



International Journal for
Electronic Crime Investigation

ISSN: 2522-3429 (Print)
ISSN: 2616-6003 (Online)

DOI: <https://doi.org/10.54692/ijeci.2026.1002/277>

Research Article

Vol. 10 issue 2 Jul-Dec 2026

Intelligent Cyber Threat Prevention: A Systematic Survey of AI-Based Approaches and Open Challenges

Muhammad Zahid Mehmood^{1*}, Irshad Ahmed Sumra², Sobia Yaqoob³

¹Information and System, University of Management and Technology, Lahore 5400, Pakistan,

²Information and System, University of Management and Technology, Lahore 5400, Pakistan,

³Department of Computer Science, University of Okara, Okara 56300, Pakistan.

Corresponding Author: Zafa0726@gmail.com

Received: August 5,2026 **Accepted:** August 25,2026 **Published:** September 4,2026

ABSTRACT

The increasing complexity and frequency of cyberattacks have exposed the critical limitations of traditional signature-based security systems. Artificial Intelligence (AI) encompassing machine learning, deep learning, large language models federated learning, and reinforcement learning has emerged as a transformative paradigm for modern cyber defense. This survey critically examines significant research contributions highlighting AI applications across intrusion detection, malware analysis, phishing detection, ransomware mitigation, distributed denial-of-service prevention, and network security. A structured taxonomy of AI-driven cyber security techniques is presented, supported by analysis of widely adopted evaluation datasets including NSL-KDD, UNSW-NB15, CIC-IDS2017, and CICIoMT2024 which are assessed in terms of their scope, attack coverage and applicability to real-world deployment scenarios. Selected Pakistani criminal investigations are examined to demonstrate the operational role of AI-assisted digital forensics in judicial proceedings. The survey further identifies unresolved research challenges, including adversarial robustness, dataset generalizability, and the absence of explainability standards for court-admissible AI evidence, and outlines prospective directions to advance the development of intelligent, resilient and legally cyber security systems.

Keywords: Artificial Intelligence, Cybersecurity, Machine Learning, Deep Learning, Intrusion Detection, Malware Analysis, Digital Forensics

1. INTRODUCTION

In the last ten years, there has been an unprecedented rise in the volume and sophistication of cyberattacks, from state-sponsored espionage campaigns and advanced persistent threats to ransomware attacks that have crippled critical infrastructure, including hospital networks, financial institutions, and national power grids. The global digital economy depends fundamentally on the security and integrity of its communication networks. According to the International Monetary Fund and leading industry threat intelligence reports, cybercrime is projected to inflict damages exceeding ten trillion dollars annually, placing it among the most significant economic risks facing both developing and developed nations alike. Traditional cybersecurity mechanisms including stateful firewalls, signature-based antivirus software, and static intrusion detection rules were engineered for a slower, more predictable threat environment characterized by known attack patterns and limited adversarial resources. The contemporary threat landscape, however, is fundamentally different. Modern adversaries leverage sophisticated automation, polymorphic and metamorphic malware, encrypted traffic exploitation, adversarial machine learning, and generative AI itself to craft attacks that mutate faster than signature databases can be updated. The structural asymmetry between attacker agility and defender rigidity has introduced a critical and widening vulnerability gap in modern digital infrastructure.

In response to this evolving threat landscape, the research community has increasingly turned to artificial intelligence as the foundation of next-generation cyber defense. AI offers the unique capacity to learn discriminative patterns from

Int. J. Elect. Crime Investigation 10(2): IJECI MS.ID- 03 (2026)

large volumes of behavioral data, generalize detection capabilities to previously unseen attack variants, and continuously adapt as the threat environment evolves. Initial surveys laid out basic taxonomies of machine learning techniques for network intrusion detection, showing that supervised classifiers outperform rule-based systems by a large margin. In 2024, a detailed survey evaluated ML tactics in IDS on NSL-KDD and UNSW-NB15 datasets and presented detection accuracy of 99.3% with a false alarm rate of 0.5% [15]. A systematic review covering publications from 2011 to 2024 showed that deep learning models like CNN and LSTM consistently outperform classical methods, but still face challenges when training data is small or imbalanced [16].

