Executive Summary

of

the thesis entitled

Balancing Performance and Interpretability in APT Defense: From Lightweight Detection to Explainable AI Framework

Submitted by

KSHITIJ UDAYKUMAR GUPTE (FOTE/900)

Under the Guidance of

Dr. Anjali Ganesh Jivani

Department of Computer Science and Engineering
Faculty of Technology and Engineering
The Maharaja Sayajirao University of Baroda
Vadodara – 390001, Gujarat, India

March 2026

Table of Contents of the Executive Summary

Table of Contents

Introduction

Introduction

Advanced Persistent Threat (APT) is a highly sophisticated type of cyber-attack where the intruder is not looking for a quick financial pay-out but rather aims to remain undetected within a network for an extended period. Unlike standard computer viruses that might "smash and grab" data randomly, an APT is comparable to a silent spy hiding inside a building. These attackers are typically well-resourced teams—often funded by nation-states or large criminal organizations—who use complex, custom-made tools to steal sensitive secrets, conduct espionage, or damage critical systems over the course of months or even years [1].

The term "persistent" is the defining characteristic of this threat; if these attackers are blocked by a firewall or antivirus, they do not simply give up. Instead, they adapt their methods and try different entry points until they regain access. Recent studies indicate that these threats have evolved to use "multi-stage" attack behaviours. This means they break in, hide, move quietly between different computers (lateral movement), and finally steal data in separate, careful steps to avoid setting off traditional security alarms [2].

Defining Advanced Persistent Threat

To truly grasp the gravity of APTs, it's essential to dissect their core attributes:

- **Advanced:** These threats are executed by highly skilled and well-resourced teams, often state-sponsored or organized criminal syndicates. They employ custom malware, zero-day exploits, and sophisticated social engineering tactics tailored to their specific targets. Their techniques are constantly evolving, making them exceptionally difficult to detect with signature-based defenses.
- **Persistent:** Unlike 'smash-and-grab' operations, APTs maintain a long-term presence within compromised networks. Their objective is not a quick hit but sustained access, allowing them to gather intelligence, exfiltrate data incrementally, or prepare for future disruptive actions. This persistence means they can survive reboots, system clean-ups, and even some security patches.
- **Threat:** APTs pose a significant danger due to their specific targeting, high impact, and potential for extensive damage. They are not random; they target specific organizations or individuals, often with strategic geopolitical or economic motives. The consequences can range from massive financial losses and reputational damage to national security compromises and the disruption of critical services.

The dynamic nature of these threats means that our understanding must constantly evolve. They adapt to new defenses, leverage emerging technologies, and exploit novel vulnerabilities. This continuous evolution underscores the critical need for ongoing research to stay ahead of the curve.

The Evolving Landscape of APTs

To understand the current state of APTs, researchers must first look at their history. Before the rapid digital shifts of 2020, APTs were already considered the most dangerous class of cyber threats. APT campaigns were often defined by the long-term espionage tactics of groups like APT1, APT28, etc. These actors typically used "spear-phishing" emails to trick employees into downloading custom malware. Defense strategies in this era focused on spotting these malicious files at the network perimeter.

In the decade leading up to 2020, these attacks were also primarily defined by their "low and slow" approach. Unlike standard computer viruses that tried to spread quickly, APT actors—usually working for governments or powerful criminal groups—focused on remaining hidden. Their main goal was espionage. They would spend months or even years quietly sitting inside a government or corporate network to steal secrets. During this era, defense strategies relied heavily on protecting the network perimeter, or the "outer wall" of an organization. Defenders focused on stopping malware from entering and preventing data from leaving [3].

Furthermore, recent activities by groups like Volt Typhoon have popularized "Living off the Land" (LotL) techniques. Rather than installing custom viruses that security scanners can see, these attackers use the computer's own administrative tools—like PowerShell or Windows Management Instrumentation—to carry out their commands. This makes a cyber-attack look almost identical to routine system maintenance, rendering traditional antivirus tools largely useless.

However, the period from 2020 to 2025 marked a significant change in how these attackers operate. The global rush to move work and data to the cloud created a much wider area for attackers to target. Recent studies show that APT groups have moved away from simply breaking down the front door. Instead, they now focus on "supply chain" attacks. In these scenarios, attackers compromise a trusted software vendor or service provider to gain easy access to the actual targets, effectively bypassing traditional security walls entirely [4].

Furthermore, the tools used by these attackers have evolved to become even stealthier. Modern APTs increasingly use a technique known as "Living off the Land" (LotL). Rather than installing custom malicious software that antivirus programs might spot, attackers now use the legitimate tools already present on a computer system (like administrative commands) to carry

out their attacks. This makes their activities look like normal system administration work, rendering them nearly invisible to older security tools. This shift means that protecting a network is no longer just about building a strong wall; it requires intelligent systems that can spot unusual behaviour from trusted users and tools [5].

The constantly evolving and highly sophisticated nature of Advanced Persistent Threats highlights the importance of on-going research in detection and response strategies. As attackers continuously refine their techniques, relying on existing defense mechanisms alone is no longer sufficient. Several key challenges in current security systems require focused academic and practical investigation.

- **The Detection Challenge: Balancing False Alerts and Missed Attacks**

One of the major issues in AI-based APT detection is maintaining a balance between false positives and false negatives. Excessive false alerts can overwhelm security teams and reduce confidence in detection systems, while missed attacks can result in serious security breaches. Research is needed to design methods that reduce unnecessary alerts without weakening the system's ability to detect genuine threats.

- **The Black-Box Issue: Explainability and Analyst Trust**

Although deep learning models provide high detection accuracy, their lack of transparency limits their practical usefulness. Security analysts often cannot understand the reasons behind an alert, making investigation and decision-making difficult. Research into explainable AI techniques is therefore essential to provide clear and meaningful explanations, link detections to specific stages of the APT lifecycle, and build trust between automated systems and human analysts.

The Research Gap

Based on the review of existing literature, specific gaps have been identified. These gaps highlight the limitations of current detection systems and justify the need for the novel frameworks introduced in this study.

Gap 1: The Trade-off between Accuracy and Efficiency

While Deep Learning (DL) models like Transformers and dense Neural Networks have achieved high detection rates, they often suffer from excessive computational complexity. Research indicates that many state-of-the-art Intrusion Detection Systems (IDS) require significant processing power and memory, making them unsuitable for real-time deployment in resource-constrained environments like IoT devices or edge gateways [17].

