International Journal of Advanced Multidisciplinary Application | IJAMA
Volume 3 Issue 5 May 2026
ISSN No: 3048-9350

Deep Learning-Based Intrusion Detection System for Industrial IoT Networks: A Comparative Evaluation of CNN-LSTM, Transformer, and Autoencoder Architectures on the CIC-IDS-2018 Dataset

Anupam Borthakur, Swapnil Deshmukh
Department of Computer Science and Engineering, Assam Engineering College, Guwahati, Assam, India

Abstract

The proliferation of Internet of Things (IoT) devices in industrial automation, smart manufacturing, and critical infrastructure — collectively termed Industrial IoT (IIoT) — has created an expansive and heterogeneous attack surface that traditional signature-based intrusion detection systems (IDS) are ill-equipped to defend. Machine-to-machine communication protocols (MQTT, OPC-UA, Modbus TCP), legacy supervisory control and data acquisition (SCADA) systems, and resource-constrained sensor nodes operating in deterministic real-time environments present unique cybersecurity challenges that demand anomaly-based detection approaches capable of identifying zero-day attacks and novel attack variants from network traffic patterns alone, without prior knowledge of attack signatures. This paper presents a systematic comparative evaluation of three deep learning architectures — a hybrid Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) model, a multi-head self-attention Transformer encoder, and a Variational Autoencoder (VAE) — for network intrusion detection on the CIC-IDS-2018 benchmark dataset (16 million traffic flows, 14 attack categories). After class-imbalance correction via Synthetic Minority Oversampling Technique (SMOTE), the CNN-LSTM model achieves macro-average F1-score of 0.9724, outperforming the Transformer (0.9681) and VAE (0.9512) on the held-out test set. Critically, the CNN-LSTM demonstrates superior detection of low-volume attacks (Infiltration: F1 = 0.912) that pose the greatest operational risk in IIoT environments. Inference latency analysis confirms that the CNN-LSTM and Transformer models satisfy real-time detection requirements (latency < 2 ms per flow on NVIDIA RTX 3060) while the VAE's reconstruction-based detection introduces unacceptable latency for real-time deployment. The paper concludes with a deployment architecture recommendation for edge-cloud hierarchical IDS in IIoT environments.

Keywords: *Industrial IoT, intrusion detection, deep learning, CNN-LSTM, Transformer, Variational Autoencoder, CIC-IDS-2018, network security, anomaly detection, SMOTE*

1. Introduction

The convergence of operational technology (OT) and information technology (IT) networks in Industry 4.0 manufacturing environments has produced interconnected systems of unprecedented complexity and attack surface breadth. A typical smart factory deploys hundreds to thousands of IIoT endpoints — programmable logic controllers (PLCs), distributed control system (DCS) nodes, industrial robots, vision inspection systems, and environmental monitoring sensors — that communicate over a combination of industrial Ethernet, wireless sensor networks, and cloud-connected gateways. The 2021 Oldsmar Water Treatment Plant intrusion, the 2022 Ukrainian power grid cyberattacks via Industroyer2 malware, and the persistent threat from nation-state actors targeting pharmaceutical and semiconductor manufacturing supply chains underscore the non-theoretical operational consequences of inadequate IIoT network security.

Traditional network intrusion detection systems rely on signature matching — comparing observed traffic against databases of known attack patterns — a fundamentally reactive approach that provides zero protection against previously unseen attack vectors. The rapid evolution of attack techniques, particularly the deployment of polymorphic malware, encrypted command-and-control channels, and low-and-slow reconnaissance strategies explicitly designed to evade signature detection, has rendered signature-based IDS an insufficient primary defence layer for IIoT environments. Anomaly-based intrusion detection, by contrast, constructs a statistical or machine learning model of normal network behaviour and flags deviations as potential intrusions — providing at least theoretical protection against novel attack categories.

Deep learning models have demonstrated strong anomaly detection performance on benchmark network traffic datasets, but the specific requirements of IIoT deployment — real-time processing of high-speed traffic, operation on resource-constrained edge hardware, tolerance for the heterogeneous traffic mix of industrial protocols coexisting with standard TCP/IP, and strict limits on false positive rates that would trigger spurious safety system activations — impose constraints not fully addressed by most published benchmark studies. This paper addresses these gaps through a rigorous comparative evaluation focused on IIoT-relevant performance metrics (detection latency, per-class F1-score for rare attack categories, resource footprint) in addition to the standard macro-average metrics reported in benchmark studies.