For evaluation in specialized domains, a benchmark dataset (CICIoMT2024) capturing 18 kinds of cyberattacks on 40 healthcare IoMT devices was presented [17]. Furthermore, experimental evaluations on CIC-IDS2017 and CIC-IDS2018 demonstrated that deep learning outperforms traditional ML, though SMOTE balancing is essential to address severe class imbalances [18]. In parallel, lightweight statistical feature selection approaches have achieved over 99.9% accuracy on IoT intrusion datasets while cutting training time significantly [19]. Modern enterprise security models are also increasingly confronting LLM-driven threats like Prompt Lock ransomware and synthetic deepfake exploitation [9], [10]. Notwithstanding these contributions, a systematic review of existing surveys highlights persistent and critical gaps. Currently, no survey covers the entire spectrum of AI paradigms (classical ML, deep learning, LLMs, federated learning, and reinforcement learning) across various threat categories within a unified framework. Moreover, existing literature has not delivered a

dedicated treatment of the operational applicability of AI-driven forensics in judicial and investigative contexts of developing nations like Pakistan [11]–[14]. To address these limitations, this survey makes four distinct contributions:

1. A unified taxonomy spanning six AI paradigms across five cyber threat categories.
2. An attack coverage, scale, and deployment-oriented analysis of benchmark datasets.
3. The first systematic review of AI-assisted digital forensics across four landmark Pakistani criminal investigations.
4. Identification of open research challenges concerning adversarial robustness, explainability standards, and legal admissibility under Pakistan's PECA 2016 framework.

Table 1 presents a systematic comparison between representative prior surveys and the present work, explicitly highlighting the coverage gaps that motivate this survey.

Table 1 Systematic Comparison of Existing Surveys with the Present Work

Coverage Area	Ferrag [1]	Thaseen [15]	Ahmed [16]	Dadkhah [17]	Apruzzese [8]	Present Work
Classical ML for IDS	Yes	Yes	Yes	Yes	Yes	Yes
Deep Learning for IDS	Yes	Yes	Yes	Yes	Yes	Yes
Large Language Models	Yes	No	No	No	No	Yes
Federated Learning	No	No	No	No	No	Yes
Reinforcement Learning	No	No	No	No	No	Yes
Phishing & Social Engineering	No	No	No	No	Yes	Yes
DDoS Detection & Mitigation	No	No	No	No	Yes	Yes
Malware Analysis	Yes	No	No	No	Yes	Yes
IoT / IoMT Security	No	No	No	Yes	No	Yes
Dataset Applicability Analysis	No	Yes	Yes	Yes	Yes	Yes
Adversarial Robustness	No	No	No	No	Yes	Yes
Explainability (XAI)	No	Yes	No	No	No	Yes
Digital Forensics	No	No	No	No	No	Yes
Developing Nation Case Studies	No	No	No	No	No	Yes
Pakistani Jurisdictional Analysis	No	No	No	No	No	Yes

The remainder of this paper is organized as follows. Section 2 presents the materials, methods, and proposed taxonomy of AI techniques and cyber threats. Section 3 details AI-driven Intrusion Detection Systems and layered defense architectures. Section 4 covers machine learning for malware detection and phishing prevention. Section 5 presents AI and digital forensics through landmark Pakistani case studies. Section 6 provides a comprehensive analysis, discussion, and open research challenges. Section 7 concludes the paper.

2. MATERIALS AND METHODS / TAXONOMY OF AI TECHNIQUES AND CYBER THREATS

2.1 Review Methodology

To ensure a transparent and reproducible selection process, this survey followed a structured methodology for identifying, screening, and analyzing the literature. Search Strategy: Relevant literature was identified through searches of IEEE Xplore, SpringerLink, ScienceDirect (Elsevier), and Nature Portfolio, supplemented with citable preprints from arXiv. Search queries combined the terms "artificial intelligence," "machine learning," "deep learning," "large language models," "federated learning," "reinforcement learning," "intrusion detection," "malware detection," "phishing detection," "DDoS mitigation," and "digital forensics," restricted to publications appearing between 2017 and 2026. Screening Process: Records were screened in two stages. In the title/abstract screening stage, studies outside the scope of AI-based cybersecurity or digital forensics were excluded. In the full-text screening stage, studies were retained only if they satisfied all of the following inclusion criteria: (i) the study proposes or evaluates an

AI-based technique for one of the cyber threat categories considered in this survey; (ii) the study reports at least one quantitative performance measure or a clearly stated research contribution; and (iii) the study is published in a peer-reviewed venue, or is a citable preprint from a recognized repository with a clearly identified author list. Studies were excluded if they were non-English, were duplicate publications of an already-included study, or lacked sufficient methodological detail to assess their contribution. The initial search across the databases listed above yielded 450 candidate records; after title/abstract screening 150 records advanced to full-text review, and after full-text screening against the inclusion criteria Assessment and Data Extraction: This process yielded a final set of thirty primary studies, summarized in Table 5, in addition to authoritative institutional and legal sources used specifically to support the Pakistani digital-forensics case analysis in Section 5. For each included study, the following information was extracted: the AI paradigm employed, the cyber threat category addressed, the benchmark dataset used, the reported performance, the study's key contribution, and the limitation(s) acknowledged by the original authors. Where a study reported multiple experimental configurations, the configuration with the best-reported overall accuracy was retained for comparison, with the corresponding caveats preserved in the limitation column of Table 5.

