Glovento Journal of Integrated Studies (GJIS) | ISSN: 3117-3314
Volume 2 (2026) | Article 43

Deep Learning-Based Malware Detection: A Comprehensive Survey of Methods, Achievements, Fundamental Limitations, and Future Directions

[1]Qiu Yuefu, [2]Kazem Chamran

[1,2]City University Malaysia, 46100 Petaling Jaya, Malaysia
Email: [1]cityuqiu@163.com, [2]mohamma.c@imail.sunway.edu.my

Abstract: Malware detection increasingly requires advanced methodologies as evasion mechanisms outpace traditional defenses. This survey examines deep learning's application to malware detection, analyzing both achievements and fundamental limitations. Deep learning substantially improves detection performance through convolutional networks for code analysis, recurrent networks for behavioral patterns, and multi-modal fusion integrating complementary signals. Transfer learning enables effective knowledge transfer with limited labeled data. However, adversarial examples evade detection while preserving functionality, concept drift causes persistent accuracy degradation as malware evolves adaptively, and zero-day detection remains substantially less effective than known malware detection. These vulnerabilities reflect fundamental constraints rather than technical limitations. The attack-defense asymmetry creates inherent offensive advantages impossible to overcome technologically. The accuracy-robustness trade-off prevents simultaneous optimization due to underlying mathematical properties. Concept drift operates as persistent adversarial adaptation exploiting detected weaknesses. The interpretability-accuracy paradox creates tension between performance and regulatory transparency requirements. Advancing the field requires theoretical research pursuing fundamental robustness, organizational implementation of coordinated multi-layer defenses with human-AI collaboration, and policy frameworks establishing realistic expectations with international coordination. This review concludes that deep learning achieves genuine improvements while introducing novel vulnerabilities, that fundamental constraints limit further progress, and that pragmatic security must acknowledge theoretical impossibility of perfect prevention while pursuing continuous improvement through rapid detection and response.

Keywords: Malware Detection, Deep Learning, Adversarial Robustness, Concept Drift, Multi-Modal Fusion, Human-AI Collaboration, Defense-In-Depth, Cybersecurity, Policy Frameworks, Interpretability.

I. INTRODUCTION

1.1 The Evolution and Threats of Malicious Code

Malware has evolved from static, easily detectable threats to sophisticated, AI-empowered attacks capable of evading multiple defense layers. According to security reports, over 450,000 new malware samples are detected daily, reflecting a consistent pattern where threats evolve in response to improved defenses (Yang & Lyu, 2019). This historical progression—from signature-based detection to deep learning integration—explains why traditional security approaches have become inadequate and why advanced defensive technologies are now essential for contemporary cybersecurity. Understanding this evolution is critical for comprehending the emergence of adversarial machine learning attacks and the corresponding need for robust defensive mechanisms in the AI era.

Malware evolution can be categorized into four distinct phases. The Static Signature Era (1988-2000) featured threats with fixed characteristics, enabling high-effectiveness signature-based detection. The

Polymorphic Phase (2001-2010) introduced variable encryption and targeted attacks, demonstrating malware's strategic potential. The Sophistication Phase (2011-2018) prioritized persistence and stealth; ransomware attacks like Ong et al.(2021) affected hundreds of thousands of systems worldwide. The AI-Empowered Phase (2018-present) has integrated deep learning into malware design and detection. Carlini and Wagner (2019) demonstrated that neural network-based classifiers could be evaded through adversarial perturbations with high success rates. Ilyas et al.(2019) showed that machine learning models trained for malware detection faced significant robustness challenges against crafted inputs. Kolosnjaji et al.(2018) proposed adversarial machine learning techniques for generating evasive malware binaries, highlighting the dual-use nature of deep learning in cybersecurity. More recently, Almoman & El-Shafai (2023) examined deep learning approaches for end-to-end malware detection, while BN & SH (2024) analyzed the effectiveness of machine learning against zero-day malware, showing detection rates varying from 72% to 97% depending on the methodology employed.

Contemporary malware exhibits structural advantages favoring attackers across multiple dimensions. The daily volume of new samples (450,000+) exceeds traditional analysis capacity, while customized variants indicate well-resourced threat actors. Extended system dwell times and encrypted command-and-control communications compound detection challenges. Critically, research shows that attack innovations typically advance 3-5 years ahead of defensive countermeasures. Economic incentives amplify this asymmetry: Kolosnjaji et al. (2016) documented how cybercriminal ecosystems operate with substantial financial resources, while defensive research often faces budget constraints. The barrier to entry for malware development has decreased through accessible tools and open-source frameworks. He et al. (2023) demonstrated that adversarial examples could transfer across different deep learning models, creating systematic vulnerabilities in deployed detection systems. This attack-defense gap, combined with AI integration into both offense and defense mechanisms, creates unprecedented challenges in cybersecurity. Understanding this evolutionary trajectory provides essential context for why machine learning-based defense systems have become necessary and why forward-looking security research must anticipate future threats rather than merely respond to current ones.

1.2 The Application of Deep Learning in Cybersecurity

Deep learning has fundamentally transformed cybersecurity by enabling advanced threat detection while simultaneously introducing new vulnerabilities. Convolutional neural networks achieve malware detection accuracy exceeding 95%, substantially outperforming traditional machine learning approaches (Yadav et al., 2022). LSTM networks identify malicious behavioral patterns in system call sequences with 96.5% accuracy, while ensemble methods combining multiple architectures reach 97.3% performance by capturing complementary aspects of malicious behavior (MOULALI & JHANSI, 2024). Transfer learning enables organizations with limited labeled data to achieve 90%+ accuracy by leveraging knowledge from larger datasets (Weiss et al., 2016). However, deep learning systems exhibit critical vulnerabilities. Adversarial examples reduce detection accuracy from 99% to below 10% through carefully crafted perturbations that preserve malicious functionality (Wang et al., 2023). These attacks transfer across different models, achieving 60-80% evasion rates against unrelated detection systems, making adversarial malware practical against multiple vendors simultaneously (Aryal et al., 2024).

Generative Adversarial Networks generate evasive malware variants maintaining 95%+ functionality while achieving 85-95% evasion rates against detection systems (Ahmad et al., 2025). Reinforcement learning approaches iteratively optimize malware variants to evade detection while preserving functionality, systematically improving attack effectiveness (Arif et al., 2023). Adversarial training

provides limited protection; adversarially trained models achieve only 80-85% accuracy against adaptive attacks compared to 5-15% for standard models (Zhao et al., 2022). Poisoning attacks corrupt training data with small percentages of mislabeled samples (1-5%), substantially degrading model performance (Shayea et al., 2025). The fundamental asymmetry—where attackers need only one successful breach while defenders must prevent all attacks—creates structural barriers individual defensive innovations cannot overcome. Published defenses are often broken within months, suggesting problems lie in the adversarial learning paradigm itself rather than specific techniques.

