning for Vehicular Ad Hoc Networks Based on ToN-IoT Dataset. *IEEE Access* **2021**, *9*, 142206–142217. [CrossRef]

24. Sowmya, T.; Mary Anita, E.A. A Comprehensive Review of AI Based Intrusion Detection System. *Meas. Sens.* **2023**, *28*, 100827. [CrossRef]
25. Salem, A.H.; Azzam, S.M.; Emam, O.E.; Abohany, A.A. Advancing Cybersecurity: A Comprehensive Review of AI-Driven Detection Techniques. *J. Big Data* **2024**, *11*, 105. [CrossRef]
26. Ofusori, L.; Bokaba, T.; Mhlongo, S. Artificial Intelligence in Cybersecurity: A Comprehensive Review and Future Direction. *Appl. Artif. Intell.* **2024**, *38*, 2439609. [CrossRef]
27. Rehman, H.M.R.U.; Liaquat, S.; Gul, M.J.; Jhandir, M.Z.; Gavilanes, D.; Masias Vergara, M.; Ashraf, I. A Systematic Literature Study of Machine Learning Techniques Based Intrusion Detection: Datasets, Models, Challenges, and Future Directions. *J. Big Data* **2025**, *12*, 264. [CrossRef]
28. Hozouri, A.; Mirzaei, A.; Effatparvar, M. A Comprehensive Survey on Intrusion Detection Systems with Advances in Machine Learning, Deep Learning and Emerging Cybersecurity Challenges. *Discov. Artif. Intell.* **2025**, *5*, 314. [CrossRef]
29. Dobler, M.; Hellwig, M.; Lopes, N.; Oakley, K.; Winterburn, M. Systematic Review and Characterisation of Malicious Industrial Network Traffic Datasets. *Int. J. Inf. Secur.* **2025**, *24*, 208. [CrossRef]
30. Alnabhan, M.Q.; Branco, P. Fake News Detection Using Deep Learning: A Systematic Literature Review. *IEEE Access* **2024**, *12*, 114435–114459. [CrossRef]
31. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Technical Report EBSE-2007-01; Keele University: Staffordshire, UK, 2007.
32. Zhu, B.; Joseph, A.; Sastry, S. A Taxonomy of Cyber Attacks on SCADA Systems. In *Proceedings of the 2011 IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing*; IEEE: Piscataway, NJ, USA, 2011; pp. 380–385.
33. Ozkan Okay, M.; Iliev, T.; Akin, E.; Aslan, O.; Kosunalp, S.; Stoyanov, I.; Beloev, I. A Comprehensive Survey: Evaluating the Efficiency of Artificial Intelligence and Machine Learning Techniques on Cyber Security Solutions. *IEEE Access* **2024**, *12*, 12229–12255. [CrossRef]
34. Wu, M.; Moon, Y.B. Taxonomy of Cross-Domain Attacks on CyberManufacturing System. *Procedia Comput. Sci.* **2017**, *114*, 367–374. [CrossRef]
35. Wu, M.; Moon, Y.B. Taxonomy for secure cybermanufacturing systems. In *Proceedings of the ASME International Mechanical Engineering Congress and Exposition Proceedings*; ASME: New York, NY, USA, 2018; Volume 2, pp. 1–10. [CrossRef]
36. Pan, Y.; White, J.; Schmidt, D.C.; Elhabashy, A.; Sturm, L.; Camelio, J.; Williams, C. Taxonomies for Reasoning About Cyber-physical Attacks in IoT-based Manufacturing Systems. *Int. J. Interact. Multimed. Artif. Intell.* **2017**, *4*, 45–54. [CrossRef]
37. Tuptuk, N.; Hailes, S. Security of smart manufacturing systems. *J. Manuf. Syst.* **2018**, *47*, 93–106. [CrossRef]
38. Wu, D.; Ren, A.; Zhang, W.; Fan, F.; Liu, P.; Fu, X.; Terpenny, J. Cybersecurity for digital manufacturing. *J. Manuf. Syst.* **2018**, *48*, 3–12. [CrossRef]
39. Yampolskiy, M.; King, W.E.; Gatlin, J.; Belikovetsky, S.; Brown, A.; Skjellum, A.; Elovici, Y. Security of additive manufacturing: Attack taxonomy and survey. *Addit. Manuf.* **2018**, *21*, 431–457. [CrossRef]
40. Elhabashy, A.E.; Wells, L.J.; Camelio, J.A.; Woodall, W.H. A cyber-physical attack taxonomy for production systems: A quality control perspective. *J. Intell. Manuf.* **2019**, *30*, 2489–2504. [CrossRef]
41. Barnum, S. *Common Attack Pattern Enumeration and Classification (CAPEC) Schema Description*; Technical Report; MITRE Corporation: McLean, VA, USA, 2008.
