GLOBAL JOURNAL
OF COMPUTER SCIENCE AND TECHNOLOGY - E: NETWORK, WEB & SECURITY
Volume 26 | 2026 Article ID: GJCST-254066 Language: English

Unsupervised Deep Autoencoder for Zero-Day Anomaly Detection in Network Traffic

Online First

Isamaliya Kajal Kanubhai[§ϑ*]
Research Scholar
*Corresponding Author

Dr. Sushma Ghode[§ξ]
Research Guide

Dr. Chaitanya Singh[§θ]
Principal

§ Vidhyadeep University, Anklav, India

RECEIVED	ACCEPTED	ONLINE PUBLISHED	PEER REVIEW
2026-03-09	2026-03-23	2026-07-10	Double Blind

Abstract

Zero-day attacks are among the most serious problems in today's network security because these attacks exploit unknown vulnerabilities and are able to evade classical signature-based intrusion detection systems. Recently, great success has been achieved in the application of deep learning, especially unsupervised deep autoencoders, in detecting anomalous patterns in the traffic data without relying on labeled attack data. The autoencoders are able to learn the representation of normal traffic and detect anomalies based on reconstruction errors. This review presents a thorough examination of the existing unsupervised deep autoencoder-based methods and their combination models proposed for zero-day anomaly detection in network traffic. The relevant work is critically analyzed from the perspective of model design, datasets, validation practices, and the ability to handle unknown anomalies. The recent advancements in the field with the evolution of convolutional, variational, temporal, and attention-driven autoencoders are also presented. Although these models demonstrated excellent results, the problem of high false positives, unstandardized zero-day validation, lack of interpretability, and practical deployability limitations exists. Finally, the discussion ends with addressing future directions toward an adaptive and interpretable autoencoder-based intrusion detection system.

Deep Autoencoder | Zero-Day Attack | Unsupervised Learning | Intrusion Detection | Network Security | Anomaly Detection

AI USE STATEMENT
No generative AI was used for analysis or results.

FUNDING
No external funding was declared for this work.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY
Not applicable for this article.

ETHICS
No ethics committee approval was required for this article type.

CONSENT
Not applicable for this article.

TRIAL REG.
Not applicable.

Crossref DOI: 10.34257/GJCSTE254066

How to Cite: Kanubhai et al. (2026). Unsupervised Deep Autoencoder for Zero-Day Anomaly Detection in Network Traffic. Global Journal of Computer Science and Technology, Ahead of Print, 1-15. DOI: 10.34257/GJCSTE254066

AR Experience Web Page DOI

Print ISSN 0975-4350
Online ISSN 0975-4172

Under the strict compliance and defined process of

METADATA CONTINUATION

AUTHOR CONTACT QR LEDGER

Isamaliya Kajal Kanubhai§θ*

Dr. Sushma Ghode§ξ

Dr. Chaitanya Singh§θ

ARCHIVAL RECORD

GJCST · Vol 26 · 2026

Article ID GJCST-254066 · DOI 10.34257/GJCSTE254066

Print ISSN 0975-4350 · Online ISSN 0975-4172

Global Journal of Computer Science and Technology - Section E
GJCST

Unsupervised Deep Autoencoder for Zero-Day Anomaly Detection in Network Traffic

Isamaliya Kajal Kanubhai[§ϑ*], Dr. Sushma Ghode[§ξ], and Dr. Chaitanya Singh[§θ]

Affiliations
§ Vidhyadeep University, Anklav, India

Qualifications / Designations
ϑ Research Scholar
ξ Research Guide
θ Principal

Abstract

Zero-day attacks are among the most serious problems in today's network security because these attacks exploit unknown vulnerabilities and are able to evade classical signature-based intrusion detection systems. Recently, great success has been achieved in the application of deep learning, especially unsupervised deep autoencoders, in detecting anomalous patterns in the traffic data without relying on labeled attack data. The autoencoders are able to learn the representation of normal traffic and detect anomalies based on reconstruction errors. This review presents a thorough examination of the existing unsupervised deep autoencoder-based methods and their combination models proposed for zero-day anomaly detection in network traffic. The relevant work is critically analyzed from the perspective of model design, datasets, validation practices, and the ability to handle unknown anomalies. The recent advancements in the field with the evolution of convolutional, variational, temporal, and attention-driven autoencoders are also presented. Although these models demonstrated excellent results, the problem of high false positives, unstandardized zero-day validation, lack of interpretability, and practical deployability limitations exists. Finally, the discussion ends with addressing future directions toward an adaptive and interpretable autoencoder-based intrusion detection system.