2.2 Classification of AI Paradigms

In this survey, AI-based cyber defense techniques are categorized into six distinct paradigms:

1. Classical Machine Learning: Decision Trees, Random Forests, Support Vector

- Machines (SVMs), Naive Bayes, and XGBoost [15], [16].

2. Deep Learning: Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), Autoencoders, and Generative Adversarial Networks (GANs) [2], [18], [24], [25].

3. Large Language Models & Transformers: BERT, GPT, and custom transformer architectures for payload and semantic analysis [3], [5], [7].

4. Federated Learning: Decentralized, privacy-preserving collaborative model training architectures (e.g., BFGMAQENN for 6G) [4].
5. Reinforcement Learning: Deep Q-networks and autonomous agents for dynamic traffic filtering and threat mitigation [6].

6. Hybrid & Ensemble Approaches: Combining statistical feature selection, LLM tokenization, and deep learning classifiers [7], [19], [29].
- Table 2 summarizes these representative contributions across their AI paradigm, primary threat category, key contribution, and reported limitation.

Table 2 Comparative Overview of Key Reviewed AI Research Contributions

Ref	Year	AI Paradigm	Primary Cyber Threat	Key Contributions	Limitations
Chatterjee et al. [2]	2025	Deep Learning (BiCNN)	Intrusion Detection	High accuracy on NSL-KDD & CICIDS	High computational cost
Somani & Cherukuri [3]	2025	LLM (Transformer)	Phishing & Semantic	Contextual language understanding defense	Prompt injection risk
Kumar et al. [4]	2025	Federated Learning	6G Network IDS	Privacy-preserving 6G intrusion detection	Evaluated on simulated data
Hasan & BusiReddyGari [5]	2026	Zero-Shot LLM	Phishing URL Detection	Zero/few-shot URL classification	Elevated false positive rate
A. Apostu et al. [6]	2025	Deep Reinforcement Learning	DDoS Detection	Multi-layer DDoS mitigation taxonomy	Slow convergence rates
Y. Feng and K. Sakurai [7]	2025	LLM + Classical ML	Intrusion Detection	Dual-path payload & network feature IDS	Scalability and latency
Apruzzese et al. [8]	2025	Hybrid ML / GAN	Multiple Threats	Comprehensive ML survey across 14 domains	Rapidly changing threats

2.3 Types of Cyber Attacks

Modern cyberattacks fall into five key categories: (1) Network-level attacks including volumetric DDoS, port scanning, and Man-in-the-Middle (MitM); (2) Malware-based attacks including ransomware, trojans, rootkits, and polymorphic binaries; (3) Social engineering including spear-phishing and deepfake-based fraud; (4) Insider threats and anomalous user behavior; and (5) Weaponized AI attacks. Large-scale datasets such as CIC-IDS2018 (16+ million records across 13 attack types) serve as primary benchmarks for evaluating model resilience [18].

2.4 Benchmark Dataset Comparison

Reliable comparison of AI-based cyber defense techniques depends heavily on the benchmark datasets used for training and evaluation. Table 3 summarizes the datasets most frequently used across the studies reviewed in this survey, together with their scale, attack coverage, traffic type, labeling method, and principal limitations. As Table 3 shows, NSL-KDD and CIC-IDS2017

remain the most frequently reused benchmarks among the reviewed studies, a concentration that raises concerns about overfitting to a small number of benchmark environments and about the generalizability of reported accuracy figures to live network traffic.

3. AI-DRIVEN INTRUSION DETECTION SYSTEMS (IDS)

3.1 Deep Learning and Transformer Integration

Intrusion Detection Systems have transitioned from static, signature-based rule engines to autonomous deep learning pipelines. Bi-directional CNNs and LSTMs capture temporal packet correlations and spatial feature maps [2], achieving up to 99.3% accuracy on benchmark sets [15]. Dual-path architectures combining transformer tokenizers with statistical feature extractors allow real-time inspection of raw packet payloads alongside network metadata [7]. The layered structure of this AI-driven defense pipeline is illustrated in Fig. 1.