42. Hansman, S.; Hunt, R. A taxonomy of network and computer attacks. *Comput. Secur.* **2005**, *24*, 31–43. [CrossRef]
43. Meyers, C.A.; Powers, S.S.; Faissol, D.M. *Taxonomies of Cyber Adversaries and Attacks: A Survey of Incidents and Approaches*. Technical Report; Lawrence Livermore National Laboratory: Livermore, CA, USA, 2009.
44. Chapman, I.M.; Leblanc, S.P.; Partington, A. Taxonomy of cyber attacks and simulation of their effects. In *Proceedings of the Military Modeling and Simulation Symposium*; The Society for Modeling and Simulation International (SCS): San Diego, CA, USA, 2011; pp. 73–80.
45. Simmons, C.B.; Shiva, S.G.; Bedi, H.; Dasgupta, D. AVOIDIT: A Cyber Attack Taxonomy. In *Proceedings of the 9th Annual Symposium on Information Assurance*; University at Albany, State University of New York: Albany, NY, USA, 2014; pp. 2–12.
46. Emmert-Streib, F.; Dehmer, M. Taxonomy of machine learning paradigms: A data-centric perspective. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2022**, *12*, e1470. [CrossRef]
47. von Rueden, L.; Mayer, S.; Beckh, K.; Georgiev, B.; Giesselbach, S.; Heese, R.; Kirsch, B.; Pfrommer, J.; Pick, A.; Bauckhage, C.; et al. Informed machine learning—A taxonomy and survey of integrating knowledge into learning systems. *arXiv* **2019**, arXiv:1903.12394. Available online: https://arxiv.org/abs/1903.12394 (accessed on 11 May 2026). [CrossRef]
48. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
49. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
50. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

51. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
52. Dietterich, T.G. Ensemble methods in machine learning. In *Multiple Classifier Systems*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15. [CrossRef]
53. Ozlem, M.; Turk, A.; Yavuz, A. A review on cyber security dataset for machine learning algorithms. In *Proceedings of the IEEE International Conference on Big Data (Big Data)*; IEEE: Piscataway, NJ, USA, 2017; pp. 2186–2193. [CrossRef]
54. Mississippi State University. *Gas Pipeline Intrusion Dataset*; Mississippi State University, Critical Infrastructure Protection Center: Starkville, MS, USA, 2020. Available online: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets (accessed on 13 January 2025).
55. Stolfo, S.; Fan, W.; Lee, W.; Prodromidis, A.; Chan, P. KDD Cup 1999 Data. UCI Machine Learning Repository. 1999. Available online: https://archive.ics.uci.edu/dataset/130/kdd+cup+1999+data (accessed on 13 February 2025).
56. Sharafaldin, I.; Lashkari, A.H.; Ghorbani, A.A. *CSE-CIC-IDS2018 Dataset*; Canadian Institute for Cybersecurity (CIC), University of New Brunswick: Fredericton, NB, Canada, 2018.
