International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 518-525
https://doi.org/10.70917/ijcisim-2026-3634
Article

An Adaptive Deep Learning Framework for Zero-Day Cyberattack Detection Using Behavioral Intelligence

Selva Birunda S[1*], Ramana R[2], Vetrivel P[3], Pradeepha S[4], Priyadharshini M[5]

[1,2,3,4,5] Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, India.
Email: [1*]selvabirunda89@gmail.com, [2]ramanaphd2020@gmail.com, [3]vetrilev17@gmail.com, [4]pradeepha@hotmail.com, [5]venkipriya2021@gmail.com
Corresponding Author*: Selva Birunda S

Abstract: The rapid development of networked systems has made the digital infrastructure more susceptible to sophisticated cyber-attacks, especially zero-day attacks, which use unseen security vulnerabilities. Conventional security systems depend largely upon the use of attack signatures, which are ineffective against unseen attacks. To overcome the inefficiency of conventional security systems, this paper proposes a framework for anomaly detection based on the use of deep learning techniques, especially the use of the Auto encoder, which learns the normal behavior of the network instead of learning the patterns of attacks. In the proposed framework, the Auto encoder, which is a type of neural network, is trained using only the normal behavior of the network. The use of the Auto encoder helps the framework to identify anomalies based on the behavior of the network. To improve the accuracy of the framework, the output of the Auto encoder is passed through a Random Forest classifier. To improve the accuracy of the framework, a probability-based threshold optimization strategy is used. Moreover, the use of the Auto encoder is combined with the use of the isolation-based anomaly detection method to improve the accuracy of the framework. In addition, federated learning is integrated to facilitate the training of the model by more than one organization without the need to share the raw network data, thus ensuring the privacy of the users. Moreover, explainable artificial intelligence is utilized, such as the SHAP method, which aids the users in getting the required information regarding the features. This helps the users understand the decision-making process of the anomaly detection method. Experimental results using the UNSW-NB15 dataset prove the efficiency of the proposed framework, where the accuracy is found to be 95.19%, thus ensuring the detection of unseen cyber-attacks.
Keywords: Zero-Day Cyber Attack, Anomaly Detection, Auto encoder, Federated Learning, Ensemble Learning, Adaptive Thresholding, Explainable AI.

1. Introduction:

Increasingly complicated networks have a greater potential for problems caused by zero-day attacks, as signature based intrusion detection systems (IDS) are only able to identify previously established threats [1]. The use of machine learning (ML) and deep learning (DL) supports detection of new attack vectors through learning adaptive methodologies. auto encoder based reconstructive models represent a form of novelty detection through establishing normal network behavior to identify anomalies when viewing an unfamiliar dataset [2][3]. This study using an unsupervised auto encoder processes normal traffic to develop latent representations to train a Random Forest classifier using probability based thresholds for enhanced detection accuracy [3][4].

In addition to developing an auto encoder, isolation forest also works to support reliability and reduces false positives via an ensemble approach based on both the auto encoder and isolation forest concepts [4]. Federated learning provides an approach to share and process code for distributed iterative global model improvements using a secured data format through the protection of individuals from public exposure of their raw data [5][6]. Furthermore, the model uses techniques such as SHAP to provide different aspects of the explanatory features, building trust of the

model's decision [7]. This synched system provides adaptability to newly created attack vectors, privacy protections, and provides effective detection methods for zero-day cyber-attacks [8].

