A Comprehensive Review of Deep Learning-Based Intrusion Detection Frameworks for High-Precision Threat Detection in Cloud Infrastructure

Dr S Boopalan, Postdoctoral Research Scholar, Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia. sribalulohith@gmail.com
Dr Sudhakar K, Department of AI & DS, Nitte Meenakshi Institute of Technology (NMIT), Nitte (Deemed-to-be University), Bengaluru, Karnataka, India. . ksudhakar.cs@gmail.com

Abstract: This review paper meticulously investigates state-of-the-art DL-driven intrusion detection systems (IDS) directed towards cloud environments, paying attention to convolutional, recurrent, attention-based, hybrid, and transformer architectures. The recent and important trends are feature extraction methods; sequence modelling, mitigation of false-positives, adversarial robustness, and real-time deployment limitations are synthesized. Research gaps in scalability, data imbalance, explainability, and cross-tenant generalization are analyzed, encouraging any future research directions and objectives towards a more robust, interpretable, and privacy-preserving cloud IDS.
Keywords: Cloud Security; Intrusion Detection System; Deep Learning; Anomaly Detection; Transformers; Adversarial Robustness.

Introduction

Due to the heterogeneous nature of cloud computing environments, it becomes difficult for high-fidelity security monitoring implementations. Classical methods of IDS techniques implementation, which are predominantly signature and rule-based, have all failed to accurately detect emerging and zero-day attacks due to their restricted adaptability and dependence on pre-defined threat signatures. The utilization of automated feature learning and feature capturing of complex but temporal and spatial patterns. These features are a consequence of deep learning technology. The patterns are from raw network flows, system logs, and a host of metrics. In addition, Convolutional Neural Networks (CNNs) can extract structural traffic features, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks capture sequential dependencies, while transformer-based models leverage self-attention to model global contextual interactions.

Literature Survey

In extensive and cloud-based infrastructures, deep learning has surfaced as a progressive approach for intrusion detection systems (IDS), especially in large-scale and cloud-based infrastructures. As opposed to traditional signature-based IDS, these deep learning models systematically extract high-feature illustrations from raw traffic data, improving responsiveness. Kimanzi et al. (2024) carried out a review by applying deep learning algorithms to intrusion detection, analyzing CNN, RNN, LSTM, autoencoder, and mixed models. These authors highlight that there is a limitation in the capacity to scale and in real-time deployment in cloud environments as compared to deep neural architectures. [1]. Likewise, an extensive survey by Buczak and Guven (2016) pointed out the drawbacks of data imbalance, which later gave rise to deep learning-based automation, creating fundamental awareness into machine learning for cyber intrusion detection [2].

Currently, in the Artificial Intelligence review (2025), DL-based IDS was progressively examined over developing technologies in cloud security applications and emphasized understandably and adversarial robustness as key research gaps [3]. Based on an extensive review by Vijayakumar et al. (2019), CNN architectures learn these traffic patterns and reduce dependency based on manual feature engineering. Moreover, CNNs, on their own, fail to capture temporal dependencies on attack patterns [4]. Yin et al (2017) focused on an LSTM- based IDS framework capable of learning long-term dependencies in network traffic order. Their study revealed improved detection rates compared to traditional RNNs. However, extended training durations and computational overhead were reported as significant limitations [5].

The authors in Sukhvinder et al (2025) applied a CNN-LSTM hybrid IDS method, which had a better performance and accuracy on the same datasets compared to individual CNN or LSTM models. Hence, hybrid frameworks are better at generalization but not when model complexity increases and latency inferences [6]. In Alshamrani (2024), a CNN-based IDS using the CSE-CICIDS2018 dataset, which is used for cloud environments. The results illustrate enhanced levels of correctness for DDoS and brute-force attacks. There is still an imbalance in class distributions [7]. The research work in Alauthman et al (2026) implements GAN-based synthetic data generation methods to create balance in the minority attack scenario. In addition, constant GAN training instability and model collapse are still the primary setbacks [8].

The work done in Farhan et al (2024) demonstrated improvements in modelling long-term dependencies with network traffic patterns via the implementation of self-attention mechanisms. This technique requires high computational power [9] for better results. In Raza et al (2024), concerns such as privacy in a multi-tenant cloud ecosystem have been handled via the implementation of federated learning-based IDS, which enables combining training across dynamic tenants without the need to share raw data. In [10], the authors discovered that convergence latency and communication overhearing were issues and acted as a limitation. Altogether, the literature illustrates that deep learning significantly enhances intrusion detection in cloud infrastructure through hierarchical feature extraction, temporal modelling, and adaptive learning. Although consistent challenges remain in scalability, adversarial robustness, data imbalance, explainability and deployment efficiency, highlighting a strong need for advanced intelligent frameworks that combine hybrid architectures, federated learning and explainable AI.