Keywords: *Deep Autoencoder, Zero-Day Attack, Unsupervised Learning, Intrusion Detection, Network Security, Anomaly Detection*

*** Corresponding Author**
Isamaliya Kajal Kanubhai

DOI
10.34257/GJCSTE254066

1. Introduction

Modern computer networks are suffering from an unprecedented escalation in both the volume and complexity of cyberattacks. Zero-day exploits, which leverage previously unknown vulnerabilities, are a significant challenge since they naturally evade traditional IDS models reliant on labeled datasets or predefined attack signatures. As networks evolve, these systems are unable to generalize to new traffic patterns and unseen attack vectors.

Deep learning has emerged as a powerful tool for modeling complicated and nonlinear data distributions. Within the domain of deep learning architectures, AEs are a prominent approach due to their unsupervised feature-learning ability. Training on normal network behaviour, an AE learns to reconstruct benign traffic accurately. Therefore, any deviation from learned patterns develops into a higher reconstruction error that indicates an anomaly.

This paper reviews recent work that leverages deep autoencoders for unsupervised anomaly detection, putting a special emphasis on their zero-day detection capability. The contribution of this review is threefold:

1. It systematizes recent works, between years 2020 and 2025, in a taxonomy of AE architectures utilized in IDS research.
2. It makes a critical comparison among the detection performance, datasets, and evaluation methodologies.
3. It highlights the main gaps in the prevailing research and further suggests directions for exploration.

2. Background

2.1. Anomaly Detection in Network Traffic

Network anomaly detection involves identifying traffic behavior that is somehow different from a model of "normal" activity. Broadly speaking, approaches fall into one of the following categories:

- Signature-Based Detection: This form of detection relies on known attack patterns.
- Supervised learning: trains classifiers on labeled benign and malicious samples.
- Unsupervised learning: It builds a model of normal traffic distribution and flags deviations from it as anomalous.

In particular, unsupervised methods have the further appeal of being immediately suited to zero-day attacks insofar as they do not rely on attack data that is labelled.

2.2. Autoencoders for Unsupervised Learning

It contains an encoding part that compresses the input, x, into a latent representation, z, and a decoder that reconstructs, x', from

Type of Architecture	Description	Advantages	Limitations
Vanilla AE	Basic dense encoder-decoder	simple, lightweight	Limited feature learning
Convolutional AE	uses CNN layers to capture features in spatial dimensions	This is excellent for image-like or flow-matrix traffic.	Requires structured inputs
VAE, Variational AE	Learn probabilistic latent variables	Model distributional uncertainty	Reconstructed details might get blurred
Recurrent AE LSTM/GRU	Captures sequential dependencies in temporal data.	Ideal for flow sequences or time series	High computation cost
Hybrid AE + SVM / AE + Clustering	Combines representation learning with discriminative models.	It has an improved separation of anomalies	Extra tuning is required

z. The model minimizes reconstruction loss, $L(x, x')$, typically mean squared error. When this is trained exclusively on normal traffic, the reconstruction errors for anomalous traffic increase drastically.

Other variants, such as CAE, VAE, and RAE, enhance feature extraction by exploiting spatial or temporal dependencies.

2.3. Evaluation Metrics

The evaluation of unsupervised intrusion detection systems needs intensive consideration when choosing evaluation metrics, especially when dealing with imbalanced classes or zero-day attacks. Anomaly detection based on reconstruction, contrasting supervised classification, uses thresholding on anomaly scores instead of predictions.

Accuracy, though frequently cited, might be misleading because the prevalence of normal traffic strongly dominates the results. Therefore, metrics related to detection reliability are preferred.

"Precision": the proportion of detected anomalies that are malicious, directly reflecting the quality of the alert and the analysis work.

"Recall" (also Detection Rate) is a measure of the system's capability to detect true attacks and has become a key aspect for zero-day threat containment.

F1-score, being the harmonic mean of precision and recall, is more robust against class imbalance than precision or recall.

A high False Positive Rate (FPR) is very important for operational IDS systems, where too many false alerts can cause the system to be considered ineligible. A few recent works have reported the results of FPR at certain points of recall, such as FPR@TPR = 0.9.

Area Under the ROC Curve (AUC) measures ranking performance irrespective of the threshold values, and it has been widely used in the test of general ability. AUC does not measure the ability of the model deployment.

In reconstruction models, reconstruction error distributions can directly serve as anomaly scores; threshold strategies, whether static, percentile, or adaptive, have a large impact on results reported. Validation of zero days requires more advanced forms of evaluation, including attack family holdout evaluation, time splitting, and cross dataset validation, which avoid information leakage and enable realistic assessment of generalization, as illustrated in [31, 42].

3. Taxonomy of Autoencoder Architectures

4. Comparison of Recent Research

In the last five years, research on unsupervised DAE for zero-day anomaly detection in network traffic has accelerated significantly. Several scholars have researched a wide spectrum of architectural enhancements, hybrid learning strategies, and evaluation protocols to improve the generalization capability of IDS models.