The CIC-IDS-2018 dataset (Canadian Institute for Cybersecurity, University of New Brunswick) was selected over the widely-used NSL-KDD dataset due to its generation from a realistic network topology with modern operating systems and applications, its inclusion of contemporary attack categories (Botnet, Infiltration, DoS attacks on Windows 10/Server 2016), and its significantly larger scale (approximately 16 million labelled flow records versus 125,000 in NSL-KDD) that better represents the data volumes encountered in production IIoT monitoring deployments.

2. Related Work and Dataset Description

2.1 Prior Deep Learning IDS Studies

Convolutional neural networks were applied to network intrusion detection by Wang et al. (2017), who transformed network flows to grayscale images and achieved 99.4% accuracy on NSL-KDD — a result widely reproduced but limited by NSL-KDD's dated attack taxonomy. Recurrent architectures (LSTM, GRU) for sequential network traffic modelling were investigated by Yin et al. (2017) and Shone et al. (2018), with the latter proposing a non-symmetric deep autoencoder (NDAE) trained in unsupervised fashion that demonstrated tolerance to label noise. Attention mechanisms, introduced to NLP through the Transformer architecture (Vaswani et al., 2017), have been applied to network traffic classification by Guo et al. (2021), who report superior handling of long-range temporal dependencies in traffic flows compared to LSTM-based approaches. Despite the rapidly growing literature, direct comparison of CNN-LSTM hybrid, Transformer, and Autoencoder architectures on the same IIoT-relevant benchmark with consistent preprocessing remains absent from published studies.

2.2 CIC-IDS-2018 Dataset Characteristics

The CIC-IDS-2018 dataset was generated over ten days of network operation (February 14 to March 2, 2018) in a controlled network environment comprising 500 machines organised into attacker and victim subnets. Traffic was captured on five capture points using CICFlowMeter and labelled with 14 traffic categories: Benign, Botnet, DoS-Hulk, DoS-GoldenEye, DoS-Slowloris, DoS-SlowHTTPTest, DoS-Heartbleed, DDoS-LOIC-HTTP, FTP-Patator, SSH-Patator, Infiltration, SQL Injection, XSS, and Brute Force. Class distribution is severely imbalanced: Benign constitutes 83.2% of all flows, while Infiltration represents only 0.003% — a class imbalance ratio of approximately 27,700:1 that poses fundamental challenges for model training and evaluation.

3. Methodology

3.1 Data Preprocessing

Raw CICFlowMeter-generated CSV files (78 features per flow record) were preprocessed through a six-stage pipeline: (i) removal of rows with NaN or infinite values (2.3% of total records); (ii) dropping constant-valued features (6 features eliminated: Bwd PSH Flags, Fwd URG Flags, Bwd URG Flags, CWE Flag Count, Fwd Avg Bytes/Bulk, Bwd Avg Bytes/Bulk); (iii) removal of highly correlated feature pairs (Pearson $|r| > 0.98$; 11 additional features eliminated); (iv) MinMax normalisation of all remaining 61 features to [0,1]; (v) SMOTE oversampling applied to training split only (70% train, 15% validation, 15% test; stratified split maintaining original class proportions in validation and test sets); and (vi) reshaping for temporal sequence models (CNN-LSTM, Transformer): flows were grouped into sliding windows of 20 consecutive flows from the same source IP, creating sequence tensors of shape (20, 61).

3.2 Model Architectures

The CNN-LSTM hybrid comprises two 1D convolutional layers (64 and 128 filters, kernel size 3, ReLU activation, batch normalisation, max-pooling stride 2) followed by a bidirectional LSTM layer (128 units, dropout 0.3) and a fully connected classification head (Dense 64, ReLU, Dense 14, Softmax). The Transformer encoder employs 4 attention heads, model

dimension d_model = 128, feedforward dimension 256, 3 encoder layers, positional encoding, and global average pooling before the classification head. The VAE uses symmetric encoder-decoder architecture (Dense 256-128-64 for encoder; 64-128-256 for decoder) with 32-dimensional latent space; anomaly scores are computed as reconstruction error and thresholded at the 99th percentile of the training set's error distribution.

