

A Comprehensive Survey of Signature-Based, Behavioral, and AI-Driven Approaches for Malware Detection

Manoj Kumar Chauhan¹, Prof.(Dr.) Sandeep Kumar Jain²

¹ Ph.D. Scholar , Department of Computer Science , Institute of Engineering & Technology , Dr. Bhimrao Ambedkar University , Agra

² Professor , Department of Computer Science, Institute of Engineering & Technology,
Dr. Bhimrao Ambedkar University , Agra

Abstract –

The rapid evolution of malware—including ransomware, trojans, spyware, botnets, worms, fileless malware, Android malware, and IoT threats—has exposed important limitations in conventional signature-based detection. Modern malware increasingly uses obfuscation, packing, polymorphism, anti-analysis mechanisms, encryption, and adversarial manipulation to evade static signatures. Machine Learning (ML) and Deep Learning (DL) therefore provide important alternatives because they can learn discriminative patterns from executable code, API calls, system calls, network behavior, binary representations, and program graphs. This survey reviews ML-based intelligent malware detection research with emphasis on the 2021–2026 period. The review organizes the literature according to malware type, platform, analysis strategy, feature representation, algorithm, dataset, evaluation metrics, and research gap. Traditional ML methods such as Random Forest, Support Vector Machine, Decision Tree, and boosting remain effective when carefully engineered features are available, while CNN, LSTM, CNN-LSTM, autoencoder, Transformer, and Graph Neural Network (GNN) models provide automated representation learning. Hybrid approaches combine static and dynamic evidence and frequently improve robustness. Emerging directions include explainable AI (XAI), federated learning, adversarially robust detection, temporal/concept-drift adaptation, and cross-platform learning. Public datasets such as EMBER, BIG2015, Malimg, Drebin, AndroZoo, CICMalDroid, VirusShare, VirusTotal, IoT-23, and Malicia are examined. The principal unresolved problems are zero-day detection, dataset bias and age, class imbalance, adversarial evasion, concept drift, explainability, computational cost, and cross-platform generalization. The paper concludes with a proposed integrated framework combining static and dynamic features, graph and sequence representations, hybrid ML/DL learning, XAI, adversarial testing, and continuous adaptation.

Key Words: malware detection; machine learning; deep learning; static analysis; dynamic analysis; hybrid analysis; explainable AI; adversarial machine learning, cybersecurity

1.INTRODUCTION

Malware is a broad class of software designed to compromise confidentiality, integrity, availability, or system resources. Its operational forms include ransomware, trojans, worms, spyware, botnets, rootkits, fileless malware, mobile malware, and IoT malware. The security problem is not only the volume of samples but also the speed with which malware families change their observable characteristics. Recent review evidence emphasizes that conventional detection becomes less effective as new variants evade antivirus and anti-malware mechanisms, motivating AI-driven detection research [1].

Traditional signature-based systems are highly effective for previously catalogued threats because signatures are compact and computationally inexpensive. Their central weakness is dependence on prior knowledge: a previously unseen family, modified binary, packed executable, or behaviorally novel sample may not match an existing signature. This limitation has encouraged statistical learning, anomaly detection, behavioral modeling, and representation learning.

ML-based detection treats malware identification as a classification or anomaly-detection problem. Features can be extracted from PE headers, imported functions, API calls, opcode sequences, byte sequences, strings, permissions, memory traces, network flows, or behavioral events. DL extends this approach by learning representations from raw or minimally processed inputs. CNNs can operate on binary images; recurrent models can learn API or opcode sequences; and GNNs can represent relationships among program components.

The objective of this survey is not to claim a universally superior algorithm. Instead, it compares the assumptions, data requirements, strengths, weaknesses, and deployment implications of the major intelligent detection families. This emphasis is consistent with recent survey literature that calls for systematic evaluation across data collection, preprocessing, feature engineering, model training, and performance evaluation [1].

II. CONTRIBUTIONS OF THIS SURVEY

The study reviews major malware detection approaches, including ML, DL, GNN, XAI, federated learning, adversarial defense, and Transformer/LLM techniques across Windows, Android, Linux, and IoT platforms. It highlights the limitations of accuracy-based evaluation and identifies key challenges such as zero-day attacks, outdated datasets, concept drift, adversarial evasion, interpretability, scalability, and cross-platform detection. Finally, it proposes an adaptive and explainable framework for future malware detection research.

