



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

AI-Driven Cybersecurity: A Comprehensive Review of Deep Learning-Based Threat Detection Techniques

Ms. Nikita Rajurkar, Ms. Mohini Masram, Ms. Kiran Gujar, Ms. Ashwini Kapile

Assistant Professor, Ranibai Agnihotri Institute of Computer Science & Information Technology, Wardha,
Maharashtra, India.

ABSTRACT: The rapid evolution of cyber threats has exposed significant limitations in traditional signature-based and rule-driven security mechanisms. With the exponential growth of network traffic, cloud services, and Internet of Things (IoT) devices, intelligent and adaptive defense mechanisms have become essential. Deep learning has emerged as a transformative technology in cybersecurity, enabling automated feature extraction, high-dimensional pattern recognition, and real-time threat detection. This review critically examines recent advancements in deep learning-based cyber threat detection, covering architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Autoencoders, Generative Adversarial Networks (GANs), Transformers, and Graph Neural Networks (GNNs). The study analyzes widely used benchmark datasets, evaluation metrics, and comparative performance trends across malware detection, intrusion detection, phishing identification, and anomaly detection tasks. Furthermore, key challenges including adversarial attacks, data imbalance, concept drift, explainability, and deployment scalability are discussed.

KEYWORDS: Deep Learning; Cybersecurity; Threat Detection; Intrusion Detection Systems (IDS); Malware Detection; Artificial Intelligence; Adversarial Machine Learning

I. INTRODUCTION

The accelerating digitization of critical infrastructures, financial systems, healthcare platforms, and industrial control environments has dramatically expanded the global cyber-attack surface. Contemporary organizations generate massive volumes of heterogeneous data through cloud computing, Internet of Things (IoT) devices, mobile platforms, and edge networks, creating complex and dynamic threat environments. Traditional cybersecurity mechanisms—primarily signature-based detection systems and rule-driven firewalls—struggle to identify sophisticated, polymorphic, and zero-day attacks. As attackers increasingly employ automation and artificial intelligence (AI) techniques, defensive systems must evolve toward intelligent, adaptive, and autonomous threat detection frameworks (1). Machine learning (ML) introduced data-driven capabilities to intrusion detection systems (IDS), malware classification, and anomaly detection. However, conventional ML approaches heavily depend on handcrafted feature engineering and domain expertise, limiting scalability and adaptability in high-dimensional environments. Deep learning (DL), a subset of ML, addresses these limitations through hierarchical representation learning and automatic feature extraction from raw data. By leveraging multi-layer neural architectures, DL models can capture spatial, temporal, and structural dependencies within cybersecurity datasets, significantly improving detection accuracy and robustness (2).

In cybersecurity applications, Convolutional Neural Networks (CNNs) have demonstrated strong performance in malware image classification and traffic pattern analysis, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks effectively model sequential behaviors in system logs and network flows. Autoencoders have proven valuable for unsupervised anomaly detection, particularly in identifying unknown or zero-day threats. More recently, advanced architectures such as Generative Adversarial Networks (GANs), Graph Neural Networks (GNNs), and Transformer-based models have enabled adversarial training, attack path modeling, and contextual threat intelligence extraction (3,4). Despite these advancements, deploying deep learning in operational cybersecurity environments presents significant challenges. Cybersecurity datasets are often imbalanced, noisy, and non-stationary, leading to concept drift and model degradation over time. Moreover, adversarial machine learning introduces vulnerabilities wherein attackers deliberately craft perturbations to mislead detection models. The black-box

nature of deep neural networks further complicates trust and regulatory compliance, emphasizing the need for explainable and interpretable AI frameworks (5). These issues necessitate comprehensive evaluation beyond accuracy metrics, incorporating robustness, latency, scalability, and computational overhead. Several large-scale benchmark datasets, including UNSW-NB15 and CICIDS2017, have facilitated empirical evaluation of DL-based intrusion detection models. Nevertheless, many studies rely on outdated or synthetic datasets that fail to reflect real-world attack diversity and encrypted traffic conditions. Furthermore, while reported detection rates are often high, cross-dataset generalization remains limited, raising concerns regarding practical deployment viability (6).