3.3 Training Protocol

All models were trained for 50 epochs with early stopping (patience=5 on validation F1-score) using Adam optimiser (learning rate 1×10^{-3}, β_1=0.9, β_2=0.999) and categorical cross-entropy loss. Batch size was set to 512 for CNN-LSTM and Transformer and 1024 for VAE. Training was conducted on NVIDIA RTX 3060 (12 GB VRAM) and Intel Core i9-11900K system. Learning rate scheduling (ReduceLROnPlateau, factor=0.5, patience=3) prevented convergence to local minima. All experiments were replicated three times with different random seeds; results report mean ± standard deviation across replications.

Table 1. Architecture Summary and Training Hyperparameters for the Three Deep Learning Models

Parameter	CNN-LSTM	Transformer	VAE
Total Parameters	2.14M	1.86M	0.98M
Input Shape	(20,61)	(20,61)	(61,)
Optimiser	Adam	Adam	Adam
Learning Rate	1e-3	1e-3	1e-3
Epochs (converged)	38	44	29
Training Time (min)	142	187	63
Inference Latency (ms/flow)	0.84	1.21	4.76

4. Experimental Results

4.1 Overall Classification Performance

Table 2 presents macro-average and per-class precision, recall, and F1-scores for the three models on the held-out test set. The CNN-LSTM achieves the highest macro-average F1 of 0.9724 ± 0.0018, followed by the Transformer (0.9681 ± 0.0024) and VAE (0.9512 ± 0.0041). The VAE's lower and higher-variance performance reflects the fundamental limitation of reconstruction-error thresholding as a multi-class classifier: the single scalar anomaly score cannot distinguish between attack categories, and the per-category assignment depends entirely on the chosen threshold, which is difficult to optimise simultaneously for 13 attack classes with widely different representations.

Infiltration detection — the most operationally critical category given its representation of insider threat and advanced persistent threat (APT) lateral movement — achieves F1 of 0.912 (CNN-LSTM), 0.884 (Transformer), and 0.741 (VAE). The CNN-LSTM's superior Infiltration detection reflects the convolutional layer's ability to extract local temporal patterns in the 20-flow sliding window that are characteristic of multi-stage infiltration behaviour (port scan followed by exploitation followed by exfiltration), patterns that the Transformer captures less efficiently due to its uniform attention weighting across the full window length.

4.2 False Positive Rate Analysis

In IIoT deployment, false positives — classifying benign traffic as an attack — trigger automated response actions (session blocking, safety system alerts, operator notifications) that impose operational costs disproportionate to the brief latency of the attack detection benefit. The false positive rate (FPR) on the Benign class is therefore a critical deployment metric. CNN-LSTM achieves FPR = 0.0034 (0.34%), Transformer achieves FPR = 0.0041, and VAE achieves FPR = 0.0098 on the held-out

test set. All three models satisfy the commonly cited 1% FPR threshold for operational IDS deployment, though the VAE's FPR is closest to this boundary and would be expected to exceed it on real-world traffic with distribution shift from the training set.

Table 2. Per-Class F1-Scores for CNN-LSTM and Transformer on CIC-IDS-2018 Test Set

Attack Category	CNN-LSTM F1	Transformer F1	VAE F1
Benign	0.9981	0.9976	0.9906
Botnet	0.9912	0.9888	0.9714
DoS-Hulk	0.9956	0.9943	0.9821
DoS-GoldenEye	0.9923	0.9901	0.9784
DoS-Slowloris	0.9874	0.9834	0.9641
FTP-Patator	0.9988	0.9981	0.9812
SSH-Patator	0.9974	0.9968	0.9798
Infiltration	0.9120	0.8840	0.7410
SQL Injection	0.9834	0.9812	0.9621
XSS	0.9741	0.9698	0.9412
Macro Average	0.9724	0.9681	0.9512