2. Literature Review:

The ongoing difficulty of detecting cyber-attacks on personal data is fueled by the inadequacy of traditional Intrusion Detection Systems (IDS) based on signature patterns to identify unknown threats. Utilizing Machine Learning (ML) and Deep Learning (DL) technologies, paradigms that can adapt to changing attack patterns have emerged. In one instance of this, Rehman et al. [9] have demonstrated that an anomaly detection method comprising the integration of an Auto encoder methodology with an Isolation Forest can provide unsupervised learning models that learn how normal network behavior occurs and how to identify new forms of attack. Federated learning will also be used to develop IDS models without transmitting personal information; for example, Olanrewaju-George and Pranggono [10] have developed a federated Auto encoder-based IDS solution for IOT devices that improves detection accuracy while safeguarding personal data [11]. Numerous research efforts involving DL based anomaly detection, notably those using Auto encoders and Recurrent Neural Networks, have established that these types of neural networks can learn complex behaviors associated with networks and therefore be able to identify previously unknown forms of attack [12]. Combining the Auto encoder with Isolation Forests for anomaly detection would also create a hybrid, or ensemble, method in its own right, providing additional robustness for use within the field of IDS with a potential for lower false alarm occurrence rates [13]. Dynamic ensemble aggregation will increase the effectiveness of federated IDS models through improved detection accuracy with a continued low false positive rate [14].Explainable AI(XAI) models, using SHAP and LIME, provide improved interpretability of models by enabling users of said model to understand why a feature contributed to the flagging of a particular anomalous event, thus developing a level of confidence in the automated decision-making of that model [15][16][17]. However, challenges facing federated IDS include high levels of communication bandwidth usage, variability in devices, and adversarial participants [18].The latest developments include combining CNNs, LSTMs, and Auto encoders into hybrid deep learning frameworks used for enhanced anomaly detection [19]. In addition, there has been research into federated IoT environments to further the effectiveness of collaborative learning while keeping user data private [20]. Most of the previous research done has been partially using one of the following techniques: unsupervised learning, federated learning, Explainability (XAI), and ensembles of these methods. Integration of these methods has not been widely researched yet, thus this work proposes to integrate Auto encoder-based feature learning, classification of Random Forest with optimized thresholds and ensemble anomaly detection using XAI together with federated learning for reliable detection of zero-day attacks while ensuring user privacy is maintained [21].

3. Methods and Methodology:

The proposed methodology describes a structured approach for detecting zero-day cyber-attacks using deep learning–based anomaly detection techniques. The system is designed to learn normal network behavior through an unsupervised learning process and to identify unusual patterns that may indicate previously unseen attacks. The overall workflow includes data pre-processing, Auto encoder-based unsupervised modeling, and adaptive threshold selection for anomaly detection, performance improvement using ensemble methods, result interpretability, and privacy preservation through federated learning.

(a) **Novelty**

This study proposes a behavior-based model for zero-day cyber-attacks through the use of unsupervised deep learning techniques, Random Forest classification with optimal thresholds, ensemble anomaly detection methods, Explainable AI, and federated learning. By recognizing the behavior of the network, the model will identify new or unknown cyber-attacks. An ensemble of the auto encoder and Isolation Forest will provide better accuracy in the model. Explainability will provide better transparency in the model, while federated learning will allow multiple organizations to collaborate.

(b) Data Pre-processing and Feature Engineering

Network traffic data is usually unstructured in nature, consisting of categorical features (like protocol, service, and connection status) and numerical features of different scales. Large-scale numerical features are likely to impact learning, making it hard for a Deep Learning algorithm to identify small-scale features, although they are significant in nature.

For this purpose, categorical features are transformed into numerical features using One-Hot Encoding, and numerical features are scaled using Min-Max Scaling to a range of 0 to 1. Training auto encoder on normal traffic data will help in learning normal behavior, while any abnormality in data, like attacks and zero-day threats, will create unique anomalies.