Recent surveys have examined AI applications in cybersecurity; however, rapid developments in transformer architectures, graph-based learning, federated learning, and edge AI demand an updated and critical synthesis of the field (7). A structured review is essential to systematically analyze deep learning architectures, datasets, evaluation metrics, and open research challenges while identifying methodological gaps and emerging research trajectories. This review therefore aims to provide a comprehensive and analytical examination of deep learning-based threat detection mechanisms in cybersecurity. The study categorizes existing approaches according to architectural paradigms, evaluates their comparative performance across attack domains, and critically discusses deployment constraints, adversarial robustness, and explainability. By synthesizing current research trends and identifying unresolved challenges, this paper seeks to guide future investigations toward building resilient, scalable, and intelligent cyber defense systems capable of adapting to the evolving threat landscape.

II. LITERATURE REVIEW

Ishtiaq et al., (2025) Recent advancements in deep learning have significantly influenced cybersecurity threat detection mechanisms across various network environments. A prominent trend involves hybrid architectures that leverage multiple neural paradigms to enhance performance and generalization. For instance, Ishtiaq et al. introduced *CST-AFNet*, a dual attention-augmented deep learning framework combining multiscale Convolutional Neural Networks (CNNs) with Bidirectional Gated Recurrent Units (BiGRUs) to extract both spatial and temporal features in IoT network traffic. Evaluated on the comprehensive Edge IoTset dataset, the model demonstrated exceptional detection accuracy and F1-scores above 99%, illustrating its effectiveness for real-time intrusion detection in heterogeneous IoT environments. (8)

Biradar et al., (2025) Hybrid models that integrate graph learning with sequential analysis have also gained traction. Biradar et al. proposed a novel architecture combining Graph Neural Networks (GNNs), Recurrent Neural Networks (RNNs), and multi-head attention to capture both topological dependencies and temporal dynamics present in network traffic. Tested using the UNSW-NB15 benchmark, this hybrid approach delivered enhanced precision and recall, particularly for advanced threats like distributed denial-of-service (DDoS) and zero-day exploits, outperforming standalone deep learning baselines. (9)

Panggabean et al., (2025) Addressing sequence modeling for specific attack types, Panggabean et al. developed a hybrid model combining Gated Recurrent Units (GRUs) with a Neural Turing Machine (NTM) to improve detection of Denial of Service (DoS) and DDoS attacks. Through sequential GRU layers and memory-augmented pattern recognition by the NTM, this approach achieved 99% classification accuracy on both the UNSW-NB15 and BoT-IoT datasets, indicating its strength in capturing long-term dependencies in traffic sequences. (10)

Kareem et al., (2024) Transformers, originally designed for natural language processing, have also been adapted for cybersecurity tasks. Kareem et al. demonstrated a transformer-based model for real-time zero-day threat detection by treating network packet sequences as analogous to language sequences. Results showed significant improvements over traditional classifiers, especially in identifying previously unseen attack patterns within high-dimensional traffic data, underscoring the relevance of attention mechanisms in anomaly detection. (11)

Deep learning-based intrusion detection for IoT networks, (2025) Traditional deep learning models like CNNs and LSTM networks continue to yield robust performance for IoT and network intrusion detection. A deep learning-based intrusion detection system employing 1D CNNs, LSTM, RNN, and multi-layer perceptrons was tested on the CIC IoT-DIAD 2024 dataset, revealing that 1D CNNs consistently achieved the highest accuracy in both anomaly detection and

multi-class classification tasks. This suggests that CNNs' spatial feature extraction capabilities remain highly beneficial for analyzing structured network data in real-time environments. (12)

Exploring structured data representations, **Jin et al. (2005)** developed a hybrid model that integrates Transformers and CNNs for IoT attack detection. Using the CIC-IoT23 dataset, the architecture effectively combined hierarchical feature learning with global contextual attention, producing accuracy of 98.75% and high F1-scores across multiple attack types. The study emphasizes the versatility of transformer-enhanced deep learning systems in multi-class threat detection within IoT frameworks. (13)

Zhang et al. (2025) investigated Graph Convolutional Network (GCN)-based intrusion detection for IoT networks, demonstrating that GNNs can model network traffic topology and structural patterns to distinguish between benign and malicious behaviors. This graph-based method effectively handles class imbalance and captures relational features among devices and traffic flows, suggesting GNNs' suitability for comprehensive IoT security frameworks. (14)

III. DEEP LEARNING ARCHITECTURES FOR THREAT DETECTION

Deep learning architectures have revolutionized cybersecurity threat detection by enabling automated feature extraction, high-dimensional data representation, and adaptive learning from dynamic network environments. Unlike traditional machine learning methods that rely on handcrafted features, deep neural networks can directly process raw traffic flows, system logs, executable binaries, and graph-based network structures. Different architectures are designed to capture specific data characteristics such as spatial correlations, temporal dependencies, contextual semantics, and relational structures. The selection of an appropriate deep learning model depends on the nature of the threat, data modality, computational constraints, and real-time deployment requirements.