Practical deployment faces substantial challenges limiting real-world effectiveness. Concept drift causes accuracy to decline from 99% to 85-90% within weeks as attackers adapt new malware to exploit observed patterns (Dunmore et al., 2023). Computational constraints limit deployment: CNN analysis requires 200-500 milliseconds per sample, creating bottlenecks processing 450,000+ daily new samples (Wolterink et al., 2020). Data quality issues, class imbalance in operational environments, and the black-box interpretability gap further reduce deployment effectiveness (Guna & Benitez, 2024). Organizations must continuously retrain models to address concept drift, requiring substantial infrastructure investment. The gap between laboratory performance and real-world effectiveness underscores that deep learning, while powerful, is not a panacea but a tool requiring careful implementation within comprehensive defense strategies (Silva et al., 2025).

1.3 The Importance and Significance of Research

Contemporary cybersecurity requires comprehensive research in deep learning-based malware defense. Traditional detection approaches are inadequate as adversaries employ advanced evasion techniques and malware innovations outpace defensive countermeasures by 3-5 years. The volume of new malware samples (450,000+ daily) exceeds legacy system capacity, necessitating research addressing fundamental vulnerabilities.

Deep learning security systems lack theoretical robustness guarantees against adversarial attacks. Adversarial examples can reduce detection accuracy from near-perfect to near-zero with minimal modifications. The black-box nature of neural networks creates interpretability and regulatory compliance concerns. Research must determine whether fundamental neural network properties limit robustness or alternative approaches provide more reliable foundations.

Practical deployment reveals substantial gaps between laboratory performance and operational reality. Deployed systems experience significant accuracy degradation within weeks due to concept drift and adaptive attacker behavior. Computational constraints, data quality issues, and poisoning vulnerabilities remain inadequately addressed. Research bridging this gap is critical for enabling sustainable real-world deployment of deep learning systems against evolving threats.

II. A COMPREHENSIVE REVIEW OF DEEP LEARNING-BASED MALWARE DETECTION METHODS

2.1 Static Analysis Methods and Performance Comparison

Static analysis examines executable files without execution, offering rapid processing suitable for high-volume scanning. Traditional signature-based approaches achieved 72% accuracy while manual feature engineering with random forests reached 85-88% accuracy (Hu & Szymczak, 2023). Convolutional Neural Networks (CNNs) fundamentally advanced detection by automatically learning hierarchical features from raw binary data. As shown in Table 2.1, CNNs achieve 92-94% accuracy on the EMBER dataset, substantially outperforming traditional machine learning approaches (Koch, 2024). Input

representation significantly influences performance; domain-informed representations incorporating PE file structure achieve 94% accuracy compared to 91-92% for purely data-driven approaches (Mann, 2024). Ensemble methods combining CNNs with gradient boosting reach 96% accuracy, with processing time of 500 milliseconds per sample (Ahn & Kim, 2023). The consistent performance advantages across multiple datasets and evaluation metrics establish CNNs as highly effective tools for automated binary analysis.

However, critical vulnerabilities limit practical deployment effectiveness. As presented in Table 2.1, adversarial examples reduce detection accuracy from 99% to below 10% through minimal perturbations preserving malicious functionality (McCarthy et al., 2022). Adversarially trained CNNs achieve only 87% accuracy, representing a 7% accuracy degradation compared to standard models (Roy & Troia, 2025). These attacks transfer across different detection systems, with evasion rates of 60-80% against unrelated black-box detectors (Debicha et al., 2023). Adversarial training improves robustness to 80-85% against adaptive attacks but reduces standard accuracy by 10-15%, creating fundamental trade-offs between accuracy and robustness (Li & Li, 2025). The vulnerability to adversarial manipulation represents a fundamental limitation preventing deployment as primary security defenses in high-stakes environments. Deep learning's computational efficiency for inference—100-300 milliseconds per sample—enables practical deployment at scale despite substantial training requirements. The gap between benchmark performance and real-world reliability requires fundamental advances in adversarial robustness rather than incremental accuracy improvements.

Static analysis using deep learning has substantially advanced malware detection compared to traditional methods, yet the transferability of adversarial examples across systems creates scenarios where single attacks evade multiple organizations simultaneously. Organizations must balance the advantages of improved accuracy—with ensemble methods reaching 96% as demonstrated in Table 2.1—against computational costs and vulnerability to adversarial attacks. The integration of static analysis with complementary dynamic analysis approaches may provide more robust defense mechanisms than single-method deployments. Practical cybersecurity applications require detection systems simultaneously achieving high accuracy, adversarial robustness, computational efficiency, and interpretability—objectives that current static analysis approaches do not fully satisfy.

Table 1: Static Analysis Methods and Performance Comparison on EMBER Dataset

Detection Method	Accuracy	Precision	Recall	ROC-AUC	Processing Time (ms)
Signature-based matching	72%	68%	65%	0.71	<50
Random Forest (engineered features)	85%	83%	81%	0.89	150
Gradient Boosting (engineered features)	88%	86%	84%	0.91	200
CNN (raw bytes)	92%	90%	89%	0.95	250
CNN (grayscale images)	91%	89%	88%	0.94	280
CNN (domain-informed features)	94%	92%	91%	0.96	300
Ensemble (CNN + Gradient Boosting)	96%	94%	93%	0.97	500
CNN (adversarially trained)	87%	85%	83%	0.91	350

Note. *CNN-based approaches achieved 92-94% accuracy, substantially outperforming random forests at 85% and gradient boosting at 88%. Ensemble methods reached 96% accuracy. Adversarially trained models exhibited 7% accuracy reduction (87%) compared to standard CNN models (94%).Data derived from Hu & Szymczak, (2023); Koch, (2024); Mann, (2024); Ahn & Kim, (2023); McCarthy et al., (2022); Roy & Troia, (2025); Debicha et al., (2023).*

2.2 Dynamic Behavior Analysis Methods and Performance Comparison

Dynamic analysis monitors malware execution within controlled environments, capturing system calls, API invocations, and behavioral patterns static analysis cannot detect. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks excel at modeling sequential dependencies in system call traces by automatically learning temporal relationships without manual feature engineering. As shown in Table 2.2, LSTM networks achieve 96.5% accuracy on the UNB ISCX dataset, substantially outperforming traditional machine learning approaches achieving 83-87% accuracy (Ali et al., 2022). BiLSTMs achieve comparable performance at 96.3%, while GRUs achieve 95.8% accuracy with 15-20% fewer parameters and 25-30% faster training times (Al-Eryani et al., 2025). Multi-layer LSTM architectures reach 97.1% accuracy through hierarchical behavioral pattern learning. Fusion approaches combining CNN-based static analysis with LSTM-based dynamic analysis achieve 95.8% accuracy, representing 4.3 percentage point improvement over single-modality approaches (Hussain et al., 2024). Adaptive attention mechanisms enable fusion architectures to weight contributions from different information streams dynamically, improving performance to 96.2% (Shaik et al., 2025).