III. REVIEW METHODOLOGY

A. Review Scope : This survey primarily focusing on studies published between 2021 and 2026 while incorporating foundational earlier works on datasets, adversarial attacks, and established malware analysis techniques. The literature matrix comprises 50 relevant survey and research studies from 2021–2026. In addition, a complementary review of AI-based ransomware detection provides a systematic perspective on datasets, preprocessing, feature engineering, model development, and evaluation [1].

B. Review Questions

- RQ1: Which ML and DL techniques are most frequently used for malware detection?
- RQ2: Which static, dynamic, and hybrid features are most informative?
- RQ3: Which datasets and platforms dominate the literature?
- RQ4: How are models evaluated, and why are reported results difficult to compare?
- RQ5: What limitations prevent deployment against zero-day and adversarial malware?
- RQ6: Which emerging methods are most promising for explainable, scalable, cross-platform detection?

C. Evidence Organization : Studies are organized using the following dimensions: malware family; platform; analysis mode; feature representation; algorithm; dataset; evaluation metric; contribution; limitation; and future direction. The approach preserves the terminology and comparative framing in the supplied research material.

IV. MALWARE TAXONOMY AND THREAT EVOLUTION

Table -1: Malware Taxonomy

Malware category	Typical objective	Representative detection evidence
Ransomware	Encrypt or deny access to data	File operations, entropy, API/system calls, network behavior
Trojan	Disguise malicious functionality	PE features, imports, behavior, API sequences

Botnet	Command-and-control and coordinated activity	Network flows, DNS, system behavior
Spyware	Steal information	API calls, permissions, process and network activity
Worm	Self-propagation	Network behavior, exploit patterns, process events
Android malware	Abuse mobile permissions/APIs	Permissions, API calls, opcode, network traffic
IoT malware	Compromise constrained connected devices	Traffic, system calls, firmware/device behavior
Fileless malware	Operate with limited disk artifacts	Memory, scripts, process and behavioral traces

V. Taxonomy of Malware Detection

Catego ry	Represent ative methods	Typical inputs	Main strengths	Main limitatio ns
Signatu re / heuristi c	Hashes, byte patterns, YARA/rules	Hashes, bytes, rule templates, predefined behaviors	Fast, low cost, effective for known threats	Weak against novel variants, polymorphism and obfuscation
Traditi onal ML	RF, SVM, DT, NB, KNN, XGBoost	PE headers, opcodes, API frequencies, n-grams, permissions	Efficient, relatively interpretable, deployable	Feature engineering; reduced adaptability
Deep learnin g	CNN, LSTM, GRU, autoencoder, CNN-LSTM	Bytes, images, API/system-call sequences, behavior logs	Automatic feature learning; strong nonlinear modeling	High data/compute needs; limited explainability
Graph learnin g	GCN, GAT,	CFG, FCG, API	Captures structural	Graph constructi

g	GNN variants	graphs, graph embeddings	relationships; useful for obfuscation-resistant analysis	on cost; benchmark scarcity
Hybrid ML/DL	Static + dynamic, ensemble, multimodal	Program structure + runtime/net work evidence	Complementary evidence and improved robustness	Higher latency and engineering complexity
Emerging AI	Transformer, LLM-assisted, federated, GAN-based	Code corpora, telemetry, semantic features, distributed updates	Contextual reasoning, transfer learning, privacy-preserving collaboration	Very high overhead; immature benchmarks and governance challenges

VI. Research Landscape: Source-Derived Quantitative Synthesis

The supplied source provides a quantitative snapshot of research focus. Deep learning represents 38% of the surveyed focus, traditional ML 26%, hybrid methods 16%, explainable AI 8%, GNNs 6%, federated learning 3%, and Transformer/LLM research 3%. These percentages form a useful descriptive picture rather than a meta-analysis effect size.