57. Sharafaldin, I.; Lashkari, A.H.; Hakak, S.; Ghorbani, A.A. Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In *Proceedings of the 2019 International Carnahan Conference on Security Technology (ICCST), Chennai, India, 1–3 October 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8. [CrossRef]
58. Damasevicius, R.; Venckauskas, A.; Grigaliunas, S.; Toldinas, J.; Morkevicius, N.; Aleliunas, T.; Smuikys, P. LITNET-2020: An annotated real-world network flow dataset for network intrusion detection. *Electronics* **2020**, *9*, 800. [CrossRef]
59. Barut, O.; Luo, Y.; Zhang, T.; Li, W.; Li, P. NetML: A challenge for network traffic analytics. *arXiv* **2020**, arXiv:2004.13006. [CrossRef]
60. Samarakoon, S.; Siriwardhana, Y.; Porambage, P.; Liyanage, M.; Chang, S.; Kim, J.; Kim, J.; Ylianttila, M. *5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network*; IEEE Dataport; IEEE: Piscataway, NJ, USA, 2022. [CrossRef]
61. Kumar, P.; Liu, J.; Tayeen, A.S.M.; Misra, S.; Cao, H.; Harikumar, J.; Perez, O. FLNET2023: Realistic Network Intrusion Detection Dataset for Federated Learning. In *Proceedings of the Proceedings of the MILCOM 2023–IEEE Military Communications Conference (MILCOM), Boston, MA, USA, 30 October–3 November 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 345–350. [CrossRef]
62. Haider, W.; Hu, J.; Slay, J.; Turnbull, B.; Xie, Y. Generating realistic intrusion detection system dataset based on fuzzy qualitative modeling. *J. Netw. Comput. Appl.* **2017**, *87*, 185–192. [CrossRef]
63. CAIDA. *The CAIDA DDoS Attack 2007 Dataset*; Center for Applied Internet Data Analysis, University of California San Diego: San Diego, CA, USA, 2007. Available online: https://www.caida.org/catalog/datasets/ddos-20070804_dataset/ (accessed on 10 May 2025).
64. BoNeSi—The DDoS Botnet Simulator. 2020. Available online: https://github.com/Markus-Go/bonesi (accessed on 26 February 2022).
65. Jonker, M.; Sperotto, A.; Pras, A. DDoSDB dataset: DDoS Mitigation—A Measurement-Based Approach. In *Proceedings of the NOMS 2020—IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 20–24 April 2022*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6. [CrossRef]
66. Samiullah, H. Application Layer DoS Attack Dataset. 2021. Available online: https://www.kaggle.com/hamzasamiullah/ml-analysis-application-layer-dos-attack-dataset (accessed on 30 August 2021).
67. Giménez, C.T.; Villegas, A.P.; Marañón, G.Á. *HTTP Data Set CSIC 2010*; Information Security Institute of CSIC (Spanish Research National Council): Madrid, Spain, 2010.
68. Raïssi, C.; Brissaud, J.; Dray, G.; Poncelet, P.; Roche, M.; Teisseire,M. Web Analyzing Traffic Challenge: Description and Results. In Proceedings of the ECML PKDD 2007 Discovery Challenge, Warsaw, Poland, 17–21 September 2007.
69. Digital Corpora. NPS-2009-Casper-Rw Dataset. 2009. Available online: https://digitalcorpora.org/corpora/disk-images/ (accessed on 11 May 2026).
70. Hostiadi, D.P.; Ahmad, T. Dataset for Botnet group activity with adaptive generator. *Data Brief* **2021**, *38*, 107334. [CrossRef] [PubMed]
71. Putra, M.A.R.; Hostiadi, D.P.; Ahmad, T. Botnet dataset with simultaneous attack activity. *Data Brief* **2022**, *45*, 108628. [CrossRef]
72. Tayfour, O.E.; Mubarakali, A.; Tayfour, A.E.; Marsono, M.N.; Hassan, E.; Abdelrahman, A.M. Adapting deep learning-LSTM method using optimized dataset in SDN controller for secure IoT. *Soft Comput.* **2023**, *27*, 1–9. [CrossRef]
73. Benign and Malicious Domains Based on DNS Logs. Benign and Malicious Domains Based on DNS Logs Dataset. 2022. Available online: https://data.mendeley.com/datasets/623sshkdrz/5 (accessed on 24 May 2022).
