International Journal of Research and Technology (IJRT)

International Open-Access, Peer-Reviewed, Refereed, Online Journal

ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529

| An ISO 9001:2015 Certified Journal |

A Systematic Review of Machine Learning Frameworks for Network Outlier Detection

Anjali

Research Scholar, Department of Computer Science and Engineering, A.N.A College of Engineering & Management, Bareilly

Dr. Vineet Agarwal

Professor, Department of Computer Science and Engineering, A.N.A College of Engineering & Management, Bareilly

ABSTRACT

The rapid expansion of networked systems and data-driven infrastructures has intensified the need for robust mechanisms to detect anomalous activities that may compromise security and performance. Network outlier detection, a critical component of intrusion detection systems, focuses on identifying patterns in network traffic that deviate from normal behavior. This systematic review examines contemporary machine learning frameworks employed for network outlier detection, encompassing supervised, unsupervised, semi-supervised, and deep learning approaches. The study analyzes the architectural components of these frameworks, including data preprocessing, feature engineering, model training, and evaluation techniques. It further compares their performance based on accuracy, scalability, and adaptability in dynamic environments such as IoT, cloud computing, and software-defined networks. Key challenges, including data imbalance, concept drift, and model interpretability, are also discussed. The review identifies significant research gaps and highlights emerging trends, providing a comprehensive foundation for future advancements in intelligent network anomaly detection systems.

Keywords: Network Outlier Detection, Machine Learning, Anomaly Detection, Intrusion Detection Systems, Deep Learning

1. INTRODUCTION

In contemporary digital ecosystems, the exponential growth of networked systems and data-intensive applications has significantly increased the complexity and vulnerability of communication infrastructures, making network security a critical area of research within cybersecurity and data science. Network outlier detection, often referred to as anomaly detection, plays a pivotal role in identifying unusual patterns or behaviors in network traffic that deviate from normal operational profiles and may indicate malicious activities such as intrusions, distributed denial-of-service (DDoS) attacks, data exfiltration, or insider threats. Traditional rule-based and signature-driven detection mechanisms, while effective against known threats, are increasingly inadequate in addressing the dynamic and evolving nature of modern cyberattacks, particularly zero-day exploits and sophisticated polymorphic threats. In this context, machine learning (ML) frameworks have emerged as powerful and adaptive solutions capable of automatically learning complex patterns from large-scale, high-dimensional network data without explicit programming. These frameworks leverage diverse

paradigms, including supervised, unsupervised, semi-supervised, and deep learning techniques, to enhance detection accuracy, scalability, and real-time responsiveness. Furthermore, the integration of advanced models such as autoencoders, clustering algorithms, and ensemble methods enables the identification of subtle and previously unseen anomalies in heterogeneous network environments, including cloud computing, Internet of Things (IoT), and software-defined networks. Despite these advancements, the design and deployment of ML-based outlier detection frameworks face several challenges, including data imbalance, concept drift, high false positive rates, lack of interpretability, and limited availability of high-quality labeled datasets. Consequently, a systematic review of existing machine learning frameworks is essential to synthesize current knowledge, evaluate methodological trends, and identify research gaps within this domain. This study aims to provide a comprehensive and structured analysis of machine learning-based network outlier detection frameworks, examining their architectures, techniques, performance metrics, and practical applications, while also highlighting emerging directions for future research and development.

2. SCOPE OF THE STUDY

This study focuses on providing a comprehensive and systematic review of machine learning frameworks developed for network outlier detection within modern cybersecurity environments. The scope encompasses a detailed examination of various machine learning paradigms, including supervised, unsupervised, semi-supervised, and deep learning approaches, applied to detect anomalous patterns in network traffic data. It covers key components of these frameworks such as data collection, preprocessing, feature engineering, model training, and evaluation metrics. The review includes analysis of widely used benchmark datasets and experimental setups to ensure comparability across studies. The study considers applications across diverse domains such as cloud computing, Internet of Things (IoT), and enterprise networks. However, it is limited to scholarly works published in recent years and primarily focuses on algorithmic and architectural aspects, excluding hardware-level implementations and non-machine learning-based detection techniques, thereby maintaining a clear emphasis on intelligent data-driven methodologies.

