Jeran Kilpack

Patricia McManus

ITAI 3377

29 April 2025

Edge AI in Industrial IoT: A Conceptual Framework for Ethical, Secure, and Feasible Deployment

Artificial Intelligence (AI) and the Industrial Internet of Things (IIoT) are transforming industrial systems, especially around predictive maintenance. While these technologies have the potential to revolutionize industries, the real question isn’t whether they can be integrated, but how they should be integrated in a way that’s both responsible and sustainable. This paper presents a conceptual framework for deploying Edge AI in IIoT-based predictive maintenance systems, focusing on how this technology can offer real-time, secure, and privacy-conscious decision-making. In doing so, it will argue that Edge AI provides a more ethically sustainable and technically feasible model compared to traditional cloud-based systems.

Recently, Edge AI has become increasingly popular for predictive maintenance, especially in time-sensitive applications that demand high reliability. One study by Zhang et al. highlights the use of Edge AI for real-time crack classification in bridges, a critical task in structural health monitoring. The framework they propose is powered by the Google Coral Dev Board, a low-power, high-performance edge device equipped with a Tensor Processing Unit (TPU). By using the MobileNetV2 model, optimized for edge devices, the system achieves over 94% classification accuracy, with inference times as fast as 2.8 milliseconds per image. This makes it ideal for continuous, high-frequency inspections without needing to rely on the cloud. The ability to process data locally ensures privacy and minimizes latency—two factors that are essential for deployment in remote or bandwidth-limited environments. The system's success in monitoring infrastructure such as bridges, pipelines, and industrial machinery demonstrates the practical benefits of Edge AI in the IIoT space.

However, this study also presents several challenges that must be addressed in future research. One issue is scalability: while the system works well for detecting cracks, it may not be easily adaptable to other types of structural degradation, such as corrosion or spalling. Another concern is model complexity. To function on resource-constrained edge devices, the model sacrifices some architectural depth, which might limit its ability to detect subtle features in complex environments. Additionally, the reliance on curated, labeled image datasets poses a bottleneck, as acquiring large-scale, accurately labeled data can be both expensive and time-consuming. To overcome these challenges, Zhang et al. suggest exploring transfer learning and on-device model updates. These approaches would allow Edge AI systems to adapt to new defect types or changing environmental conditions without requiring full retraining. Incorporating semi-supervised learning or synthetic data generation could also help expand datasets without the heavy burden of manual annotation. Further research into multi-modal Edge AI frameworks, which combine visual, acoustic, and vibrational data, could improve predictive maintenance across different IIoT sectors.

In the context of smart manufacturing, Edge AI can be used to build a layered architecture for predictive maintenance. In this architecture, AI-powered edge devices monitor the condition of machinery and production assets in real-time, detecting potential failures before they occur. By enabling local decision-making, this setup minimizes system latency, reduces network congestion, and ensures that operations can continue even during cloud outages. The system begins with the sensor layer, where key manufacturing equipment—such as CNC machines, conveyor systems, motors, and robotic arms—are equipped with a range of IIoT sensors. These sensors might include accelerometers for vibration analysis, thermocouples for temperature monitoring, and cameras for visual inspections. By constantly collecting operational data, the sensors can detect signs of wear, misalignment, or impending failure. Despite the promise of Edge AI, integrating it into IIoT systems introduces several security challenges. Manufacturing environments often rely on mission-critical assets—such as robotic arms, CNC machines, and programmable logic controllers (PLCs) which cannot afford unscheduled downtime or malicious interference. Among the key security risks are data tampering, where compromised edge devices might generate false maintenance reports; device spoofing, in which attackers impersonate legitimate devices to inject malicious updates; and model manipulation, where adversarial inputs or poisoned training data could distort the AI’s ability to detect faults. There are also vulnerabilities related to over-the-air (OTA) updates, where attackers might push malicious code directly to edge devices, or denial-of-service (DoS) attacks, which could overwhelm devices and render them unavailable for real-time analysis.

As with any emerging technology, the use of Edge AI in IIoT systems raises ethical questions. One concern is worker displacement: as AI begins to automate maintenance decisions, human technicians may find their roles diminished. This could lead to deskilling, where operators lose critical diagnostic skills, and issues of accountability, particularly in cases where AI-driven systems fail. If a failure occurs due to a machine-predicted error, who is held responsible—the engineer, the software vendor, or the AI algorithm itself? Another ethical concern is the potential for workplace surveillance. Many IIoT systems rely on cameras and microphones to detect anomalies, but these sensors could unintentionally monitor workers, raising privacy issues. Finally, there is the question of bias in AI models. If training data is limited or drawn from specific types of equipment, the resulting models may underperform in underrepresented scenarios, potentially leading to unsafe or inequitable maintenance outcomes.

Edge AI devices offer many advantages, such as low power consumption and portability, but they are also constrained by limited processing power, memory, and thermal tolerance compared to cloud-based systems. These limitations affect the ability of Edge AI to run large-scale deep learning models or process multiple sensor inputs simultaneously. Furthermore, industrial environments are often harsh, with conditions such as extreme temperatures, dust, and vibration, which can impact the reliability of edge devices. Industrial-grade edge devices, like the NVIDIA Jetson Xavier NX or the Google Coral Dev Board, help alleviate some of these issues by offering specialized hardware, such as GPUs and TPUs, to accelerate processing and withstand demanding conditions. On the software side, challenges remain in model compression, as transforming complex models into smaller, more efficient formats (e.g., TensorFlow Lite or ONNX) requires significant expertise. Interoperability is another hurdle, as many manufacturing systems still rely on legacy control systems (such as PLCs and SCADA) that are not always compatible with modern AI models. Managing a fleet of distributed edge devices across different environments also requires sophisticated MLOps practices to handle model versions, updates, and deployments.

In conclusion, AI-driven predictive maintenance in IIoT offers transformative potential for manufacturing industries, allowing businesses to predict failures before they happen, optimize resources, and extend the lifespan of valuable assets. While challenges related to hardware, security, and data integration remain, Edge AI provides a viable solution by enabling real-time data processing and decision-making. The future of predictive maintenance will not only depend on refining AI models but also on improving system interoperability, enhancing cybersecurity, and developing scalable solutions that can adapt to diverse industrial environments. As Edge AI continues to evolve, its success will hinge on advances in hardware, software optimization, and industry-specific customizations.

Works Cited

Zhang, Shijie, et al. *Real-Time Structural Health Monitoring using Edge AI: A Crack Classification Case Study on Bridges*. arXiv, 7 Apr. 2023, <https://arxiv.org/abs/2304.03718>.

**IEEE Access.** “Edge AI: A Survey on Architecture, Applications, and Future Directions.” *IEEE Access*, 2020, <https://ieeexplore.ieee.org/document/9075673>.

**arXiv.** “Quantizing Deep Convolutional Networks for Efficient Inference on Mobile Devices.” *arXiv*, 2018, <https://arxiv.org/abs/1802.05668>.

**Future Generation Computer Systems.** “Towards Industry 4.0: Edge Computing for IoT-Based Manufacturing.” *Future Generation Computer Systems*, 2020, <https://www.sciencedirect.com/science/article/pii/S0167739X1932410X>.