74. Garcia, S.; Grill, M.; Stiborek, J.; Zunino, A. An empirical comparison of botnet detection methods. In *Proceedings of the 2014 IEEE 32nd International Conference on Performance, Computing and Communications Conference (IPCCC)*; Elsevier: Amsterdam, The Netherlands, 2014; pp. 1–8. [CrossRef]
75. Wang, W.; Zhu, M.; Zeng, X.; Ye, X.; Sheng, Y. Malware traffic classification using convolutional neural network for representation learning. In *2017 International Conference on Information Networking (ICOIN)*; IEEE: Piscataway, NJ, USA, 2017; pp. 712–717. [CrossRef]

76. Meidan, Y.; Bohadana, M.; Mathov, Y.; Mirsky, Y.; Shabtai, A.; Breitenbacher, D.; Elovici, Y. N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders. *IEEE Pervasive Comput.* **2018**, *17*, 12–22. [CrossRef]
77. Pa, Y.M.P.; Suzuki, S.; Yoshioka, K.; Matsumoto, T.; Kasama, T.; Rossow, C. IoTPOT: Analysing the rise of IoT compromises. In Proceedings of the 9th USENIX Workshop on Offensive Technologies (WOOT), Washington, DC, USA, 10–11 August 2014.
78. García, S.; Shuvaev, S.; Uritskaya, A. *IoT-23: A Labeled Dataset with Malicious and Benign IoT Network Traffic*; Stratosphere Laboratory, Czech Technical University: Prague, Czech Republic, 2020.
79. Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. *IEEE Access* **2022**, *10*, 40281–40306. [CrossRef]
80. Dadkhah, S.; Mahdikhani, H.; Danso, P.K.; Zohourian, A.; Truong, K.A.; Ghorbani, A.A. Towards the development of a realistic multidimensional IoT profiling dataset. In *Proceedings of the 19th Annual International Conference on Privacy, Security and Trust (PST)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–11. [CrossRef]
81. Neto, E.C.P.; Dadkhah, S.; Ferreira, R.; Zohourian, A.; Lu, R.; Ghorbani, A.A. CicIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment. *Sensors* **2023**, *23*, 5941. [CrossRef] [PubMed]
82. Hamza, A.; Gharakheili, H.H.; Benson, T.A.; Sivaraman, V. UNSW IoT Traffic Attack Dataset. In *Proceedings of the 2019 ACM Symposium on SDN Research (SOSR)*; ACM: New York, NY, USA, 2019; pp. 36–48. [CrossRef]
83. Aramini, A.; Arazzi, M.; Facchinetti, T.; Ngankem, L.S.; Nocera, A. Distributed IoT Traffic Attack Dataset. In *Proceedings of the 2022 IEEE 18th International Conference on Factory Communication Systems (WFCS)*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–8. [CrossRef]
84. Emec, M. ROUT-4-2023: RPL Based Routing Attack Dataset for IoT. IEEE Dataport. 2023. Available online: https://ieee-dataport.org/documents/rout-4-2023-rpl-based-routing-attack-dataset-iot (accessed on 14 June 2024).
85. Mirsky, Y.; Doitshman, T.; Elovici, Y.; Shabtai, Y. Kitsune: An ensemble of autoencoders for online network intrusion detection. *arXiv* **2018**, arXiv:1802.09089. [CrossRef]
86. Samson, K. Wi-Fi Association and Disassociation Dataset. 2023. Available online: https://github.com/samsonkg/Wi-Fi-Association_Disassociation-Dataset (accessed on 26 August 2023).
87. Pazul, K. Controller Area Network (CAN) Basics, 1999. Available online: https://cika.com/soporte/Information/Microchip/AnalogInterface/CAN/AppNotes/AN713(DS00713a).pdf (accessed on 20 May 2025).