The following section critically summarizes and compares the findings of fifteen key contributions from the period between 2020 and 2025. These papers have been selected based on their relevance, methodology, quality of datasets considered, and unsupervised/semi-supervised anomaly detection concepts involving an autoencoder-based model.

4.1. Overview of Important Related Research

Narrative Summary for Remaining Papers:

A few more studies, which have not been included in above table, support such views. Some of these studies include [8, 14, 16, 17, 20, 24-27, 30, 33-38, 40, 41, 43] which deal with architectural optimization, threshold adaptation, drift issues, adversarial abilities, and online learning, respectively. Even though they achieve very good detection accuracy, most of these studies employ randomly chosen training and testing sets, which tend to be biased and render zero-day performance poor and unreliable in actuality. Recent book reviews, such as [12] and [45], establish this view and establish a taxonomy along these lines.

4.2. Comparative Insights

A close comparison of the aforementioned forty-five reviewed studies highlights some consistent and statistically significant trends across autoencoder-based intrusion detection research.

1. **Hybrid architectures have always provided better robustness and enhanced precision**. It is shown in a large fraction of recent works that the integration of autoencoder-based representation learning with secondary decision mechanisms—be it clustering algorithms, support vector machines, or ensemble voting—significantly enhances anomaly discrimination and reduces the false-positive rates. Hybrid AE–SVM, AE–clustering, and ensemble-based frameworks obtain far more stable decision boundaries in the latent space compared to standalone reconstruction-based autoencoders under heterogeneous and class-imbalanced traffic conditions.

2. **Temporal modeling yields better detection for streaming and sequential traffic.** Explicitly modeling temporal dependencies by using LSTM-based auto-encoders, CNN-RNN hybrids, and transformer-based architecture consistently outperforms static dense auto-encoders in detecting low-frequency and stealthy attacks. Such methodologies explicitly modeling temporal dependencies are highly effective for real-time monitoring of attack behaviors, which evolve over time rather than as single point events.

3. **Correlation-aware and structured latent learning improves generalization.** Meanwhile, several works emphasize the significance of maintaining relationships between features in latent-space learning. Autoencoders that utilize kernels, mutual information constraints, graph models, and contrastive learning improvements give better results compared to traditional mean squared error-based autoencoders

Ref. Group	Model Type	Dataset(s)	Evaluation Protocol	Combined Key Findings
[1,6,12]	Dense / Denoising AE	NSL-KDD, CICIDS2017	Benign-only training	Simple AEs detect gross anomalies but suffer high FP rates and poor generalization
[3,10,18]	AE + SVM / Rule-based	UNSW-NB15, CICIDS2017	Latent-space classification	Hybrid models reduce FP rate and improve decision boundaries
[4,9]	LSTM / CNN-RNN AE	CAN, CICIDS2017	Sequential modeling	Temporal dependencies significantly improve stealth attack detection
[7,32,44]	Attention / Explainable AE	CICIDS2017, TON-IoT	Feature attribution	Attention improves interpretability and detection robustness
[15,39]	Transformer-VAE	IoT, 5G datasets	Probabilistic latent modeling	Strong zero-day performance but high computational cost
[19,21]	Variational / Graph AE	UNSW-NB15	Distribution modeling	Better uncertainty estimation and structural awareness
[22,29,34]	Lightweight / Quantized AE	IoT / Edge datasets	Resource-aware testing	Suitable for edge deployment with accuracy trade-offs
[28]	Federated AE	Multi-site IDS	Privacy-preserving training	Maintains performance while preserving data privacy
[31,42]	Cross-dataset Evaluation	CICIDS → UNSW	True zero-day testing	Reveals performance drop under realistic conditions
[45]	Review Paper	—	Meta-analysis	Highlights evaluation inconsistencies and research gaps

due to their ability to model correlated and relational patterns in traffic, which often signal either coordinated or multi-stage attacks.

4. **Zero-day evaluation practices remain inconsistent across the literature.** On the contrary, despite the high reported accuracy in detection, only a few of these reviewed works use some strict zero-day evaluation strategies such as attack-family holdout, cross-dataset testing, or time-based splits. Most of these reviewed studies rely on random data partitioning. This means that during training, there is unintended leakage that exposes their models to the general attack characteristics, thereby inflating their performance metrics. This inconsistency underlines the requirement of standardized benchmarking protocols that can actually evaluate the true zero-day generalization.

5. **Attention and transformer-based autoencoders represent a growing research frontier.** Some recent works have increasingly integrated attention mechanisms and transformer encoders into autoencoder frameworks to enhance both detection accuracy and interpretability, including [7], [15], [32] [39], and [44]. These models dynamically emphasize salient traffic features with limited explanatory insights into anomaly decisions, thus becoming particularly appropriate for IoT, cloud, and 5G environments with high-dimensional and non-stationary traffic.

