VOL. 07 NO. 9,
FEBRUARY, 2025
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

A Review of Outlier Detection Techniques in Cybersecurity: A Machine Learning Perspective

Fatima Rilwan Ododo[1]; & Ridwan Rahmat Sadiq[2]

[1]Department of Computer Science, Montana State University Bozeman, MT 59717, USA. [2]Department of Computer Science, Nasarawa State University Keffi, Nigeria.
Corresponding Author: fatimaododo@montana.edu
DOI: https://doi.org/10.70382/ajsitr.v7i9.034

Abstract

Outlier detection has emerged as a critical component of modern cybersecurity systems, enabling the timely identification of anomalous behavior to prevent breaches, detect insider threats, and mitigate zero-day attacks. Traditional rule-based systems are proving inadequate for the increasing scale and complexity of cyber threats, prompting the integration of machine learning (ML) techniques to enhance detection accuracy and adaptability. This review paper synthesizes existing studies from 2000 to 2025, identified using keyword-based searches in Scopus, IEEE Xplore, and Google Scholar. The selection criteria focused on relevance, recency, and applications of ML-based outlier detection in cybersecurity. We categorize outlier detection methods into statistical, distance-based, density-based, clustering-based, and ML-driven approaches, and discuss their applications in intrusion detection, malware analysis, phishing detection, and Internet of Things (IoT) security. Additionally, the paper addresses commonly used datasets and evaluation metrics, challenges such as class imbalance and concept drift, and future research directions, including explainable AI and adversarial robustness. By synthesizing the current landscape and identifying research gaps, this review aims to guide the development of intelligent, scalable, and interpretable outlier detection systems for cybersecurity.

Keywords: Outlier Detection, Cybersecurity, Machine Learning, Intrusion Detection Systems, Malware Detection, Phishing Detection, Explainable AI, Deep Learning, Adversarial Robustness.

Introduction

In an era defined by ubiquitous connectivity and rapid digital transformation, cybersecurity has become an indispensable pillar of digital infrastructure (M¨oller, 2023). As organizations and individuals increasingly rely on interconnected systems, the volume, variety, and velocity of cyber threats have simultaneously surged (Jada and Mayayise, 2024; Kalonde et al., 2024a). Traditional defense mechanisms, particularly signature-based intrusion detection systems, are often limited in their ability to recognize new, evolving, or subtle attacks (Prabhakaran et al., 2025). This growing complexity has spurred interest in more intelligent and adaptive approaches—chief among them, machine learning-based outlier detection (Talaei Khoei and Kaabouch, 2023; Kalonde et al., 2024b).

The central problem addressed in this review is the lack of an integrated understanding of how different machine learning-driven outlier detection techniques perform across cybersecurity domains with real-world data constraints. Although many studies have explored individual techniques, few provide a holistic view of their applicability, limitations, and challenges in evolving cyber threat landscapes.

Outlier detection, often interchangeably referred to as anomaly detection, involves identifying patterns in data that deviate significantly from expected behavior (Ododo and Addotey, 2025b). These deviations may correspond to intrusions, fraudulent transactions, insider threats, or zero-day attacks (Al-Sadi, 2018). Unlike signature-based methods that require prior knowledge of attack patterns, outlier detection offers the advantage of identifying previously unseen threats, making it a powerful tool in the cybersecurity arsenal (Ododo and Addotey, 2025b).

In cybersecurity, outliers may manifest as unusual login times, irregular network traffic, unauthorized access to sensitive files, or abnormal user behavior (Ododo and Addotey, 2025b). Detecting these anomalies in real-time is crucial for maintaining the integrity, confidentiality, and availability of information systems (Diro et al., 2024). However, the nature of cyber data—high dimensionality, class imbalance, and the lack of labeled malicious samples—poses significant challenges to conventional detection systems (Diro et al., 2024).

To address these challenges, machine learning (ML) provides a flexible and data-driven framework for building robust detection systems (Ododo and Addotey, 2025a). MLbased outlier detection techniques can be broadly categorized into supervised, unsupervised, and semi-supervised learning models (Ododo and Addotey, 2025a). Supervised models rely on labeled datasets, which are often scarce in cybersecurity; unsupervised models, such as clustering and density-based approaches, identify outliers without prior labeling; while semi-supervised models learn the boundary of normal behavior and flag deviations (da Costa Brito, 2023).

Recent advances in deep learning, ensemble methods, and adaptive systems have further expanded the landscape of anomaly detection in cybersecurity (Ododo and Addotey, 2025a). Tools like autoencoders, isolation forests, and local outlier factors (LOF) have demonstrated efficacy in detecting complex and subtle threats in both network-based and host-based environments (Ododo and Addotey, 2025a).