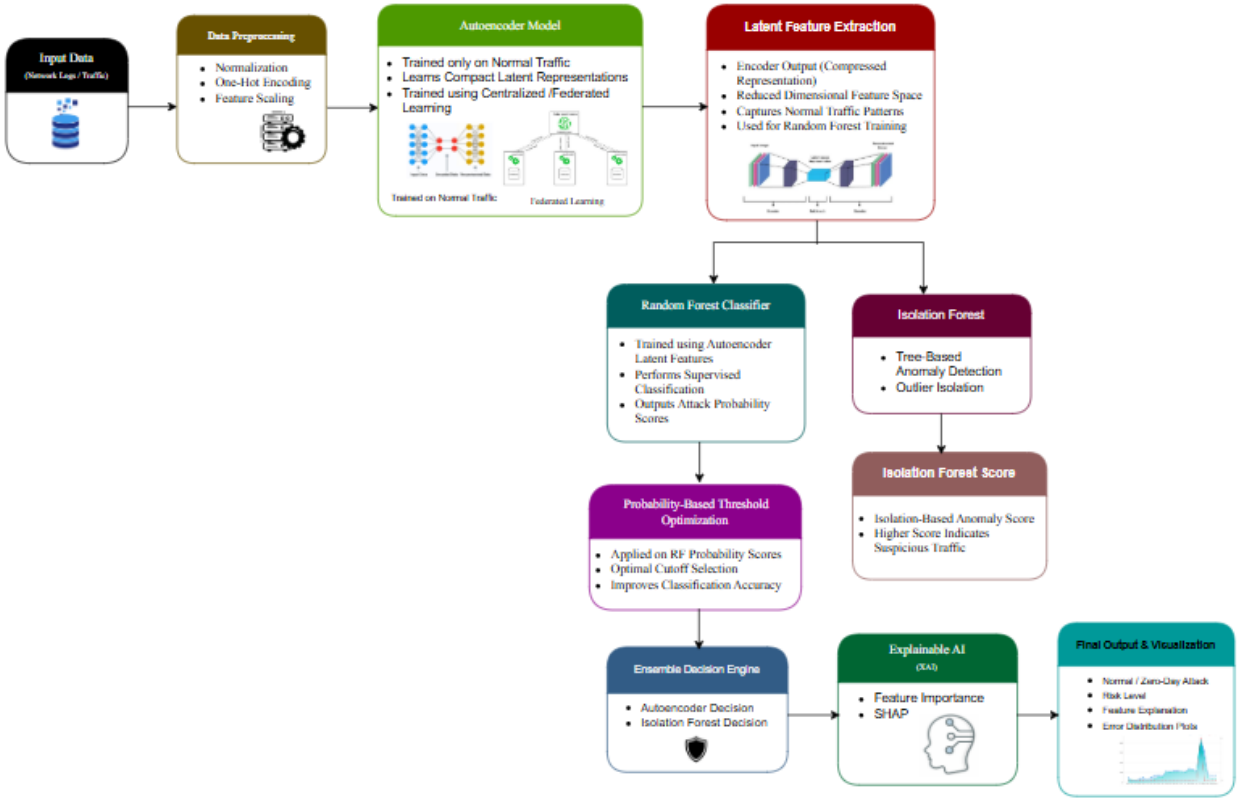

Figure 1 Architecture of the Proposed Anomaly-Based Intrusion Detection Framework

(c) Auto encoder Model for Anomaly Detection

Auto encoder is a neural network that maps input data onto a condensed representation and then reconstructs the input data using a decoder network. The encoder part of the network maps the high-dimensional network traffic data to a condensed representation while preserving the most important information in the condensed representation. The decoder then reconstructs the input data, and the network is trained to reconstruct the input data with minimal reconstruction error using only normal traffic data:

$$\hat{x} = D(E(x)) \qquad (1)$$

Normal patterns are accurately reconstructed by the Auto encoder because it only encounters normal traffic during the training process. However, any new or malicious patterns, like zero-day attacks, produce a high reconstruction error that can be used to identify potential anomalies.

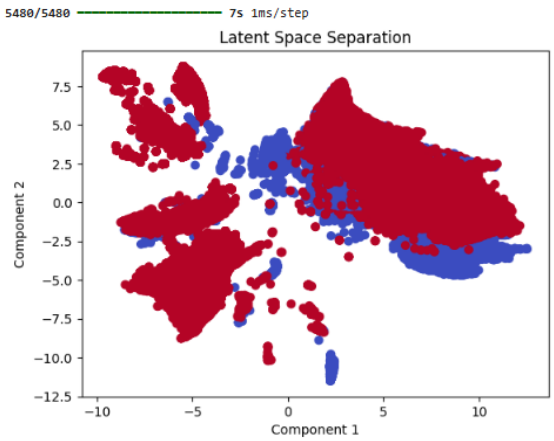

Figure 2: PCA Visualization of auto encoder Latent Space Separation between Normal and Attack Traffic

(d) Reconstruction Error Computation

For testing, each network sample is passed through the trained Auto encoder model. The reconstruction error is computed using Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \qquad (2)$$

If the value is low, it is considered normal traffic. If the value is large, it is considered abnormal or malicious traffic.