6. **Architectural choices are based on practical deployment considerations.** While complex architectures achieve high detection performance, several studies emphasize trade-offs between accuracy, computational overhead, and real-time feasibility. Lightweight, quantized, and pruned autoencoders ([22], [29], [34]) are favored for edge and IoT deployments, whereas deeper temporal and transformer-based models are more suitable for centralized or cloud-based IDS infrastructures.

4.3. Critical Evaluation

While most recent autoencoder-based intrusion detection research reports detection accuracies of over 98%, a closer look at their experimental approach indicates that such metrics often present an inflated view of real-world performance. In many cases, this involves random train–test splits. Such a method inadvertently exposes the model to attack traffic characteristics during training. Only a few studies employ proper zero-day validation strategies, such as excluding attack families, time-aware splitting, or cross-dataset testing ([1], [6], [15], [31], [42]). Cross-dataset evaluation, such as training on CICIDS2017 and testing on UNSW-NB15 or TON-IoT, remains rare despite offering the most realistic assessment of generalization capability.

One of the persistent and widely reported limitations in the reviewed literature is the trade-off between detection sensitivity and false-positive rates. The deep autoencoders, while being trained on exclusively benign traffic, tend to overfit normal data distributions and may reconstruct subtle or slowly evolving attacks with low reconstruction error. This behavior leads to increased false negatives or forces practitioners to adopt conservative thresholds, thereby increasing false-positive rates and reducing operational usability. Hybrid frameworks that embed clustering, SVMs, ensemble voting, or adaptive thresholding have partially mitigated this issue but at the cost of increased complexity and further tuning requirements.

Interpretability remains another significant weakness of current autoencoder-based intrusion detection. The majority of the reviewed studies consider the autoencoder as a black-box model that provides very limited insight into which traffic features or latent representations contribute to decisions about anomalies. Only a few recent works ([7], [32], [44]) try to fill this gap by using attention mechanisms, feature attribution techniques, or disentangled latent spaces. In the absence of explainability, analyst trust is significantly degraded, as well as forensic investigation and regulatory compliance, in critical sectors such as finance, health, and industrial control systems.

From the deployment viewpoint, most of the top-performing architectures - including deep LSTM autoencoders and transformer-based variational autoencoders - bear high computation and memory overheads as per [15, 39, 41]. In this regard, such models will be unsuitable for real-time detection at the edge of the network or in resource-constrained IoT settings. On the contrary, lightweight and quantized autoencoders, such as those proposed in [22, 29, 34], tend to be efficient with degraded detection accuracy and adaptability.

In an overall sense, this critical review of benchmark accuracy demonstrates that mere high values are not enough to achieve real-world zero-day intrusion detection. From now on, any future autoencoder-based IDS research should stress the issues related to zero-day evaluations, sensitivity and specificity trade-offs, explainability, and computational efficiency. Overcoming these challenges is crucial for the future development of adaptive

and trustworthy intrusion detection systems that work reliably in large-scaled dynamic network environments.

5. Discussion and Research Trends

5.1. Evolution of Unsupervised Deep Autoencoders in IDS

From 2020 to 2025, there has been a notable shift in focus from simple feed-forward autoencoders to hybrid and attention-enhanced architectures. Early works ([12], [14]) developed dense AEs for dimensionality reduction and unsupervised anomaly detection. By 2022–2025, hierarchical, variational, and transformer-based architectures ([9], [15]) that could model nonlinear and context-dependent traffic features started to attract attention.

The primary motivation for this evolution is the need for zero-day robustness—ensuring IDS systems can detect anomalies unseen during training through probabilistic modelling in VAE, temporal context in LSTM-AE, and feature correlation in kernelized AEs.

5.2. Architectural Innovation Trends

1. **Variational and Probabilistic AEs:** VAEs model the distributions of the latent space, enabling probabilistic thresholds of anomaly scores. The most recent work in this paper by Khalaf et al. [15] extended VAEs with transformer encoders, achieving better results in IoT environments with non-stationary data streams.

2. **Hybrid AE Frameworks:** Sharper decision boundaries in the latent space are obtained when autoencoders are integrated with discriminative classifiers, such as LS-SVMs [3] or clustering modules [8]. Such models are computationally heavier but deliver better separation between normal and abnormal samples.

3. **Temporal and Attention-Based Models:** Sequential dependencies in traffic data are efficiently captured by LSTM-AEs [4] and CNN-RNN hybrids [9]. Attention mechanisms further enhance these models by assigning adaptive weights to features most relevant for anomaly identification [7].