1. Convolutional Neural Networks (CNNs): CNNs are highly effective for extracting spatial features from structured cybersecurity data such as malware binaries converted into grayscale images or packet payload representations. Their convolutional filters automatically learn hierarchical feature maps that distinguish malicious patterns from benign behavior. CNNs are widely used in malware detection and traffic classification tasks. However, they are less effective in modeling long-term temporal dependencies in sequential network data.

2. Recurrent Neural Networks (RNNs): RNNs are designed to process sequential data by maintaining hidden states that capture temporal information across time steps. In cybersecurity, RNNs analyze system logs, API call sequences, and network flow records to detect anomalous behavioral patterns. They are suitable for intrusion detection systems (IDS) where attack signatures unfold over time. However, traditional RNNs suffer from vanishing gradient problems in long sequences.

3. Long Short-Term Memory (LSTM) Networks: LSTMs are an advanced form of RNN that address long-term dependency issues through memory cells and gating mechanisms. They are particularly effective in detecting advanced persistent threats (APTs) and multi-stage attacks where temporal correlations are critical. LSTMs outperform basic RNNs in sequential anomaly detection but require higher computational resources.

4. Autoencoders (AEs): Autoencoders are unsupervised neural networks used primarily for anomaly detection. They learn compressed representations of normal network behavior and flag deviations as potential threats. Variants such as Variational Autoencoders (VAEs) enhance probabilistic modeling of anomalies. Autoencoders are effective for zero-day attack detection but may generate high false positives in noisy environments.

5. Generative Adversarial Networks (GANs): GANs consist of a generator and discriminator network competing in a minimax framework. In cybersecurity, GANs are used for adversarial training, synthetic data generation, and improving detection robustness. They help mitigate class imbalance by generating realistic attack samples. However, GAN training is complex and may suffer from instability issues.

6. Transformer-Based Models: Transformers leverage self-attention mechanisms to model long-range dependencies without relying on recurrence. They are highly effective in analyzing large-scale network logs and encrypted traffic

sequences. Their parallel processing capability improves scalability and performance in real-time systems. However, transformers demand significant computational power and large training datasets.

7. Graph Neural Networks (GNNs): GNNs model relational and topological structures within network traffic and device communication graphs. They are well-suited for detecting coordinated attacks, botnets, and lateral movement in enterprise networks. By learning node and edge representations, GNNs capture complex dependencies beyond flat feature vectors. Their main limitation lies in computational overhead for large-scale dynamic graphs.

IV. COMPARATIVE ANALYSIS OF DEEP LEARNING ARCHITECTURES FOR THREAT DETECTION

Below is a structured comparative table evaluating major deep learning architectures used in AI-driven cybersecurity threat detection. The comparison is based on modeling capability, application suitability, computational complexity, and practical deployment considerations.

Table: Comparative Analysis of Deep Learning Models in Cybersecurity

Architecture	Primary Data Type	Key Strengths	Limitations	Best Use Cases	Computational Complexity
CNN	Structured traffic data, malware images	Automatic spatial feature extraction; high accuracy in classification tasks; minimal feature engineering	Poor temporal modeling; limited sequential memory	Malware detection, packet classification	Moderate
RNN	Sequential logs, traffic flows	Captures temporal dependencies; suitable for behavioral modeling	Vanishing gradient problem; unstable for long sequences	Intrusion detection, log analysis	Moderate
LSTM	Long sequential network events	Handles long-term dependencies; robust for multi-stage attacks	Higher training time; resource-intensive	APT detection, anomaly detection	High
Autoencoder	Unlabeled traffic data	Effective for unsupervised anomaly detection; detects zero-day threats	High false positives; sensitive to noise	Anomaly detection, insider threat detection	Low-Moderate
GAN	Imbalanced attack datasets	Synthetic data generation; improves adversarial robustness	Training instability; complex tuning	Data augmentation, adversarial defense	High
Transformer	Large-scale traffic sequences	Captures long-range dependencies; parallel processing; scalable	Requires large datasets; high memory consumption	Encrypted traffic analysis, zero-day detection	Very High
GNN	Network topology graphs	Models relational dependencies; effective for botnet detection	Scalability issues in large dynamic graphs	Botnet detection, lateral movement analysis	High

V. CHALLENGES IN AI-BASED CYBERSECURITY

Although deep learning has significantly improved threat detection accuracy and automation, deploying AI-driven cybersecurity systems in real-world environments remains complex and challenging. Cyber threat landscapes are dynamic, adversarial, and highly imbalanced, which affects model reliability and generalization. Furthermore,

operational constraints such as computational overhead, privacy concerns, and regulatory requirements complicate large-scale deployment. Addressing these challenges is essential to ensure robustness, scalability, and trustworthiness of AI-based cybersecurity frameworks.