3. BACKGROUND OF NETWORK SECURITY AND ANOMALY DETECTION

The rapid digitization of communication systems and the proliferation of interconnected devices have significantly transformed the landscape of network security, making it a critical domain within modern information technology. As organizations increasingly rely on complex network infrastructures to support cloud computing, Internet of Things (IoT), and distributed applications, the risk of cyber threats such as unauthorized access, malware propagation, and distributed denial-of-service (DDoS) attacks has grown substantially. Network security, therefore, encompasses a set of policies, practices, and technologies designed to protect data integrity, confidentiality, and availability across these systems. Within this context, anomaly detection has emerged as a fundamental technique for identifying irregular patterns in network traffic that deviate from established normal behavior. Unlike traditional signature-based methods, which depend on predefined attack patterns, anomaly detection approaches are capable of identifying previously unseen or zero-day attacks by modeling normal network

behavior and flagging deviations. This capability is particularly important in dynamic and high-dimensional network environments where attack vectors continuously evolve. Machine learning techniques have further enhanced anomaly detection by enabling automated pattern recognition, adaptive learning, and real-time analysis of large-scale network data. These approaches utilize statistical, clustering, and neural network models to detect subtle and complex anomalies that may not be easily identifiable through manual or rule-based systems.

4. DEFINITION OF OUTLIERS AND NETWORK ANOMALIES

Outliers, in the context of data analysis and network security, refer to data instances or observations that significantly deviate from the established normal patterns within a dataset. These deviations may arise due to errors, rare events, or meaningful but unusual behaviors, and their identification is crucial for ensuring data integrity and system reliability. In network environments, such outliers are commonly termed as network anomalies, representing irregular or suspicious activities in network traffic that differ from expected operational behavior. Network anomalies can manifest in various forms, including sudden spikes in traffic volume, unusual packet structures, unexpected communication between hosts, or deviations in protocol usage patterns. These anomalies are broadly categorized into point anomalies, where a single data instance is abnormal; contextual anomalies, where behavior is abnormal in a specific context such as time or location; and collective anomalies, where a group of related data points collectively indicate unusual activity. The detection of such anomalies is essential for identifying potential security threats, including intrusion attempts, malware infections, insider threats, and distributed denial-of-service (DDoS) attacks. Unlike traditional approaches that rely on predefined signatures, anomaly-based detection focuses on modeling normal behavior and identifying deviations, making it more effective in detecting novel and unknown threats. Machine learning techniques have significantly advanced this domain by enabling automated and adaptive identification of outliers in high-dimensional and dynamic network datasets, thereby enhancing the efficiency, accuracy, and scalability of modern network security systems.

5. IMPORTANCE OF OUTLIER DETECTION IN CYBERSECURITY

Outlier detection is a cornerstone capability in modern cybersecurity because it enables the identification of deviations from baseline network behavior that often correspond to malicious or policy-violating activities. Unlike signature-based defenses that depend on known attack patterns, outlier (anomaly) detection models establish profiles of normal traffic and flag statistically significant deviations, thereby providing resilience against zero-day exploits, polymorphic malware, and advanced persistent threats (APTs). This is particularly critical in high-velocity, high-dimensional environments such as cloud-native architectures, Internet of Things (IoT) ecosystems, and software-defined networks, where attack surfaces are broad and continuously evolving. By surfacing rare events—such as unusual lateral movement, privilege escalation sequences, data exfiltration patterns, or atypical protocol usage—outlier detection strengthens intrusion detection and prevention systems (IDS/IPS) and supports early-stage threat discovery before damage escalates.