This review paper provides a comprehensive examination of outlier detection techniques in cybersecurity through a machine learning lens. It categorizes detection methods, evaluates their applications across different threat domains, discusses the challenges associated with imbalanced and high-dimensional data, and explores promising research directions. By synthesizing the current body of knowledge, this work aims to guide researchers and practitioners in developing more effective, scalable, and intelligent cybersecurity solutions.

Research Methodology

This review is based on a narrative synthesis of peer-reviewed articles from 2000 to 2025, retrieved using keyword-based queries in Scopus, IEEE Xplore, and Google Scholar. The search terms included combinations of “outlier detection,” “anomaly detection,” “machine learning,” and “cybersecurity.” Studies were included based on their relevance to cybersecurity, the use of ML-driven outlier detection, and their publication in reputable venues.

Preference was given to recent publications and those that addressed evolving threats such as zero-day attacks, insider threats, and phishing. While a PRISMA diagram was not included, this methodology ensures a comprehensive and up-to-date overview of the field.

Literature Review

The field of outlier detection in cybersecurity has garnered substantial attention due to the evolving nature of cyber threats and the inadequacy of traditional security mechanisms. As cyberattacks become more sophisticated, anomaly detection—

particularly through machine learning (ML)—has emerged as a vital strategy for identifying unknown or zero-day attacks. This section reviews existing research that explores various outlier detection methodologies, their applications, and associated challenges within cybersecurity contexts.

Bou Nassif et al. (2021) provided a comprehensive systematic literature review (SLR) of ML models for anomaly detection, analyzing 290 studies from 2000–2020. Their review categorized anomalies into point, contextual, and collective, and highlighted that unsupervised methods have been adopted more widely due to the scarcity of labeled datasets in real-world cybersecurity scenarios (Nassif et al., 2021). The study emphasized the increasing use of models such as autoencoders, One-Class SVMs, and Isolation Forests, which can detect abnormal behaviors without requiring labeled attack data (Nassif et al., 2021).

Devineni et al. (2023) reinforced the importance of anomaly detection in the protection of data integrity and network infrastructure, highlighting how machine learning enables realtime and adaptive threat identification. They outlined a taxonomy of ML-based detection methods, including supervised, unsupervised, and semi-supervised learning, and discussed their application across domains like fraud detection, IT systems monitoring, and network security (Devineni et al., 2023).

Chan et al. (2003) offered early but foundational insights into anomaly detection, distinguishing it from signature-based detection. Their study explored rule learning and clustering algorithms for detecting outliers, particularly focusing on constructing behavioral models in the absence of attack-labeled data (Chan et al., 2003). This work underscored the importance of customizing normalcy models to specific environments, given the contextual nature of what constitutes “normal” behavior in different systems (Chan et al., 2003).

Buczak and Guven (2015) surveyed data mining and machine learning techniques for intrusion detection, drawing attention to the critical need for hybrid models that combine anomaly-based and misuse-based detection to balance detection accuracy and false positive rates. Their review acknowledged the inherent limitations of standalone anomaly detection techniques, especially under imbalanced class distributions common in intrusion detection systems (IDS) (Buczak and Guven, 2015).

Al-Shehari et al. (2024) contributed significantly to this area by evaluating the DensityBased Local Outlier Factor (DBLOF) algorithm in detecting insider threat. Their study focused on the CERT r4.2 dataset, a highly imbalanced dataset reflective of real-world cybersecurity environments, and demonstrated that DBLOF achieved a notable F1-score of 98%, validating the effectiveness of density-based techniques in skewed (Al-Shehari et al., 2024).

Jiang et al. (2020) explored outlier detection in the Internet of Things (IoT), where resource constraints and heterogeneous data present additional challenges. Their review categorized ML techniques into clustering-based, classification-based, and hybrid approaches, and highlighted the need for lightweight and energy-efficient algorithms tailored for IoT environments (Jiang et al., 2020).

Lastly, Sarker et al. (2020) discussed the broader integration of data science and machine learning in cybersecurity, proposing a multi-layered framework for intelligent security modeling. Their work argued for moving beyond static rule-based defenses toward adaptive, data-driven detection systems that can learn and evolve alongside emerging threats (Sarker et al., 2020).

In summary, the literature reflects a growing consensus that ML-based outlier detection offers essential advantages in detecting unknown threats across varied cybersecurity domains. However, challenges such as class imbalance, high false alarm rates, lack of labeled data, and model interpretability persist. These gaps continue to drive the need for novel methodologies, improved datasets, and the integration of explainability into detection frameworks.