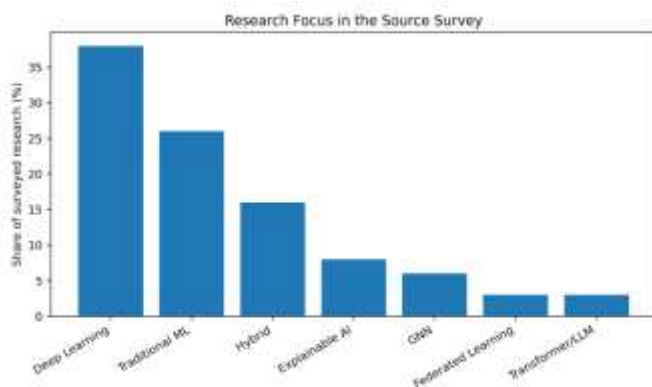


Fig. 1. Distribution of research focus reported in the supplied source survey.

The source also states that supervised learning accounts for 81.8% of primary studies. This dominance reflects the availability of labeled malware/benign datasets and the relative simplicity of supervised evaluation. Nevertheless, reliance on labels can limit discovery of novel threats and can amplify dataset bias.

VII. Dataset and Platform Analysis

Dataset	Primary platform/domain	Typical representation	Use in detection research	Key caution
EMBER	Windows PE	Static PE-derived features	Large-scale classical ML/DL benchmarking	Temporal drift and benchmark overfitting
BIG2015	Windows PE	Binary-derived/static features	Malware family classification	Historical dataset; limited modern coverage
Maling	Windows malware	Malware images	CNN/image-based classification	Image representation may discard semantic context
Drebin	Android	Permissions, API/features	Android malware classification	Older Android ecosystem
AndroZoo	Android	APK/static and metadata	Large-scale Android research	Access, labeling and temporal consistency
CICMaIDroid	Android	Behavior/network/system evidence	Android malware and traffic studies	Dataset-specific conditions
VirusShare	Multi-family malware	Malware samples	Research collection and benchmarking	Label quality and legal/operational handling
VirusT	Multi-	Multi-engine scan	Threat intelligence	Not a homogen

otal	platform	intelligence	nce and validati on	eous benchma rk; access constrain ts
Malicia	Windows	Malware samples/featur es	Malwar e analysis and ML studies	Dataset age and family coverage

Dataset choice strongly affects reported performance. A model trained and tested on randomly partitioned samples from the same collection can benefit from family similarity, duplicated artifacts, or temporally correlated features.

VIII. Feature Engineering and Representation

Feature family	Exempl es	Analysis mode	Advantag es	Risks / limitations
PE / executabl e metadata	Header s, sections , imports , entropy , sizes	Static	Fast and scalable	Packing/obf uscation can distort features
Opcode / byte sequence	Opcode s, byte n- grams, raw bytes	Static	Fine- grained program informati on	High dimensionali ty; evasion possible
API / system calls	Call frequen cy, sequenc es, argume nts	Static/Dy namic	Behaviora l semantics	Runtime cost and sandbox sensitivity
Network behavior	DNS, flows, destinat ions, packet statistic s	Dynamic	Captures communi cation behavior	Environment dependence and encryption
Permissi ons / compone	Androi d permiss ions,	Static	Low-cost mobile features	Permission overuse may not imply maliciousnes

nts	intents, compon ents			s
Program graphs	CFG, FCG, API graph	Static/Dy namic	Structural relationsh ips and obfuscatio n resistance	Graph extraction and scalability
Memory / runtime traces	Process es, registry , files, memor y artifacts	Dynamic	Rich behaviora l evidence	Expensive and potentially evasive
Multimo dal represent ations	Static + dynami c + graph + networ k	Hybrid	Complem entary evidence	Complex fusion and deployment

IX. Comparative Assessment of ML, DL, Hybrid and Emerging Approaches

Criterion	Traditio nal ML	Deep learning	Hybrid / multimod al	Graph / Transforme r / federated
Feature engineerin g	High	Low– medium	Medium– high	Medium
Training cost	Low– medium	High	High	High–very high
Inference cost	Low– medium	Medium– high	High	Medium– high
Interpretab ility	General ly higher	Generally lower	Medium	Low– medium; improving with XAI
Zero-day potential	Moderate	Moderate –high	High potential	High potential
Obfuscatio n resilience	Moderate	Moderate	Higher potential	High potential for structural methods

Data requirement	Medium	High	High	High/distributed
Edge deployment	Good for compact models	Challenging	Challenging	Model-dependent
Maturity	High	High	Medium	Emerging
Main research need	Better features and temporal validation	Robustness and explainability	Standardized multimodal benchmarks	Reliable, efficient, trustworthy deployment

X. PROPOSED INTEGRATED FRAMEWORK

Based on the reviewed evidence, this paper proposes an Integrated Adaptive Explainable Malware Detection Framework (IAEMDF). The framework is conceptual rather than experimentally validated in this survey.