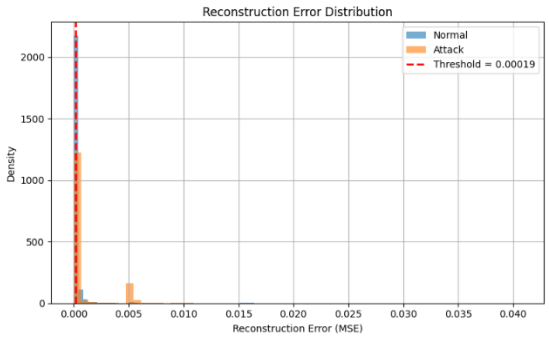

Figure 3: Reconstruction Error Distribution of Normal and Attack Traffic on UNSW-NB15 Dataset

(e) Probability-Based Classification and Threshold Optimization

Latent features obtained from the Auto encoder are then passed to a Random Forest classifier to improve detection accuracy. A probability score is obtained:

$$D(x) = \begin{cases} 1, & \text{if } P\_attack(x) > T \\ 0, & \text{otherwise} \end{cases} \qquad (3)$$

Where P_attack(x) is the probability of an attack, and T is an optimized threshold value. Optimizing the threshold value improves the accuracy of the detection process with fewer false alarm instances.

(f) Ensemble Enhancement Using Isolation Forest

To improve reliability, the Auto encoder is combined with Isolation Forest, which identifies anomalies as statistical outliers via random decision trees. Isolation Forest computes anomaly scores:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \qquad (4)$$

Traffic is flagged as anomalous if either model detects it. This ensemble reduces false negatives and improves detection robustness.

$$D(x) = \begin{cases} 1, & \text{if } AE(x) = 1 \ \vee \ IF(x) = 1 \\ 0, & \text{if } AE(x) = 0 \ \wedge IF(x) = 0 \end{cases} \quad (5)$$

(g) Explainable AI (XAI) Integration

SHAP (Shapley Additive Explanations) is used to understand the reason for the detection of anomalies. A contribution score is assigned to each of the features:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \quad (6)$$

Where positive φ_i represents the contribution of the feature to the detection of the anomaly, and negative φ_i represents the normal behavior.

(h) Federated Learning-Based Distributed Training

Federated learning enables multiple organizations to cooperatively learn the models without the need to share the raw traffic. Each client trains an auto encoder locally and sends the parameters to the central server. The server aggregates the parameters using the Federated Averaging method:

$$w_{t+1} = \sum_{k=1}^{k} \frac{n_k}{N} w_t^k \quad (7)$$

wt+1= k=1knkNwtk

The global model is sent to the clients for further training. This maintains the privacy of the clients while improving the detection of new attacks.

4. Results

The framework was tested using the UNSW-NB15 dataset with 82,332 samples under the training set and 175,341 samples under the test set with 10 categories of attacks. The experimental results verify the effectiveness of the proposed architecture in detecting network intrusion attacks, including zero-day attacks.

Auto encoder was successful in learning the patterns of normal network traffic and creating a set of feature representation with a reduced dimensionality of 32. This was used to classify normal and anomalous network traffic using the Random Forest classifier. The framework was able to achieve an accuracy of 95.19% and a ROC-AUC score of 0.9853, which verifies the discriminative ability of the framework between normal and anomalous network traffic.

The ensemble model was able to achieve a recall of 96.76%, which verifies the ability of the model to correctly classify actual attacks. The federated learning model was able to achieve an accuracy of 88.93% with 3 distributed clients over 5 communication rounds. Figure 6 summarizes the performance comparison across all model components.

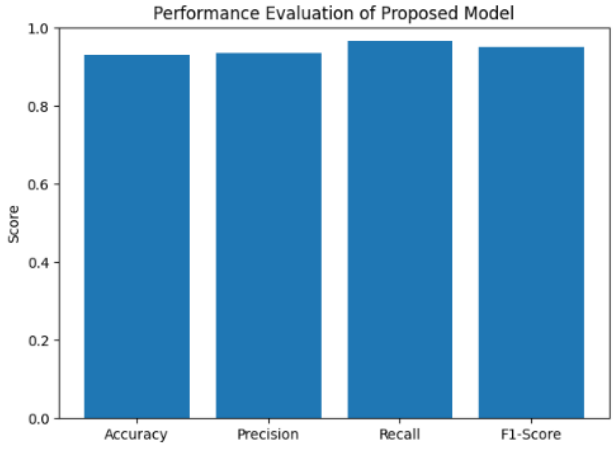

Figure 4: Performance Evaluation of the Proposed Auto encoder-Based Intrusion Detection Framework on UNSW-NB15 Dataset