Taxonomy of Outlier Detection Techniques

Outlier detection techniques in cybersecurity can be broadly categorized into statistical, distance-based, density-based, clustering-based, and machine learning-driven approaches. Each category offers distinct methods for identifying deviations in behavior, depending on assumptions about data distribution, structure, or label availability. This section presents a detailed taxonomy of these approaches, with a focus on their relevance, mechanisms, and applicability in cybersecurity contexts.

Statistical Methods

Statistical methods operate under the assumption that normal data conforms to a known probability distribution (Garba et al., 2019). Data points that significantly deviate from this distribution are flagged as outliers (Ododo and Addotey, 2025b). These methods are among the earliest developed and are often used when the underlying distribution of data can be reasonably estimated (Davis and McCuen, 2005).

- Parametric methods (e.g., Z-score, Gaussian models) assume a specific distribution and calculate how far each point is from the mean, typically using standard deviations (Thatcher et al., 2005).
- Non-parametric methods (e.g., histogram-based, kernel density estimators) do not assume any predefined distribution (Thatcher et al., 2005).

These techniques are simple and interpretable but often fall short in high-dimensional or non-linear domains common in cybersecurity applications

Distance-Based Methods

Distance-based methods determine outliers by calculating the spatial separation between data points (Knorr et al., 2000). A data point is considered an outlier if its distance from most other points exceeds a defined threshold (Knorr et al., 2000).

- k-Nearest Neighbors (k-NN): Calculates the distance of a point from its k nearest neighbors (Banu and Praveen, 2016). If this average distance is high, the point is labeled as an outlier (Banu and Praveen, 2016).
- Mahalanobis Distance: A multivariate technique that accounts for correlations in the dataset to measure distance more effectively (Leys et al., 2018).

Distance-based methods are sensitive to the choice of distance metric and do not perform well when data is sparse or noisy, which is typical in cybersecurity network traffic data.

Density-Based Methods

Density-based methods identify outliers as data points located in regions of low density compared to their neighbors.

- Local Outlier Factor (LOF): Measures the local deviation of density around a data point relative to its neighbors. Points in regions with significantly lower density than their neighbors are considered outliers (Alghushairy et al., 2020).
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Detects outliers as noise points that do not belong to any cluster due to insufficient density around them (Sharma et al., 2016).

These methods are highly effective in identifying local outliers and are especially suited to detecting insider threats, where anomalous behavior is subtle but contextually significant.

Clustering-Based Methods

Clustering-based approaches first group the data into clusters and then identify outliers as points that do not fit well into any cluster.

- k-Means Clustering: After clustering the data, points that lie far from their cluster centroids may be considered anomalies (Mu¨nz et al., 2007).
- Hierarchical Clustering: Constructs a hierarchy of clusters and isolates data points that do not strongly belong to any group (Farrelly et al., 2017).

Clustering techniques are intuitive but often struggle with high-dimensional data and require careful tuning of the number of clusters, which is often non-trivial in cybersecurity domains.

Machine Learning Approaches

Machine learning techniques have emerged as powerful tools for outlier detection due to their ability to model complex patterns and generalize from data. These methods can be categorized as supervised, unsupervised, or semi-supervised.

- Supervised Learning: Requires labeled examples of both normal and anomalous behavior. Techniques like Random Forests, Support Vector Machines (SVM), and Deep Neural Networks fall under this category. However, labeled anomalies are often rare or unavailable in practice (Ododo and Addotey, 2025a; Chalapathy and Chawla, 2019).
- Unsupervised Learning: Identifies anomalies without labels by assuming that normal instances are far more frequent. Popular models include Autoencoders, Isolation Forests, and k-Means. Autoencoders, for instance, are trained to reconstruct input data, and high reconstruction errors signal potential anomalies (Ododo and Addotey, 2025a; Chalapathy and Chawla, 2019).
- Semi-Supervised Learning: Uses labeled normal data to define a boundary (e.g., OneClass SVM). Any deviation from this boundary is flagged as an anomaly. This is effective when normal behavior is well understood but labeled attacks are scarce (Ododo and Addotey, 2025a; Chalapathy and Chawla, 2019).

Ensemble Methods

Ensemble approaches combine multiple models or detection techniques to improve accuracy, robustness, and generalization.

- Voting-based ensembles: Combine outputs of various models to reach a consensus on whether a data point is anomalous (Hisham et al., 2022).
- Model fusion: Merges multiple methods, such as combining supervised and unsupervised models, to compensate for individual weaknesses (Hisham et al., 2022).

Ensemble methods are particularly useful in cybersecurity, where attack patterns are diverse, evolving, and context-dependent.

Applications in Cybersecurity

Outlier detection techniques have become central to the development of intelligent, automated cybersecurity systems. They enable the identification of abnormal

behaviors or patterns that often indicate cyberattacks, policy violations, or insider misuse. This section explores how various outlier detection techniques are applied in key cybersecurity domains.