88. Song, H.M.; Woo, J.; Kim, H.K. In-vehicle network intrusion detection using deep convolutional neural network. *Veh. Commun.* **2020**, *21*, 100198. [CrossRef]
89. Nataraj, L.; Karthikeyan, S.; Jacob, G.; Manjunath, B.S. Malware images. In *Proceedings of the 8th International Symposium on Visualization for Cyber Security (VizSec), Pittsburgh, PA, USA, 20 July 2011*; ACM: New York, NY, USA, 2011; pp. 1–7. [CrossRef]
90. Ronen, R.; Radu, M.; Feuerstein, C.; Yom-Tov, E.; Ahmadi, M. Microsoft Malware Classification Challenge. *arXiv* **2018**, arXiv:1802.10135. [CrossRef]
91. Bozkir, A.S.; Cankaya, A.O.; Aydos, M. Utilization and Comparison of Convolutional Neural Networks in Malware Recognition. In *Proceedings of the 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 24–26 April 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4. [CrossRef]
92. Nappa, A.; Rafique, M.Z.; Caballero, J. The MALICIA dataset: Identification and analysis of drive-by download operations. *Int. J. Inf. Secur.* **2015**, *14*, 15–33. [CrossRef]
93. Drebin: Android Malware Dataset. 2014. Available online: https://drebin.mlsec.org/ (accessed on 16 July 2025).
94. VX Heavens. 2021. Available online: https://vx-underground.org/Archive (accessed on 6 July 2021).
95. VirusShare Dataset. 2021. Available online: https://virusshare.com/ (accessed on 6 July 2021).
96. VirusTotal. *VirusTotal: Free Online Virus, Malware and URL Scanner*. Available online: https://www.virustotal.com/ (accessed on 11 May 2026).
97. Ullah, I.; Ahmad, J.; Ahmed, I.; Amin, R.; Imran, M. CIC-MalMem-2022: Malware detection in memory dumps using machine learning. In *Proceedings of the 2022 International Conference on Cyber Security and Resilience (CSR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 153–159. [CrossRef]
98. Maniriho, P.; Mahmood, A.N.; Chowdhury, M.J.M.C. MeMalDet: A Memory Analysis-Based Malware Detection Framework Using Deep Autoencoders and Stacked Ensemble under Temporal Evaluations. *Comput. Secur.* **2024**, *142*, 103864. [CrossRef]
99. Canadian Institute for Cybersecurity (CIC). CIC-CMD-2024: Command and Control Malware Dataset. 2024. Available online: https://www.kaggle.com/datasets/datasetengineer/cybertec-iiot-malware-dataset-cimd-2024 (accessed on 11 May 2026).
100. Hopkins, M.; Reeber, E.; Forman, G.; Suermondt, J. Spambase Dataset. UCI Machine Learning Repository. 1999. Available online: https://archive.ics.uci.edu/dataset/94/spambase (accessed on 11 May 2026).
101. Biswas, B. Email Spam Classification Dataset CSV. 2020. Available online: https://www.kaggle.com/balaka18/email-spam-classification-dataset-csv (accessed on 5 May 2022).
102. Nitisha. Email Spam Dataset. 2020. Available online: https://www.kaggle.com/nitishabharathi/email-spam-dataset (accessed on 1 May 2022).

103. Naidu, C. Spam Classification for Basic NLP. 2021. Available online: https://kaggle.com/chandramoulinaidu/spam-classification-for-basic-nlp (accessed on 15 January 2022).
104. Murthy, M.Y.B.; Mastanbi, S.; Sujitha, B.; Babu, K.R. Evaluating deep learning algorithms for natural language processing. In *Algorithms for Intelligent Systems*; Springer Nature: Singapore, 2023; pp. 709–720.
105. Kaggle. Phishing Email Collection. 2020. Available online: https://www.kaggle.com/datasets/akashsurya156/phishing-paper1 (accessed on 11 May 2026).
106. Jagtap, S. Kaggle Bot Account Detection Dataset. Available online: https://www.kaggle.com/datasets/shriyashjagtap/kaggle-bot-account-detection/data (accessed on 11 May 2026).
107. MITRE. Sharing Threat Intelligence Just Got a lot Easier. 2018. Available online: https://oasis-open.github.io/cti-documentation/stix/intro (accessed on 31 December 2022).
108. Zeng, V.; Baki, S.; Aassal, A.E.; Verma, R.; Moraes, L.F.T.D.; Das, A. Diverse datasets and a customizable benchmarking framework for phishing. In *Proceedings of the Proceedings 6th International Workshop on Security and Privacy Analytics, New Orleans, LA, USA, 18 March 2020*; ACM: New York, NY, USA, 2020; pp. 35–41.