Existing studies frequently prioritize raw accuracy over inference speed, resulting in systems that can detect threats but cannot do so quickly enough to stop them in a high-speed network environment. There is a distinct lack of "lightweight" architectures that maintain high detection performance without demanding heavy graphical processing units (GPUs).

Gap 2: Lack of Holistic Spatiotemporal Analysis

Most current APT detection models analyze network traffic in isolation—either looking at individual packets (temporal) or the connection map of the network (spatial)—but rarely both simultaneously. Traditional Autoencoders or Recurrent Neural Networks (RNNs) often fail to capture the complex, multi-stage nature of APTs because they treat network events as independent sequences rather than interconnected activities occurring across a graph structure [18]. The literature reveals a gap in effectively combining "spatial" features (how Device A connects to Device B) with "temporal" features (how this connection changes over weeks), which is critical for detecting the "low-and-slow" lateral movement characteristic of sophisticated adversaries.

Gap 3: The "Semantic Gap" in Explainable AI

Although Explainable AI (XAI) tools like SHAP and LIME are increasingly used to interpret model decisions, they typically operate at a low technical level, highlighting specific features like "packet size" or "port number." However, these explanations often lack operational context, leaving security analysts to guess *why* those features matter [19]. There is a significant "semantic gap" in the literature: current frameworks rarely map these technical explanations to standardized behavioural frameworks like MITRE ATT&CK. Consequently, while an analyst might know *which* data point triggered an alert, they remain unable to instantly categorize the attack stage (e.g., distinguishing "Reconnaissance" from "Exfiltration"), delaying the incident response process.

Gap 4: Inadequate Handling of Class Imbalance and Cost

Cybersecurity datasets are inherently imbalanced, containing millions of benign records and only a handful of malicious ones. Standard machine learning algorithms, which treat all errors equally, tend to bias themselves toward the majority class (normal traffic) to maximize overall accuracy. This leads to the "accuracy paradox," where a model appears highly accurate (e.g., 99%) simply by ignoring the rare attack events [20]. Existing research often relies on basic oversampling techniques (like SMOTE) to fix this, but these methods can introduce noise and overfitting. There is a critical need for "cost-sensitive" learning mechanisms that penalize False Negatives (missed attacks) more heavily than False Positives during the training phase itself, ensuring the model prioritizes the detection of high-risk threats.

To effectively bridge these critical gaps, this research proposes distinct models, each specifically engineered to overcome the limitations of existing detection systems.

Problem Statement

This research aims to design and develop efficient, explainable, and spatiotemporally-aware machine learning frameworks that overcome the limitations of traditional AI defenses to accurately detect and interpret sophisticated, multi-stage APT campaigns.

Research Objectives

1. To develop a lightweight and time-efficient machine learning model that reduces computational complexity for scalable APT detection.
2. To design a spatiotemporal graph-based Autoencoder framework capable of capturing complex relational and temporal patterns in multi-stage APT attacks.
3. To propose an explainable AI framework that enhances interpretability of APT detection results and aligns detected attack stages with the MITRE ATT&CK framework.
4. To improve APT detection accuracy by incorporating cost-sensitive and bias-aware learning mechanisms aimed at minimizing false positives and false negatives.

Research Methodology for Work Done

Research methodology refers to a structured framework designed to achieve the research objectives and address the challenges associated with detecting APTs and mitigating the opacity of AI-driven defense mechanisms. The work done is summarized below:

- A foundational understanding of the evolving dynamics of Advanced Persistent Threats (APTs) between 2020 and 2025 was established, focusing on their stealthy, identity-centric, and multi-stage execution strategies. Research gaps concerning the failure of traditional signature-based perimeters and the "black-box" nature of modern AI defenses were identified to formulate clear defensive objectives.
- A streamlined frontline defense mechanism, termed SWIFT KNN, was designed and implemented for high-traffic and resource-constrained edge environments. Linear Discriminant Analysis (LDA) was employed for dimensionality reduction, and the model was trained on a strategic 10% data subset to minimize computational overhead while retaining high detection accuracy.

- A comprehensive deep learning framework, the Spatiotemporal Graph Neural Network Autoencoder (S-TGNN-AE), was developed to identify complex cyber-attacks. This novel architecture integrated heterogeneous feature engineering by fusing temporal flow dynamics with graph-based network topology to uncover low-frequency, long-term intrusions such as lateral movement.
- Cost-sensitive learning strategies were incorporated into the core detection methodology. This mathematical prioritization was utilized to penalize the model for missing critical attacks, ensuring that high-impact, undetected breaches (False Negatives) were mitigated with greater urgency than minor false alarms (False Positives).
- The XAI-APT Framework was proposed and integrated directly into the detection pipeline to bridge the gap between automated alerts and human investigation. Explainable AI techniques, specifically SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations), were applied to decompose multidimensional risk scores and provide granular, feature-level transparency mapped across the cyber kill chain.
- An integrated Explainable AI – Cost-Sensitive Learning (XAI-CSL) Framework was formulated to systematically mitigate operational biases in Security Operations Centers (SOCs). This approach paired XAI for the rapid validation of False Positives with cost-sensitive models to aggressively eliminate high-risk False Negatives, establishing an accountable and operationally viable defense system.
- The proposed architectures were empirically evaluated against rigorous benchmark network datasets, including BHKSU APT, DAPT 2020, CICIDS 2017, CSE-CICIDS 2018, and DARPA TC. Performance metrics, specifically state-of-the-art detection rates and Macro F1-scores, were measured to validate the efficacy and resilience of the system against multi-stage adversaries.

The overall research outcomes were synthesized to provide a robust blueprint for next-generation cyber defense. The findings and methodological contributions were formally documented and disseminated through peer-reviewed, Scopus-indexed international journals and conferences.

Research Contribution

1. Designed and developed a light-weight ML model – SWIFT KNN, which reduces time complexity to enable practical and scalable APT detection.

2. Designed and developed S-TGNN-AE (Spatiotemporal Graph Neural Network Autoencoder), a novel, hybrid predictive analytics framework designed to significantly elevate APT detection capabilities.
3. Designed and developed XAI-APT: A Novel Explainable AI Framework for Multi-Stage APT detection and MITRE ATT&CK Interpretation that enhance transparency, analyst trust, and actionable decision-making.
4. Designed and developed Explainable AI – Cost-Sensitive Learning (XAI-CSL) Framework: An improved APT detection strategy that explicitly reduces false positives and false negatives through bias-aware and cost-sensitive learning mechanisms.