Intrusion Detection Systems (IDS)

Intrusion Detection Systems are one of the most critical applications of outlier detection in cybersecurity (Jabez and Muthukumar, 2015). IDS monitor and analyze system or network behavior to identify potential security breaches, including unauthorized access and misuse (Jabez and Muthukumar, 2015).

- Anomaly-based IDS use statistical or ML-based models to learn patterns of normal network activity and flag deviations as potential intrusions (Jabez and Muthukumar, 2015).
- For example, autoencoders and One-Class SVMs are frequently used to model baseline behavior in host-based and network-based IDS.
- Datasets: Common datasets used include NSL-KDD, UNSW-NB15, and CICIDS2017, which contain both normal and malicious activity (Choudhary and Kesswani, 2020).

The advantage of anomaly-based IDS is their potential to detect zero-day attacks, but they often suffer from high false alarm rates due to the dynamic nature of normal behavior.

Table 1: Summary of Outlier Detection Techniques in Cybersecurity

Technique	Description	Advantages	Limitations
Statistical Methods	Use probability distributions to define normal behavior (e.g., Zscore, Grubbs' Test)	Simple, interpretable, good for low-dimensional data	Poor performance in high-dimensional or non-Gaussian data
Distance-Based Methods	Measure distance to other points; outliers are far from others (e.g., k-NN, Maha-lanobis Distance)	No assumption on data distribution, intuitive	Sensitive to distance metric, less effective in sparse or highdimensional data
Density-Based Methods	Detects low-density regions (e.g., LOF, DB-SCAN)	Good at finding local outliers, handles noise well	High computational cost, parameter sen-sitivity
Clustering-Based Methods	Outliers are points that do not fit into clusters (e.g., K-means, hierarchical clustering)	Unsupervised, can reveal structure in data	Requires setting number of clusters, poor with complex cluster shapes
Supervised Learning	Trained on labeled normal and attack data (e.g., Random Forest, SVM)	High accuracy with sufficient labeled data	Requires labeled attack data, not effective for zero-day threats
Unsupervised Learning	Learns patterns without labels (e.g., Autoencoders, Isolation Forest)	Suitable for unlabeled data, detects novel attacks	Higher false positives, sensitive to noise
SemiSupervised Learning	Trained on only normal data (e.g., One-Class SVM)	Effective when attack labels are unavailable	Relies heavily on accurate normal data
Ensemble Methods	Combines multiple models (e.g., voting, fusion)	Improved robustness, reduces variance	Increased complexity and computational cost

Malware Detection

Outlier detection also supports malware identification by discovering malicious behaviors that deviate from the norms observed in benign software.

- Supervised learning methods, such as Random Forests or deep neural networks, are used when labeled data is available (Kong and Yu, 2018).
- Unsupervised methods, including Isolation Forests and LOF, are employed to detect unknown or polymorphic malware, which may not match known signatures (Kong and Yu, 2018).

These techniques are especially useful in dynamic malware analysis, where behavioral logs or system call traces are used to distinguish malicious software from benign applications.

Insider Threat Detection

Insider threats are one of the most challenging issues in cybersecurity, as the malicious activity originates from legitimate users with authorized access. Such behavior can be subtle and often mimics normal user activity.

- Density-based techniques like LOF have shown promise in detecting insider threats by identifying anomalies in behavioral patterns within imbalanced datasets (Al-Shehari et al., 2024).
- Insider detection systems rely on modeling user behavior over time, using features such as login time, access frequency, and file interaction history (Al-Shehari et al., 2024).

Due to the imbalance of benign vs. malicious user instances, semi-supervised and unsupervised approaches are typically preferred.

Phishing and Social Engineering Detection

Phishing attacks exploit human behavior and social engineering tactics to steal sensitive information (Al-Otaibi and Alsuwat, 2020). These attacks often manifest as abnormal communication patterns via email or messaging platforms.

- Natural Language Processing (NLP) combined with anomaly detection can identify emails with unusual syntax, suspicious links, or uncommon senders (Sharma and Arjunan, 2023).
- Outlier detection models are trained to recognize deviations in language structure or metadata (e.g., frequency of links, sender domain reputation) (Ododo and Addotey, 2025b).

While effective, such systems require frequent retraining due to the rapid evolution of phishing tactics.

Internet of Things (IoT) Security

The Internet of Things introduces a highly heterogeneous environment with limited computational resources and varied data modalities, making outlier detection more complex. Lightweight models, such as clustering-based or tree-based methods (e.g., k-Means, Isolation Forests), are often used for outlier detection in resource-constrained devices (Samara et al., 2022).