From an operational perspective, effective outlier detection reduces mean time to detect (MTTD) and mean time to respond (MTTR), enabling faster incident triage and containment. It also improves risk management by uncovering insider threats and misconfigurations that traditional controls may overlook. Machine learning-driven approaches further enhance this capability through adaptive learning, handling class imbalance, and scaling to large telemetry streams while maintaining acceptable false positive rates. Additionally, outlier detection contributes to compliance and auditability by providing evidence of anomalous access or data flows.

6. LITERATURE REVIEW

The field of anomaly detection has emerged as a critical research domain in data mining, cybersecurity, artificial intelligence, and network monitoring due to the rapid growth of digital infrastructures and cyber threats. Early foundational studies established the theoretical basis for identifying abnormal patterns that significantly deviate from expected behavior. Chandola, Banerjee, and Kumar (2009) presented one of the most influential surveys on anomaly detection, categorizing anomalies into point, contextual, and collective anomalies while discussing statistical, distance-based, density-based, clustering, and classification approaches. Their work highlighted that anomaly detection techniques vary according to application domains such as fraud detection, network intrusion detection, healthcare monitoring, and industrial fault analysis. Similarly, Aggarwal (2017) provided a comprehensive exploration of outlier analysis methodologies, emphasizing the mathematical and computational principles underlying anomaly detection systems. The author discussed proximity-based models, kernel methods, subspace analysis, and high-dimensional anomaly detection, noting that increasing data dimensionality creates challenges in accurately distinguishing normal from abnormal patterns. In the cybersecurity domain, Garcia-Teodoro et al. (2009) examined anomaly-based network intrusion detection systems and explained how behavioral monitoring can identify previously unknown attacks compared to signature-based methods that rely on predefined attack databases. Their research emphasized adaptability and the capability of anomaly-based systems to detect zero-day threats. Ahmed, Mahmood, and Hu (2016) further extended this discussion by reviewing network anomaly detection techniques, categorizing them into statistical, knowledge-based, and machine learning approaches. The authors emphasized that modern communication networks generate massive and heterogeneous traffic data, making traditional rule-based systems insufficient for detecting sophisticated attacks. Bhuyan, Bhattacharyya, and Kalita (2014) also analyzed various network anomaly detection tools and frameworks, highlighting the increasing integration of machine learning algorithms into intrusion detection systems due to their ability to learn evolving traffic behaviors. These studies collectively established anomaly detection as a multidisciplinary research field requiring intelligent computational techniques capable of adapting to dynamic environments and large-scale datasets.

The development of benchmark datasets significantly contributed to the advancement and evaluation of anomaly detection and intrusion detection systems. Tavallaee et al. (2009) critically analyzed the widely used KDD Cup 99 dataset and identified several limitations such

as redundant records, imbalance issues, and unrealistic traffic distributions that negatively affected classifier performance evaluation. Their work demonstrated that many machine learning algorithms achieved artificially high accuracy due to repeated records rather than genuine learning capability. To overcome these limitations, newer datasets were introduced to represent realistic modern network traffic patterns. Moustafa and Slay (2015) developed the UNSW-NB15 dataset, which included contemporary attack scenarios, normal network behaviors, and diverse traffic features generated using modern network simulation tools. The dataset was designed to address deficiencies in earlier datasets and provide a reliable benchmark for evaluating intrusion detection techniques. Similarly, Sharafaldin, Lashkari, and Ghorbani (2018) introduced the CICIDS2017 dataset, which incorporated updated attack profiles including brute force attacks, denial-of-service attacks, botnets, and web attacks within realistic network traffic environments. Their research emphasized the importance of capturing benign and malicious traffic under real-world operational conditions to improve model generalization. These datasets enabled researchers to evaluate both supervised and unsupervised anomaly detection models using comprehensive traffic characteristics and labeled attack categories. In addition to dataset development, Sommer and Paxson (2010) critically examined the application of machine learning techniques in network intrusion detection systems. They argued that despite the popularity of machine learning, practical deployment remains challenging because real-world operational environments contain noisy, evolving, and highly imbalanced data. Their work emphasized the “closed world assumption,” where many research systems are evaluated in static environments that do not reflect practical network conditions. This critique motivated researchers to develop adaptive, robust, and context-aware anomaly detection systems capable of handling concept drift, evolving attack patterns, and operational uncertainties. Consequently, dataset realism and deployment feasibility became essential considerations in the design of anomaly detection frameworks.