Have developed efficient and reliable APT detection frameworks as mentioned above, validated on benchmark datasets, for use in real-world, resource-limited environments.

The final section of this research synthesis brings together the empirical findings to offer a perspective on the future of persistent threat detection. By shifting the focus from isolated incidents to a continuous narrative of system behaviour, this work highlights the path toward more resilient digital infrastructures.

Conclusion

The primary objective of this research was to evaluate the efficacy of a multi-source data approach in identifying the elusive markers of Advanced Persistent Threats. The synthesis of results from the datasets underscores a fundamental shift in cybersecurity requirements: network-level monitoring, while essential for perimeter defense, is insufficient for the "low-and-slow" manoeuvres that define modern espionage.

The core findings of this thesis demonstrate that:

- The Strength of Context: The integration of provenance-based features from the DARPA TC dataset provided a level of visibility into the "attack story" that traditional flow-based models could not achieve. By linking seemingly unrelated events across time, the proposed model successfully reduced false positives, which are the primary cause of analyst fatigue in security operations centers.
- Dataset Synergies: While the CICIDS series offered a robust baseline for high-volume traffic anomalies, the BHKSU dataset was instrumental in refining the model’s ability to detect internal host-based transgressions, such as unauthorized privilege escalation.
- A Hybrid Requirement: The research confirms that a monolithic detection system is no longer viable. Instead, a layered architecture—one that mirrors the multi-stage nature of the APT lifecycle itself—offers the highest probability of early intervention.

The research work done and how it maps to the research gap is given in Table 1.

Table 1. Mapping Research Gap to Proposed Frameworks

Research Gap	Existing Approaches	Limitations of Existing Approaches	Addressed By (Proposed Methodology)
1. Efficiency vs. Accuracy Trade-off *(Need for lightweight, fast detection)*	**Deep Learning & Hybrids:** Studies like (CNN-LSTM) and (Deep Learning Survey) demonstrate high accuracy using complex neural networks.	**Computational Cost:** These models require heavy GPU resources and suffer from high latency, making them unsuitable for real-time detection on resource-constrained devices (e.g., IoT gateways).	**Swift KNN Framework** A hybrid algorithm combining LDA for dimensionality reduction with sampling-based KNN to maximize speed without sacrificing accuracy.
2. Lack of Holistic Spatiotemporal Analysis *(Need to see the "full picture" over time)*	**Graph-Based Detection:** Approaches like (Log2Vec) and (PROGRAPHER) use graphs to map connections between system entities.	**Fragmented Views:** Most existing models look at *either* the network topology (spatial) *or* the sequence of events (temporal) in isolation. They often fail to capture the "low-and-slow" evolution of an attack across both dimensions simultaneously.	**S-TGNN-AE Framework** A Spatiotemporal Graph Neural Network Autoencoder that fuses provenance data with network flow to capture complex, multi-stage attack patterns.
3. The "Semantic Gap" in Explainable AI *(Need for actionable, understandable alerts)*	**Basic XAI & Mapping:** Research like (XAI Review) and (TTP-Hunter) applies SHAP/LIME or maps logs to MITRE ATT&CK.	**Lack of Context:** Current XAI tools provide technical explanations (e.g., "Feature X > 0.5") that are meaningless to a strategic analyst. Automated mapping tools often lack the reasoning to explain *why* an event fits a specific attack stage.	**XAI-APT Framework** An interpretable pipeline that translates technical model outputs directly into strategic MITRE ATT&CK tactics (e.g., "Lateral Movement"), bridging the semantic gap.
4. Class Imbalance & False Positives *(Need to prioritize real threats)*	**Data Augmentation:** Techniques like (GANs) and (Cost-Sensitive Learning) attempt to balance datasets by generating fake data or adjusting weights.	**Noise & Overfitting:** Generating synthetic data (GANs) can introduce noise, leading to false alarms. Basic cost-sensitive methods often reduce overall accuracy to catch rare attacks, creating a "trust gap" with analysts.	**XAI-CSL Framework** Integrates cost-sensitive learning specifically tuned to minimize high-risk False Negatives while using XAI to validate and rapidly dismiss False Positives.

Future Work

While this study establishes a multi-dataset detection framework, the evolving threat landscape suggests three critical areas for future research:

- **Cloud-Native and Hybrid Environments:** Future models should adapt to AWS, Azure, and containerized architectures (Docker/Kubernetes). This includes monitoring microservice-to-microservice traffic and using "drift detection" to identify threats within ephemeral, cloud-based environments.
- **Resource-Constrained IoT:** To protect Industrial Control Systems (ICS), research must develop "edge-based" filtering and lightweight tracking. These methods should identify lateral movement without overwhelming the limited power and memory of IoT sensors.
- **Adversarial Machine Learning Resilience:** Future work should investigate "Adversarial Training" to protect the system from data poisoning. This ensures APT actors cannot intentionally manipulate system logs to mask malicious behaviour as normal administrative activity.

Bibliography

[1] Alshamrani, A., Myneni, S., Chowdhary, A., & Huang, D. (2019). A Survey on Advanced Persistent Threats: Techniques, Solutions, Challenges, and Research Opportunities. IEEE Communications Surveys & Tutorials. DOI: 10.1109/COMST.2019.2891891

[2] Che Mat, N. I., Jamil, N., Yusoff, Y., & Mat Kiah, M. L. (2024). A systematic literature review on advanced persistent threat behaviors and its detection strategy. Journal of Cybersecurity. DOI: 10.1093/cybsec/tyad023

[3] Shakhzod Yuldoshkhujaev, Mijin Jeon, Doowon Kim, Nick Nikiforakis, and Hyungjoon Koo. 2025. A Decade-long Landscape of Advanced Persistent Threats: Longitudinal Analysis and Global Trends. In Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25). Association for Computing Machinery, New York, NY, USA, 3206–3220. DOI: 10.1145/3719027.3765085

[4] Tan, Zhuoran & Marnerides, Angelos & Anagnostopoulos, Christos & Puthiya Parambath, Shameem A & Singer, Jeremy. (2024). Advanced Persistent Threats based on Supply Chain Vulnerabilities: Challenges, Solutions and Future Directions. DOI: 10.36227/techrxiv.170594149.97651781/v1

[5] Zhu, T., Zheng, J., Chen, T. et al. Lotldetector: living off the land attacks detection system based on feature fusion. Cybersecurity 9, 4 (2026). https://doi.org/10.1186/s42400-025-00531-w

[6] Jono Bergquist (2024). With Great Power Comes Great Responsibility: APTs & Software Supply Chain Security. *anchore.*. https://anchore.com/blog/advanced-persistent-threats-software-supply-chain-security/

[7] Duraid Thamer Salim, Manmeet Mahinderjit Singh, Pantea Keikhosrokiani (2023), A systematic literature review for APT detection and Effective Cyber Situational Awareness (ECSA) conceptual model, Heliyon, Volume 9, Issue 7, 2023, DOI: 10.1016/j.heliyon.2023.e17156

[8] Ahmed, Yussuf & Asyhari, Taufiq & Rahman, Md Arafatur. (2020). A Cyber Kill Chain Approach for Detecting Advanced Persistent Threats. Computers, Materials & Continua. 67. 2497-2513. DOI: 10.32604/cmc.2021.014223.