These methods help identify faulty sensors, malicious firmware behavior, or network anomalies in IoT ecosystems (Samara et al., 2022).
The open nature of IoT networks makes them vulnerable to both internal and external anomalies, increasing the demand for distributed, edge-deployable detection solutions.

Datasets and Evaluation Metrics

The effectiveness of outlier detection techniques in cybersecurity depends significantly on the quality of datasets and the robustness of the evaluation metrics used to assess model performance. This section outlines widely used datasets and standard performance measures employed in anomaly detection research.

Commonly Used Datasets

- **NSL-KDD:** An improved version of the KDD'99 dataset, addressing issues like redundancy and imbalance. It includes 41 features with 5 classes (normal and 4 types of attacks), making it suitable for anomaly-based intrusion detection systems (Bala and Nagpal, 2019).
- **CICIDS2017:** A realistic dataset that simulates up-to-date benign and malicious network traffic. It includes detailed flow features and attack types such as DDoS, brute force, and infiltration (Sharafaldin et al., 2018).
- **UNSW-NB15:** Generated with IXIA PerfectStorm, this dataset comprises 49 features across various modern attack types, such as exploits, fuzzers, and backdoors, capturing a broader attack surface than older datasets (Moustafa and Slay, 2015).
- **CERT Insider Threat Dataset (r4.2):** Designed for insider threat research, this dataset provides comprehensive user activity logs over months, including both benign and malicious behavior in enterprise settings (Micklitz et al., 2013).
- **ADFA-LD:** Created by the Australian Defence Force Academy, this dataset includes system call traces for normal and attack scenarios on Linux systems, emphasizing host-based intrusion detection (Creech and Hu, 2013).

Evaluation Metrics

Due to class imbalance and the critical nature of detecting rare events, standard accuracy is insufficient. Below are metrics commonly used for performance evaluation in outlier detection:

- **Accuracy:** Measures overall correctness of the model (Saito and Rehmsmeier, 2015): Accuracy = $\frac{TP + TN}{TP + TN + FP + FN}$
- **Precision:** Indicates the proportion of true positives among all predicted positives (Davis and Goadrich, 2006):
  Precision = $\frac{TP}{TP + FP}$
- **Recall (Sensitivity):** Reflects the ability to identify actual anomalies (Davis and Goadrich, 2006):
  Recall = $\frac{TP}{TP + FN}$
- **F1-Score:** Harmonic mean of precision and recall, useful for imbalanced datasets (Powers, 2020):
  F1-Score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
- **ROC-AUC:** Receiver Operating Characteristic curve and its area under the curve. AUC close to 1.0 indicates excellent model performance (Fawcett, 2006).
- **Confusion Matrix:** Summarizes true/false positives and negatives, providing a complete view of classification results (Kohavi, 1998).
- **Matthews Correlation Coefficient (MCC):** Balanced metric suitable for imbalanced data (Kohavi, 1998)
  MCC = $\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$

Selecting appropriate datasets and evaluation metrics is essential for fair and comprehensive assessment of outlier detection techniques, especially in cybersecurity domains where attacks are rare but critical.

Challenges and Limitations

Despite the effectiveness of machine learning-based outlier detection in cybersecurity, several challenges and limitations hinder their practical deployment and performance. These challenges span data quality, algorithm scalability, real-time processing requirements, and interpretability.

High Dimensionality of Security Data

Cybersecurity datasets, particularly those from network traffic and system logs, are often high-dimensional (Chen et al., 2022). This “curse of dimensionality” degrades

the performance of distance-based and clustering-based methods and increases computational complexity (Chen et al., 2022). Dimensionality reduction techniques such as Principal Component Analysis (PCA) or feature selection methods are often employed, but improper use may result in the loss of valuable security-related information (Chen et al., 2022).

Class Imbalance

Cybersecurity datasets typically exhibit severe class imbalance, where normal activities vastly outnumber anomalous or malicious events. This imbalance causes learning algorithms to bias toward the majority class, increasing false negatives (Chawla et al., 2002). Although oversampling techniques such as SMOTE or cost-sensitive learning approaches are used to mitigate this issue, they may introduce synthetic noise or overfitting.

Lack of Labeled and Realistic Data

The availability of labeled and realistic cybersecurity datasets is limited due to privacy, legal, and ethical concerns (Ring et al., 2019). Many public datasets are outdated or synthetically generated and may not reflect modern attack vectors (Ring et al., 2019). This restricts the ability to generalize models to real-world scenarios (Ring et al., 2019). The absence of unified benchmarks also hampers cross-comparison between different methods (Ring et al., 2019).