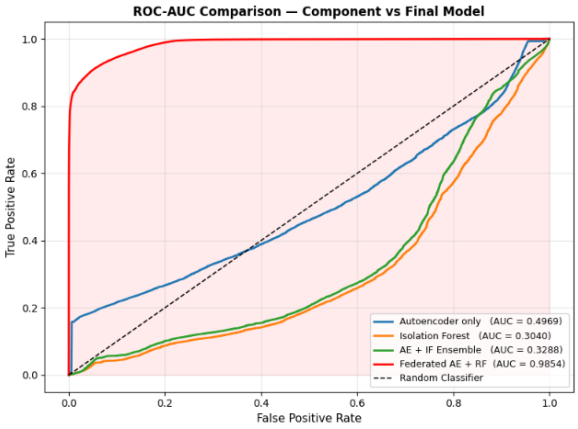

Figure 5: ROC-AUC Curve Comparison of auto encoder, Isolation Forest, Ensemble, and Federated Models

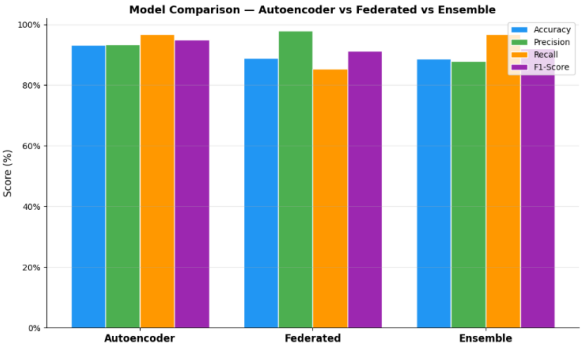

Figure 6: Performance Comparison of Proposed Model Components on the UNSW-NB15 Dataset

5. Discussion

This experiment proves the efficacy of the proposed approach in terms of anomaly-based intrusion detection. This is because the probability-driven classification approach with optimized thresholds improves the detection performance with fewer false positives compared to other approaches like using techniques based on reconstruction errors. Moreover, using the Isolation Forest approach improves the robustness with a high rate of recall at 96.76%, thus reducing false negatives for different types of attacks. Federated learning enabled the model to be trained without sharing raw traffic data. Federated averaging was applied to average the model parameters. Despite the relatively lower accuracy of 88.93%, the proposed model offers a good trade-off between accuracy and privacy preservation.

Explainable AI with the help of SHAP analysis also demonstrated the effectiveness of the presented framework for zero-day attacks. With the help of the presented framework, the issues of zero-day attacks, privacy preservation, false positives, and false negatives were effectively addressed.