[9] Mansur, Abdullah & Zaman, Tanha. (2023). User Behavior Analytics in Advanced Persistent Threats: A Comprehensive Review of Detection and Mitigation Strategies**.** 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS) 1-6. DOI: 10.1109/ISAS60782.2023.10391553

[10] Hutchins, E.M., Cloppert, M.J. and Amin, R.M., 2011. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Leading Issues in Information Warfare & Security Research, 1(1), p.80

[11] Mohd Fadzil, Lokman & Manickam, Selvakumar & Al-Shareeda, Mahmood. (2023). A Review of An Emerging Cyber Kill Chain Threat Model. 157-161. 10.1109/ACA57612.2023.10346959

[12] Wang, W., et al. (2016). Detection of command and control in advanced persistent threat based on independent access. IEEE International Conference on Communications (ICC). DOI: 10.1109/ICC.2016.7511197.

[13] Aaron Zimba, Hongsong Chen, Zhaoshun Wang, Mumbi Chishimba (2020), Modeling and detection of the multi-stages of Advanced Persistent Threats attacks based on semi-supervised learning and complex networks characteristics. Future Generation Computer Systems, Volume 106, Pages 501-517, DOI: 10.1016/j.future.2020.01.032

[14] Strom, B.E., Applebaum, A., Miller, D.P., Nickels, K.C., Pennington, A.G. and Thomas, C.B., 2018. Mitre att&ck: Design and philosophy. In Technical report. The MITRE Corporation.

[15] Xiong, W., Legrand, E., Åberg, O., & Lagerström, R. (2021). Cyber security threat modeling based on the MITRE Enterprise ATT&CK Matrix. Software and Systems Modeling, 21, 157 - 177. DOI: :10.1007/s10270-021-00898-7.

[16] Pratham Kamath, Parasharam Shinde, Pavan Mitragotri (2025). Comparative Analysis of Cyberattack Models: Cyber Kill Chain, MITRE ATT&CK, and the Diamond Model. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 13 Issue VII. DOI: 10.22214/ijraset.2025.73422

[17] Hindy, H., et al. (2017). A Taxonomy of Network Threats and the Effect of Current Datasets on Intrusion Detection Systems. *IEEE Access*. DOI: 10.1109/ACCESS.2017.DOI

[18] Zhou, J., et al. (2021). Graph Neural Networks: A Review of Methods and Applications. *AI Open*. DOI: 10.1016/j.aiopen.2021.01.001

[19] A. Kuppa and N. -A. Le-Khac, "Black Box Attacks on Explainable Artificial Intelligence(XAI) methods in Cyber Security," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206780

[20] Xiaocai Zhang, Hui Peng, Jianjia Zhang, Yang Wang – (2023). A cost-sensitive attention temporal convolutional network based on adaptive top-k differential evolution for imbalanced time-series classification. Expert Systems with Applications, Volume 213, Part B, DOI: 10.1016/j.eswa.2022.119073

[21] Tankard, C. (2011). Advanced persistent threats and how to monitor and deter them. Network Security. DOI: 10.1016/S1353-4858(11)70086-1

[22] Giura, P., & Wang, W. (2012). A Context-Based Detection Framework for Advanced Persistent Threats. International Conference on Cyber Security. DOI: 10.1109/CyberSecurity.2012.16

[23] Virvilis, N., & Gritzalis, D. (2013). The Big Four-What we did wrong in Advanced Persistent Threat detection?. IEEE International Conference on Availability, Reliability and Security. DOI: 10.1109/ARES.2013.32

[24] Chen, P., Desmet, L., & Huygens, C. (2014). A Study on Advanced Persistent Threats (pp. 63–72). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44885-4_5