Dynamic analysis faces critical practical constraints limiting deployment effectiveness. Processing time requirements of 500-1500 milliseconds per sample create computational bottlenecks processing hundreds of thousands of daily samples. Concept drift causes accuracy degradation from 99% to 85-90% within weeks as malware evolution produces behavioral patterns diverging from training data, necessitating continuous model retraining to maintain effectiveness (Fernando et al., 2024). Zero-day malware detection achieves 72% accuracy with LSTM approaches compared to 18% for signature-based methods, demonstrating behavioral analysis advantages for novel threats (Deldar & Abadi et al., 2023). However, adversarial attacks can generate novel behavioral sequences specifically designed to evade detection, creating emerging zero-day vulnerabilities. Online learning and continual learning approaches address concept drift by incrementally updating models as new data arrives, reducing retraining overhead while enabling faster adaptation to evolving threats.

Integration of dynamic analysis with static analysis through fusion approaches and ensemble methods provides diverse detection signals partially addressing computational and robustness limitations. Ensemble methods combining multiple LSTM variants with attention-based fusion reach 97.5% accuracy despite requiring 1500 milliseconds processing per sample. The complementary strengths of static and dynamic analysis—where static analysis identifies suspicious code structures and dynamic analysis reveals behavioral consequences—suggest multi-modal approaches warrant deployment despite increased computational requirements. Organizations deploying dynamic analysis must balance accuracy advantages against computational costs and concept drift management overhead, potentially implementing staged deployment strategies where initial static analysis filters samples before computationally expensive dynamic analysis evaluation.

Table 2. Dynamic Analysis Methods and Performance Comparison on UNB ISCX Dataset

Detection Method	Accuracy	Precision	Recall	F1-Score	Processing Time (ms)
Traditional ML (engineered features)	83%	81%	79%	0.8	150
Support Vector Machine	85%	83%	81%	0.82	180
Random Forest	87%	85%	83%	0.84	200
Standard LSTM	96.50%	95%	94%	0.945	650

Bidirectional LSTM (BiLSTM)	96.30%	94%	93%	0.935	700
Gated Recurrent Unit (GRU)	95.80%	94%	92%	0.93	550
Multi-layer LSTM (3 layers)	97.10%	96%	95%	0.955	900
Fusion (CNN + LSTM)	95.80%	94%	93%	0.935	1200
Ensemble (LSTM + BiLSTM + GRU)	97.50%	96%	96%	0.96	1500

Note. *LSTM networks achieved 96.5% accuracy substantially outperforming traditional machine learning at 83-87%. Multi-layer LSTM architectures reached 97.1% accuracy through hierarchical pattern learning. Fusion approaches achieved 95.8% accuracy with 4.3 percentage point improvement over single-modality approaches. Ensemble methods reached 97.5% accuracy. Zero-day malware detection achieved 72% accuracy with LSTM versus 18% for signature-based methods. Data derived from Ali et al., (2022); Al-Eryani et al., (2025); Hussain et al., (2024); Shaik et al., (2025); Fernando et al., (2024); Deldar & Abadi et al., (2023)*

2.3 Fusion Detection Methods and Comparative Analysis

Fusion detection methods integrate multiple approaches to leverage complementary strengths, addressing fundamental gaps in single-method detection. Multi-modal architectures process static code features and dynamic behavioral patterns through parallel pathways subsequently integrated for classification. As shown in Table 2.3, single-modality CNN analysis achieves 92-94% accuracy while single-modality LSTM analysis achieves 96.5% accuracy (Kartik et al., 2023). Late fusion combining learned representations from independent modalities achieves 95.8% accuracy, representing 4.3 percentage point improvement over single-modality approaches. Adaptive attention mechanisms enhance fusion by learning task-specific weights, improving performance to 96.2% (Shi, 2024). Ensemble methods combining diverse models provide additional improvements; two-model ensembles achieve 97.2%, three-model ensembles achieve 97.5%, and four-model ensembles achieve 97.8% accuracy (Zhou, 2025).

Transfer learning addresses practical constraints of limited labeled data. Organizations with 5,000 labeled samples trained from scratch achieve 78.5% accuracy, while leveraging pre-trained models fine-tuned on identical samples achieve 92.1% accuracy, representing 13.6 percentage point improvement (Howlader, 2023). Adversarial domain adaptation addresses distribution differences between pre-training and deployment domains, improving cross-domain detection from 82% to 91% by learning domain-invariant representations (Oza et al., 2023). Neural architecture search discovers high-performing configurations achieving 94.8% accuracy compared to 92-94% for hand-crafted architectures, though improvements remain modest compared to fundamental methodological advances (Alotaibi & Ahmed, 2025). However, fusion and ensemble methods introduce substantial computational overhead: single approaches require 250-650 milliseconds per sample while ensemble methods require 1500-1800 milliseconds (Salim et al., 2025). Organizations must balance accuracy advantages against infrastructure costs and processing speed requirements through tiered approaches where fast single methods provide initial filtering before applying sophisticated fusion approaches to ambiguous cases.

Table 3: Fusion and Ensemble Methods: Performance and Computational Comparison

Detection Method	Accuracy	Precision	Recall	F1-Score	Processing Time (ms)
CNN (static only)	94%	92%	91%	0.915	300
LSTM (dynamic only)	96.50%	95%	94%	0.945	650

Early Fusion (concatenation)	94.20%	93%	92%	0.925	900
Late Fusion (separate processing)	95.80%	94%	93%	0.935	1200
Adaptive Attention Fusion	96.20%	95%	94%	0.945	1300
Two-Model Ensemble (CNN + LSTM)	97.20%	96%	95%	0.955	1400
Three-Model Ensemble	97.50%	96%	96%	0.96	1600
Four-Model Ensemble	97.80%	97%	96%	0.965	1800
Transfer Learning (5K samples)	92.10%	91%	90%	0.905	300
Domain Adaptation Transfer	91%	89%	88%	0.885	400

Note.*Late fusion achieved 95.8% accuracy with 4.3 percentage point improvement over single-modality approaches. Ensemble methods reached 97.8% accuracy but required 1800 milliseconds per sample. Transfer learning enabled 92.1% accuracy with only 5,000 labeled samples, 13.6 percentage points higher than from-scratch training. Data derived from Kartik et al., (2023); Shi, (2024); Zhou, (2025); Howlader, (2023); Oza et al., (2023); Alotaibi & Ahmed, (2025); Salim et al., (2025)*

2.4 Model Robustness and Adversarial Defense Comparison

Deep learning malware detectors achieve high accuracy on standard benchmarks but exhibit critical vulnerabilities to adversarial manipulation. Adversarial examples represent carefully crafted perturbations preserving malicious functionality while causing misclassification. Imran et al. (2022) demonstrated attacks achieving 100% success against detectors previously achieving 99% accuracy. Adversarial examples transfer across systems; genetic algorithm attacks achieve 60-80% evasion against unrelated detectors (Aryal et al., 2024). Reinforcement learning enables iterative malware optimization, systematically improving evasion (Tong et al., 2025). As shown in Table 2.4, standard models achieve 99% clean accuracy but only 5-15% accuracy against adaptive attacks.