Table 1 Summary of key referenced works, their primary architectures, and principal limitations

Reference	**Architecture / Focus**	**Principal Limitation**
Kimanzi et al. [1] (2024)	CNN / LSTM / AE review	Scalability in real-time cloud environments
Buczak & Guven [2] (2016)	ML-IDS survey baseline	Class imbalance; manual feature engineering
Neto, Euclides & Iqbal [3] (2025)	DL-IDS emerging technologies	Adversarial robustness; interpretability
Vinayakumar et al. [4] (2019)	Deep CNN	Temporal dependency blind spot
Yin et al. [5] (2017)	LSTM-based IDS	High training time and inference latency
Sukhvinder et al. [6] (2025)	Hybrid CNN-LSTM	Model complexity vs. latency trade-off
Aljuaid & Alshamrani [7] (2024)	Cloud-specific CNN-IDS	Class imbalance sensitivity

Reference	Architecture / Focus	Principal Limitation
Alauthman, et al. [8] (2026)	GAN-augmented IDS	GAN training instability; mode collapse
Farhan et al [9] (2024)	Transformer-based IDS	Computational overhead at scale
Raza et al. [10] (2024)	Federated DL-IDS	Communication overhead; slow convergence

Research Gaps

Existing studies highlight several ongoing gaps, along with limited scalability of deep models in real-time cloud environments, inadequate handling of rare and zero-day attacks due to dataset inconsistency, weak cross-cloud generalization, vulnerability to adversarial evasion and insufficient support for encrypted traffic analysis. Furthermore, most IDS frameworks function as black box systems, omitting explainable AI mechanisms, leading to a reduction of trust and regulatory compliance. Privacy-distributed learning methods are still insufficiently explored, and the inadequacies of hybrid architectures joining and integrating spatial-temporal modelling, adversarial defence, federated learning, and interpretability into a comprehensive cloud-ready solution.

- **Scalability and Real-Time Processing Constraints**
  Many deep learning–based IDS models struggle to efficiently handle high-volume, real-time traffic in dynamic cloud environments.
- **Class Imbalance and Zero-Day Attack Detection**
  A significant disparity between normal and malicious traffic reduces the system's ability to reliably identify rare and previously unknown threats.
- **Adversarial Vulnerability**
  Deep learning IDS frameworks are susceptible to adversarial manipulation, enabling attackers to evade detection.
- **Lack of Explainability and Interpretability**
  Most models function as black boxes, which lowers analyst confidence and restricts forensic and regulatory compliance.
- **Limited Cross-Environment Generalization**
  Already existing models often fail to generalize different cloud platforms, tenants, and workload distributions.

Problem Statement

Detecting intrusion in cloud infrastructure remains an advanced and unresolved challenge, because cloud environments are dynamic and multi-tenant. While deep learning models such as CNNs, LSTMs, autoencoders and transformers have advanced detection accuracy, recent frameworks still find it challenging to achieve high precision as well as low false-positive rates, real-time scalability, adversarial robustness across cloud platforms. Moreover, significant class imbalance, the growing encrypted traffic, and the limited explainability of deep learning models continue to restrict their effective real-world deployment.