Stage	Function	Candidate methods
1. Data acquisition	Collect multi-platform malware and benign samples	EMBER, AndroZoo, IoT-23, VirusShare
2. Preprocessing	Deduplicate, normalize, label, temporal split	Hashing, family-aware split
3. Static extraction	Extract structural indicators	PE, imports, opcodes, bytes
4. Dynamic extraction	Observe runtime behavior	API/system calls, files, network
5. Graph construction	Model program relationships	CFG, FCG, API graph
6. Representation learning	Learn complementary embeddings	CNN, LSTM, Transformer, GNN
7. Fusion	Combine modalities	Attention/ensemble
8. Robustness	Test evasive manipulation	Adversarial evaluation

9. XAI	Explain decisions	SHAP, graph explanations
10. Adaptation	Handle temporal change	Continual/federated learning
11. Deployment	Real-time classification	Edge/cloud hybrid

XI. Challenges and Research Gaps

Gap	Observed problem	Research implication	Recommended direction
Benchmark standardization	Datasets differ in age, labels, family overlap and feature availability	Scores are difficult to compare	Temporal, family-disjoint and cross-dataset benchmarks
Zero-day generalization	Models often learn benchmark-specific patterns	High test accuracy may not transfer	Leave-family-out and future-sample evaluation
Concept drift	Malware and benign software evolve	Model degradation over time	Continual learning and drift monitoring
Adversarial robustness	Feature manipulation and sandbox evasion can fool detectors	Security of detector itself is uncertain	Adversarial testing and robust training
Explainability	DL/GNN/Transformer decisions can be opaque	Limits analyst trust and incident response	XAI, feature attribution, counterfactual explanations
Class imbalance	Benign/malware proportions and family distributions vary	Accuracy can hide minority failure	PR-AUC, MCC, balanced sampling and cost-sensitive learning
Dataset quality	Labels can be noisy or based on imperfect engines	Models may learn label artifacts	Multi-source validation and label auditing

Deployment cost	DL and hybrid systems can be expensive	Difficult for mobile/IoT	Compression, distillation, quantization and edge inference
Privacy	Threat intelligence may contain sensitive telemetry	Centralized learning can create leakage	Federated learning with privacy/security controls
Reproducibility	Code, preprocessing and splits are inconsistently shared	Results cannot be independently reproduced	Open protocols, versioned datasets and complete reporting

XII. CONCLUSIONS

Machine learning and deep learning have transformed malware detection from predominantly signature-driven matching toward adaptive pattern recognition. The surveyed evidence shows strong growth in deep learning, continued dominance of supervised learning, persistent reliance on static features, and increasing interest in hybrid, graph, explainable, federated and Transformer-based approaches. Yet the most important unresolved issue is not whether intelligent models can achieve high benchmark accuracy; it is whether they can maintain reliable detection against unseen malware, temporal drift, adversarial manipulation, obfuscation, heterogeneous platforms and operational constraints.

A Scopus-oriented research contribution should therefore move beyond another benchmark-specific accuracy claim. The stronger scientific position is to establish reproducible, temporal and cross-dataset evaluation; compare predictive performance with robustness and computational cost; and integrate explainability and deployment considerations. The resulting research agenda is centered on trustworthy, multimodal, adaptive and resource-aware malware detection.

ACKNOWLEDGEMENT

The authors sincerely acknowledge their institution, faculty members, researchers, and cybersecurity professionals for their valuable support and scholarly contributions. Published studies, datasets, tools, and technical resources provided the foundation for this survey of signature-based, behavioral, and AI-driven malware detection approaches. The authors also appreciate colleagues, friends, and family for their encouragement.