4. **Latent Learning with Correlation Awareness:** Roy & McNeely [5] enforced mutual-information constraints within the loss function, which granted the guarantee that the encoder would preserve the inter-feature correlations that are usually ignored when using a reconstruction-only objective.

5.3. Dataset and Evaluation Practices

This is a recurring problem with most of the studies: their lack of uniformity in evaluation protocols.

- Most of them have used popular datasets such as CICIDS2017, NSL-KDD, and UNSW-NB15, which are already acknowledged to possess redundant or outdated attack types.

- Only a few papers include newer IoT and 5G datasets such as TON-IoT (2023) or Edge-IIoT (2024), which can better represent modern network infrastructures.

- True zero-day validation (that is, excluding whole categories of attack from training) is done inconsistently.

This will be future work; cross-dataset testing should be performed along with time-based splits to ensure temporal generalization. The recent benchmarking proposals combine datasets, such as training on CICIDS2017 and testing on UNSW-NB15, to mimic real-world deployment scenarios.

5.4. Industrial and Real-World Applicability

While real-time detection remains one of the big challenges,

- Temporal models - the LSTM/Transformer AEs offer accuracy but at higher computational costs.

- Lightweight AEs can run on either edge or IoT devices and have poor precision.

It is important to balance latency with computational cost and the reliability of detection in deploying AE-based IDSs in production systems.

In domains such as finance, healthcare, and IoT, explainability is also crucial. Organisations require explanations for why a packet or flow was classified as anomalous. Their application to AEs represents an area of emerging research where visualization tools and feature-attribution methods (e.g., SHAP or gradient-based saliency) are adapted to AEs in order to bridge this gap.

5.5. Directions for Future Research

Several exciting directions have emerged recently:

- **Adversarially Robust Autoencoders:** Integrate adversarial training to make them resistant to evasion attacks by leveraging the vulnerabilities of the models themselves.

- **Self-Supervised Learning:** Using pretext tasks to enhance feature representations without labeled data.

- **Federated Learning for IDS:** The training of distributed AE models in different organizations while preserving the privacy.

- **Explainable Autoencoders (XAE):** The proposal of interpretable latent factors to make the explanation of anomalies transparent to the analysts.

Taken together, these approaches portend a direction toward explainable, trustworthy, and decentralized intrusion detection systems.

6. Open Challenges and Future Directions

Despite the remarkable achievements made so far, a number of challenges have not been resolved in autoencoder-based zero-day detection systems. Careful understanding of these challenges will usher in the next phase of research and practical deployment.

6.1. Insufficient True Zero-Day Testing

Most of the published works make use of random data splits, mixing the known attack traces inadvertently into both training and test sets, inflating accuracy and failing to reflect conditions that are realistic for zero-day.

Future work: standardize attack-family holdout protocols and conduct more cross-dataset testing, such as training on CICIDS 2017 and testing on UNSW-NB15/TON-IoT, to estimate true generalization.

6.2. Overfitting and Poor Generalization

Large deep autoencoders can memorize normal traffic patterns, allowing the reconstruction of malicious flows with low error.

Future direction: Introduce regularization that includes dropout, sparsity, or variational noise, introducing domain adaptation techniques so that the models stay robust when any change in network topology or protocol mix arrives.

6.3. Explainability and Interpretability

Most of the AE-based IDSs are black boxes in their behavior. Analysts want interpretable results to understand the alerts.

Future direction: Develop XAEs, which attribute the latent variables to specific network features, and provide visualizations of reconstruction discrepancies using SHAP, Grad-CAM, or latent attention heatmaps.

6.4. Adversarial Robustness

Attackers can craft adversarial packets that bypass anomaly detectors by slightly perturbing features.

Future direction: Use adversarial training and robust loss formulations-e.g., min-max objectives-to resist such evasion strategies.

6.5. Computational Efficiency and Deployment

Complex architectures such as LSTM-AE and Transformer-VAE yield high accuracy but are inappropriate for edge or real-time environments.

Future direction: design lightweight, quantized, or pruning-based AEs deployable on routers, IoT gateways, or SDN controllers without dependence on any GPUs.

6.6. Data privacy and federated learning

Centralized IDS training violates data-sharing regulations quite often.

Future direction: utilize Federated AEs in which multiple organizations collaborate in training of the local models and share only the gradients, ensuring confidentiality with generalization enhancement.

6.7. Standardization of Assessments

Disparate metrics and inconsistent baselines impede comparison.

A proposed future direction: use unified benchmarks-report AUC, FPR@TPR=0.9, and latency-release reproducible code and models, ensure transparency.