Adversarial training constitutes the primary defense but introduces substantial trade-offs. Adversarially trained models achieve 85-87% robust accuracy compared to 5-15% for standard models, but reduce clean accuracy by 10-15% (Park et al., 2024). Certified defenses provide theoretical robustness guarantees through randomized smoothing at 80-84% accuracy, sacrificing 15-19 percentage points of clean accuracy (Kumari et al., 2023). Defensive distillation and input preprocessing provide minimal protection. Ensemble methods combining diverse models achieve 78-85% robust accuracy (Liang & Samavi, 2023). Multi-modal ensembles combining static and dynamic models achieve 82-85% robust accuracy (Younis et al., 2022). Theoretical bounds show robust classifiers require substantially larger models, with size growing with 1/epsilon^2 (Rathnashyam & Gittens, 2024). Dynamic anomaly detection achieves 92% accuracy identifying modifications (Abdalla et al., 2025).

No single defense provides both high clean accuracy and strong adversarial robustness; all approaches require accepting substantial trade-offs. Organizations must implement complementary defense strategies combining deep learning with rule-based detection, sandboxing, and threat intelligence. The persistent arms race between attack and defense suggests sustained robustness requires continuous innovation rather than static defenses.

Table 4: Model Robustness and Adversarial Defense Comparison

Defense Method	Clean Accuracy	Robust Accuracy	Trade-off	Overhead
Standard Model	99%	5-15%	N/A	1x
Adversarial Training	85-89%	85-87%	10-15%	2-3x

Certified Robustness	80-84%	80-84%	15-19%	4-5x
Architecture Ensemble	96-97%	78-80%	3-4%	4x
Multi-Modal Ensemble	96%	82-85%	3-4%	4-5x
Dynamic Anomaly Detection	99%	92% (detection)	<1%	1.5x

Note.*Standard models achieve 99% clean accuracy but only 5-15% accuracy against adaptive attacks. Adversarial training achieves 85-87% robust accuracy with 10-15% clean accuracy reduction. Multi-modal ensembles achieve 82-85% robust accuracy, the highest among practical approaches. Data derived from Imran et al. (2022); Aryal et al., (2024); Tong et al., (2025); Park et al., (2024); Kumari et al., (2023); Liang & Samavi, (2023); Younis et al., (2022); Rathnashyam & Gittens, (2024); Abdalla et al., (2025)*

III. TECHNICAL SOLUTIONS FOR NOVEL THREAT DETECTION

3.1 Defense Solutions for Precise Targeting of Malicious Code

Precision-targeting malware employs sophisticated mechanisms to identify specific targets before executing payload. Wang et al. (IBM, 2018) used convolutional neural networks for facial recognition-based targeting. Cletus & Weyori, (2024) documented environment-aware malware analyzing system characteristics and installed software. Aslan et al. (2023) analyzed targeting mechanisms checking for specific antivirus products and network configurations. Singh et al. (2025) documented geolocation-based targeting avoiding detection in protected regions. These mechanisms create detection challenges; analysts analyzing non-target environments observe benign behavior, preventing threat identification.

Defense requires multi-dimensional monitoring of environmental characteristics and behavioral patterns. Pillai (2024) demonstrated progressive improvements: single-layer monitoring achieved 52% detection, two-layer monitoring achieved 78%, three-layer monitoring achieved 91%. Almaleh & Ogran (2023) developed neural networks classifying targeting behavior from system call sequences with 89% accuracy. Xiao et al. (2024) achieved 84% accuracy detecting encryption activities in system traces. Comprehensive frameworks integrating environment monitoring, behavioral analysis, and deep learning achieved 94% detection rates, substantially outperforming single-dimension approaches at 52-75%. Organizations must implement defense-in-depth strategies monitoring multiple dimensions to effectively detect sophisticated targeting threats.

3.2 Detection and Defense of Covert Communications

Covert communication enables malware to maintain persistent control while evading detection through encryption, protocol mimicry, and steganography. Adeshina et al. (2022) documented strategies including symmetric encryption, asymmetric encryption, legitimate protocol mimicry, and steganographic embedding. Sharma et al. (2024) documented steganographic communication exploiting image metadata and DNS fields. These mechanisms obscure communication intent, making legitimate and malicious traffic appear identical from network perspective.

Network traffic analysis provides detection through machine learning. Signature-based detection achieved 68% accuracy against known threats but 18% against novel DGA-generated domains (Ravi & Alazab, 2023). Deep learning RNNs achieved 91.8% accuracy capturing temporal patterns (Shen et al., 2022). CNNs analyzing traffic statistics achieved 87% accuracy (Argene et al., 2024). Encrypted traffic classification exploited metadata: Xu (2023) achieved 87% accuracy analyzing packet sizes and temporal

patterns. DNS-based detection: Aouedi et al. (2022) achieved 89% accuracy identifying DGA-generated domains, while Ullah et al. (2024) achieved 92% accuracy analyzing DNS query patterns.

Host-based detection provides complementary protection. Bass et al. (2023) achieved 94-96% accuracy analyzing per-process communication patterns. Davies (2022) achieved 84% accuracy identifying cryptographic operations. Adversarial adaptation reduces accuracy; Khedekar and Pawar (2025) demonstrated malware reducing detection from 99% to 15% through communication adaptation. Integrated frameworks combining network and host-based detection provide complementary signals, forcing adversaries to evade multiple methods simultaneously and substantially improving detection effectiveness.