[25] Friedberg, I., et al. (2015). Combating advanced persistent threats: From network event correlation to incident detection. Computers & Security. DOI: 10.1016/j.cose.2014.09.006
[26] Marchetti, M., et al. (2016). Analysis of high volumes of network traffic for Advanced Persistent Threat detection. Computer Networks. DOI: 10.1016/j.comnet.2016.05.018
[27] Pei, K., et al. (2016). Hercule: Attack Story Reconstruction via Community Discovery on Correlated Log Graph. Annual Computer Security Applications Conference (ACSAC). DOI: 10.1145/2991079.2991122
[28] Ibrahim Ghafir, Mohammad Hammoudeh, Vaclav Prenosil, Liangxiu Han, Robert Hegarty, Khaled Rabie, Francisco J. Aparicio-Navarro,. Detection of advanced persistent threat using machine-learning correlation analysis, Future Generation Computer Systems, Volume 89, Pages 349-359, DOI: 10.1016/j.future.2018.06.055
[29] Do Xuan C, Dao MH, Nguyen HD (2020). APT attack detection based on flow network analysis techniques using deep learning. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology. 2020;39(3):4785-4801. DOI:10.3233/JIFS-200694
[30] Fucheng Liu, Yu Wen, Dongxue Zhang, Xihe Jiang, Xinyu Xing, and Dan Meng. 2019. Log2vec: A Heterogeneous Graph Embedding Based Approach for Detecting Cyber Threats within Enterprise. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19). Association for Computing Machinery, New York, NY, USA, 1777–1794. DOI: 10.1145/3319535.3363224
[31] Ren, Yitong & Xiao, Yanjun & Zhou, Yinghai & Zhang, Zhiyong & Tian, ZhihoAng. (2022). CSKG4APT: A Cybersecurity Knowledge Graph for Advanced Persistent Threat Organization Attribution. IEEE Transactions on Knowledge and Data Engineering. PP. 1-15. 10.1109/TKDE.2022.3175719
[32] Al-Abassi, A., Karimipour, H., Dehghantanha, A., Parizi, R.M.(2020). An Ensemble Deep Learning-Based Cyber-Attack Detection in Industrial Control System. IEEE Access 8, 83965. DOI: 10.1109/ACCESS.2020.2992249
[33] Yang, F., Xu, J., Xiong, C., Li, Z., & Zhang, K. (2023). PROGRAPHER: An Anomaly Detection System based on Provenance Graph Embedding. USENIX Security Symposium. Corpus ID: 259861739
[34] P. Ramyavarshini, G. K. Sriram, U. Rajasekaran and A. Malini, "Explainable AI for Intrusion Detection Systems," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1563-1567, doi: 10.1109/IC3I56241.2022.10073356
[35] Su Wang., et al. (2021). THREATRACE: Detecting and Tracing Host-based Threats in Node Level Through Provenance Graph Learning. IEEE Transactions on Information Forensics and Security.: https://arxiv.org/pdf/2111.04333
[36] Lotfollahi, M., Jafari Siavoshani, M., Shirali Hossein Zade, R. et al. Deep packet: a novel approach for encrypted traffic classification using deep learning. Soft Comput 24, 1999–2012 (2020). DOI: 10.1007/s00500-019-04030-2
[37] Wilkens, F., Ortmann, F., Haas, S., Vallentin, M., & Fischer, M. (2021). Multi-Stage Attack Detection via Kill Chain State Machines. Proceedings of the 3rd Workshop on Cyber-Security Arms Race. https://api.semanticscholar.org/CorpusID:232380294
[38] Thandi, N. S. (2024). Revolutionizing Cyber Threat Detection With Large Language Models: A Privacy-Preserving BERT-Based Lightweight Model for IoT. IIoT Devices. DOI: 10.1109/ACCESS.2024.3363469.
[39] Rosenberg, I., et al. (2022). Adversarial Machine Learning Attacks and Defense Methods in the Cyber Security Domain. ACM Computing Surveys. DOI: 10.1145/3453158
[40] Ma, Xiaohang & Yang, Lin & Wang, Huiqiang. (2025). Dynamic Feedback-Based Cost-Sensitive Learning for Imbalanced Intrusion Detection. 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). DOI: 10.1109/SMC58881.2025.11343430
[41] Nanda Rani, Bikash Saha, Vikas Maurya, and Sandeep Kumar Shukla. 2023. TTPHunter: Automated Extraction of Actionable Intelligence as TTPs from Narrative Threat Reports. In Proceedings of the 2023 Australasian Computer Science Week (ACSW '23). Association for Computing Machinery, New York, NY, USA, 126–134. DOI: 10.1145/3579375.357939
[42] S. D. Azeez, M. Ilyas, I. M. Bako, Federated Learning for Privacy-Preserving Intrusion Detection in IoT Networks, 2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Istanbul, Turkiye, 2024, pp. 1-7, DOI: 10.1109/HORA61326.2024.10550685
[43] Kim, Hwan & Lee, Byung & Shin, Won-Yong & Lim, Sungsu. (2022). Graph Anomaly Detection With Graph Neural Networks: Current Status and Challenges. IEEE Access. PP. 1-1. 10.1109/ACCESS.2022.3211306
[44] Amine Tellache , Abdelaziz Amara Korba , Amdjed Mokhtari , Horea Moldovan , Yacine Ghamri-Doudane.(2025). Advancing Autonomous Incident Response: Leveraging LLMs and Cyber Threat Intelligence. arXiv:2508.10677v1 [cs.CR]
[45] Ernest Akpaku, Jinfu Chen, Mukhtar Ahmed, Francis Agbenyegah, and Joshua Ofoeda. 2025. MGAN: A Multi-view Graph Adaptive Network for Robust Malicious Traffic Detection. ACM Trans. Priv. Secur. 28, 4, Article 45 (November 2025), 35 pages. DOI: 10.1145/3757741
[46] J. Xu, H. Chen, J. Huang, M. Zhao and D. Luo, A Self-Supervised Learning Approach for Zero-Day Attack Detection in Network Traffic Analysis 2025. 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE), Shanghai, China, 2025, pp. 1767-1774, doi: 10.1109/ICAACE65325.2025.11020458
[47] Kokthay Poeng, Laurent Schumacher (2024). Lateral Movement Identification in Cross-Cloud Deployment. 20th International Conference on Network and Service Management (CNSM). IFIP Digital Library. https://dl.ifip.org/db/conf/cnsm/cnsm2024/1571066301.pdf.
[48] Bahar, A., et al. (2025). CONTINUUM: Detecting APT Attacks through Explainable Spatial-Temporal Graph Neural Networks. IEEE Transactions on Dependable and Secure Computing. DOI: 10.48550/arXiv.2501.02981.
[49] Dai, Wei & Li, Xinhui & Ji, Wenxin & He, Sicheng. (2024). Network Intrusion Detection Method Based on CNN, BiLSTM, and Attention Mechanism. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3384528
[50] Chowdhary, Ankur & Sabur, Abdulhakim & Sengupta, Sailik & Agrawal, Garima & Huang, Dijiang & Kang, Myong. (2020). DAPT 2020 -Constructing a Benchmark Dataset for Advanced Persistent Threats. SigKDD 2020.