Concept Drift and Evolving Threats

Cyber threats are dynamic in nature. As attackers continuously evolve their strategies, models trained on historical data may become obsolete—a phenomenon known as concept drift (Gama et al., 2014). Addressing this challenge requires the integration of online learning or adaptive modeling techniques that can update themselves as new data arrives.

Interpretability of Detection Models

Black-box models, particularly deep learning-based techniques, often lack interpretability (Ribeiro et al., 2016). In cybersecurity, understanding why a certain event is flagged as anomalous is critical for response and audit purposes. Post-hoc explanation tools such as SHAP or LIME are emerging, but building inherently interpretable models remains an open research challenge.

Real-Time Detection and Scalability

Real-time threat detection is essential in operational environments (Sommer and Paxson, 2010). However, many sophisticated anomaly detection models are computationally intensive and unsuitable for deployment in real-time or resource-constrained settings (e.g., IoT networks). This necessitates trade-offs between detection accuracy and system latency.

In summary, addressing these challenges requires continued research in areas such as efficient model design, synthetic data generation, explainable AI, and adaptive learning. Overcoming these limitations is essential for the widespread adoption of machine learning-based outlier detection in cybersecurity.

Future Directions

As cyber threats continue to evolve in scale, complexity, and frequency, there is a growing need for more robust, adaptive, and interpretable outlier detection systems in cybersecurity.

This section outlines several promising research directions and technological advancements that can enhance anomaly detection capabilities in the near future.

Integration with Deep Learning Techniques

Deep learning offers a powerful framework for modeling complex, non-linear data distributions commonly found in cybersecurity. Future research may explore advanced architectures such as Convolutional Neural Networks (CNNs) for spatial data, Recurrent Neural Networks (RNNs) for sequential system logs, and Transformer models for context-aware detection. Deep generative models like Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) also present opportunities for detecting sophisticated attack patterns and synthesizing high-quality training data.

Explainable AI (XAI)

Interpretability is crucial in high-stakes cybersecurity environments where understanding why a model flagged an event as an anomaly is as important as the detection itself. Future work should prioritize the integration of explainable AI frameworks, such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and attention mechanisms. This will aid security analysts in validating alerts, understanding model decisions, and ensuring accountability.

Federated and Privacy-Preserving Learning

Cybersecurity data is often distributed across multiple sources, such as endpoints, servers, and organizations. Federated Learning (FL) enables collaborative model training without centralizing data, thus preserving privacy. Future research can explore the application of FL for outlier detection in scenarios like distributed intrusion detection systems, where privacy concerns and data heterogeneity are paramount.

Adversarial Robustness

Adversarial machine learning poses a threat to anomaly detection models by generating carefully crafted inputs that evade detection. Enhancing the robustness of detection systems against adversarial examples will be vital. This includes incorporating adversarial training, robust optimization techniques, and regularization strategies to build resilient models.

Continual and Online Learning

Given the dynamic nature of cyber threats, static models rapidly become outdated. There is a need for continual learning techniques that allow models to adapt to evolving data distributions without catastrophic forgetting. Online learning algorithms that update incrementally as new data arrives will also be crucial for real-time applications.

Unified Evaluation Frameworks and Benchmarking

To facilitate reproducibility and fair comparison of anomaly detection models, there is a pressing need for standardized evaluation frameworks, benchmark datasets, and agreedupon metrics. Future initiatives should focus on creating shared repositories and collaborative testbeds that reflect real-world attack scenarios.

In summary, addressing current limitations and embracing these future directions will be key to building next-generation cybersecurity systems. Research efforts should emphasize robustness, scalability, interpretability, and collaboration to stay ahead of emerging and adaptive threats.

Conclusion

Outlier detection plays a pivotal role in strengthening cybersecurity defenses by enabling the identification of unknown, rare, and sophisticated threats that may evade traditional signature-based systems. This paper has presented a comprehensive review of outlier detection techniques from a machine learning perspective, with a focus on

their application in cybersecurity contexts such as intrusion detection, malware analysis, insider threat monitoring, phishing detection, and IoT security.

We explored various categories of outlier detection methods, including statistical, distancebased, density-based, clustering-based, and machine learning-driven approaches. Each method offers unique strengths and trade-offs depending on the nature of the data, the availability of labels, and the operational constraints. Furthermore, we examined widely used datasets and evaluation metrics that facilitate the training and benchmarking of anomaly detection models.

Despite significant advancements, several challenges persist, including data imbalance, concept drift, interpretability issues, and the lack of high-quality labeled datasets. These limitations underscore the need for continued research in areas such as explainable AI, online learning, adversarial robustness, federated learning, and deep learning integration.

In conclusion, machine learning-based outlier detection represents a promising and evolving frontier in cybersecurity. By addressing current limitations and embracing emerging methodologies, researchers and practitioners can develop intelligent, scalable, and trustworthy systems capable of defending against an ever-changing threat landscape.