4.3 Computational Resource Analysis

Inference latency per flow (Table 1) reveals a critical distinction between the CNN-LSTM (0.84 ms) and Transformer (1.21 ms) versus the VAE (4.76 ms). At a typical IIoT gateway traffic rate of 50,000 flows/min, the CNN-LSTM and Transformer can process flows in real-time on NVIDIA RTX 3060-class GPU hardware (theoretical throughput: 71,400 flows/min for CNN-LSTM). The VAE's 4.76 ms latency limits throughput to approximately 12,600 flows/min — insufficient for real-time processing of all traffic at the gateway level. However, the VAE's parameter count (0.98M versus 2.14M for CNN-LSTM) makes it attractive for edge microcontroller deployment where the latency constraint is relaxed (batch processing of archived traffic rather than real-time inline analysis).

5. Discussion and Recommended Deployment Architecture

The CNN-LSTM model's consistent outperformance across macro-average F1, Infiltration detection F1, and false positive rate metrics positions it as the primary recommendation for IIoT network intrusion detection. However, the deployment architecture for production environments must address several challenges not captured by benchmark dataset evaluation: concept drift (the gradual shift in network traffic distributions as IIoT device inventory and protocols evolve over time), adversarial attacks targeting the IDS itself (traffic crafted to exploit classifier blind spots), and the heterogeneity of real IIoT networks compared to the controlled CIC-IDS-2018 capture environment.

A hierarchical edge-cloud deployment architecture is proposed: at the edge gateway level, a compressed and quantised CNN-LSTM model (INT8 quantisation reducing model size by 4× with < 1% F1-score degradation in preliminary experiments) processes all ingress/egress flows in real-time, flagging high-confidence attack classifications for automated response and uncertain samples for cloud-level analysis. The cloud layer maintains the full-precision CNN-LSTM and periodically retrains on labelled traffic samples from edge deployments, implementing federated learning to enable model updating without centralising sensitive operational data. Concept drift detection using ADWIN (Adaptive Windowing) monitors distribution shift in feature statistics and triggers retraining when drift is detected.

6. Conclusion

This paper has presented a rigorous comparative evaluation of CNN-LSTM, Transformer, and Variational Autoencoder architectures for network intrusion detection in Industrial IoT environments, evaluated on the CIC-IDS-2018 benchmark dataset with careful attention to class imbalance correction, per-category performance analysis, and inference latency measurement. The CNN-LSTM hybrid architecture achieves the best overall performance (macro-F1 = 0.9724, FPR = 0.34%) with inference latency (0.84 ms/flow) compatible with real-time deployment at IIoT gateway nodes. Critical operational findings include: the VAE's unsuitability for real-time inline deployment due to latency, the superior Infiltration detection of the CNN-LSTM reflecting its local temporal pattern extraction capability, and the 1% FPR boundary achievability by all architectures under benchmark conditions. Future work should evaluate robustness to adversarial traffic crafted specifically to evade these detection models and extend evaluation to industrial protocol traffic (Modbus, DNP3, IEC 61850) datasets more representative of pure OT environments.

References

[1] Guo, L., Wu, Q., Liu, S., Duan, M., Li, H., & Sun, J. (2021). Deep learning-based real-time anomaly detection for industrial IoT. IEEE Internet of Things Journal, 8(8), 6926-6934.

[2] Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. Proceedings of NDSS Symposium 2018.

[3] Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. IEEE Transactions on Emerging Topics in Computational Intelligence, 2(1), 41-50.

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[5] Wang, W., Zhu, M., Zeng, X., Ye, X., & Sheng, Y. (2017). Malware traffic classification using convolutional neural network for representation learning. Proceedings of ICOIN 2017, 712-717.

[6] 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.

[7] Zhang, Y., Chen, X., Jin, L., Wang, X., & Guo, D. (2019). Network intrusion detection: Based on deep hierarchical network and original flow data. IEEE Access, 7, 37004-37016.

[8] Sharafaldin, I., Habibi Lashkari, A., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proceedings of ICISSP 2018, 108-116.

[9] Pahl, M. O., & Aubet, F. X. (2018). All eyes on you: Distributed multi-dimensional IoT microservice anomaly detection. Proceedings of IFIP/IEEE NOMS 2018, 1-9.

[10] Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761-768.

[11] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.