1. Adversarial Attacks: Deep learning models are vulnerable to adversarial examples, where attackers intentionally manipulate input data to evade detection. Small perturbations in network traffic or malware binaries can mislead classifiers without noticeable changes to human analysts. This exposes AI systems to evasion, poisoning, and model inversion attacks. Ensuring adversarial robustness requires defensive training strategies and secure model validation mechanisms.

2. Data Imbalance: Cybersecurity datasets are typically imbalanced, with benign traffic vastly outnumbering malicious samples. This leads to biased models that favor majority classes and fail to detect rare but critical attacks. Techniques such as oversampling, synthetic data generation (e.g., GANs), and cost-sensitive learning are used to mitigate this issue. However, imbalance remains a persistent challenge in real-world deployments.

3. Concept Drift: Cyber threats continuously evolve, resulting in distributional shifts in network traffic patterns over time. Models trained on historical data may degrade in performance when exposed to new attack variants. Continuous learning, adaptive retraining, and online learning strategies are necessary to address concept drift. Without adaptation, detection systems become obsolete quickly.

4. Explainability and Interpretability: Deep neural networks operate as black-box systems, making it difficult to interpret their decisions. In cybersecurity, explainability is crucial for incident response, forensic analysis, and regulatory compliance. Lack of transparency reduces trust among security analysts and organizations. Explainable AI (XAI) techniques are being explored but remain computationally intensive and domain-specific.

5. Scalability and Computational Overhead: Advanced architectures such as Transformers and GNNs demand substantial memory and processing power. High-throughput networks generate massive real-time traffic, making low-latency detection challenging. Deploying resource-heavy models in edge or IoT environments further complicates scalability. Efficient model compression and hardware acceleration are required for practical implementation.

6. Privacy and Data Security Concerns: Training deep learning models often requires access to sensitive network logs and user data. Centralized data aggregation increases privacy risks and regulatory challenges. Techniques such as federated learning and privacy-preserving computation are emerging solutions. However, balancing detection accuracy with privacy compliance remains difficult.

7. Lack of High-Quality and Realistic Datasets: Many existing benchmark datasets are outdated or synthetically generated, failing to represent modern encrypted traffic and advanced persistent threats. This limits model generalization and real-world effectiveness. There is a strong need for updated, diverse, and real-world datasets to evaluate AI-based security systems rigorously.

VI. DISCUSSION

The comparative analysis of deep learning architectures and the identified operational challenges reveal that AI-driven cybersecurity is progressing toward highly adaptive and context-aware threat detection systems; however, practical deployment remains constrained by robustness, scalability, and trust issues. While CNNs and LSTMs demonstrate strong performance in malware classification and sequential intrusion detection, their effectiveness is often dataset-dependent and limited by generalization gaps. Emerging architectures such as Transformers and Graph Neural Networks provide superior contextual and relational modeling capabilities, particularly for detecting coordinated attacks and encrypted traffic anomalies, yet they introduce significant computational overhead and infrastructure requirements. A critical observation across recent studies is the over-reliance on benchmark datasets that may not accurately reflect evolving real-world attack patterns. High reported accuracies frequently decline when models are evaluated across different datasets, highlighting issues related to overfitting and limited cross-domain adaptability. Moreover, adversarial machine learning presents a systemic risk, as attackers can exploit model vulnerabilities through evasion and poisoning strategies. This underscores the necessity of incorporating adversarial robustness testing into

standard evaluation protocols. Data imbalance and concept drift further complicate model reliability, especially in dynamic enterprise and IoT environments where attack distributions change over time. Continuous learning frameworks, federated learning paradigms, and edge-based intelligence are emerging as promising solutions to enhance adaptability while preserving privacy. Additionally, explainability remains a central concern; without interpretable decision mechanisms, security analysts may hesitate to trust automated detection outputs, particularly in high-stakes domains such as finance, healthcare, and critical infrastructure. The discussion suggests that future AI-based cybersecurity systems must prioritize robustness, explainability, scalability, and real-world validation over isolated accuracy improvements. Integrating hybrid architectures, privacy-preserving mechanisms, and adaptive learning strategies will be essential to building resilient, autonomous, and trustworthy cyber defense ecosystems capable of countering sophisticated and evolving threat landscapes.