REFERENCES

- [1] J. Ferdous, R. Islam, A. Mahboubi, and M. Z. Islam, "AI-Based Ransomware Detection: A Comprehensive Review," IEEE Access, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3461965.
- [2] J. Ferdous, R. Islam, A. Mahboubi, and M. Z. Islam, "A review of state-of-the-art malware attack trends and defense mechanisms," IEEE Access, vol. 11, pp. 121118–121141, 2023, doi: 10.1109/ACCESS.2023.3328351.
- [3] B. A. S. Al-Rimy, M. A. Maarof, and S. Z. M. Shaid, "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions," Computers & Security, vol. 74, pp. 144–166, 2018, doi: 10.1016/j.cose.2018.01.001.
- [4] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, and M. K. Khan, "Ransomware: Recent advances, analysis, challenges and future research directions," Computers & Security, vol. 111, Art. no. 102490, 2021, doi: 10.1016/j.cose.2021.102490.
- [5] S. Razauulla et al., "The age of ransomware: A survey on the evolution, taxonomy, and research directions," IEEE Access, vol. 11, pp. 40698–40723, 2023, doi: 10.1109/ACCESS.2023.3268535.
- [6] T. McIntosh et al., "Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions," ACM Computing Surveys, vol. 54, no. 9, pp. 1–36, 2022, doi: 10.1145/3479393.
- [7] I. Almomani, A. Alkhayer, and W. El-Shafai, "E2E-RDS: Efficient end-to-end ransomware detection system based on static-based ML and vision-based DL approaches," Sensors, vol. 23, no. 9, p. 4467, 2023, doi: 10.3390/s23094467.
- [8] E. B. Karbab, M. Debbabi, and A. Derhab, "SwiftR: Cross-platform ransomware fingerprinting using hierarchical neural networks on hybrid features," Expert Systems with Applications, vol. 225, Art. no. 120017, 2023, doi: 10.1016/j.eswa.2023.120017.
- [9] H. S. Anderson and P. Roth, "EMBER: An open dataset for training static PE malware machine learning models," arXiv:1804.04637, 2018.
- [10] Q. M. Yaseen, "The effect of the ransomware dataset age on the detection accuracy of machine learning models," Information, vol. 14, no. 3, p. 193, 2023, doi: 10.3390/info14030193.
- [11] B. A. S. Al-Rimy, M. A. Maarof, and S. Z. M. Shaid, "Crypto-ransomware early detection model using novel incremental bagging with enhanced semi-random subspace