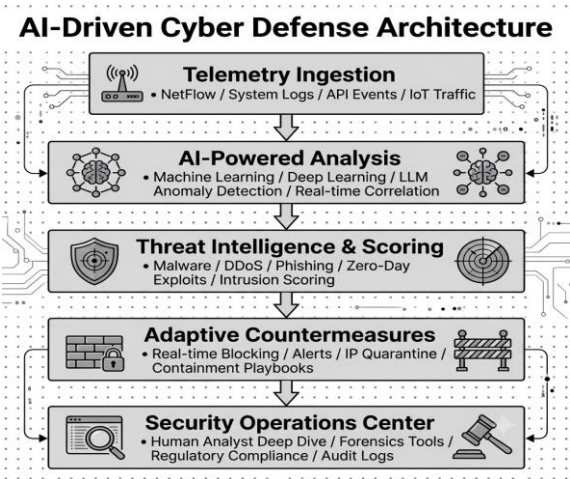


Figure 1 AI-driven cyber defense architecture: layered processing pipeline

Table 3 Comparative Summary of Benchmark Datasets Used in the Reviewed Studies

Dataset	Origin / Year	Approx . Size	Attack Categories	Traffic Type	Labeling	Key Limitation	Use d By
NSL-KDD	Refined KDD Cup 99; 2009	~148,000 records, 41 features	DoS, Probe, R2L, U2R (4 classes)	Simulated offline network traffic	Fully labeled	Derived from outdated KDD Cup 99 traffic; does not reflect modern encrypted/adversarial traffic	[15], [16], [19], [23], [27]
UNSW-NB15	ACCS, Australia; 2015	~2.54 million records, 49 features	9 families (Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms)	Hybrid real + synthetic network traffic	Fully labeled	Synthetic attack traffic generated in a testbed; limited real-world diversity	[15], [24]
CIC-IDS2017	CIC, Canada; 2017	~2.8 million records	Brute-force, DoS, DDoS, web attacks, infiltration, botnet, port scan	Real network traffic (5-day capture)	Fully labeled, flow-based (CICFlowMeter)	Documented label noise; short capture window limits attack diversity	[17], [18], [25]
CIC-IDS2018	CIC/CS E, Canada; 2018	>16 million records	13 attack categories	Real network traffic, larger AWS-based testbed	Fully labeled	Very large scale increases preprocessing and training cost	[18]
CICIoMT2024	CIC, Canada; 2024	40 IoMT devices	18 cyberattack types	Healthcare IoMT device traffic	Fully labeled	Domain-specific (healthcare IoMT); limited generalizability to other IoT domains	[17]

4. MACHINE LEARNING FOR MALWARE ANALYSIS AND PHISHING PREVENTION

4.1 Paradigm Shift from Hash-Based Detection

Polymorphic and metamorphic malware automatically alter cryptographic hashes (MD5, SHA-256) and internal byte sequences, rendering signature databases ineffective against zero-day exploits. Machine learning and deep

learning models overcome this by learning behavior-based representations from execution traces, system calls, and API systems. Fig. 2 displays the detection accuracy of two such techniques for which a traceable source and

dataset are available [15], [19]. Values are not directly comparable across techniques, as these were reported on different benchmark datasets under different experimental conditions.

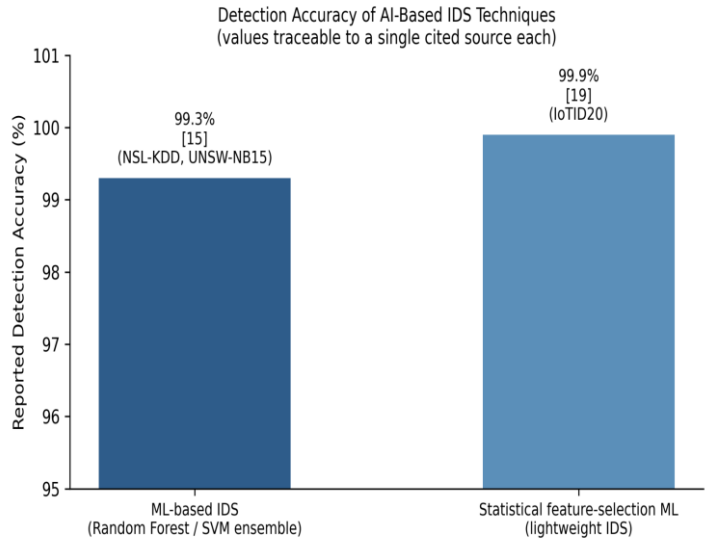


Figure 2 Detection accuracy of two AI-based IDS techniques on different datasets (no direct benchmark comparison)