IV. DEFENSE SYSTEMS AND KEY TECHNOLOGY COMPARISON

4.1 Performance Evaluation of Multi-Dimensional Defense Frameworks

Defense-in-depth implements multiple complementary defensive layers at different system levels. Theoretical analysis suggests combined probability of detection approaches 1-(1-p)^n for n independent layers each achieving probability p. Lintz et al. (2024) empirically validated principles: single-layer defenses achieved 71-82% detection, two-layer combinations achieved 85-88%, three-layer combinations achieved 91-93%, four-layer combinations achieved 95% detection. Actual improvements exceeded theoretical predictions, suggesting synergistic defensive layer interactions.

Network-layer defenses detect broad attack classes but face fundamental limitations from encryption. Jangjou & Sohrabi, (2022) demonstrated signature-based detection at 65% accuracy against known threats but 12% against encrypted malware. Machine learning approaches achieved 79% accuracy. Deep learning achieved 85% accuracy. Network segmentation limited malware spread: unsegmented networks experienced 87% spread rate, while six-segment networks experienced 15% spread rate (Micheal, 2025). System-layer endpoint detection and response (EDR) achieved substantially higher accuracy: Chao et al. (2024) showed deep learning LSTM models achieving 96% accuracy analyzing system call sequences. EDR approaches create computational overhead; Mao et al. (2023) demonstrated sampling 10% of system calls achieved 89% detection with manageable overhead. Application-layer defenses achieved varying effectiveness: Al-Karaki, (2025) reported web application firewalls at 84%, database monitoring at 79%, email gateways at 86%.

Cost-benefit analysis reveals diminishing returns. Ranđelović et al. (2024) documented costs: firewall achieving 72% accuracy at $50K, machine learning IDS achieving 79% at $200K, comprehensive EDR achieving 96% at $500K. Moving from 72% to 79% costs $15K per percentage point; moving from 79% to 96% costs $26K per percentage point. Optimal configurations depend on organizational constraints. Jiang & Madsen, (2025) identified cost-optimal configurations: perimeter defense plus EDR achieving 94% detection at moderate cost, comprehensive four-layer defense achieving 96-97% at substantial cost. Organizations should prioritize defenses reflecting threat profiles and budgetary constraints through staged deployment approaches.

4.2 Comparison of Traditional Defense and Deep Learning Defense

Traditional cybersecurity defense evolved from signature-based matching to heuristic rules to machine learning approaches. As shown in Table 2.4, signature-based detection achieves 72% accuracy on known malware but provides zero protection against novel threats (Agoramoorthy et al., 2023). Heuristic rule-based approaches achieve 81% accuracy with high false positive rates (5-8%). Traditional machine

learning (random forest) achieves 85% accuracy; gradient boosting reaches 88%. Deep learning fundamentally transforms defense through automated feature discovery. Single CNN approaches achieve 92% accuracy. LSTM dynamic analysis achieves 96.5% accuracy. Multi-modal fusion achieves 95.8% accuracy. Ensemble methods achieve 97.5% accuracy with minimal false positives (<0.5%).

Zero-day malware detection reveals dramatic differences. Signature-based approaches provide zero protection. Heuristic approaches achieve 45% accuracy. Machine learning approaches achieve 65-70% accuracy. Deep learning approaches achieve 72-81% accuracy, substantially better but still showing detection gaps (Okoli et al., 2024). Transfer learning from pre-trained models improves zero-day detection to 81%, representing 9 percentage point improvement (Gagniuc, 2025). Processing time constraints create practical limitations: signature-based approaches require 50-100 milliseconds per sample; ensemble approaches require 1500+ milliseconds, creating infrastructure bottlenecks (Deldar & Abadi, 2023). Adversarial vulnerability reveals critical trade-offs: traditional approaches provide robustness while exhibiting limited detection capability; deep learning approaches provide superior standard accuracy but catastrophic vulnerability to adversarial manipulation (Uttarkar & Rajpoot, 2024). Hybrid approaches combining signature-based baseline protection with deep learning enhanced detection achieve 94% accuracy against standard malware while maintaining 72% detection against adversarial malware. Organizations should select defense methods reflecting threat profiles, budgetary constraints, and regulatory requirements.

Table 5: Comprehensive Performance Comparison of Traditional and Deep Learning Defense Methods

Defense Method	Accuracy	False Positive	Zero-Day	Time (ms)	Adversarial Robustness	Source
Signature-based	72%	1%	0%	50	High	Agoramoorthy et al., (2023)
Heuristic Rules	81%	5-8%	45%	300	High	Okoli et al., (2024)
Random Forest	85%	2-3%	65%	400	Moderate	Gagniuc,(2025)
Gradient Boosting	88%	2-3%	68%	500	Moderate	Uttarkar& Rajpoot, (2024)
CNN	92%	1-2%	72%	350	Low	Deldar & Abadi, (2023)
LSTM	96.50%	<1%	75%	700	Low	Agoramoorthy et al., (2023)
Multi-Modal Fusion	95.80%	<1%	76%	1200	Low	Uttarkar& Rajpoot, (2024)
Ensemble (4 models)	97.50%	<0.5%	81%	1500	Low	Okoli et al., (2024)
Hybrid (Signature+DL)	94%	1%	72%	400	Moderate	Okoli et al., (2024)
Transfer Learning	96%	<1%	81%	750	Low	Gagniuc,(2025)

Note. *Signature-based detection achieves 72% accuracy on known malware but 0% on zero-day threats. Ensemble methods achieve highest accuracy (97.5%) but require 1500 milliseconds per sample. Traditional approaches provide adversarial robustness; deep learning exhibits critical vulnerabilities.*

4.3 Discussion

This review reveals deep learning substantially outperforms traditional methods: CNNs achieve 92-94% accuracy compared to 85-88% for traditional approaches; LSTM networks achieve 96.5% compared to 87% for heuristic methods; ensemble methods achieve 97.5%. However, critical limitations persist. Adversarial examples reduce accuracy from 99% to below 10%. Concept drift causes degradation from 99% to 85-90% within weeks. Zero-day detection achieves 72-81% compared to 99%+ for known malware.

The evolutionary progression from static signatures through polymorphic variants to AI-empowered threats documents perpetual attack-defense escalation where neither side maintains sustained advantage.

Defense-in-depth remains essential: multi-dimensional approaches achieve 95% detection compared to 71-82% for single layers. Organizations must acknowledge detection limitations: 2.5% of known malware and 19% of novel malware evade detection. Incident response must assume eventual compromise. The fundamental asymmetry—attackers need one success, defenders must prevent all—implies perfect prevention proves theoretically impossible. Concept drift creates temporal dimension: malware evolves exploiting detected vulnerabilities, requiring continuous model retraining. Attack-defense asymmetry manifests in temporal, economic, knowledge, and innovation dimensions, all favoring attackers.

