A first approach for Explainable Anomaly Detection in Smart Manufacturing through Deep Reinforcement Learning and Lightweight Federated Learning

Viet Hieu TRAN[1,2,*], Kim Duc TRAN[1,2], Sébastien Thomassey[1], Kim Phuc TRAN[1]

[1] *Univ. Lille, ENSAIT, ULR 2461 - GEMTEX - Génie et Matériaux Textiles, F-59000 Lille, France*
[2] *International Research Institute for Artificial Intelligence and Data Science, Dong A University, Danang, Vietnam*
* *viet-hieu.tran@ensait.fr, hieutv@donga.edu.vn*

Keywords: *Smart Manufacturing, Anomaly Detection, Federated Learning, Reinforcement Learning, XAI*

In smart manufacturing, anomaly detection plays a crucial role in maintaining product quality, optimizing operations, and minimizing unexpected downtime. Smart manufacturing systems face increasing challenges in detecting and interpreting anomalies that can affect production quality and system reliability. One major limitation is the lack of adaptability to dynamic and constantly changing production environments, where fixed detection models may quickly become obsolete. Traditional approaches often struggle with low accuracy when identifying complex or subtle anomaly patterns, particularly in high-dimensional and sparse industrial data. Moreover, they typically function as “black boxes”, offering little to no explainability, which makes it difficult for engineers and operators to trust and act upon the system’s outputs [1]. Another significant issue is the reliance on centralized data collection, which not only raises data privacy concerns but also increases the burden of data transmission in large-scale, distributed factory settings. Many traditional models are computationally intensive, making them unsuitable for edge devices with limited processing power and energy resources. Additionally, these systems often lack support for real-time feedback and continuous learning, preventing them from adapting to new anomaly types or shifts in operational patterns. Addressing these challenges requires more intelligent, adaptive, and explainable solutions that integrate decentralized learning, real-time adaptation, and transparent decision-making. In this context, artificial intelligence (AI) has emerged as a powerful technology for the implementation of smart manufacturing [2], and more especially for anomaly detection [1]. In this context, Deep Reinforcement Learning (DRL) offers the ability to continuously learn from the manufacturing environment, enhancing the accuracy in detecting complex anomaly patterns and adapting to changing conditions [3]. Meanwhile, lightweight federated learning (LFL) enables distributed model training across multiple edge devices in factories without centralizing sensitive data, ensuring security and data transmission efficiency [4].

This research proposes an explainable anomaly detection framework to address the limitations found in traditional anomaly detection methods. This framework not only focuses on identifying anomalies but also provides clear explanations for the causes of these anomalies. This helps operators and engineers understand the issues in the manufacturing process, allowing them to make accurate and timely decisions to improve performance and product quality. By enabling decentralized model training across distributed manufacturing nodes, federated learning enhances privacy by keeping sensitive data local and reduces data transfer needs. The framework incorporates deep reinforcement learning (DRL) for anomaly detection, allowing the system to dynamically learn and adapt to complex manufacturing conditions. DRL optimizes anomaly detection by continuously updating detection strategies based on real-time feedback from production environments. This enables the system to improve accuracy by autonomously adjusting to variations in production flows, machine states, and operational patterns. This enhances the framework's resilience to evolving anomalies and new operational contexts. To support efficient, low-latency processing on edge devices, quantization techniques are employed to reduce the computational load of deep learning models without compromising accuracy. Furthermore, Field Programmable Gate Arrays (FPGAs) are utilized to accelerate DRL operations, providing a flexible, low-power hardware solution ideal for real-time applications in resource-constrained environments [5]. The framework also integrates explainable artificial intelligence (XAI) methods, ensuring transparency by enabling operators to understand the root causes of detected anomalies and make data-driven adjustments. The study evaluates the framework’s performance, adaptability, and interpretability through extensive testing in simulated and real-world smart factory environments, with a focus on optimized anomaly detection, resource efficiency, and system responsiveness. The findings are expected to contribute to smart factory optimization by delivering a robust, privacy-preserving, and interpretable anomaly detection solution, bridging the gap between advanced AI and practical deployment in Industry 5.0.

References:

[1] Elía, I. & Pagola, M. (2025). Anomaly detection in the smart manufacturing era: A review. Engineering Applications of Artificial Intelligence, 139, 109578. Elsevier.
[2] Tran, K. P. (2021). Artificial intelligence for smart manufacturing: methods and applications. Sensors, 21(16), 5584. MDPI.
[3] Li, C., Zheng, P., Yin, Y., Wang, B., & Wang, L. (2023). Deep reinforcement learning in smart manufacturing: A review and prospects. CIRP Journal of Manufacturing Science and Technology, 40, 75–101. Elsevier.
[4] Qi, P., Chiaro, D., & Piccialli, F. (2024). Small models, big impact: A review on the power of lightweight Federated Learning. Future Generation Computer Systems, 107484. Elsevier.
[5] Seng, K. P., Lee, P. J., & Ang, L. M. (2021). Embedded intelligence on FPGA: Survey, applications and challenges. Electronics, 10(8), 895. MDPI.
[6] Orabi, M., Tran, K. P., Egger, P., & Thomassey, S. (2024). Anomaly detection in smart manufacturing: An Adaptive Adversarial Transformer-based model. Journal of Manufacturing Systems, 77, 591-611. Elsevier.