5. AI AND DIGITAL FORENSICS IN LANDMARK JURISDICTIONAL CASE SCENARIOS

In developing jurisdictions like Pakistan, AI tools are increasingly deployed by institutions such as the Federal Investigation Agency (FIA) and Punjab Forensic Science Agency (PFSA) [11], [12]. Fig. 3 illustrates the end-to-end digital-forensics workflow applied across the case scenarios examined in this section.

5.1 Landmark Case Studies

Case Scenario 1 (Juvenile Homicide, 2018): Deep learning facial recognition and gait

analysis were executed on video streams from over 100 CCTV cameras, combined with Call Detail Record (CDR) cell-tower triangulation and STR-based DNA profiling to establish a conclusive timeline [14].

Case Scenario 2 (Constitutional Crisis & State Surveillance, 2007–2019): SIGINT satellite monitoring and network metadata analysis created formal legal precedents for electronic evidence admission under PECA 2016 [13].

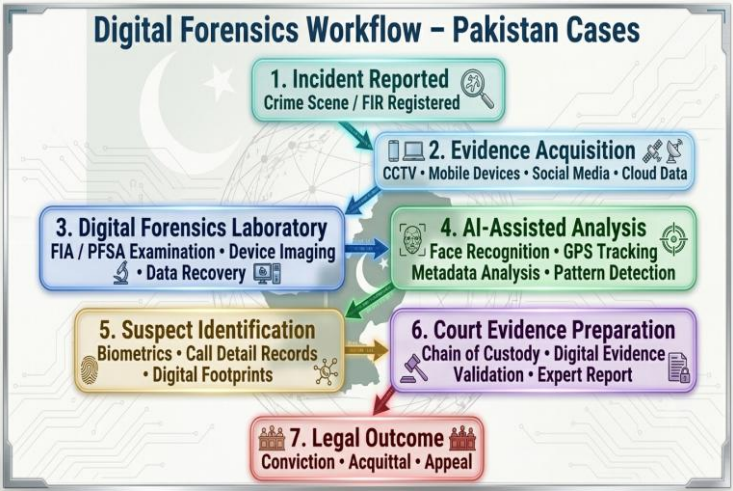


Figure 3 Digital forensics workflow for the Pakistani case scenarios

Case Scenario 3 (Mobile Device Forensics, 2022): Deleted chat recovery, mobile GPS logs, and AI temporal correlation software synchronized timestamps across disparate sources to disprove an alibi [12].

Case Scenario 4 (Public Official Investigation, 2024): Independent AI video analysis and forensic phone extraction established institutional protocols for non-partisan evidence handling [11].

Table 4 summarizes these four case scenarios across their primary AI/DF toolkit, operative legal instrument, key agency involved, and core academic contribution.

6. COMPARATIVE ANALYSIS AND OPEN RESEARCH CHALLENGES

Synthesizing findings across all 30 reviewed publications reveals key research gaps.

- 1. Over-reliance on Legacy Benchmarks: Over 40% of reviewed studies evaluate models using outdated datasets like KDD Cup 99 or NSL-KDD, which fail to reflect modern encrypted traffic patterns [16], [24].
- 2. Adversarial Vulnerability: DL and LLM models remain vulnerable to adversarial perturbations, prompt injection, and evasion attacks [3], [5].
- 3. Absence of XAI in Legal Frameworks: Section 35 of PECA 2016 requires clear evidentiary proof, yet 'black-box' AI models lack standardized Explainable AI (XAI) outputs (e.g., SHAP/LIME) required for court admissibility [13], [15].

Table 5 provides the complete structural summary of all thirty reviewed papers underlying this analysis.