Traditional anomaly detection techniques primarily relied on statistical and distance-based algorithms before the emergence of deep learning approaches. Breunig et al. (2000) introduced the Local Outlier Factor (LOF) algorithm, which became a landmark density-based anomaly detection method. LOF identifies anomalies by measuring the local density deviation of a data point relative to its neighbors, allowing the detection of local anomalies that may not appear globally abnormal. This approach proved highly effective in multidimensional datasets where anomalies occur within localized regions. Later, Liu, Ting, and Zhou (2008) proposed the Isolation Forest algorithm, which revolutionized anomaly detection through an efficient tree-based isolation mechanism. Unlike distance-based methods, Isolation Forest isolates anomalies using random partitioning, where anomalous instances require fewer partitions to separate due to their rarity and distinctiveness. The algorithm demonstrated superior computational efficiency and scalability for large datasets, making it highly suitable for real-time intrusion detection applications. Goldstein and Uchida (2016) conducted a comparative evaluation of unsupervised anomaly detection algorithms across multivariate datasets and highlighted the strengths and limitations of different techniques under varying data distributions and noise levels. Their findings indicated that no single algorithm consistently outperformed others

across all scenarios, emphasizing the importance of selecting methods according to data characteristics and application requirements. Statistical methods were found effective for simple distributions, while density-based and ensemble approaches performed better in complex and high-dimensional environments. These traditional methods laid the groundwork for modern anomaly detection systems and continue to serve as important baselines in contemporary research. However, researchers observed limitations in handling highly nonlinear relationships, feature complexity, and unstructured data. As network environments became increasingly dynamic and data-intensive, the demand for more advanced representation learning techniques accelerated the transition toward deep learning-based anomaly detection frameworks.

Recent research has increasingly focused on deep learning techniques due to their ability to automatically extract complex hierarchical representations from high-dimensional data. Chalapathy and Chawla (2019) provided a comprehensive survey of deep learning approaches for anomaly detection, discussing autoencoders, recurrent neural networks, convolutional neural networks, generative adversarial networks, and deep belief networks. Their study emphasized that deep learning models can learn latent representations of normal behavior without extensive feature engineering, thereby improving anomaly detection accuracy in complex environments. Pang et al. (2021) further expanded this area by reviewing modern deep anomaly detection techniques and categorizing them into supervised, semi-supervised, hybrid, and self-supervised approaches. The authors highlighted that deep learning models significantly outperform traditional methods when handling large-scale, heterogeneous, and nonlinear datasets such as network traffic, image streams, and sensor data. However, they also noted challenges including interpretability, training complexity, high computational cost, and dependence on large datasets. Ruff et al. (2020) proposed the Deep One-Class Classification framework, which extended one-class classification into deep neural architectures. Their approach aimed to learn compact representations of normal samples while maximizing separation from anomalous patterns. This method demonstrated strong performance in image and cybersecurity anomaly detection tasks, particularly when labeled anomaly data were scarce. Deep learning models have also enabled the integration of temporal and spatial information, making them suitable for detecting sophisticated cyberattacks, insider threats, and evolving malicious behaviors. Despite these advancements, researchers continue to face challenges related to adversarial attacks, class imbalance, model explainability, and real-time deployment in resource-constrained environments. Current research trends focus on hybrid frameworks combining statistical analysis, machine learning, and deep learning to enhance detection accuracy and robustness. Furthermore, the integration of explainable artificial intelligence, federated learning, edge computing, and adaptive learning mechanisms is expected to shape the future of anomaly detection systems.