- selection,” *Future Generation Computer Systems*, vol. 101, pp. 476–491, 2019, doi: 10.1016/j.future.2019.06.005.
- [12] Z.-G. Chen, H.-S. Kang, S.-N. Yin, and S.-R. Kim, “Automatic ransomware detection and analysis based on dynamic API calls flow graph,” *Proc. Int. Conf. Research in Adaptive and Convergent Systems*, pp. 196–201, 2017, doi: 10.1145/3129676.3129704.
- [13] Y. A. Ahmed et al., “A system call refinement-based enhanced minimum redundancy maximum relevance method for ransomware early detection,” *Journal of Network and Computer Applications*, vol. 167, Art. no. 102753, 2020, doi: 10.1016/j.jnca.2020.102753.
- [14] S. Homayoun et al., “DRTHIS: Deep ransomware threat hunting and intelligence system at the fog layer,” *Future Generation Computer Systems*, vol. 90, pp. 94–104, 2019, doi: 10.1016/j.future.2018.07.045.
- [15] H. Zhang et al., “Classification of ransomware families with machine learning based on N-gram of opcodes,” *Future Generation Computer Systems*, vol. 90, pp. 211–221, 2019, doi: 10.1016/j.future.2018.07.052.
- [16] S. Poudyal and D. Dasgupta, “Analysis of crypto-ransomware using ML-based multi-level profiling,” *IEEE Access*, vol. 9, pp. 122532–122547, 2021, doi: 10.1109/ACCESS.2021.3109260.
- [17] D. W. Fernando, N. Komninos, and T. Chen, “A study on the evolution of ransomware detection using machine learning and deep learning techniques,” *IoT*, vol. 1, no. 2, pp. 551–604, 2020, doi: 10.3390/iot1020030.
- [18] J. Hwang, J. Kim, S. Lee, and K. Kim, “Two-stage ransomware detection using dynamic analysis and machine learning techniques,” *Wireless Personal Communications*, vol. 112, no. 4, pp. 2597–2609, 2020, doi: 10.1007/s11277-020-07166-9.
- [19] D. W. Fernando and N. Komninos, “FeSA: Feature selection architecture for ransomware detection under concept drift,” *Computers & Security*, vol. 116, Art. no. 102659, 2022, doi: 10.1016/j.cose.2022.102659.
- [20] B. Kolosnjaji et al., “Adversarial malware binaries: Evading deep learning for malware detection in executables,” *Proc. EUSIPCO*, pp. 533–537, 2018, doi: 10.23919/EUSIPCO.2018.8553214.
- [21] D. Li and Q. Li, “Adversarial deep ensemble: Evasion attacks and defenses for malware detection,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3886–3900, 2020, doi: 10.1109/TIFS.2020.3003571.
- [22] H. Rathore, A. Sasan, S. K. Sahay, and M. Sewak, “Defending malware detection models against evasion based adversarial attacks,” *Pattern Recognition Letters*, vol. 164, pp. 119–125, 2022, doi: 10.1016/j.patrec.2022.10.010.
- [23] K. Shaukat, S. Luo, and V. Varadharajan, “A novel machine learning approach for detecting first-time-appeared malware,” *Engineering Applications of Artificial Intelligence*, vol. 131, Art. no. 107801, 2024, doi: 10.1016/j.engappai.2023.107801.
- [24] F. Mercaldo, “A framework for supporting ransomware detection and prevention based on hybrid analysis,” *Journal of Computer Virology and Hacking Techniques*, vol. 17, no. 3, pp. 221–227, 2021, doi: 10.1007/s11416-021-00388-w.
- [25] S. K. Shaukat and V. J. Ribeiro, “RansomWall: A layered defence system against cryptographic ransomware attacks using machine learning,” *Proc. 10th Int. Conf. Communication Systems & Networks*, pp. 356–363, 2018, doi: 10.1109/COMSNETS.2018.8328219.
- [26] F. Khan et al., “A digital DNA sequencing engine for ransomware detection using machine learning,” *IEEE Access*, vol. 8, pp. 119710–119719, 2020, doi: 10.1109/ACCESS.2020.3003785.
- [27] E. Berrueta, D. Morato, E. Magaña, and M. Izal, “Crypto-ransomware detection using machine learning models in file-sharing network scenarios with encrypted traffic,” *Expert Systems with Applications*, vol. 209, Art. no. 118299, 2022, doi: 10.1016/j.eswa.2022.118299.
- [28] N. Lachtar, D. Ibdah, H. Khan, and A. Bacha, “RansomShield: A visualization approach to defending mobile systems against ransomware,” *ACM Transactions on Privacy and Security*, vol. 26, no. 3, pp. 1–30, 2023, doi: 10.1145/3579822.
- [29] S. K. Shaukat, S. Luo, and V. Varadharajan, “A novel method for improving the robustness of deep learning-based malware detectors against adversarial attacks,” *Engineering Applications of Artificial Intelligence*, vol. 116, Art. no. 105461, 2022, doi: 10.1016/j.engappai.2022.105461.
- [30] J. Chen et al., “ELAMD: An ensemble learning framework for adversarial malware defense,” *Journal of Information Security and Applications*, vol. 75, Art. no. 103508, 2023, doi: 10.1016/j.jisa.2023.103508.
- [31] X. Zhang, J. Wang, and S. Zhu, “Dual generative adversarial networks based unknown encryption ransomware attack detection,” *IEEE Access*, vol. 10, pp. 900–913, 2022, doi: 10.1109/ACCESS.2021.3128024.
- [32] L. Yang, A. Ciptadi, I. Laziuk, A. Ahmadzadeh, and G. Wang, “BODMAS: An open dataset for learning based temporal analysis of PE malware,” *Proc. IEEE Security and*