References

Abebe Diro, Shahriar Kaisar, Athanasios V Vasilakos, Adnan Anwar, Araz Nasirian, and Gaddisa Olani. Anomaly detection for space information networks: A survey of challenges, techniques, and future directions. *Computers & Security*, 139:103705, 2024.

Abeer F Al-Otaibi and Emad S Alsuwat. A study on social engineering attacks: Phishing attack. *Int. J. Recent Adv. Multidiscip. Res*, 7(11):6374–6380, 2020.

Abhinesh Prabhakaran, Rhoda Gasue, Abdul-Majeed Mahamadu, Colin A Booth, Patrick Manu, Kofi Agyekum, and Promise D Nukah. Applications of immersive technology in architecture, engineering and construction. *Applications of Immersive Technology in Architecture, Engineering and Construction: A Handbook*, 2025.

Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, and Fatima Mohamad Dakalbab. Machine learning for anomaly detection: A systematic review. *Ieee Access*, 9:78658–78700, 2021.

Aliyu Garba, Sandip Rakshit, and Fatima Rilwan. Detection and sentiment analysis of hate speech on twitter in nigerian politics. In *Proceedings of: 2nd International Conference of the IEEE Nigeria*, page 285, 2019.

Allen P Davis and Richard H McCuen. Statistical methods for data analysis. *Stormwater Management for Smart Growth*, pages 37–62, 2005.

Anna L Buczak and Erhan Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*, 18(2): 1153–1176, 2015.

Arvind Sharma, RK Gupta, and Akhilesh Tiwari. Improved density based spatial clustering of applications of noise clustering algorithm for knowledge discovery in spatial data. *Mathematical Problems in Engineering*, 2016(1):1564516, 2016.

Azzat Ahmed Ali Al-Sadi. *Towards an Effective Approach of Insider Attacks Detection Using the Human Physiological Signals*. PhD thesis, King Fahd University of Petroleum and Minerals (Saudi Arabia), 2018.

Christophe Leys, Olivier Klein, Yves Dominicy, and Christophe Ley. Detecting multivariate outliers: Use a robust variant of the mahalanobis distance. *Journal of experimental social psychology*, 74:150–156, 2018.

Colleen M Farrelly, Seth J Schwartz, Anna Lisa Amodeo, Daniel J Feaster, Douglas L Steinley, Alan Meca, and Simona Picariello. The analysis of bridging constructs with hierarchical clustering methods: An application to identity. *Journal of Research in Personality*, 70:93–106, 2017.

David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020.

Dietmar PF M¨oller. Cybersecurity in digital transformation. In *Guide to cybersecurity in digital transformation: Trends, methods, technologies, applications and best practices*, pages 1–70. Springer, 2023.

Edwin M Knorr, Raymond T Ng, and Vladimir Tucakov. Distance-based outliers: algorithms and applications. *The VLDB Journal*, 8(3):237–253, 2000.

FATIMA ODODO and NICHOLAS ADDOTEY. Advancements and challenges in deep learning for cyber threat detection. *International Journal of Science Research and Technology*, 2025a.

FATIMA ODODO and NICHOLAS ADDOTEY. Understanding the influence of outliers on machine learning model interpretability. *International Journal of African Sustainable Development Research*, 2025b.

Francisco da Costa Brito. Anomaly detection in cybersecurity through semi-supervision. Master's thesis, Universidade do Porto (Portugal), 2023.

Gerhard Mu¨nz, Sa Li, and Georg Carle. Traffic anomaly detection using k-means clustering. In *Gi/itg workshop mmbnet*, volume 7, 2007.

Gideon Creech and Jiankun Hu. Generation of a new ids test dataset: Time to retire the kdd collection. In *2013 IEEE wireless communications and networking conference (WCNC)*, pages 4487–4492. IEEE, 2013.

Gilbert Kalonde, Samuel Boateng, Lateefat Sanni, Silas Chotwe, and Fatima Ododo. Artificial intelligence and special education: The use and the integration. In *Society for Information Technology & Teacher Education International Conference*, pages 1926–1932. Association for the Advancement of Computing in Education (AACE), 2024b.

Gilbert Kalonde, Samuel Boateng, Silas Chotwe, Lateefat Sanni, and Fatima Ododo. Integration of computer science in rural math classroom: A case study. In *Society for Information Technology & Teacher Education International Conference*, pages 1757–1761. Association for the Advancement of Computing in Education (AACE), 2024a.

Honggang Chen, Xiaohai He, Linbo Qing, Yuanyuan Wu, Chao Ren, Ray E Sheriff, and Ce Zhu. Real-world single image super-resolution: A brief review. *Information Fusion*, 79:124–145, 2022.

