2026 International Nuclear Atlantic Conference – INAC 2026
Escola de Guerra Naval, Rio de Janeiro, RJ, August 24 - 28, 2026

ISBN: 978-65-272-2398-6

Towards Deep Learning-Driven Cybersecurity in Nuclear Instrumentation and Control Systems

J. C. Garcia[1], F. A. Renteria del Toro[2], A. M. Silva[3], A. J. G. Rosa[4], D. R. Araoz Prado[5], E. F. Luca[6], G. V. Santa Cruz Mérida[7], J. I. Otomo[8], O. V. Bustillos[9], D. A. Andrade[10], and E. A. Rodrigues[11]

julia@cpic.com.br[1]; florenciaren@gmail.com[2]; alan@cpic.com.br[3]; aldo@cpic.com.br[4]; quwaraozpradodr@gmail.com[5]; efluca@gmail.com[6]; santacruz.viviana@usp.br[7]; julianaikebe@gmail.com[8]; oscar@ipen.br[9]; delvonei@ipen.br[10]; elainearodrigues@usp.br[11]

1. Introduction

The integration of emerging technologies has accelerated the global energy transition, positioning artificial intelligence (AI) as a central element in the digitalisation of the nuclear sector [1]. This modernisation has transformed power plants into cyber-physical systems, in which digital Instrumentation and Control (I&C) networks deliver substantial gains in operational precision, remote monitoring, and predictive capability — while demanding institutional and organisational adaptation to keep these systems safe in safety-critical environments. Advances in machine learning and autonomous systems are already opening new frontiers in diagnostics and decision support, progressively raising operational standards across the sector [2].

However, this highly interconnected environment expands the attack surface available to malicious actors: the very tools that enhance operational efficiency also broaden exposure to cyber threats. In nuclear facilities, where a systemic failure may carry severe radiological consequences, data integrity and the security of control networks are paramount.

Incorporating AI into digital I&C systems adds further challenges of model reliability, algorithmic transparency, and the detection of increasingly sophisticated threats. Against these constraints, deep learning has emerged as a promising avenue for strengthening cybersecurity in nuclear environments [3, 4]. Accordingly, this study examines deep learning approaches for intrusion and anomaly detection in nuclear I&C networks, characterising dominant architectures and performance benchmarks, and identifying the technical, regulatory, and certification barriers that constrain deployment in licensed nuclear facilities.

2. Methodology

A search was conducted in March 2026 across the Web of Science and Scopus repositories, employing a Boolean string combining the following key descriptors: ("nuclear power plant*" OR "nuclear facilit*" OR "nuclear I&C" OR "nuclear instrumentation and control" OR "nuclear reactor") AND ("cybersecurity" OR "intrusion detection" OR "anomaly detection" OR "cyberattack*") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural network*" OR "CNN" OR "LSTM" OR "federated learning").

A total of 92 records were retrieved, comprising peer-reviewed journal articles, conference proceedings, and technical publications. Titles and abstracts were screened to identify each study's application domain (I&C/CPS-IDS), AI/deep-learning technique, experimental setting (simulators, testbeds), reported performance and limitations, and cybersecurity and V&V considerations for AI components in safety-critical systems. Studies focused exclusively on non-civilian applications, or on domains without

direct relevance to cybersecurity, were excluded. The remaining studies were analysed to characterise the dominant architectures, consolidate performance benchmarks, and map the technical, regulatory, and certification barriers to deployment in licensed nuclear facilities.

3. Results and Discussion

The evidence base for AI-based intrusion detection targeting nuclear I&C networks remains far more limited than the broader Industrial Control Systems (ICS) security literature might suggest: a systematic review identified only 23 peer-reviewed studies explicitly addressing AI-based cybersecurity in nuclear power plants [5], underscoring the nascent state of the fiel. Among architectures evaluated in nuclear-relevant settings, Bidirectional Long Short-Term Memory (Bi-LSTM) networks have detected simulated cyberattacks before plant protection systems are activated, with post hoc explainability identified as essential for regulatory acceptance [6]. This interpretability requirement is not merely technical: nuclear regulatory frameworks require operators to articulate the causal chain from sensor observation to safety-critical decision, a condition opaque black-box models are structurally unable to meet [7].