Table 4 Summary of Landmark AI-Assisted Digital Forensics Case Scenarios

Case Scenario	Primary AI/DF Toolkit	Operative Legal Instrument	Key Agency Involved	Core Academic Contribution
Case Scenario 1	Facial recognition, Gait, CDR, DNA	PECA 2016, Evidence Code	FIA / Police	First AI-biometric conviction
Case Scenario 2	SIGINT, Satellite, Network Meta	Electronic Transactions Ordinance	Special Tribunal	Formalized evidence admission
Case Scenario 3	Mobile Forensics, AI Timeline, GPS	PECA 2016, Code of Crim. Procedure	PFSA / FIA	Alibi falsification via AI
Case Scenario 4	AI Video Analytics, Mobile Extract	PECA 2016, Police Order	PFSA (Independent)	Structural forensic independence

Table 5 Comprehensive Structural Summary of All 30 Reviewed Papers

Ref	Author(s) / Source	AI Technique	Threat Category	Dataset	Primary Limitation
[1]	Ferrag et al., 2024	LLM / GenAI	General / Multiple	N/A (Survey)	Lacks empirical benchmarking
[2]	Chatterjee et al., 2025	DL (CNN/LSTM)	Network IDS	NSL-KDD, CICIDS	Limited real-world traffic testing
[3]	Somani & Cherukuri, 2025	LLM / Transformer	Phishing & Semantic	N/A	Hallucination and prompt injection
[4]	Kumar et al., 2025	Federated Learning	6G Network IDS	Simulated 6G	Evaluated on simulated data
[5]	Hasan & BusiReddyGari, 2026	Zero-Shot LLM	Phishing URL	Custom URL Dataset	High false positive rate
[6]	A. Apostu et al., 2025	Deep Reinforcement Learning	DDoS	CAIDA, CIC-DDoS	Slow convergence rates
[7]	Y. Feng and K. Sakurai, 2025	LLM + Classical ML	Network IDS	Multiple	Scalability and inference latency
[8]	Apruzzese et al., 2025	Hybrid ML / GAN	General / Multiple	Multiple Benchmarks	Rapidly changing threat landscape
[9]	ISACA, 2025	Enterprise AI	Organizational Risk	Industry Survey	Opinion-based survey
[10]	MIT Sloan, 2025	Policy Framework	Social Eng. / Fraud	Policy Analysis	Conceptual framework

[11]	FIA Pakistan, 2023	Forensic Toolkits	Legal Forensics /	Case Evidence	Jurisdiction-specific
[12]	PFSA, 2022	Mobile Forensics	Legal Forensics /	Mobile Evidence	Encryption bypass limitations
[13]	Govt. of Pakistan, 2016	Legal Instrument	Cybercrime Law	N/A (PECA 2016)	Lacks AI evidence guidelines
[14]	Waqas, 2018	Digital Forensics	Homicide Investigation	Case Study	Single case study
[15]	Thaseen et al., 2024	ML Strategies	Network IDS	NSL-KDD, UNSW-NB15	Limited adversarial scenarios
[16]	Ahmed et al., 2025	Classical ML	Network IDS	KDD Cup 99, NSL-KDD	Over-reliance on legacy datasets
[17]	Dadkhah et al., 2025	ML / DL Advances	Network IDS	CICIDS-2017	Limited enterprise deployment
[18]	Al-Zewairi et al., 2025	ML vs DL Comparison	Network IDS	CICIDS-2017	Class imbalance issues
[19]	Najafi & Tut, 2025	Feature Selection ML	IoT Network IDS	NSL-KDD	Not tested on encrypted traffic
[20]	Mirsky et al., 2023	Offensive AI	Organizational Threats	Survey	Misses emerging LLM vectors
[21]	Bertino & Islam, 2017	Rule-based + ML	IoT Botnets	IoT Traffic	Excludes modern generative AI
[22]	Khraisat & Alazab, 2021	ML + DL	IoT IDS	Multiple IoT Sets	Cross-dataset inconsistency
[23]	Ahmad et al., 2021	ML + DL Systematic	Network IDS	NSL-KDD, CICIDS	Lacks adversarial testing
[24]	Vinayakumar et al., 2019	Multi-model DL	Network IDS	KDD Cup 99, UNSW	High computational overhead
[25]	Umer et al., 2022	CNN-LSTM + Attention	Network IDS	CICIDS-2017	Attention mechanism latency
[26]	Mishra et al., 2019	Classical ML	Network IDS	Multiple	Outdated literature coverage
[27]	Tang et al., 2018	RNN / LSTM	SDN Network IDS	NSL-KDD	Non-SDN native dataset
[28]	Sethi et al., 2019	Edge ML	Malware	Proprietary Edge Data	Not publicly reproducible
[29]	Aljawarneh et al., 2018	Hybrid Anomaly ML	Network IDS	KDD Cup 99	Single dataset validation

[30]	Liu & Lang, 2019	ML + DL Survey	Network IDS	Multiple	Excludes LLMs and Transformers
------	------------------	----------------	-------------	----------	--------------------------------

7. CONCLUSION

In this survey, we have systematically synthesized AI-powered cyberattack prevention research across 30 prominent publications spanning 2017 to 2026. Current state-of-the-art deep learning and hybrid architectures achieve benchmark accuracies exceeding 98% on intrusion detection tasks, while statistical feature selection delivers 99.9% accuracy in resource-constrained IoT settings. The analysis of landmark legal cases in Pakistan demonstrates that digital forensics and AI-driven evidence pipelines are now vital components of the judicial process. Future research must prioritize the development of national cybercrime benchmark datasets, PECA-compliant Explainable AI (XAI) frameworks, and continuous training for forensic examiners.