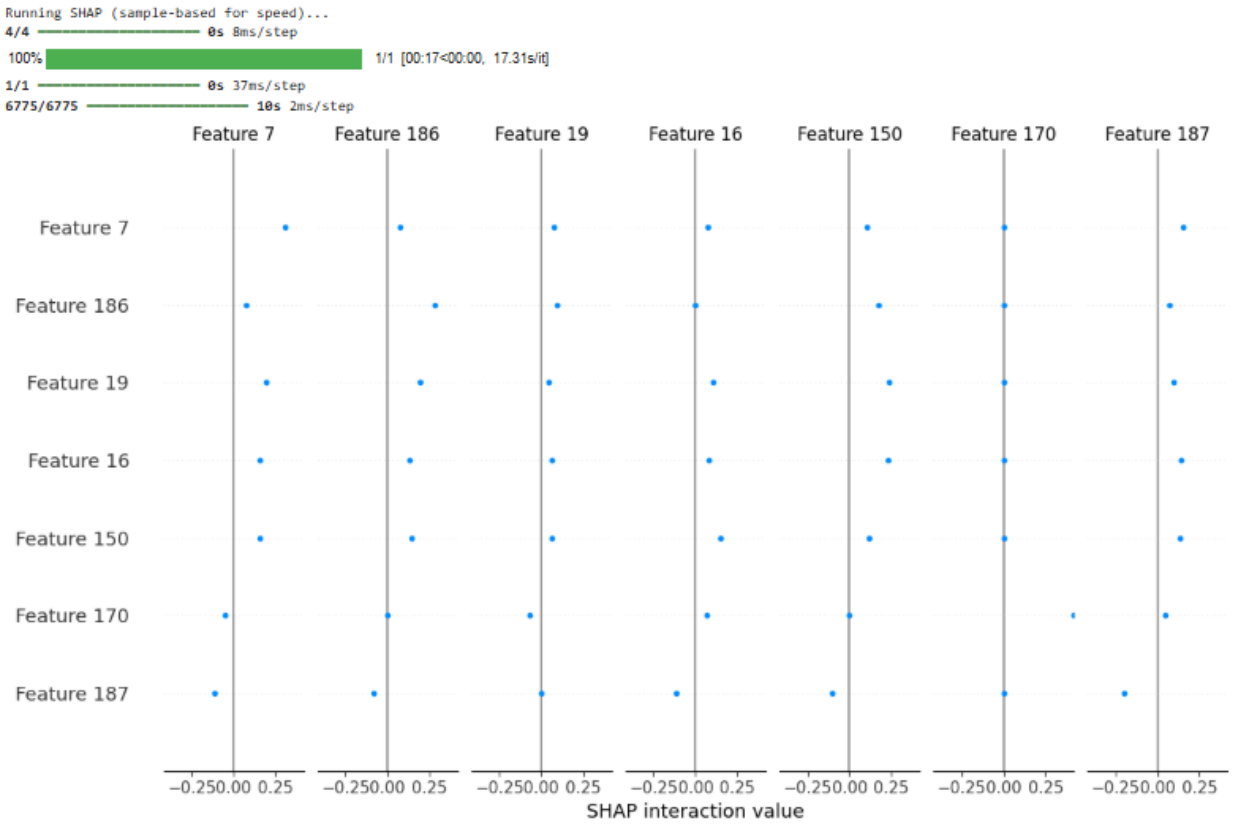

Figure 7: SHAP Feature Importance Analysis for Anomaly Detection on UNSW-NB15 Dataset

6. Conclusion

This project introduced a deep learning-based framework to carry out zero-day cyber-attack detection using Auto encoder-based anomaly detection, Random Forest classification, and ensemble learning with Isolation Forest. Federated learning was used to preserve the privacy of the data during the distributed learning process, while Explainable AI was used to improve the model's interpretability. The experimental results showed that the Auto encoder had the best performance, while the Isolation Forest improved the detection performance by joint usage with the Auto encoder. Federated learning was used to preserve the privacy of the data without affecting the performance of the model. The proposed framework provides a scalable, private, and interpretable intrusion detection system that can solve the problem of zero-day cyber-attacks and can be used as a base to carry out future research in the field.

7. Funding Statement:

The authors did not receive financing for the development of this research

8. Data Availability:

No new data were created in this study.

9. Conflict of interest:

The authors declare that there is no conflict of interest.

References

1. Y. Zhang, R. C. Muniyandi, and F. Qamar, "A review of deep learning applications in intrusion detection systems: overcoming challenges in spatiotemporal feature extraction and data imbalance," *Appl. Sci.*, vol. 15, no. 3,p.1552,2025. https://doi.org/10.3390/app15031552
2. S. Oluwadare and Z. ElSayed, "A survey of unsupervised learning algorithms for zero-day attacks in intrusion detection systems," in *Proc. Int. FLAIRS Conf.*, vol. 36, May 2023. https://doi.org/10.32473/flairs.36.133182