Proposed Solution

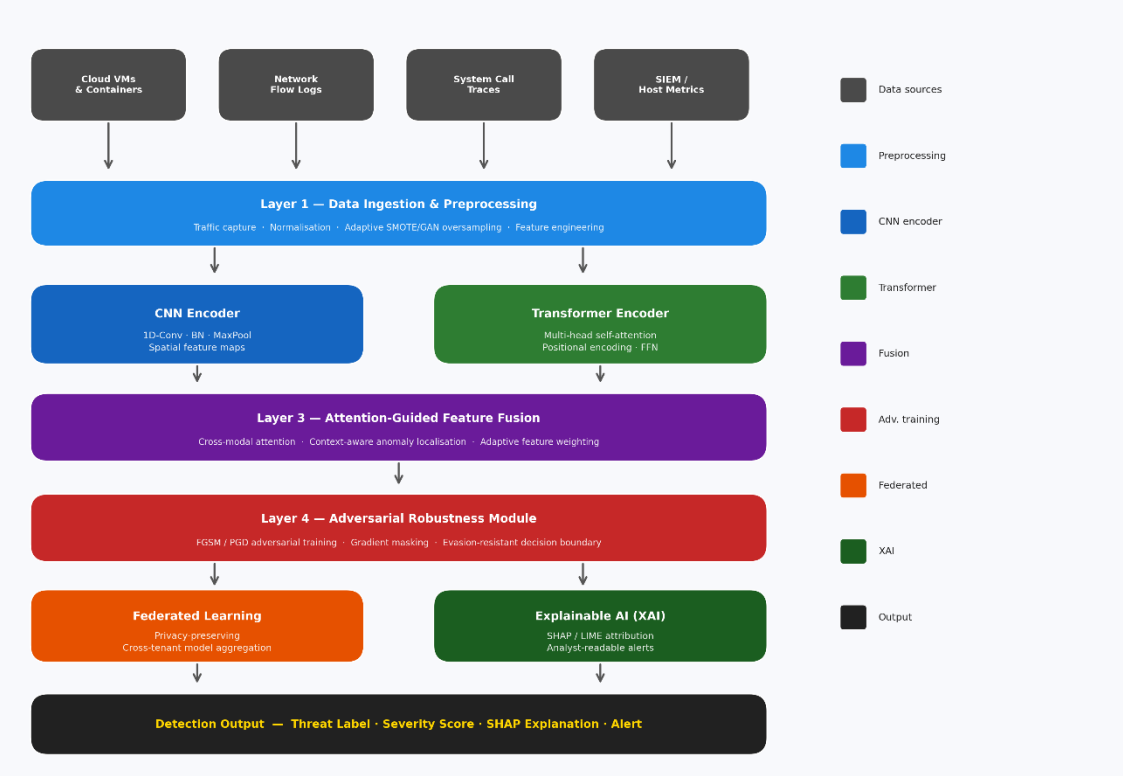

Fig. 1 Proposed hybrid deep learning IDS architecture.

Sophisticated hybrid deep learning-based IDS architectures as shown in fig.1 that are made up of spatial-temporal modelling with the integration of CNN and various transformer frameworks to handle traffic representation attention mechanisms for unpredicted localization, and adversarial training to improve robustness against threats and attacks. These present advantages, such as the reduction of false positives, improve scalability and strengthen security in real-time cloud environments.

Conclusion

The implementation of spatiotemporal modelling and automated feature extraction (Deep Learning techniques) has enhanced the accuracy of intrusion detection in a cloud computing environment. There are still some limitations, which include scalability, class imbalance, adversarial weakness, encrypted traffic handling, and, most importantly, the lack of interpretability, which restricts real-world deployment. The focus for future research should aim at creating scalable and robust as well as explainable deep learning frameworks to ensure high-precision threat detection and strengthened security in modern cloud infrastructures.

References

1. Kimanzi, R., Kimanga, P., Cherori, D., & Gikunda, P.K. (2024). Deep Learning Algorithms Used in Intrusion Detection Systems - A Review. ArXiv, abs/2402.17020.
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.

3. Neto, Euclides & Iqbal, Shahrear & Buffett, Scott & Sultana, Madeena & Taylor, Adrian. (2025). Deep learning for intrusion detection in emerging technologies: a comprehensive survey and new perspectives. Artificial Intelligence Review. 58. 10.1007/s10462-025-11346-z.
4. Vinayakumar, R., Alazab, M., Soman, K. P., et al. (2019). Deep Learning Approach for Intelligent Intrusion Detection System. *IEEE Access*, 7, 41525–41550.
5. 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.
6. Sukhvinder Singh Bamber, Aditya Vardhan Reddy Katkuri, Shubham Sharma, Mohit Angurala, A hybrid CNN-LSTM approach for intelligent cyber intrusion detection system, Computers & Security, Volume 148, 2025, 104146, ISSN 0167-4048,
7. Aljuaid, W. H., & Alshamrani, S. S. (2024). A Deep Learning Approach for Intrusion Detection in Cloud Computing. *Applied Sciences*.
8. Alauthman, M., Aslam, N., Al-Qerem, A. et al. Generative Adversarial Networks for Intrusion Detection Systems: A Comprehensive Survey of Applications, Challenges, and Research Directions. Arab J Sci Eng 51, 179–203 (2026).
9. Farhan Ullah, Shamsher Ullah, Gautam Srivastava, Jerry Chun-Wei Lin, IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic, Digital Communications and Networks, Volume 10, Issue 1, 2024, Pages 190-204, ISSN 2352-8648, https://doi.org/10.1016/j.dcan.2023.03.008.
10. M. Raza, M. Jasim Saeed, M. B. Riaz and M. Awais Sattar, "Federated Learning for Privacy-Preserving Intrusion Detection in Software-Defined Networks," in IEEE Access, vol. 12, pp. 69551-69567, 2024, doi: 10.1109/ACCESS.2024.3395997.