Hybrid and physics-informed architectures have emerged as particularly relevant responses to the dual demand of detection performance and physical consistency. Physics-informed Generative Adversarial Networks (GANs) embedding reactor-relevant constraints have outperformed standard GAN baselines in anomaly detection, for both automated sensor data and manually collected surveillance data [8]. An orthogonal, model-independent approach based on dynamic noise analysis of I&C signals from an all-digital research reactor has been proposed as an alternative that avoids the adversarial vulnerabilities of trained classifiers and provides deterministic interpretability meeting nuclear safety certification standards [9]. In ICS settings directly relevant to the nuclear sector, hybrid CNN-LSTM architectures exploiting the spatial and temporal features of process data have achieved high classification accuracy across multiple benchmark datasets [10, 11], while lighter convolutional models with dimensionality reduction show that strong detection need not entail prohibitive computational overhead [12, 13].

Raw process data cannot be centralised across nuclear facilities, owing to national security classifications, commercial confidentiality, and regulatory constraints, which is itself a fundamental barrier to conventional model training. Federated deep learning schemes that combine CNN-GRU intrusion-detection models with cryptographically secured parameter exchange have been proposed as privacy-preserving alternatives for distributed industrial environments [14], and are conceptually transferable to multi-site nuclear fleets. GAN-based data augmentation has been explored to address the severe class imbalance of operational traffic, in which attack instances form only a small fraction of recorded data [15, 16]. The adversarial fragility of trained classifiers, however, remains one of the most acute obstacles to deployment: constrained perturbations compliant with ICS communication protocols can evade ML-based detectors [17, 18], raising serious questions about the robustness of learning-based defences in safety-critical environments.

Taken together, the reviewed evidence indicates that deep learning for intrusion and anomaly detection in nuclear I&C networks is still at an early stage of operational maturity. Promising detection performance has been reported in simulated and ICS testbed environments, but deployment in licensed operating facilities is rare, largely because of limited access to real process and network-traffic data for training and validation. The absence of consolidated standards for the verification and validation (V&V) of AI-based safety components [7, 19] - compounded by unresolved adversarial-robustness concerns and the limited interpretability of the highest-performing architectures - is the principal barrier to near-term integration into the protection systems of operating plants. Addressing these gaps in trust, standardisation, and regulatory guidance is a prerequisite for the responsible advancement of AI-driven cybersecurity in the nuclear sector.

4. Conclusions

This study characterised the dominant deep learning architectures proposed for intrusion and anomaly detection in nuclear I&C networks, consolidated available empirical performance benchmarks, and

mapped the principal barriers to operational deployment in licensed facilities. Whilst Bi-LSTM, hybrid CNN-LSTM, physics-informed GANs, and federated learning schemes have demonstrated promising detection performance in simulated and ICS testbed environments, deployment in operating nuclear plants remains rare. Three interrelated barriers account for this gap: the scarcity of real nuclear process and network traffic data for model training and validation; the adversarial fragility of trained classifiers under evasion attacks compliant with ICS communication constraints; and the absence of consolidated regulatory frameworks defining V&V and certification requirements for AI-based safety components in licensed facilities.

Progress on these fronts will require sustained collaboration among research institutions, plant operators, and international and national regulatory bodies, including the IAEA and national regulatory authorities, to establish reliable cybersecurity benchmarks, develop standardised V&V methodologies for AI components in nuclear I&C systems, and enable the protected exchange of threat intelligence across licensed facilities. Only through such coordinated international effort can the demonstrable potential of deep learning be responsibly translated into strengthened cybersecurity for the nuclear sector, without compromising the rigorous safety standards upon which public trust in nuclear energy depends.

Acknowledgements

The authors are grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico [National Council for Scientific and Technological Development] (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil [Brazilian Federal Agency for Support and Evaluation of Graduate Education] (CAPES) for their support. We also gratefully acknowledge the institutions to which our team members are affiliated for their infrastructure and continuous support throughout this work: the Nuclear and Energy Research Institute (IPEN-CNEN), C&P Engineering and Automation Ltd., the Environmental Research Institutes (IPA), and the Research and Development Council - CNEN.