3. L. Segurola-Gil, M. Moreno-Moreno, I. Irigoien, and A. M. Florez-Tapia, “Unsupervised anomaly detection approach for cyberattack identification,” *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 11, pp. 5291–5302, 2024. https://doi.org/10.1007/s13042-024-02237-w
4. S. Sridharan, S. Patil, T. Shobha, and P. Pai, “Hybrid machine learning–based intrusion detection for zero-day attack prevention in digital education networks,” *Int. J. Safety & Security Eng.*, vol. 15, no. 8, pp. —, 2025. https://doi.org/10.18280/ijsse.150815
5. B. Buyuktanir, Ş. Altinkaya, G. Karatas Baydogmus, and K. Yildiz, “Federated learning in intrusion detection: advancements, applications, and future directions,” *Cluster Computing*, vol. 28, no. 7, p. 473, 2025. https://doi.org/10.1007/s10586-025-05325-w
6. T. Ohtani, R. Yamamoto, and S. Ohzahata, “IDAC: Federated learning-based intrusion detection using autonomously extracted anomalies in IoT,” *Sensors*, vol. 24, no. 10, p.3218,2024. https://doi.org/10.3390/s24103218
7. D. Krishnan, S. Singh, and V. Sugumaran, “Explainable AI for zero-day attack detection in IoT networks using attention fusion model,” *Discover Internet of Things*, vol. 5, no. 1, p. 83, 2025. https://doi.org/10.1007/s43926-025-00184-8
8. S. A. Alansary, S. M. Ayyad, F. M. Talaat, and M. M. Saafan, “Emerging AI threats in cybercrime: a review of zero-day attacks via machine, deep, and federated learning,” *Knowledge and Information Systems*, vol. 67, no. 11, pp. 10951–10987, 2025. https://doi.org/10.1007/s10115-025-02556-6
9. M. U. Rehman, M. Zita, M. Abrar, M. Kazim, and S. Khalid, “Zero-day attack detection system using autoencoders and isolation forest: An unsupervised machine learning approach,” in *Proc. Int. Conf. Neural Comput. Adv. Appl.*, Singapore, Jul. 2025, pp. 245–258.
10. B. Olanrewaju-George and B. Pranggono, “Federated learning-based intrusion detection system for the Internet of Things using unsupervised and supervised deep learning models,” *Cyber Secur. Appl.*, vol. 3, p. 100068,2025. https://doi.org/10.1016/j.csa.2024.100068
11. V. T. Nguyen and R. Beuran, “FedMSE: Federated learning for IoT network intrusion detection,” *arXiv preprint*, Oct. 2024. https://doi.org/10.1016/j.cose.2025.104337
12. Z. Xu, Y. Wu, S. Wang, J. Gao, T. Qiu, Z. Wang, H. Wan, and X. Zhao, “Deep Learning-Based Intrusion Detection Systems: A Survey,” arXiv preprint, Apr. 2025. https://doi.org/10.48550/arXiv.2504.07839
13. M. Bachar, A. Khiat, and K. El Guemmat, “Hybrid Autoencoder and Isolation Forest for IoT Anomaly Detection with a Novel Model,” Eng. Technol. Appl. Sci. Res., vol. 16, no. 1, pp. 31123–31129, Feb. 2026, https://doi.org/10.48084/etasr.15288
14. M. Baich and N. Sael, “A Federated Learning-Based Intrusion Detection System Using Dynamic Ensemble Aggregation for IoT Networks,” IEEE Access, vol. 13, pp. 205826–205839, 2025, https://doi.org/10.1109/ACCESS.2025.3640521
15. R. Taheri, R. Jafari, A. Gegov, F. Arabikhan, and A. Ichtev, “Explainable AI for Federated Learning-Based Intrusion Detection Systems in Connected Vehicles,” Electronics, vol. 14, no. 22, art. no. 4508, Nov. 2025, https://doi.org/10.3390/electronics14224508
16. H.-T. Pai, Y.-H. Kang, and W. C. Chung, “An interpretable generalization mechanism for accurately detecting anomaly and identifying networking intrusion techniques,” IEEE Trans. Inf. Forensics Secur., vol. 19, pp. 10302–10313, 2024, https://doi.org/10.1109/TIFS.2024.3488967
17. M. Keshk, N. Koroniotis, N. Pham, N. Moustafa, B. Turnbull, and A. Y. Zomaya, “An explainable deep learning-enabled intrusion detection framework in IoT networks,” Inf. Sci., vol. 639, art. no. 119000, Aug. 2023, https://doi.org/10.1016/j.ins.2023.119000
18. N. Latif, W. Ma, and H. B. Ahmad, “Advancements in securing federated learning with IDS: A comprehensive review of neural networks and feature engineering techniques for malicious client detection,” Artif. Intell. Rev., vol. 58, no. 3, art. no. 91, Jan. 2025, https://doi.org/10.1007/s10462-024-11082-w
19. R. Almuhanna and S. Dardouri, “A deep learning/machine learning approach for anomaly based network intrusion detection,” Front. Artif. Intell., vol. 8, art. no. 1625891, Sep. 2025, https://doi.org/10.3389/frai.2025.1625891
20. N. Albanbay, Y. Tursynbek, K. Graffi, R. Uskenbayeva, Z. Kalpeyeva, Z. Abilkaiyr, and Y. Ayapov, “Federated learning-based intrusion detection in IoT networks: Performance evaluation and data scaling study,” J. Sens. Actuator Netw., vol. 14, no. 4, art. no. 78, Jul. 2025, https://doi.org/10.3390/jsan14040078
21. S. Birunda, M. Kaliappan, and R. Ramana, “An explainable modified convolutional mixer neural network-based deep learning framework for accurate brain tumor detection and classification,” Neurol. Res., pp. 1–16, 2025. https://doi.org/10.1080/01616412.2025.2574357