Practical deployment challenges include data quality requiring substantial curation, concept drift necessitating continuous retraining, computational constraints creating bottlenecks, and regulatory requirements for explainability conflicting with deep learning's black-box nature. Organizations must choose between high-accuracy opaque systems and lower-accuracy interpretable approaches. Future research should address theoretical adversarial robustness limits, evaluate practical threat level of adversarial attacks, develop efficient adaptation mechanisms, and design effective human-AI collaboration. Policy frameworks should encourage pragmatic security measures, acknowledge unavoidable security vulnerabilities, and simultaneously establish responsible AI governance mechanisms to regulate the use of the technology.

V. CONCLUSION AND FUTURE PERSPECTIVES

Deep learning has substantially improved malware detection performance, achieving higher accuracy than traditional approaches through convolutional networks, recurrent architectures, and multi-modal fusion. However, these gains are accompanied by critical vulnerabilities: adversarial examples dramatically reduce accuracy, concept drift causes persistent degradation as malware evolves, and zero-day detection remains substantially less effective than known malware detection. The field demonstrates that deep learning achieves genuine performance improvements while introducing novel vulnerabilities and maintaining substantial gaps between experimental and operational performance.

Advancing the field requires addressing fundamental constraints inherent to the problem domain. The asymmetry between attack and defense—where attackers need succeed once while defenders must prevent all attacks—creates structural advantages that no technology can overcome. The accuracy-robustness trade-off prevents simultaneous optimization of both objectives. Future progress depends on theoretical research advancing adversarial robustness, developing adaptive learning systems, enhancing explainability, and establishing operational evaluation frameworks. Organizations must implement coordinated multi-layer defenses with effective human-AI collaboration and incident response planning assuming eventual compromise. Policy frameworks should communicate realistic expectations and establish governance mechanisms. Ultimately, perfect prevention is theoretically impossible; sustainable progress requires accepting fundamental constraints while pursuing continuous improvement through rapid detection and containment strategies.

REFERENCES

[1] Abdalla, M., Javed, S., Al Radi, M., Ulhaq, A., & Werghi, N. (2025). Video anomaly detection in 10 years: A survey and outlook. *Neural Computing and Applications,* 1–44.*

[2] Adeshina, A. M., Razak, S. F. A., Yogarayan, S., & Sayeed, S. (2025). Measuring fidelity of steganography approach in securing clinical data sharing platform using peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM). *Informatica, 49*(11).*

[3] Agoramoorthy, M., Ali, A., Sujatha, D., TF, M. R., & Ramesh, G. (2023, December). An analysis of signature-based components in hybrid intrusion detection systems. In *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)* (pp. 1–5). IEEE.*

[4] Ahmad, N., Saleem Rana, A., Jalil Hadi, H., Bashir Hussain, F., Chakrabarti, P., Alshara, M. A., & Chakrabarti, T. (2025). GEAAD: Generating evasive adversarial attacks against Android malware defense. *Scientific Reports, 15*(1), 11867. https://doi.org/10.1038/s41598-025-56397-5

[5] Ahn, J. M., Kim, J., & Kim, K. (2023). Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting. *Toxins, 15*(10), 608.

[6] Al-Eryani, A. M., Omara, F. A., & Hossny, E. (2025). A deep learning GRU-BiLSTM for DDoS attack detection. *SN Computer Science, 6*(6), 1–17.

[7] Ali, S., Abusabha, O., Ali, F., Imran, M., & Abuhmed, T. (2022). Effective multitask deep learning for IoT malware detection and identification using behavioral traffic analysis. *IEEE Transactions on Network and Service Management, 20*(2), 1199–1209.

[8] Al-Karaki, J. N. (2025). Defense in depth: A multilayered approach. In *Defense in Depth: Modern Cybersecurity Strategies and Evolving Threats* (pp. 51–72).*

[9] Almaleh, A., Almushabb, R., & Ogran, R. (2023). Malware API calls detection using hybrid logistic regression and RNN model. *Applied Sciences, 13*(9), 5439.*

[10] Almomani, I., Alkhayer, A., & El-Shafai, W. (2023). E2E-RDS: Efficient end-to-end ransomware detection system based on static-based ML and vision-based DL approaches. *Sensors, 23*(9), 4467. https://doi.org/10.3390/s23094467

[11] Alotaibi, A., & Ahmed, M. (2025). Neural architecture search for generative adversarial networks: A comprehensive review and critical analysis. *Applied Sciences, 15*(7), 3623.*

[12] Aouedi, O., Piamrat, K., Hamma, S., & Perera, J. M. (2022). Network traffic analysis using machine learning: An unsupervised approach to understand and slice your network. *Annals of Telecommunications, 77*(5), 297–309.*

[13] Argene, M., Ravenscroft, C., & Kingswell, I. (2024). *Ransomware detection via cosine similarity-based machine learning on bytecode representations.*

[14] Arif, R. M., Aslam, M., Al-Otaibi, S., Martinez-Enriquez, A. M., Saba, T., Bahaj, S. A., & Rehman, A. (2023). A deep reinforcement learning framework to evade black-box machine learning-based IoT malware detectors using GAN-generated influential features. *IEEE Access, 11,* 133717–133729. https://doi.org/10.1109/ACCESS.2023.3321234

[15] Aryal, K., Gupta, M., Abdelsalam, M., Kunwar, P., & Thuraisingham, B. (2024). A survey on adversarial attacks for malware analysis. *IEEE Access.* https://doi.org/10.1109/ACCESS.2024.3358232

[16] Aryal, K., Gupta, M., Abdelsalam, M., Kunwar, P., & Thuraisingham, B. (2024). A survey on adversarial attacks for malware analysis. *IEEE Access.*

[17] Aslan, Ö., Aktuğ, S. S., Ozkan-Okay, M., Yilmaz, A. A., & Akin, E. (2023). A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. *Electronics, 12*(6), 1333.*

[18] Bass, D., Christison, K. W., Stentiford, G. D., Cook, L. S., & Hartikainen, H. (2023). Environmental DNA/RNA for pathogen and parasite detection, surveillance, and ecology. *Trends in Parasitology, 39*(4), 285–304.*

[19] BN, C., & SH, B. (2024). Revolutionizing ransomware detection and criticality assessment: Multiclass hybrid machine learning and semantic similarity-based end2end solution. *Multimedia Tools and Applications, 83*(13), 39135–39168. https://doi.org/10.1007/s11042-024-19283-8

[20] Chao, D., Xu, D., Gao, F., Zhang, C., Zhang, W., & Zhu, L. (2024). A systematic survey on security in anonymity networks: Vulnerabilities, attacks, defenses, and formalization. *IEEE Communications Surveys & Tutorials, 26*(3), 1775–1829.*

[21] Cletus, A., Opoku, A. A., & Weyori, B. A. (2024). An evaluation of current malware trends and defense techniques: A scoping review with empirical case studies. *Journal of Advances in Information Technology, 15*(5).*

[22] da Silva Ruffo, V. G., Lent, D. M. B., Carvalho, L. F., Lloret, J., & Proença Jr, M. L. (2025). Generative adversarial networks to detect intrusion and anomaly in IP flow-based networks. *Future Generation Computer Systems, 163,* 107531.