References

[1] IAEA, "Considerations for Deploying Artificial Intelligence Applications in the Nuclear Power Industry," *IAEA Nuclear Energy Series No. NR-T-1.26*, IAEA, Vienna (2025).

[2] B. Blakely, Y. Li, A. Dave, D. Kultgen, R. Vilim, "AI-Driven Cybersecurity Testbed for Nuclear Infrastructure: Comprehensive Evaluation Using METL Operational Data," *arXiv preprint* (2025).

[3] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–444 (2015).

[4] M. Macas, C. Wu, "Review: Deep Learning Methods for Cybersecurity and Intrusion Detection Systems," *arXiv preprint*, arXiv:2012.02891 (2020).

[5] M. Lezzi, L. Martino, E. Damiani, C. Y. Yeun, "A Systematic Literature Review on AI-Based Cybersecurity in Nuclear Power Plants," *Journal of Cybersecurity and Privacy*, vol. 5, no. 4, p. 79 (2025).

[6] A. Chaudhary, J. Han, S. Kim, A. Kim, S. Choi, "Anomaly Detection and Analysis in Nuclear Power Plants," *Electronics*, vol. 13, no. 22, p. 4428 (2024).

[7] U.S. Nuclear Regulatory Commission (NRC), "Artificial Intelligence Strategic Plan: Fiscal Years 2023–2027," NRC, Washington, DC (2023).

[8] E. Gursel *et al.*, "SPIDARman: System-Level Physics-Informed Detection of Anomalies in Reactor Collected Data Considering Human Errors," *Nuclear Technology*, vol. 210, no. 12, pp. 2299–2311 (2024).

[9] K. Gkouliaras, V. Theos, Z. Dahm, W. Richards, K. Vasili, S. Chatzidakis, "False Data Injection Detection in Nuclear Systems Using Dynamic Noise Analysis," *IEEE Access*, vol. 12, pp. 94936–94949 (2024).

[10] H. C. Altunay, Z. Albayrak, "A Hybrid CNN+LSTM-Based Intrusion Detection System for Industrial IoT Networks," *Engineering Science and Technology, an International Journal*, vol. 38, p. 101322 (2023).

[11] C. Eddine Ben Ncir, M. A. Ben HajKacem, M. Alattas, "Enhancing Intrusion Detection Performance Using Explainable Ensemble Deep Learning," *PeerJ Computer Science*, vol. 10, e2289 (2024).

[12] U. K. Lilhore *et al.*, "HIDM: Hybrid Intrusion Detection Model for Industry 4.0 Networks Using an Optimized CNN-LSTM with Transfer Learning," *Sensors*, vol. 23, no. 18, p. 7856 (2023).

[13] P. Sinha *et al.*, "A High Performance Hybrid LSTM CNN Secure Architecture for IoT Environments Using Deep Learning," *Scientific Reports*, vol. 15, p. 9786 (2025).

[14] B. Li, Y. Wu, J. Song, R. Lu, T. Li, L. Zhao, "DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5615–5624 (2021).

[15] J. Li, W. Zong, Y.-W. Chow, W. Susilo, "Mitigating Class Imbalance in Network Intrusion Detection with Feature-Regularized GANs," *Future Internet*, vol. 17, no. 5, p. 216 (2025).

[16] Y. N. Rao, K. S. Babu, "An Imbalanced Generative Adversarial Network-Based Approach for Network Intrusion Detection in an Imbalanced Dataset," *Sensors*, vol. 23, no. 1, p. 550 (2023).

[17] J. Vella, C. Colombo, "Adversarial Attacks on Time-Series Intrusion Detection for Industrial Control Systems," in *Proc. IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, Guangzhou, China, pp. 1102–1110 (2020).

[18] E. Anthi, L. Williams, M. Rhode, P. Burnap, A. Wedgbury, "Adversarial Attacks on Machine Learning Cybersecurity Defences in Industrial Control Systems," *Journal of Information Security and Applications*, vol. 58, p. 102717 (2021).

[19] IAEA, "Computer Security of Instrumentation and Control Systems at Nuclear Facilities," *IAEA Nuclear Security Series No. 33-T*, IAEA, Vienna (2018).