Literature Summary

S. No.	Author(s) & Year	Methodology / Technique	Key Findings	Research Contribution

1	Chandola, Banerjee, and Kumar (2009)	Survey of anomaly detection methods including statistical, clustering, and classification approaches	Categorized anomalies into point, contextual, and collective anomalies	Provided a comprehensive theoretical foundation for anomaly detection research
2	Aggarwal (2017)	Outlier analysis using proximity-based and high-dimensional techniques	Discussed challenges of anomaly detection in multidimensional datasets	Advanced computational methods for outlier detection systems
3	Ahmed, Mahmood, and Hu (2016)	Review of network anomaly detection techniques	Compared statistical, knowledge-based, and machine learning methods	Improved understanding of cybersecurity anomaly detection frameworks
4	Bhuyan, Bhattacharyya, and Kalita (2014)	Analysis of network anomaly detection systems and tools	Highlighted the effectiveness of intelligent intrusion detection models	Contributed toward adaptive network security mechanisms
5	Garcia-Teodoro et al. (2009)	Anomaly-based intrusion detection analysis	Demonstrated advantages over signature-based systems	Improved detection of unknown cyberattacks
6	Sommer and Paxson (2010)	Evaluation of machine learning for intrusion detection	Identified limitations of machine learning in practical deployment	Emphasized realistic evaluation of intrusion detection systems
7	Tavallaee et al. (2009)	Analysis of KDD Cup 99 dataset	Found redundancy and imbalance issues in the dataset	Motivated development of more realistic benchmark datasets
8	Moustafa and Slay (2015)	Development of UNSW-NB15 dataset	Introduced realistic and modern network traffic scenarios	Enhanced evaluation of intrusion detection systems
9	Sharafaldin, Lashkari, and Ghorbani (2018)	Creation of CICIDS2017 intrusion detection dataset	Included updated attack profiles and realistic traffic patterns	Improved benchmarking for cybersecurity research

10	Goldstein and Uchida (2016)	Comparative evaluation of unsupervised anomaly detection algorithms	Showed that no single method performs best for all datasets	Assisted researchers in selecting suitable anomaly detection models
11	Liu, Ting, and Zhou (2008)	Isolation Forest algorithm	Efficiently isolated anomalies using random partitioning	Introduced scalable anomaly detection for large datasets
12	Breunig et al. (2000)	Local Outlier Factor (LOF) density-based method	Detected local anomalies using density deviation	Improved local anomaly identification in multidimensional data
13	Chalapathy and Chawla (2019)	Survey of deep learning-based anomaly detection	Reviewed autoencoders, GANs, and neural network methods	Highlighted the role of deep learning in anomaly detection
14	Pang et al. (2021)	Review of deep anomaly detection techniques	Discussed supervised, unsupervised, and hybrid deep models	Provided a detailed overview of modern deep learning approaches
15	Ruff et al. (2020)	Deep One-Class Classification framework	Learned compact representations of normal data	Enhanced anomaly detection performance in limited-label environments

7. RESEARCH METHODOLOGY

1. Supervised Learning Approaches

Supervised learning approaches for outlier detection rely on labeled datasets in which network traffic instances are pre-classified as normal or anomalous. These methods learn explicit decision boundaries that distinguish between benign and malicious behavior, making them highly effective when high-quality labeled data is available. Decision Trees are widely used due to their interpretability and ability to model non-linear relationships through hierarchical rule structures. Support Vector Machines (SVM) are particularly effective in high-dimensional spaces, where they construct optimal separating hyperplanes to maximize classification margins. Logistic Regression, although simpler, provides probabilistic outputs and is useful for binary classification tasks in intrusion detection systems. Despite their accuracy, supervised methods are limited by the scarcity of labeled anomaly data and their inability to generalize well to unseen attack patterns, especially in dynamic network environments.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

2. Unsupervised Learning Approaches

Unsupervised learning approaches do not require labeled data and instead detect anomalies by identifying deviations from the inherent structure of network traffic. These methods are particularly valuable in real-world scenarios where labeling is expensive or impractical. K-Means Clustering partitions data into clusters based on similarity, and data points that fall far from cluster centroids are considered anomalies. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) identifies dense regions in the data and labels low-density points as anomalies, making it effective for detecting irregular patterns. Isolation Forest is a tree-based method that isolates anomalies through random partitioning, offering high efficiency and scalability for large datasets. One-Class SVM models the boundary of normal data and flags any deviations as anomalies. While unsupervised techniques are flexible and capable of detecting unknown threats, they may suffer from lower precision due to the absence of labeled guidance.