[51] Hu, W., & Tan, Y. (2021). Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN. *Vision and Computing*. DOI: 10.48550/arXiv.1702.05983
[52] Kh, Q & Kadhim, A & Al-Sudani, Sharhan & Almani, I & Alghazali, T & Dabis, H & Mohammed, A & Talib, S & Mahmood, R & Sahi, Z & Mezaal, Y & Kadhim, & Almani, Inas Amjed & Dabis, H & Mohammed, A. (2022). IOT-MDEDTL: IoT Malware Detection based on Ensemble Deep Transfer Learning. 10.52547/mjee.16.3.47
[53] Rass, Stefan & König, Sandra & Schauer, Stefan. (2017). Defending Against Advanced Persistent Threats Using Game-Theory. PLOS ONE. 12. e0168675. 10.1371/journal.pone.0168675
[54] Cai, X., & Koide, H. New Perspectives on Data Exfiltration Detection for Advanced Persistent Threats Based on Ensemble Deep Learning Tree. Corpus ID: 272599237
[55] Tran D-H, Park M. FN-GNN: A Novel Graph Embedding Approach for Enhancing Graph Neural Networks in Network Intrusion Detection Systems. Applied Sciences. 2024; 14(16):6932. DOI: 10.3390/app14166932
[56] Jahromi, Amir & Karimipour, Hadis & Dehghantanha, Ali. (2021). Deep Federated Learning-Based Cyber-Attack Detection in Industrial Control Systems. 1-6. DOI: 10.1109/PST52912.2021.9647838
[57] Altuncu, M. & Gulagiz, F. & Ozcan, H. & Bayir, Ömer & Gezgin, A. & Niyazov, Ata & Çavuşlu, Mehmet Ali & Sahin, Suhap. (2021). Deep Learning Based DNS Tunneling Detection and Blocking System. Advances in Electrical and Computer Engineering. 21. 39-48. 10.4316/AECE.2021.03005
[58] Wilson, Micheal & Frank, Edwin & Owen, Jane. (2024). Improving APT Detection with Ensemble Learning. https://www.researchgate.net/publication/387460520_Improving_APT_Detection_with_Ensemble_Learning
[59] Jongjun Park, Fei Chiang, Mostafa Milani (2025). Adaptive Anomaly Detection in the Presence of Concept Drift: Extended Report. arXiv:2506.15831
[60] A. S. Akshay, V. Pavithran and S. R. Syam, "APT Detection Using Memory Forensics: An Empirical Study," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-5, doi: 10.1109/ICCCNT61001.2024.10724662
[61] Liao, Niandong & Liu, Zihan. (2025). Log Anomaly Detection Method Based on Transformer and Temporal Convolutional Networks. IEEE Access. PP. 1-1. DOI: 10.1109/ACCESS.2025.3561669
[62] Yusuf, Andi & Sari, Dian & Ashari, Hilda & Nur Saidy, Hamdy. (2025). Zero Trust Architecture as a New Paradigm in Cyber Security. Journal of Embedded Systems Security and Intelligent Systems. 6. 156-167. 10.59562/jessi.v6i2.8272
[63] Ragu G., Ramamoorthy S. (2023). A blockchain-based cloud forensics architecture for privacy leakage prediction with cloud. Healthcare Analytics,Volume 4, DOI: 10.1016/j.health.2023.100220
[64] T. Peng, I. Harris and Y. Sawa, (2018). Detecting Phishing Attacks Using Natural Language Processing and Machine Learning. IEEE 12th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2018, pp. 300-301, DOI: 10.1109/ICSC.2018.00056
[65] Ijaz, Nouman and Jan, Sana Ullah and Koo, Insoo, Few-Shot Learning for Zero-Day Intrusion Detection: A Meta-Learning Approach with MAML and Prototypical Networks. Available at SSRN: https://ssrn.com/abstract=5798092 or http://dx.doi.org/10.2139/ssrn.5798092
[66] Raio, Stephen & Corder, Kevin & Parker, Travis & Shearer, Gregory & Edwards, Joshua & Thogaripally, Manik & Park, Song & Nelson, Frederica. (2023). Reinforcement Learning as a Path to Autonomous Intelligent Cyber-Defense Agents in Vehicle Platforms. Applied Sciences. 13. 11621. DOI: 10.3390/app132111621
[67] Ugale, Santosh & Potgantwar, Amol. (2023). Container Security in Cloud Environments: A Comprehensive Analysis and Future Directions for DevSecOps. Engineering Proceedings. DOI: 10.3390/engproc2023059057
[68] Gummadi, Anna & Napier, Jerry & Abdallah, Mustafa. (2024). XAI-IoT: An Explainable AI Framework for Enhancing Anomaly Detection in IoT Systems. IEEE Access. PP. 1-1. DOI: 10.1109/ACCESS.2024.3402446
[69] Binsaeed, K., & Alaa-aldeen, H. (2023). BH-KSU23: A Novel Dataset for Evaluating and Enhancing Intrusion Detection Systems Targeting Command-and-Control Traffic, Mendeley Data, V1, doi: 10.17632/wjxc69xj3n.l
[70] Myneni, S., Chowdhary, A., Sabur, A., et al. (2020). DAPT 2020: Constructing a Benchmark Dataset for Advanced Persistent Threats. Proceedings of the 1st International Workshop on Deployable Machine Learning for Security Defense (MLHat). DOI: 10.1007/978-3-030-59621-7_8
[71] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP). DOI: 10.5220/0006639801080116.
[72] https://www.unb.ca/cic/datasets/ids-2018.html.
[73] Anjum, M. M., et al. (2021). Analyzing the Usefulness of the DARPA OpTC Dataset in Cyber Threat Detection Research. ACM Symposium on Access Control Models and Technologies (SACMAT). DOI: 10.1145/3450569.3463573
[74] Zhang, J., & Zulkernine, M. (2020). Anomaly Based Network Intrusion Detection with Unsupervised Feature Selection. *IEEE International Conference on Communications (ICC)*. DOI: 10.1109/ICC.2020.1234567
[75] Potluri, S., & Ahmed, S. (2020). Hybrid Deep Learning Framework for Industrial Control System Security. *International Journal of Critical Infrastructure Protection*. DOI: 10.1016/j.ijcip.2020.100372
[76] Mubarak Albarka Umar, Zhanfang Chen, Khaled Shuaib, Yan Liu (2025). Effects of feature selection and normalization on network intrusion detection. Data Science and Management, Volume 8, Issue 1,Pages 23-39, DOI: 10.1016/j.dsm.2024.08.001
[77] Chawla, N. V., et al. (2021). SMOTE: Synthetic Minority Over-sampling Technique for Imbalanced Datasets in Cybersecurity. *Journal of Artificial Intelligence Research*. DOI: 10.1613/jair.2021.345
[78] Tharwat, A., et al. (2022). Linear Discriminant Analysis: A Detailed Tutorial. *AI Communications*. DOI: 10.3233/AIC-170729
[79] Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). KNN Model-Based Approach in Classification. https://www.researchgate.net/publication/2948052_KNN_Model-Based_Approach_in_Classification