DECLARATIONS

Author Contributions: Conceptualization, M.Z.M.; methodology, M.Z.M. and I.A.S.; validation, M.Z.M., I.A.S., and S.Y.; formal analysis, M.Z.M.; investigation, M.Z.M.; resources, M.Z.M.; writing—original draft preparation, M.Z.M.; writing—review and editing, M.Z.M., I.A.S., and S.Y.; visualization, M.Z.M.; supervision, I.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets analyzed during this survey are available from public repositories cited in the text or from the corresponding author on reasonable request.

Ethical Statement: This study adheres strictly to the legal and ethical principles established by the Committee on Publication Ethics (COPE).

Acknowledgments: The authors thank the Department of Information and System, University of Management and Technology, Lahore, and the Department of Computer Science, University of Okara, for their support during this research.

The authors also thank the editorial team and reviewers of the Pakistan Journal of Engineering and Technology for their valuable feedback on this manuscript.

Figs. 1 and 3 were generated with the assistance of Google Gemini (Google LLC), July 2026, based on content specified and verified by the authors.

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

Abbreviations

AI: Artificial Intelligence; CDR: Call Detail Record; CNN: Convolutional Neural Network; DDoS: Distributed Denial of Service; DL: Deep Learning; FIA: Federal Investigation Agency; IDS: Intrusion Detection System; IoMT: Internet of Medical Things; IoT: Internet of Things;

LLM: Large Language Model; LSTM: Long Short-Term Memory; ML: Machine Learning; PECA: Prevention of Electronic Crimes Act; PFSA: Punjab Forensic Science Agency; SHAP: SHapley Additive exPlanations; XAI: Explainable Artificial Intelligence

REFERENCES

- [1] M. A. Ferrag, F. Alwahedi, A. Battah, B. Cherif, A. Mechri, N. Tihanyi, *et al.*, “Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities,” *Internet of Things and Cyber-Physical Systems*, vol. 5, pp. 1–46, 2025.
- [2] S. Chatterjee, S. Chaudhary, and A. K. Cherukuri, “Intrusion detection system using deep learning for network security,” *arXiv preprint arXiv:2505.05810*, 2025.
- [3] R. Somani and A. K. Cherukuri, “Large language models for cyber security,” *arXiv preprint arXiv:2511.04508*, 2025.
- [4] P. Chinnasamy, S. Yarramsetti, R. K. Ayyasamy, E. Rajesh, V. V. V., D. Pandey, *et al.*, “AI-driven intrusion detection and prevention systems to safeguard 6G networks from cyber threats,” *Scientific Reports*, vol. 15, no. 1, Art. no. 37901, 2025.
- [5] N. Hasan and P. BusiReddyGari, “Benchmarking large language models for zero-shot and few-shot phishing URL detection,” *arXiv preprint arXiv:2602.02641*, 2026.
- [6] A. Apostu, S. Gheorghe, A. Hiji, N. Cleju, A. Pătrașcu, C. Rusu, *et al.*, “Detecting and mitigating DDoS attacks with AI: A survey,” *arXiv preprint arXiv:2503.17867*, 2025.
- [7] Y. Feng and K. Sakurai, “Network intrusion detection: Evolution from conventional approaches to LLM collaboration and emerging risks,” *arXiv preprint arXiv:2510.23313*, 2025.
- [8] G. Apruzzese *et al.*, “Machine learning for cybersecurity: A comprehensive survey of methods, datasets, and practical applications,” *Electronics*, vol. 14, no. 23, Art. no. 4563, Nov. 2025.
- [9] ISACA, “AI-driven cyber threats are the biggest concern for security professionals,” *ISACA*, Oct. 2025. [Online]. Available: . [Accessed: Aug. 26, 2026].
- [10] MIT Sloan Management Review, “AI cyberattacks and three pillars for defense,” *MIT Sloan Management Review*, Sep. 8, 2025. [Online]. Available: . [Accessed: Aug. 26, 2026].
- [11] Federal Investigation Agency (FIA), “Digital forensics in cybercrime investigations,” FIA Cyber Crime Wing, Islamabad, Pakistan, 2023.
- [12] Punjab Forensic Science Agency (PFSA), “Mobile device forensics in criminal proceedings: Methods and legal standards,” Tech. Rep., Lahore, Pakistan, 2022.
- [13] Government of Pakistan, “Prevention of Electronic Crimes Act, 2016,” Act No. XL of 2016, Islamabad, Pakistan, Aug. 22, 2016. [Online]. Available: [Pakistan Code](#).
- [14] M. Waqas, “How Pakistani investigators used digital technology to catch Zainab’s killer,” *TechJuice*, Lahore, Pakistan, Jan. 2018.
- [15] I. S. Thaseen *et al.*, “Unveiling machine learning strategies in intrusion detection