109. Chiew, K.L.; Chang, E.H.; Tan, C.L.; Abdullah, J.; Yong, K.C. Building standard offline anti-phishing dataset for benchmarking. *Int. J. Eng. Technol.* **2018**, *7*, 71–74. [CrossRef]
110. Ariyadasa, S.; Fernando, S.; Fernando, S. Phishing Websites Dataset. Mendeley Data. 2021. Available online: https://data.mendeley.com/datasets/n96ncsr5g4/1 (accessed on 10 May 2025).
111. Bahnsen, A.C.; Bohorquez, E.C.; Villegas, S.; Vargas, J.; Gonzalez, F.A. Classifying phishing URLs using recurrent neural networks. In *Proceedings of the Proceedings APWG Symposium on Electronic Crime Research (eCrime), Scottsdale, Arizona, USA, 25–27 April 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 1–8.
112. Mukherjee, A.; Venkataraman, V.; Liu, B.; Glance, N. What yelp fake review filter might be doing? In *Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM)*; AAAI: Palo Alto, CA, USA, 2013; pp. 409–418.
113. Rayana, S.; Akoglu, L. Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*; ACM: New York, NY, USA, 2015; pp. 985–994. [CrossRef]
114. Wang, W.; Harrou, F.; Bouyeddou, B.; Senouci, S.M.; Sun, Y. A stacked deep learning approach to cyber-attacks detection in industrial systems: Application to power system and gas pipeline systems. *Clust. Comput.* **2022**, *25*, 561–578. [CrossRef]
115. Mathur, A.P.; Tippenhauer, N.O. SWaT: A water treatment testbed for research and training on ICS security. In *Proceedings of the 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*; IEEE: Piscataway, NJ, USA, 2016. [CrossRef]
116. Mubarak, S.; Habaebi, M.H.; Islam, M.R.; Balla, A.; Tahir, M. Industrial datasets with ICSs testbed and attack detection using machine learning techniques. *Intell. Autom. Soft Comput.* **2022**, *31*, 1345–1360. [CrossRef]
117. Quincozes, S.E.; Albuquerque, C.; Passos, D.G.; Mossé, D. ERENO: A framework for generating realistic IEC-61850 intrusion detection datasets for smart grids. *IEEE Trans. Dependable Secur. Comput.* **2023**, *21*, 3851–3865. [CrossRef]
118. Xu, Z. *IEEE 118-Bus, 300-Bus and 3266-Bus System Dataset for Unit Commitment*; IEEE DataPort; IEEE: Piscataway, NJ, USA, 2020. [CrossRef]
119. Software Engineering Institute (CERT Division), Carnegie Mellon University. Insider Threat Test Dataset (Versions r4–r6). Data Set, 2020. Synthetic Insider Threat Logs, Including Releases r4.x Through r6.x. Available online: https://kilthub.cmu.edu/articles/dataset/Insider_Threat_Test_Dataset/12841247/1 (accessed on 10 May 2025).
120. Udacity. An Open Source Self-Driving Car. 2016. Available online: https://www.udacity.com/ (accessed on 10 May 2025).
121. Stallkamp, J.; Schlipsing, M.; Salmen, J.; Igel, C. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Netw.* **2012**, *32*, 323–332. [CrossRef] [PubMed]
122. Unmanned Aerial Vehicle (UAV) Intrusion Detection. 2020. Available online: https://archive.ics.uci.edu/dataset/564/unmanned+aerial+vehicle+uav+intrusion+detection (accessed on 10 May 2025).
123. Consumer Complaint Database. 2019. Available online: https://catalog.data.gov/dataset/consumer-complaint-database (accessed on 10 May 2025).
124. TensorFlow Speech Recognition Challenge. 2019. Available online: https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/data (accessed on 11 January 2025).
125. IMDB Dataset of 50K Movie Reviews. 2019. Available online: https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews (accessed on 1 May 2024).
126. Krizhevsky, A. *Learning Multiple Layers of Features from Tiny Images*; Technical report; Citeseer: University Park, PA, USA, 2009.