7. Summary Highlights

- **The prominent architectures:** include CNN-AE, VAE, LSTM-AE, and Hybrid AE-SVM models that dominate the literature of 2020–2025. An investigation on transformer-based and graph-structured autoencoders should be pursued where there is a complex topology.
- **Datasets:** CICIDS 2017 and UNSW-NB15 remain the norm. TON-IoT and Edge-IIoT are two new state-of-the-art datasets for IoT/5G networks.
- **Performance:** Most models report > 98% under relaxed conditions. Enforce genuine zero-day validation and cross-dataset testing.
- **Explainability:** Emerging attention-based AEs improve transparency slightly. Need interpretable latent features and visualization tools.
- **Deployment:** High-accuracy models are often too heavy for real-time usage. Light-weight and energy-efficient AEs shall be developed targeting edge environments.
- **Future Research:** Integration of AEs with federated, self-supervised, and adversarially-robust training frameworks. Multimodal AEs fusing network, host, and behavioral data.

These insights align directly with the Ph.D. research focus of Kajal Patel: extending unsupervised deep autoencoder frameworks that generalize across datasets and detect truly unseen zero-day threats.

8. Conclusion

Unsupervised deep autoencoders have indeed emerged as one of the most promising paradigms towards zero-day anomaly detection. Theoretically, they are capable of modeling normal network behavior without labeled data and revealing unknown attacks. Significant architectural advances have been made over the last five years: from vanilla dense AEs to hybrid, probabilistic, and attention-driven models. However, there are still a number of gaps that remain to be bridged in realistic zero-day evaluation, interpretability, adversarial defense, and deployment efficiency. Cross-domain benchmarking, explainable latent representations, and privacy-preserving federated learning are crucial directions in future IDS research. Addressing these challenges will enable trustworthy, adaptive, and scalable intrusion detection systems fit for next-generation network infrastructures.

REFERENCES

[1] M. H. Javed, S. Hameed, and A. Khan, "AECNN: A convolutional autoencoder-based intrusion detection system," *IEEE Access*, vol. 12, pp. 44521–44534, 2024.

[2] V. Q. Nguyen, T. T. Huong, and J. Kim, "Deep nested clustering autoencoder for network anomaly detection," *Computers & Security*, vol. 124, Art. no. 102972, 2023.

[3] C. Zhang, W. Wei, and H. Wang, "Hybrid sparse weighted autoencoder and LS-SVM for intrusion detection," *Expert Systems with Applications*, vol. 213, Art. no. 118903, 2023.

[4] K. K. Afify, M. M. Fouda, and A. S. Taha, "LSTM-based autoencoder for CAN bus intrusion detection," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6121–6133, 2023.

[5] P. Roy and J. McNeely, "Correlation-aware kernelized autoencoders for network anomaly detection," *Neural Computing and Applications*, vol. 35, pp. 18921–18936, 2023.

[6] R. Sharma and M. Grover, "Comparative analysis of autoencoder and isolation forest for unsupervised intrusion detection," *Applied Intelligence*, vol. 54, pp. 1176–1192, 2024.

[7] S. S. Khan and A. B. Mailewa, "Attention-enhanced autoencoder for network intrusion detection," *Sensors*, vol. 24, no. 3, Art. no. 911, 2024.

[8] C. Zhang, Y. Song, and S. Liu, "Autoencoder-based clustering for high-dimensional anomaly detection," *Information Sciences*, vol. 560, pp. 332–346, 2021.

[9] E. Fotiadou, M. R. Mosavi, and A. Vakili, "CNN–RNN autoencoder for sequential network traffic anomaly detection," *IEEE Access*, vol. 10, pp. 115487–115500, 2022.

[10] M. Usama, J. Qadir, and A. R. Baig, "Hybrid autoencoder with feature fusion for intrusion detection," *Computers & Security*, vol. 105, Art. no. 102241, 2021.

[11] A. Dutta, S. Ghosh, and P. K. Singh, "Deep autoencoder-based network traffic analysis," *Procedia Computer Science*, vol. 198, pp. 78–85, 2022.

[12] M. Ahmed, A. Mahmood, and J. Hu, "A survey of network anomaly detection techniques using deep autoencoders," *Journal of Information Security and Applications*, vol. 58, Art. no. 102782, 2021.

[13] J. Zhu, H. Xiao, and R. Wang, "Chebyshev-bounded autoencoder with SVM for intrusion detection," *Future Internet*, vol. 15, no. 2, Art. no. 67, 2023.

[14] A. Kumar, R. Shukla, and S. Tripathi, "Hierarchical autoencoder ensemble for intrusion detection," *IEEE Access*, vol. 8, pp. 174820–174832, 2020.

[15] S. Khalaf, A. Alazab, and M. Mahmoud, "Transformer-based variational autoencoder for IoT intrusion detection," *Future Generation Computer Systems*, vol. 146, pp. 73–86, 2025.