VII.CONCLUSION

Artificial intelligence, particularly deep learning, has emerged as a transformative force in modern cybersecurity by enabling intelligent, adaptive, and automated threat detection mechanisms. Unlike traditional signature-based systems, deep learning models can extract complex spatial, temporal, and relational patterns from high-dimensional network traffic, system logs, and malware artifacts. Architectures such as CNNs, LSTMs, Autoencoders, Transformers, and Graph Neural Networks have demonstrated strong capabilities across intrusion detection, malware classification, anomaly detection, and zero-day attack identification. However, despite promising experimental results, significant challenges remain in terms of adversarial robustness, data imbalance, concept drift, computational scalability, and explainability. The effectiveness of AI-driven cybersecurity solutions depends not only on model accuracy but also on generalization across diverse and evolving real-world environments. Future research must focus on developing resilient, interpretable, and privacy-preserving frameworks capable of continuous adaptation to emerging threats. Hybrid architectures, federated learning, adversarial defense strategies, and edge-based deployment models represent promising directions for enhancing operational feasibility. Ultimately, the integration of robust deep learning methodologies with domain-aware cybersecurity strategies will be critical in building autonomous, scalable, and trustworthy cyber defense systems for next-generation digital infrastructures.

REFERENCES

1. Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2020). On the effectiveness of adversarial examples against machine learning-based intrusion detection systems. *IEEE Transactions on Information Forensics and Security*, 15, 189–203. <https://doi.org/10.1109/TIFS.2019.2928184>
2. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153–1176. <https://doi.org/10.1109/COMST.2015.2494502>
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
4. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
5. 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>
6. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *Proceedings of the International Conference on Information Systems Security and Privacy*, 108–116.
7. Zhang, Y., Chen, X., Li, L., & Zhang, H. (2022). Deep learning in cybersecurity: A comprehensive survey. *IEEE Access*, 10, 22254–22276. <https://doi.org/10.1109/ACCESS.2022.3151234>
8. Biradar, J., Shah, S., & Naik, T. (2025). *Attention augmented GNN RNN-Attention models for advanced cybersecurity intrusion detection*. arXiv.
9. Deep learning-based intrusion detection for IoT networks: a scalable and efficient approach. (2025). *EURASIP Journal on Information Security*.
10. Hybrid deep learning model for attack detection in IoT environment: Convolutional Neural Network with Transformer approach. (2025). *Journal of Visualized Experiments*.

11. Ishtiaq, W., Zannat, A., Parvez, A. H. M. S., Hossain, M. A., Kanchan, M. H., & Tarek, M. M. (2025). *CST-AFNet: A dual attention-based deep learning framework for intrusion detection in IoT networks*. arXiv.
12. Kareem, S. A., Sachan, R. C., Malviya, R. K., & Wadhvani, P. (2024). *Neural transformers for zero-day threat detection in real-time cybersecurity network traffic analysis*. International Journal of Geographical Information Science.
13. Optimizing IoT intrusion detection—A graph neural network approach with attribute-based graph construction. (2025). *Information*.
14. Panggabean, C., Venkatachalam, C., Shah, P., John, S., Devi, R. D. P., & Venkatachalam, S. (2025). *Intelligent DoS and DDoS detection: A hybrid GRU-NTM approach to network security*. arXiv.
15. Almseidin, M., Alzubi, M., Kovacs, S., & Alkasassbeh, M. (2017). Evaluation of machine learning algorithms for intrusion detection system. *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, 000277–000282. <https://doi.org/10.1109/SISY.2017.8080566>
16. Doshi, R., Apthorpe, N., & Feamster, N. (2018). Machine learning DDoS detection for consumer Internet of Things devices. *2018 IEEE Security and Privacy Workshops (SPW)*, 29–35. <https://doi.org/10.1109/SPW.2018.00013>
17. Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications*, 41(4), 1690–1700. <https://doi.org/10.1016/j.eswa.2013.08.066>
18. Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. *Network and Distributed System Security Symposium (NDSS)*. <https://doi.org/10.14722/ndss.2018.23204>
19. Saxe, J., & Berlin, K. (2015). Deep neural network based malware detection using two-dimensional binary program features. *2015 10th International Conference on Malicious and Unwanted Software (MALWARE)*, 11–20. <https://doi.org/10.1109/MALWARE.2015.7413680>
20. Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, 5, 21954–21961. <https://doi.org/10.1109/ACCESS.2017.2762418>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details