Iman Sharafaldin, Arash Habibi Lashkari, Ali A Ghorbani, et al. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, 1(2018): 108–116, 2018.

Iqbal H Sarker, ASM Kayes, Shahriar Badsha, Hamed Alqahtani, Paul Watters, and Alex Ng. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, 7:1–29, 2020.

Irshaad Jada and Thembekile O Mayayise. The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, 8(2):100063, 2024.

Ja Jabez and BJPCS Muthukumar. Intrusion detection system (ids): Anomaly detection using outlier detection approach. *Procedia Computer Science*, 48:338–346, 2015.

Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning*, pages 233–240, 2006.

Jinfang Jiang, Guangjie Han, Lei Shu, Mohsen Guizani, et al. Outlier detection approaches based on machine learning in the internet-of-things. *IEEE Wireless Communications*, 27 (3):53–59, 2020.

Jo˜ao Gama, Indre˙ Zliobaite˙,˘ Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46 (4):1–37, 2014.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes, and Andreas Hotho. A survey of network-based intrusion detection data sets. *Computers & security*, 86:147–167, 2019.

Mustafa Al Samara, Ismail Bennis, Abdelhafid Abouaissa, and Pascal Lorenz. A survey of outlier detection techniques in iot: Review and classification. *Journal of Sensor and Actuator Networks*, 11(1):4, 2022.

Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357, 2002.

Nour Moustafa and Jill Slay. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In *2015 military communications and information systems conference (MilCIS)*, pages 1–6. IEEE, 2015.

Omar Alghushairy, Raed Alsini, Terence Soule, and Xiaogang Ma. A review of local outlier factor algorithms for outlier detection in big data streams. *Big Data and Cognitive Computing*, 5(1):1, 2020.

Philip K Chan, Matthew V Mahoney, and Muhammad H Arshad. A machine learning approach to anomaly detection. 2003.

Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

Ritu Bala and Ritu Nagpal. A review on kdd cup99 and nsl nsl-kdd dataset. *International Journal of Advanced Research in Computer Science*, 10(2), 2019.

Robert W Thatcher, D North, and C Biver. Parametric vs. non-parametric statistics of low resolution electromagnetic tomography (loreta). *Clinical EEG and neuroscience*, 36(1): 1–8, 2005.

Robin Sommer and Vern Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE symposium on security and privacy*, pages 305–316. IEEE, 2010.

Ron Kohavi. Glossary of terms. *Machine learning*, 30:271–274, 1998.

Sabri Hisham, Mokhairi Makhtar, and Azwa Abdul Aziz. Combining multiple classifiers using ensemble method for anomaly detection in blockchain networks: A comprehensive review. *International Journal of Advanced Computer Science and Applications*, 13(8), 2022.

Sarika Choudhary and Nishtha Kesswani. Analysis of kdd-cup'99, nsl-kdd and unsw-nb15 datasets using deep learning in iot. *Procedia Computer Science*, 167:1561–1573, 2020.

Siva Karthik Devineni, Satish Kathiriya, and Abhishek Shende. Machine learning-powered anomaly detection: Enhancing data security and integrity. *Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-198. DOI: doi. org/10.47363/JAICC/2023 (2)*, 184:2–9, 2023.

Stephan Micklitz, Martin Ortlieb, and Jessica Staddon. " i hereby leave my email to...": Data usage control and the digital estate. In *2013 IEEE Security and Privacy Workshops*, pages 42–44. IEEE, 2013.

Suresh Sharma and Tamilselvan Arjunan. Natural language processing for detecting anomalies and intrusions in unstructured cybersecurity data. *International Journal of Information and Cybersecurity*, 7(12):1–24, 2023.

Syeda Khaja Momina Banu and P Praveen. A novel approach for k-nn on unsupervised distance-based outlier detection. *International Journal For Technological Research In Engineering*, 4(3):505–508, 2016.

Taher Ali Al-Shehari, Domenico Rosaci, Muna Al-Razgan, Taha Alfakih, Mohammed Kadrie, Hammad Afzal, and Raheel Nawaz. Enhancing insider threat detection in imbalanced cybersecurity settings using the density-based local outlier factor algorithm. *IEEE Access*, 12:34820–34834, 2024.

Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3): e0118432, 2015.

Tala Talaei Khoei and Naima Kaabouch. A comparative analysis of supervised and unsupervised models for detecting attacks on the intrusion detection systems. *Information*, 14(2):103, 2023.

Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

Yunchuan Kong and Tianwei Yu. A deep neural network model using random forest to extract feature representation for gene expression data classification. *Scientific reports*, 8 (1):16477, 2018.