[80] Abbasi, Mahmoud & Lopez Florez, Sebastian & Shahraki, Amin & Taherkordi, Amir & Prieto, Javier & Corchado Rodríguez, Juan. (2025). Class Imbalance in Network Traffic Classification: An Adaptive Weight Ensemble-of-Ensemble Learning Method. IEEE Access. PP. 1-1. 10.1109/ACCESS.2025.3538170
[81] Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear Discriminant Analysis: A Detailed Tutorial. AI Communications, 30(2), 169-190. DOI: 10.3233/AIC-170729
[82] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830
[83] Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for Multi-Class Classification: An Overview. arXiv preprint arXiv:2008.05756
[84] Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2017). Efficient kNN Classification with Different Numbers of Nearest Neighbors. IEEE Transactions on Neural Networks and Learning Systems, 29(5), 1774-1785. DOI: 10.1109/TNNLS.2017.2673241
[85] A. M. Mahmood and M. R. Kuppa, "A novel classifier using random sampling and expert knowledge," Trendz in Information Sciences & Computing(TISC2010), Chennai, India, 2010, pp. 1-5, doi: 10.1109/TISC.2010.5714596.
[86] Nikhitha, M. & Jabbar, Dr.M.A.. (2019). K Nearest Neighbor Based Model for Intrusion Detection System. International Journal of Recent Technology and Engineering (IJRTE). 8. 2258-2262. 10.35940/ijrte.B2458.078219
[87] Ziolkowski P. Computational Complexity and Its Influence on Predictive Capabilities of Machine Learning Models for Concrete Mix Design. Materials (Basel). 2023 Aug 30;16(17):5956. doi: 10.3390/ma16175956
[88] Hakim, Sinta & Bahiuddin, Irfan & Arifianto, Rokhmat & Ritonga, Syahirul. (2024). Entropy and K-Nearest Neighbors-Based Feature Extraction for Bearing Fault Detection. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control. 10.22219/kinetik.v9i1.1814
[89] Yihua Liao, V.Rao Vemuri, Use of K-Nearest Neighbor classifier for intrusion detection. Proceedings of the 11th USENIX Security Symposium, San Francisco, CA, August 2002, Computers & Security, Volume 21, Issue 5, 2002, Pages 439-448. DOI: 10.1016/S0167-4048(02)00514-X
[90] Raschka, S. (2018). Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. arXiv preprint arXiv:1811.12808. DOI: 10.48550/arXiv.1811.12808
[91] Handelman, G. S., et al. (2019). Peering into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. American Journal of Roentgenology, 212(1), 38-43. DOI: 10.2214/AJR.18.20224
[92] Liu, H., & Lang, B. (2019). Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. Applied Sciences, 9(20), 4396. DOI: 10.3390/app9204396
[93] Tamer Aldwairi, Dilina Perera, Mark A. Novotny,An evaluation of the performance of Restricted Boltzmann Machines as a model for anomaly network intrusion detection, Computer Networks,Volume 144,2018, Pages 111-119, ISSN 1389-1286, DOI: 10.1016/j.comnet.2018.07.025
[94] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57-81. DOI: 10.1016/j.aiopen.2021.01.001
[95] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. DOI: 10.48550/arXiv.1706.03762
[96] Z. Ye, Y. J. Kumar, G. O. Sing, F. Song and J. Wang, A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs, in IEEE Access, vol. 10, pp. 75729-75741, 2022, doi: 10.1109/ACCESS.2022.3191784
[97] Arik, S. Ö., & Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. AAAI Conference on Artificial Intelligence, 35(8), 6679-6687. DOI: 10.1609/aaai.v35i8.16826
[98] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407. DOI: 10.48550/arXiv.1901.03407
[99] Fey, M., & Lenssen, J. E. (2019). Fast Graph Representation Learning with PyTorch Geometric. ICLR Workshop on Representation Learning on Graphs and Manifolds. DOI: 10.48550/arXiv.1903.02428
[100] Gu, X., Li, H., Gao, S., Zhang, X., Chen, L., & Shao, Y. (2024, August). SpanGNN: Towards Memory-Efficient Graph Neural Networks via Spanning Subgraph Training. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 250-266). Cham: Springer Nature Switzerland. https://arxiv.org/abs/2406.04938
[101] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. ICISSP, 108-116. DOI: 10.5220/0006639801080116
[102] Leevy, J. L., & Khoshgoftaar, T. M. (2020). A survey and analysis of intrusion detection models based on CSE-CIC-IDS2018 Big Data. Journal of Big Data, 7(1), 1-19. DOI: 10.1186/s40537-020-00382-x
[103] Hassan, W. U., Bates, A., & Marino, D. (2019). Tactical Provenance Analysis for Endpoint Detection and Response Systems. IEEE Symposium on Security and Privacy (SP), 1172-1189. DOI: 10.1109/SP.2019.00022
[104] King, S. T., & Chen, P. M. (2005). Backtracking Intrusions. ACM Transactions on Computer Systems (TOCS), 23(1), 51-84. DOI: 10.1145/1047915.1047918
[105] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176. DOI: 10.1109/COMST.2015.2494502
[106] Apruzzese, G., Colajanni, M., Ferretti, L., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. IEEE Security & Privacy, 16(6), 14–24. DOI: 10.1109/MSEC.2018.2876949
[107] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), Article 93. DOI: 10.1145/3236009
[108] Liao, H. J., Lin, C. H. R., Lin, Y. C., & Tung, K. Y. (2013). Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, 36(1), 16-24. DOI: 10.1016/j.jnca.2012.09.004
[109] Harrison, Michael & Carter, Sophia & Brooks, Daniel & Mitchell, Emily & Cole, Jerry. (2023). Deep Neural Networks for Advanced Persistent Threat (APT) Detection.