[16] Y. Lin, X. Wang, and L. Chen, "Denoising autoencoder for robust network anomaly detection," *IEEE Access*, vol. 8, pp. 163292–163303, 2020.

[17] J. Wang, Y. Yang, and X. Zhou, "Sparse contractive autoencoder for network intrusion detection," *Computers & Security*, vol. 100, Art. no. 102084, 2021.

[18] S. Aljawarneh, M. B. Yassein, and M. Alawneh, "Hybrid autoencoder-rule-based intrusion detection system," *Journal of Network and Computer Applications*, vol. 175, Art. no. 102890, 2021.

[19] H. Kim, J. Park, and S. Cho, "Variational autoencoder-based anomaly detection for network traffic," *Neurocomputing*, vol. 469, pp. 79–91, 2022.

[20] A. Verma and V. Ranga, "Autoencoder-based intrusion detection for software-defined networks," *IEEE Transactions on Network and Service Management*, vol. 19, no. 2, pp. 1690–1703, 2022.

[21] Y. Li, Z. Tian, and J. Sun, "Graph autoencoder-based intrusion detection," *IEEE Access*, vol. 10, pp. 84566–84578, 2022.

[22] M. Noor, A. Khan, and F. Masood, "Energy-efficient autoencoder-based IDS for edge networks," *Sensors*, vol. 22, no. 18, Art. no. 7021, 2022.

[23] A. Hassan, S. Rahman, and M. Alazab, "Adversarial attacks against autoencoder-based intrusion detection systems," *Computers & Security*, vol. 123, Art. no. 102925, 2023.

[24] R. Silva, J. Mendes, and P. Pereira, "Transfer learning-based autoencoder for network intrusion detection," *Expert Systems with Applications*, vol. 212, Art. no. 118695, 2023.

[25] K. Patel, N. Shah, and S. Mehta, "Scalable autoencoder-based intrusion detection for cloud networks," *Future Internet*, vol. 15, no. 7, Art. no. 241, 2023.

[26] T. Brown, J. Miller, and K. Lee, "Ensemble reconstruction-error-based anomaly detection," *Pattern Recognition Letters*, vol. 169, pp. 22–30, 2023.

[27] J. Yoon, S. Lee, and H. Kim, "Self-supervised representation learning for intrusion detection using autoencoders," *IEEE Access*, vol. 11, pp. 43821–43834, 2023.

[28] M. Iqbal, A. R. Baig, and J. Qadir, "Federated autoencoder-based intrusion detection system," *Applied Intelligence*, vol. 53, pp. 9124–9140, 2023.

[29] L. Chen, X. Huang, and Z. Wang, "Lightweight quantized autoencoder for IoT intrusion detection," *Sensors*, vol. 24, no. 3, Art. no. 114, 2024.

[30] M. Rahman, S. Islam, and A. Karim, "Concept drift-aware autoencoder for online intrusion detection," *Knowledge-Based Systems*, vol. 281, Art. no. 110998, 2024.

[31] D. López, R. Martínez, and J. Gómez, "Cross-dataset evaluation of autoencoder-based IDS," *Computers & Security*, vol. 131, Art. no. 103273, 2024.

[32] J. Park, H. Kim, and S. Cho, "Explainable autoencoder-based intrusion detection using SHAP," *IEEE Access*, vol. 12, pp. 60214–60228, 2024.

[33] R. Mehta, P. Desai, and S. Kulkarni, "Multimodal autoencoder for cyber intrusion detection," *Information Fusion*, vol. 95, pp. 240–253, 2024.

[34] P. Singh, A. Mishra, and R. Kumar, "Model compression of deep autoencoders for intrusion detection," *Neural Computing and Applications*, vol. 36, pp. 1791–1806, 2024.

[35] X. Zhao, Y. Liu, and Z. Chen, "Contrastive learning-enhanced autoencoder for anomaly detection," *Pattern Recognition*, vol. 146, Art. no. 109998, 2024.

[36] A. Omar, M. Hassan, and S. Almotairi, "Encrypted traffic anomaly detection using autoencoders," *Future Generation Computer Systems*, vol. 149, pp. 318–331, 2024.

[37] S. Gupta, A. Jain, and R. Bansal, "Graph vs. CNN autoencoders for lateral movement detection," *IEEE Access*, vol. 12, pp. 78112–78126, 2024.

[38] Y. Tan, L. Zhou, and M. Chen, "Adaptive thresholding for reconstruction-based intrusion detection," *Applied Soft Computing*, vol. 149, Art. no. 111103, 2024.

[39] M. Rivera, J. Torres, and L. Garcia, "Transformer autoencoder for 5G core network security," *IEEE Communications Magazine*, vol. 63, no. 2, pp. 88–95, 2025.