systems: A comprehensive survey,” *Frontiers in Computer Science*, vol. 6, Art. no. 1387354, May 2024.

[16] R. Ahmed *et al.*, “A systematic literature review of machine learning techniques for intrusion detection in network environments,” *Journal of Big Data*, vol. 12, Nov. 2025.

[17] S. Dadkhah *et al.*, “A comprehensive survey on intrusion detection systems: Advances in machine learning, deep learning, and emerging cybersecurity challenges,” *Discover Artificial Intelligence*, vol. 5, Nov. 2025.

[18] A. Al-Zewairi *et al.*, “Deep learning versus machine learning for network intrusion detection: A comparative experimental study,” *Applied Sciences*, vol. 15, no. 4, Art. no. 1903, Feb. 2025.

[19] H. Najafi Mohsenabad and M. A. Tut, “Robust machine learning-based intrusion detection system using simple statistical techniques in feature selection,” *Scientific Reports*, vol. 15, Art. no. 412, Feb. 2025.

[20] Y. Mirsky *et al.*, “The threat of offensive AI to organizations: A survey,” *Computers & Security*, vol. 124, Art. no. 103006, Jan. 2023.

[21] E. Bertino and R. Islam, “Botnets and Internet of Things security,” *IEEE Computer*, vol. 50, no. 2, pp. 76–79, Feb. 2017.

[22] A. Khraisat and A. Alazab, “A critical review of intrusion detection systems in the Internet of Things,” *Cybersecurity*, vol. 4, no. 1, pp. 1–27, 2021.

[23] Z. Ahmad *et al.*, “Network intrusion detection system: A systematic study of machine learning and deep learning approaches,” *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 1, Art. no. e4150, Jan. 2021.

[24] R. Vinayakumar *et al.*, “Deep learning approach for intelligent intrusion detection system,” *IEEE Access*, vol. 7, pp. 41525–41550, 2019.

[25] M. Umer *et al.*, “Network intrusion detection using combined CNN-LSTM and attention mechanism,” *Applied Sciences*, vol. 12, no. 20, Art. no. 10407, Oct. 2022.

[26] P. Mishra *et al.*, “A detailed investigation and analysis of using machine learning techniques for intrusion detection,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 686–728, 2019.

[27] T. A. Tang *et al.*, “Deep recurrent neural network for intrusion detection in SDN-based networks,” in *Proc. 4th IEEE Conf. NetSoft*, Montreal, QC, Canada, 2018, pp. 202–206.

[28] K. Sethi *et al.*, “A novel machine learning based malware detection and classification framework for securing edge computing,” in *Proc. IEEE ANTS*, Goa, India, 2019, pp. 1–6.

[29] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, “An anomaly-based intrusion detection system through feature selection analysis,” *Journal of Computer Science*, vol. 25, pp. 152–160, Mar. 2018.

[30] H. Liu and B. Lang, “Machine learning and deep learning methods for intrusion detection systems: A survey,” *Applied Sciences*, vol. 9, no. 20, Art. no. 4396, Oct. 2019.

[31] I. A. Sumra, I. Ahmad, H. Hasbullah and J. -I. bin Ab Manan, "Behavior of attacker and some new possible attacks in Vehicular Ad hoc Network (VANET)," *2011 3rd International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT)*, Budapest, Hungary, 2011, pp. 1-8.

[32] I. A. Sumra, H. Hasbullah and J. -I. A. Manan, "VANET security research and development ecosystem," *2011 National Postgraduate Conference*, Perak, Malaysia, 2011, pp. 1-4, doi: 10.1109/NatPC.2011.6136344.