3. Semi-Supervised Learning

Semi-supervised learning serves as a bridge between supervised and unsupervised approaches by utilizing a small amount of labeled data, typically normal instances, along with a large volume of unlabeled data. These methods assume that anomalous behavior deviates significantly from learned normal patterns. Hybrid anomaly detection models in this category combine statistical techniques with machine learning algorithms to improve detection accuracy and robustness. This approach is particularly suitable for network security applications where labeled anomalies are rare but normal traffic data is abundant. Semi-supervised models offer a balanced trade-off between detection performance and data dependency, making them increasingly popular in practical implementations.

4. Deep Learning Approaches

Deep learning approaches leverage advanced neural network architectures to model complex, high-dimensional patterns in network data. Autoencoders are commonly used for anomaly detection by learning compressed representations of normal data and identifying anomalies through high reconstruction errors. Recurrent Neural Networks (RNN), including variants like LSTM, are effective for analyzing sequential and time-series data, capturing temporal dependencies in network traffic. Convolutional Neural Networks (CNN) can extract spatial features from structured representations of network data, improving detection accuracy. Transformer-based models, which utilize attention mechanisms, are capable of capturing long-range dependencies and have shown promising results in large-scale anomaly detection tasks. Although deep learning methods offer superior performance, they require significant computational resources and large datasets for effective training.

8. ARCHITECTURE OF MACHINE LEARNING FRAMEWORKS FOR NETWORK OUTLIER DETECTION

1. Data Collection (Network Traffic, Logs, IoT Streams)

The architecture of machine learning frameworks for network outlier detection begins with comprehensive data collection from diverse sources within the network ecosystem. These sources include raw network traffic (packet-level and flow-level data), system and application

logs, and data streams generated by Internet of Things (IoT) devices. The objective is to capture a holistic view of network behavior, encompassing parameters such as IP addresses, ports, protocols, timestamps, and payload characteristics. Effective data collection ensures the availability of high-quality, representative datasets necessary for building robust anomaly detection models. However, challenges such as data heterogeneity, volume, and velocity must be managed through efficient data acquisition and storage mechanisms.

2. Data Preprocessing (Cleaning, Normalization, Feature Selection)

Once data is collected, preprocessing is performed to enhance its quality and suitability for machine learning algorithms. This stage involves cleaning the data by removing duplicates, handling missing values, and filtering irrelevant or corrupted entries. Normalization or scaling is applied to ensure that features with different ranges do not disproportionately influence the model. Feature selection techniques are employed to identify the most relevant attributes, reducing redundancy and improving computational efficiency. Proper preprocessing is crucial for minimizing noise and bias, thereby improving the accuracy and reliability of subsequent detection models.

3. Feature Engineering and Dimensionality Reduction (e.g., PCA)

Feature engineering involves transforming raw data into meaningful representations that better capture the underlying patterns of network behavior. This may include creating new features such as traffic rates, session durations, or protocol distributions. Given the high dimensionality of network data, dimensionality reduction techniques are often applied to simplify the feature space while preserving essential information. Methods such as Principal Component Analysis (PCA) help reduce computational complexity and mitigate the curse of dimensionality. This stage enhances model performance by improving generalization and reducing overfitting.

4. Model Training and Validation

In this phase, machine learning models are trained using the processed dataset to learn patterns of normal and anomalous behavior. Depending on the paradigm, this may involve supervised, unsupervised, or semi-supervised learning techniques. The dataset is typically divided into training and testing subsets to evaluate model performance. Validation techniques such as cross-validation are used to ensure that the model generalizes well to unseen data. Hyperparameter tuning is also conducted to optimize model performance. This stage is critical for building accurate and reliable anomaly detection systems.