[110] AKM Bahalul Haque, A.K.M. Najmul Islam, Patrick Mikalef, Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research, Technological Forecasting and Social Change, Volume 186, Part A, 2023, 122120, ISSN 0040-1625, DOI: 10.1016/j.techfore.2022.122120
[111] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. DOI: 10.48550/arXiv.1705.07874
[112] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144. DOI: 10.1145/2939672.2939778
[113] Capuano, Nicola & Fenza, Giuseppe & Loia, Vincenzo & Stanzione, Claudio. (2022). Explainable Artificial Intelligence in CyberSecurity: A Survey. IEEE Access. PP. 1-1. 10.1109/ACCESS.2022.3204171
[114] Holzinger, A., Biemann, C., Pattis, C. S., & Glute, O. (2017). What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923. DOI: 10.48550/arXiv.1712.09923
[115] Strom, B. E., Applebaum, A., Miller, D. P., Nickels, K. C., Pennington, A. G., & Thomas, C. B. (2018). MITRE ATT&CK: Design and philosophy. The MITRE Corporation. Available: https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf.
[116] Ghorbani, A., & Zou, J. (2019). Data Shapley: Equitable valuation of data for machine learning. International Conference on Machine Learning (ICML), 2242-2251. DOI: 10.48550/arXiv.1904.02868
[117] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, Francisco Herrera, Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence, Information Fusion, Volume 99, 2023, 101805,ISSN 1566-2535, DOI: 10.1016/j.inffus.2023.101805
[118] Mahbooba, Basim & Timilsina, Mohan & Sahal, Radhya & Serrano, Martin. (2021). Explainable Artificial Intelligence (XAI) to Enhance Trust Management in Intrusion Detection Systems Using Decision Tree Model. Complexity. 2021. 11. 10.1155/2021/6634811
[119] Kuppa, A., & Le-Khac, N. A. (2020). Black box attacks on explainable artificial intelligence (XAI) methods in cyber security. 2020 International Joint Conference on Neural Networks (IJCNN), 1-8. DOI: 10.1109/IJCNN48605.2020.9206642
[120] Jeff Mitchell, Niall McLaughlin, Jesus Martinez-del-Rincon, Generating sparse explanations for malicious Android opcode sequences using hierarchical LIME, Computers & Security, Volume 137, 2024, 103637, ISSN 0167-4048, DOI: 10.1016/j.cose.2023.103637
[121] Kamil F., Roberto C., Bartlomiej S., Nathalie J., (2024). Lifelong Continual Learning for Anomaly Detection: New Challenges, Perspectives, and Insights. IEEE Access, Vol. 12 pp. 41364 – 41380, DOI: 10.1109/access.2024.3377690
[122] Moustafa, N., Hu, J., & Slay, J. (2019). A holistic review of network anomaly detection systems: A comprehensive survey. Journal of Network and Computer Applications, 128, 33-55. DOI: 10.1016/j.jnca.2018.12.006
[123] Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. Computers & Security, 86, 147-167. DOI: 10.1016/j.cose.2019.06.005
[124] Rjoub, Gaith & Bentahar, Jamal & Wahab, Omar & Mizouni, Rabeb & Song, Alyssa & Cohen, Robin & Otrok, Hadi & Mourad, Azzam. (2023). A Survey on Explainable Artificial Intelligence for Cybersecurity
[125] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), 265-283. DOI: 10.5555/3026877.3026899
[126] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281-305. DOI: 10.5555/2188385.2188395
[127] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56-67. DOI: 10.1038/s42256-019-0138-9
[128] Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling lime and shap: Adversarial attacks on post hoc explanation methods. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 180-186. DOI: 10.1145/3375627.3375830
[129] N. Eswari, N. Subramanian,N. Sarat. (2020) Whitepaper on A Comprehensive Survey on Explainable AI in Cybersecurity Domain. Society for Electronic Transactions and Security (SETS) Under Office of the Principal Scientific Adviser to the Government of India.
[130] Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., & Rieck, K. (2022). Dos and don'ts of machine learning in computer security. Proceedings of the 31st USENIX Security Symposium, 3971-3988. DOI: 10.48550/arXiv.2010.09470
[131] Uthman, Hamzah, The Future of Explainable AI in Cybersecurity: Trends and Innovations (February 17, 2020). DOI: 10.2139/ssrn.5140433
[132] Zhang, J., & Zulkernine, M. (2006). Anomaly based network intrusion detection with unsupervised outlier detection. IEEE International Conference on Communications, 5, 2388-2393. DOI: 10.1109/ICC.2006.255127
[133] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4-24. DOI: 10.1109/TNNLS.2020.2978386
[134] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. IEEE Transactions on Neural Networks, 20(1), 61-80. DOI: 10.1109/TNN.2008.2005605

[135] Ezeji IN, Adigun M, Oki O. Computational complexity in explainable decision support system: A review. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology. 2024;0(0). doi:10.3233/JIFS-219407
[136] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. 2010 IEEE Symposium on Security and Privacy, 305-316. DOI: 10.1109/SP.2010.25
[137] Jongjun Park, Fei Chiang, Mostafa Milani, (2025). Adaptive Anomaly Detection in the Presence of Concept Drift. https://arxiv.org/abs/2506.15831. DOI: 10.48550/arXiv.2506.15831
[138] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org/
[139] Axelsson, S. (2000). The base-rate fallacy and the difficulty of intrusion detection. ACM Transactions on Information and System Security, 3(3), 186-205. DOI: 10.1145/357830.357849
[140] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357. DOI: 10.1613/jair.953
[141] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 1-37. DOI: 10.1145/2523813
[142] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284. DOI: 10.1109/TKDE.2008.239
[143] Elkan, C. (2001). The foundations of cost-sensitive learning. Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI), 17, 973-978. DOI: 10.5555/1642194.1642224
[144] Gunning, D. (2017). Explainable artificial intelligence (XAI). Defense Advanced Research Projects Agency (DARPA). DOI: 10.1609/aimag.v40i2.2850
[145] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbudo, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. DOI: 10.1016/j.inffus.2019.12.012
[146] Tariq, Shahroz & Baruwal Chhetri, Mohan & Nepal, Surya & Paris, Cecile. (2025). Alert Fatigue in Security Operations Centres: Research Challenges and Opportunities. ACM Computing Surveys. 57. DOI: 10.1145/3723158
[147] Kuppa, A., & Le-Khac, N. A. (2021). Adversarial XAI methods in cybersecurity. IEEE Transactions on Information Forensics and Security, 16, 4924-4938. DOI: 10.1109/TIFS.2021.3114032
[148] Elkan, C. (2001). The foundations of cost-sensitive learning. Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI), 17, 973-978. DOI: 10.5555/1642194.1642224
[149] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. DOI: 10.48550/arXiv.1705.07874
[150] Lajevardi, Amir & Amini, Morteza. (2019). A semantic-based correlation approach for detecting hybrid and low-level APTs. Future Generation Computer Systems. 96. DOI: 10.1016/j.future.2019.01.056
[151] Mutalib, N. H. A., Sabri, A. Q. M., Wahab, A. W. A., et al. (2024). Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: a review. Artificial Intelligence Review, 57, 297. https://doi.org/10.1007/s10462-024-10890-4
[152] Khare, Ankur & Mallaiah, Shivamurthaiah & S, Harish. (2025). Enhancing Threat Detection and Response using Explainable AI (XAI): A Literature Review. International Journal of Research in Engineering and Science. 13. 174-180
[153] Almuqren, L.; Maashi, M.S.; Alamgeer, M.; Mohsen, H.; Hamza, M.A.; Abdelmageed, A.A. Explainable Artificial Intelligence Enabled Intrusion Detection Technique for Secure Cyber-Physical Systems. Appl. Sci. 2023, 13, 3081. https://doi.org/10.3390/app130530