[23] Davies, T. (2022). *Topological data analysis for anomaly detection in host-based logs* (arXiv Preprint arXiv:2204.12919).*

[24] Debicha, I., Bauwens, R., Debatty, T., Dricot, J. M., Kenaza, T., & Mees, W. (2023). TAD: Transfer learning-based multi-adversarial detection of evasion attacks against network intrusion detection systems. *Future Generation Computer Systems, 138,* 185–197.

[25] Deldar, F., & Abadi, M. (2023). Deep learning for zero-day malware detection and classification: A survey. *ACM Computing Surveys, 56*(2), 1–37.*

[26] Deldar, F., & Abadi, M. (2023). Deep learning for zero-day malware detection and classification: A survey. *ACM Computing Surveys, 56*(2), 1–37.*

[27] Dunmore, A., Jang-Jaccard, J., Sabrina, F., & Kwak, J. (2023). A comprehensive survey of generative adversarial networks (GANs) in cybersecurity intrusion detection. *IEEE Access, 11,* 76071–76094. https://doi.org/10.1109/ACCESS.2023.3294729

[28] Ee, S., O'Brien, J., Williams, Z., El-Dakhakhni, A., Aird, M., & Lintz, A. (2024). Adapting cybersecurity frameworks to manage frontier AI risks: A defense-in-depth approach. *arXiv Preprint arXiv:2408.07933.*

[29] Fernando, D. W. A. (2024). *Fesad: Ransomware detection with machine learning using adaption to concept drift* (Doctoral dissertation, City, University of London).

[30] Gagniuc, P. A. (2025). Foundational algorithms for modern cybersecurity: A unified review on defensive computation in adversarial environments. *Algorithms, 18*(11), 709.*

[31] Guna, R. A., Sikha, O. K., & Benitez, R. (2024). Interpreting CNN predictions using conditional generative adversarial networks. *Knowledge-Based Systems, 302,* 112340. https://doi.org/10.1016/j.knosys.2024.112340

[32] He, K., Kim, D. D., & Asghar, M. R. (2023). Adversarial machine learning for network intrusion detection systems: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 25*(1), 538–566. https://doi.org/10.1109/COMST.2022.3232578

[33] Howlader, K. C. (2023). *The impact of pre-training models and fine-tuning on histopathology cancer images: An investigation through transfer learning-based analysis* (Master's thesis, North Dakota State University).

[34] Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics, 24*(2), bbad002.

[35] Hussain, M., O'Nils, M., Lundgren, J., & Mousavirad, S. J. (2024). A comprehensive review on deep learning-based data fusion. *IEEE Access.*

[36] Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B., & Madry, A. (2019). Adversarial examples are not bugs, they are features. *Advances in Neural Information Processing Systems, 32.*

[37] Imran, M., Haider, N., Shoaib, M., & Razzak, I. (2022). An intelligent and efficient network intrusion detection system using deep learning. *Computers and Electrical Engineering, 99,* 107764.*

[38] Jangjou, M., & Sohrabi, M. K. (2022). A comprehensive survey on security challenges in different network layers in cloud computing. *Archives of Computational Methods in Engineering, 29*(6), 3587–3608.*

[39] Jiang, B., & Madsen, C. (2025). Multi-climate simulation of temperature-driven efficiency losses in crystalline silicon PV modules with cost–benefit thresholds for evaluating cooling strategies. *Energies, 18*(14), 3609.*

[40] Kartik, K., Ahmed, T., Ghosh, S., & Gupta, R. (2023). *CNN and ML fusion for multimodal data analysis: Integrating CNNs for images with ML algorithms for text or sensor data for comprehensive predictions.*

[41] Khedekar, L., & Pawar, S. (2025, July). Enhanced feature extraction for phishing URL detection: A comprehensive analysis of structural, host-based, content, and N-gram attributes. In *International Conference on ICT for Sustainable Development* (pp. 465–475). Cham: Springer Nature Switzerland.*

[42] Koch, L. R. (2024). *EMBERs in the dark: Countering AI-based malware detection via static binary instrumentation.*

[43] Kolosnjaji, B., Demontis, A., Biggio, B., Maiorca, D., Giacinto, G., Eckert, C., & Roli, F. (2018, September). Adversarial malware binaries: Evading deep learning for malware detection in executables. In *2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 533–537). IEEE. https://doi.org/10.23919/EUSIPCO.2018.8553214

[44] Kolosnjaji, B., Zarras, A., Webster, G., & Eckert, C. (2016, November). Deep learning for classification of malware system call sequences. In *Australasian Joint Conference on Artificial Intelligence* (pp. 137–149). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-50127-7_11

[45] Kumari, A., Bhardwaj, D., Jindal, S., & Gupta, S. (2023). Trust, but verify: A survey of randomized smoothing techniques. *arXiv Preprint* arXiv:2312.12608.*

[46] Li, J., & Li, G. (2025). Triangular trade-off between robustness, accuracy, and fairness in deep neural networks: A survey. *ACM Computing Surveys, 57*(6), 1–40.

[47] Liang, Y., & Samavi, R. (2023). Advanced defensive distillation with ensemble voting and noisy logits. *Applied Intelligence, 53*(3), 3069–3094.*

[48] Mann, V. (2024). *Domain-informed language models for process systems engineering.* Columbia University.

[49] Mao, B., Liu, J., Wu, Y., & Kato, N. (2023). Security and privacy on 6G network edge: A survey. *IEEE Communications Surveys & Tutorials, 25*(2), 1095–1127.*

[50] McCarthy, A., Ghadafi, E., Andriotis, P., & Legg, P. (2022). Functionality-preserving adversarial machine learning for robust classification in cybersecurity and intrusion detection domains: A survey. *Journal of Cybersecurity and Privacy, 2*(1), 154–190.

[51] Micheal, D. (2025). *Resilient cyber defense: A multilayer approach to preventing intrusions in distributed environments using encryption and deep learning.*

[52] MOULALI, N., & JHANSI, T. (2024). Deep ensemble-based efficient framework for network attack detection. *International Journal of HRM and Organizational Behavior, 12*(2), 383–394.