[40] M. Abbas, S. Khan, and A. Alazab, "Adversarially robust autoencoder for intrusion detection," *Computers & Security*, vol. 136, Art. no. 103550, 2025.

[41] R. Nair, P. Menon, and V. Namboodiri, "Continual learning autoencoders for evolving network environments," *Neurocomputing*, vol. 566, pp. 126–138, 2025.

[42] H. Holm, T. Olovsson, and E. Jonsson, "Standardized benchmarking of zero-day intrusion detection systems," *IEEE Access*, vol. 13, pp. 11245–11259, 2025.

[43] U. Farooq, A. Rehman, and S. Hussain, "Causal representation learning using autoencoders for intrusion detection," *Artificial Intelligence Review*, vol. 58, pp. 4123–4145, 2025.

[44] S. Das, P. Roy, and S. Banerjee, "Explainable latent disentanglement in autoencoder-based intrusion detection," *Pattern Recognition*, vol. 147, Art. no. 110075, 2025.

[45] J. Moreno, L. Sánchez, and R. Pérez, "A comprehensive survey of autoencoder-based intrusion detection systems," *ACM Computing Surveys*, vol. 57, no. 1, Art. no. 12, 2025.

[46] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," *IEEE Symposium on Security and Privacy*, pp. 305–316, 2010.

[47] T. Kim, Y. Kim, and H. Kim, "Anomaly detection in network traffic using autoencoder and reconstruction probability," *IEEE Access*, vol. 9, pp. 148104–148116, 2021.

[48] J. P. A. Aljawarneh, M. B. Yassein, and M. Alawneh, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *Journal of Computational Science*, vol. 25, pp. 152–160, 2018.

[49] M. Ring, S. Wunderlich, D. Grüdl, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," *Computers & Security*, vol. 86, pp. 147–167, 2019.

[50] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," *Military Communications and Information Systems Conference*, pp. 1–6, 2015.

[51] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSP*, pp. 108–116, 2018.

[52] F. Erlacher and F. Dressler, "On high-dimensional anomaly detection using autoencoders," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2278–2282, 2020.

[53] M. Conti, T. Dargahi, and A. Dehghantanha, "Cyber threat intelligence: Challenges and opportunities," *IEEE Security & Privacy*, vol. 16, no. 5, pp. 54–61, 2018.

[54] A. H. Lashkari et al., "Characterization of Tor traffic using time-based features," *ICISSP*, pp. 253–262, 2017.

[55] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE TPAMI*, vol. 35, no. 8, pp. 1798–1828, 2013.

[56] G. Pang, C. Shen, L. Cao, and A. van den Hengel, "Deep learning for anomaly detection: A review," *ACM Computing Surveys*, vol. 54, no. 2, Art. no. 38, 2021.

[57] S. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv preprint arXiv:1901.03407*, 2019.

[58] L. Ruff et al., "Deep one-class classification," *International Conference on Machine Learning*, pp. 4393–4402, 2018.

[59] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[60] K. H. Abdulraheem and N. A. Ali, "Evaluation metrics for intrusion detection systems: A survey," *Journal of Information Security*, vol. 12, no. 4, pp. 215–234, 2021.

[61] S. García, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *Computers & Security*, vol. 45, pp. 100–123, 2014.

[62] H. Hindy et al., "A taxonomy of network threats and the effectiveness of machine learning in intrusion detection," *IEEE Access*, vol. 8, pp. 97438–97455, 2020.

[63] A. Zoppi, S. Ceccarelli, and A. Bondavalli, "Evaluation of IDS performance under realistic traffic conditions," *Dependable Systems and Networks*, pp. 1–12, 2018.

[64] S. Marchal et al., "Off-the-shelf machine learning classifiers: A benchmark study," *Computer Communications*, vol. 143–144, pp. 86–101, 2019.

[65] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[66] J. Ma and S. Perkins, "Online novelty detection on temporal sequences," *ACM SIGKDD*, pp. 613–618, 2003.

[67] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *IEEE CISDA*, pp. 1–6, 2009.

[68] Y. Zhou and C. Qian, "A survey on adversarial machine learning in intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 507–532, 2023.

[69] J. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018.

[70] M. A. Ferrag et al., "Security for 5G and IoT networks: A survey," *IEEE Access*, vol. 8, pp. 36100–36136, 2020.

[71] S. Alshamrani et al., "An overview of intrusion detection techniques for IoT," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1190–1213, 2021.

[72] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *International Conference on Learning Representations*, 2014.

[73] I. Goodfellow et al., "Explaining and harnessing adversarial examples," *ICLR*, 2015.

[74] M. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," *ACM SIGKDD*, pp. 1135–1144, 2016.

[75] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *NeurIPS*, pp. 4765–4774, 2017.