5. Ensemble and Hybrid Frameworks (Voting-Based Models, Boosting and Bagging Approaches)

Ensemble and hybrid frameworks combine multiple models to improve detection performance, robustness, and generalization. Voting-based models aggregate predictions from different classifiers to produce a final decision, reducing individual model bias. Boosting techniques, such as AdaBoost, iteratively focus on misclassified instances to improve accuracy, while bagging methods like Random Forest reduce variance by training multiple models on different data subsets. Hybrid frameworks may integrate statistical, machine learning, and deep learning techniques to leverage their complementary strengths. These approaches are highly effective in complex environments but require careful design and increased computational resources.

6. Deep Learning Frameworks (End-to-End Pipelines, AutoML-Based Frameworks)

Deep learning frameworks represent the most advanced category, utilizing neural networks to automatically learn hierarchical representations of network data. End-to-end anomaly detection pipelines integrate data preprocessing, feature extraction, and detection within a single architecture, reducing manual intervention. Techniques such as autoencoders, recurrent neural networks, and convolutional neural networks are commonly used. Additionally, AutoML-based frameworks automate model selection, hyperparameter tuning, and pipeline optimization, making advanced anomaly detection accessible and efficient. While these frameworks offer superior performance and scalability, they require large datasets, significant computational power, and may lack interpretability compared to traditional methods.

9. CONCLUSION

This systematic review has comprehensively examined the landscape of machine learning frameworks for network outlier detection, highlighting their critical role in strengthening modern cybersecurity infrastructures. With the exponential growth of networked systems, including cloud computing, Internet of Things (IoT), and smart environments, the detection of anomalous behavior has become increasingly complex and essential. The study explored a wide range of machine learning paradigms, including supervised, unsupervised, semi-supervised, and deep learning approaches, each offering unique advantages in terms of accuracy, adaptability, and scalability. It also analyzed the architectural components of these frameworks, from data collection and preprocessing to model training and real-time detection, emphasizing their importance in building robust and efficient detection systems. Furthermore, the review provided a detailed taxonomy of existing frameworks and evaluated widely used benchmark datasets, enabling a clearer understanding of comparative performance and research trends. Despite significant advancements, several challenges persist, such as handling high-dimensional and imbalanced data, ensuring model interpretability, and adapting to evolving threat landscapes. The study also identified emerging directions, including explainable artificial intelligence, federated learning, and real-time analytics, which have the potential to address current limitations.

REFERENCES

1. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882
2. Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-47578-3
3. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016
4. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network anomaly detection: Methods, systems and tools. IEEE Communications Surveys & Tutorials, 16(1), 303–336. https://doi.org/10.1109/SURV.2013.052213.00046

5. Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28(1–2), 18–28. https://doi.org/10.1016/j.cose.2008.08.003
6. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. IEEE Symposium on Security and Privacy, 305–316. https://doi.org/10.1109/SP.2010.25
7. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD Cup 99 dataset. IEEE Symposium on Computational Intelligence for Security and Defense Applications, 1–6. https://doi.org/10.1109/CISDA.2009.5356528
8. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. Military Communications and Information Systems Conference (MilCIS), 1–6. https://doi.org/10.1109/MilCIS.2015.7348942
9. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSP, 108–116. https://doi.org/10.5220/0006639801080116
10. Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS ONE, 11(4), e0152173. https://doi.org/10.1371/journal.pone.0152173
11. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. IEEE International Conference on Data Mining, 413–422. https://doi.org/10.1109/ICDM.2008.17
12. Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. ACM SIGMOD Record, 29(2), 93–104. https://doi.org/10.1145/335191.335388
13. Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.
14. Pang, G., Shen, C., Cao, L., & Hengel, A. V. D. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys, 54(2), 1–38. https://doi.org/10.1145/3439950
15. Ruff, L., Vandermeulen, R. A., Görnitz, N., Binder, A., Müller, E., Müller, K.-R., & Kloft, M. (2020). Deep one-class classification. International Conference on Machine Learning, 4393–4402.