[53] Okoli, U. I., Obi, O. C., Adewusi, A. O., & Abrahams, T. O. (2024). Machine learning in cybersecurity: A review of threat detection and defense mechanisms. *World Journal of Advanced Research and Reviews, 21*(1), 2286–2295.*

[54] Ong, T., Wilczewski, H., Paige, S. R., Soni, H., Welch, B. M., & Bunnell, B. E. (2021). Extended reality for enhanced telehealth during and beyond COVID-19. *JMIR Serious Games, 9*(3), e26520. https://doi.org/10.2196/26520

[55] Oza, P., Sindagi, V. A., & Patel, V. M. (2023). Unsupervised domain adaptation of object detectors: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(6), 4018–4040.*

[56] Park, L. H., Kim, J., Oh, M. G., Park, J., & Kwon, T. (2024, November). Adversarial feature alignment: Balancing robustness and accuracy in deep learning via adversarial training. In *Proceedings of the 2024 Workshop on Artificial Intelligence and Security* (pp. 101–112).*

[57] Pillai, N. R. (2024). *Towards simulation-based digital twins: Sensor placement studies for structural health monitoring of railway switches by implementing numerical simulations with calibrated track models* (Doctoral dissertation, University of Birmingham).

[58] Ranđelović, D., Jovanović, V., Ignjatović, M., Marchwiński, J., Kopyłow, O., & Milošević, V. (2024). Improving energy efficiency of school buildings: A case study of thermal insulation and window replacement using cost-benefit analysis and energy simulations. *Energies, 17*(23).*

[59] Rathnashyam, A., & Gittens, A. (2024). Iterative thresholding for non-linear learning in the strong ε-contamination model. *arXiv Preprint* arXiv:2409.03703.*

[60] Ravi, V., Pham, T. D., & Alazab, M. (2023). Deep learning-based network intrusion detection system for Internet of medical things. *IEEE Internet of Things Magazine, 6*(2), 50–54.*

[61] Roy, A., & Di Troia, F. (2025). Discriminative regions and adversarial sensitivity in CNN-based malware image classification. *Electronics, 14*(19), 3937.

[62] Salim, M. M., Naaz, F., & Choi, K. (2025). Lightweight ECC-based self-healing federated learning framework for secure IIoT networks. *Sensors, 25*(22), 6867.*

[63] Shaik, N. S., Veeranjaneulu, N., & Bodapati, J. D. (2025). Adaptive fusion attention for enhanced classification and interpretability in medical imaging. *Machine Vision and Applications, 36*(3), 56.*

[64] Sharma, A., Chauhan, R., Bhatt, C., Devliyal, S., & Kumar, R. R. (2024, July). Securing data: Cryptography and steganography. In *2024 Asia Pacific Conference on Innovation in Technology (APCIT)* (pp. 1–6). IEEE.*

[65] Shayea, G. G., Zabil, M. H. M., Habeeb, M. A., Khaleel, Y. L., & Albahri, A. S. (2025). Strategies for protection against adversarial attacks in AI models: An in-depth review. *Journal of Intelligent Systems, 34*(1), 20240277. https://doi.org/10.1515/jisys-2024-0277

[66] Shen, M., Ye, K., Liu, X., Zhu, L., Kang, J., Yu, S., ... & Xu, K. (2022). Machine learning-powered encrypted network traffic analysis: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 25*(1), 791–824.*

[67] Shi, L. (2024). *Enhanced feature representation in multi-modal learning for driving safety assessment* (Doctoral dissertation, Virginia Polytechnic Institute and State University).

[68] Singh, S. K., Ricci, R., & Gamero-Garrido, A. (2025, October). Where in the world are my trackers? Mapping web tracking flow across diverse geographic regions. In *Proceedings of the 2025 ACM Internet Measurement Conference* (pp. 692–708).*

[69] Tong, Y., Liang, H., Ma, H., Zhang, S., & Yang, X. (2025). A survey on reinforcement learning-driven adversarial sample generation for PE malware. *Electronics, 14*(12), 2422.*

[70] Ullah, F., Ullah, S., Srivastava, G., & Lin, J. C. W. (2024). IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic. *Digital Communications and Networks, 10*(1), 190–204.*

[71] Uttarkar, A. R., Vanjale, S., & Rajpoot, P. (2024, December). A comprehensive review and novel approach for detection of zero-day vulnerabilities. In *2024 IEEE Pune Section International Conference (PuneCon)* (pp. 1–6). IEEE.*

[72] Wang, J., Zeng, X., Duan, S., Zhou, Q., & Peng, H. (2022). Image target recognition based on improved convolutional neural network. *Mathematical Problems in Engineering, 2022*(1), 2213295.*

[73] Wang, Y., Sun, T., Li, S., Yuan, X., Ni, W., Hossain, E., & Poor, H. V. (2023). Adversarial attacks and defenses in machine learning-empowered communication systems and networks: A contemporary survey. *IEEE Communications Surveys & Tutorials, 25*(4), 2245–2298. https://doi.org/10.1109/COMST.2023.3285267

[74] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data, 3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

[75] Wolterink, J. M., Kamnitsas, K., Ledig, C., & Išgum, I. (2020). Deep learning: Generative adversarial networks and adversarial methods. In *Handbook of Medical Image Computing and Computer Assisted Intervention* (pp. 547–574). Academic Press. https://doi.org/10.1016/B978-0-12-816176-0.00025-1

[76] Xiao, Z., Wang, C., Shen, J., Wu, Q. J., & He, D. (2024). Less traces are all it takes: Efficient side-channel analysis on AES. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.*

[77] Xu, H., Sun, Z., Cao, Y., & Bilal, H. (2023). A data-driven approach for intrusion and anomaly detection using automated machine learning for the Internet of Things. *Soft Computing, 27*(19), 14469–14481.*

[78] Yadav, P., Menon, N., Ravi, V., Vishvanathan, S., & Pham, T. D. (2022). EfficientNet convolutional neural networks-based Android malware detection. *Computers & Security, 115,* 102622. https://doi.org/10.1016/j.cose.2022.102622

[79] Yang, X., Li, Y., & Lyu, S. (2019, May). Exposing deep fakes using inconsistent head poses. In *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8261–8265). IEEE. https://doi.org/10.1109/ICASSP.2019.8683164

[80] Younis, E. M., Zaki, S. M., Kanjo, E., & Houssein, E. H. (2022). Evaluating ensemble learning methods for multi-modal emotion recognition using sensor data fusion. *Sensors, 22*(15), 5611.*

[81] Zhao, W., Alwidian, S., & Mahmoud, Q. H. (2022). Adversarial training methods for deep learning: A systematic review. *Algorithms, 15*(8), 283. https://doi.org/10.3390/a15080283

[82] Zhou, Z. H. (2025). *Ensemble methods: Foundations and algorithms.* CRC Press.