- Privacy Workshops, pp. 78–84, 2021, doi: 10.1109/SPW53761.2021.00020.
- [33] A. O. Almarshhadani et al., “A multi-classifier network-based crypto ransomware detection system: A case study of Locky ransomware,” *IEEE Access*, vol. 7, pp. 47053–47067, 2019, doi: 10.1109/ACCESS.2019.2907485.
- [34] A. O. Almarshhadani, D. Carlin, M. Kaiiali, and S. Sezer, “MFM CNS: A multi-feature and multi-classifier network-based system for ransomworm detection,” *Computers & Security*, vol. 121, Art. no. 102860, 2022, doi: 10.1016/j.cose.2022.102860.
- [35] K. Jethva et al., “Multi-layer ransomware detection using grouped registry key operations, file entropy and file signature monitoring,” *Journal of Computer Security*, vol. 28, no. 3, pp. 337–373, 2020, doi: 10.3233/JCS-191346.
- [36] S. Sharmeen et al., “Avoiding future digital extortion through robust protection against ransomware threats using deep learning based adaptive approaches,” *IEEE Access*, vol. 8, pp. 24522–24534, 2020, doi: 10.1109/ACCESS.2020.2970466.
- [37] H. Daku, P. Zavorsky, and Y. Malik, “Behavioral-based classification and identification of ransomware variants using machine learning,” *Proc. IEEE TrustCom/BigDataSE*, pp. 1560–1564, 2018, doi: 10.1109/TrustCom/BIGDATASE.2018.00224.
- [38] M. Egele, T. Scholte, E. Kirda, and C. Kruegel, “A survey on automated dynamic malware-analysis techniques and tools,” *ACM Computing Surveys*, vol. 44, no. 2, pp. 1–42, 2008, doi: 10.1145/2089125.2089126.
- [39] S. Banescu et al., “A framework for empirical evaluation of malware detection resilience against behavior obfuscation,” *Proc. MALWARE*, pp. 40–47, 2015, doi: 10.1109/MALWARE.2015.7413683.
- [40] A. Aghakhani et al., “When malware is packin’ heat; limits of machine learning classifiers based on static analysis features,” *NDSS*, 2020, doi: 10.14722/ndss.2020.24310.
- [41] R. Vinayakumar et al., “Robust intelligent malware detection using deep learning,” *IEEE Access*, 2019.
- [42] Ö. Aslan and R. Samet, “A comprehensive review on malware detection approaches,” *IEEE Access*, 2020.
- [43] S. Li, Q. Zhou, R. Zhou, and Q. Lv, “Intelligent malware detection based on graph convolutional network,” *Journal of Supercomputing*, 2021, doi: 10.1007/s11227-021-04020-y.
- [44] M. J. A. Awan et al., “Image-based malware classification using VGG19 network and spatial convolutional attention,” *Electronics*, vol. 10, no. 19, p. 2444, 2021.
- [45] R. Raff et al., “Classifying sequences of extreme length with constant memory applied to malware detection,” *Proc. AAAI*, 2021.
- [46] R. Kumar and G. Subbiah, “Zero-day malware detection and effective malware analysis using Shapley ensemble boosting and bagging approach,” *Sensors*, vol. 22, no. 7, p. 2798, 2022.
- [47] L. Cui et al., “API2Vec: Learning representations of API sequences for malware detection,” *Proc. ACM*, 2023.
- [48] F. Deldar and M. Abadi, “Deep learning for zero-day malware detection and classification: A survey,” *ACM Computing Surveys*, 2023.
- [49] N. Gorment et al., “Machine learning algorithm for malware detection: Taxonomy, current challenges, and future directions,” *IEEE*, 2023.
- [50] M. Gaber, M. Ahmed, and H. Janicke, “Malware detection with artificial intelligence: A systematic literature review,” *ACM*, 2024.
- [51] Shokouhinejad et al., “Recent Advances in Malware Detection: Graph Learning and Explainability,” *arXiv:2502.10556*, 2025.
- [52] Gibert, “Machine Learning for Windows Malware Detection,” *arXiv*, 2024.
- [53] Bensoud and Kalita, “A Survey of Malware Detection Using Deep Learning,” *Neurocomputing*, 2024.
- [54] Maganur et al., “Feature-Centric Android Malware Analysis: A Survey,” *arXiv:2509.10709*, 2025.
- [55] Bean and Torri, “Survey of Malware Detection through Neural Networks,” *ICCSI Proceedings*, 2024.
- [56] Venčkauskas et al., “Systematic Literature Review on Windows Malware Detection,” 2025.
- [57] AlShoulie et al., “Survey of deep learning approaches for malware detection,” 2025.
- [58] Hermosilla et al., “Systematic review of malware classification using machine learning,” 2025.
- [59] He et al., “Survey of adversarial machine learning defenses for malware detection,” 2025.
- [60] Dodiya, “Machine Learning-Based Malware Detection: A Systematic Review,” 2026.