IJRESM
International Journal of Research in Engineering, Science and Management
Volume 9, Issue 1, January 2026
www.ijresm.com | E-ISSN: 2581-5792 | RESAIM Publishing | www.resaim.com

Using Tiny-ML Models on Edge Devices to Improve Industrial Control Systems in Real Time

Kadari Achwak*
University of Science and Technology Mohamed Boudiaf, Algeria

***Abstract*: Tiny Machine Learning (Tiny-ML) lets machine learning models run on edge devices with very few resources. This opens up new possibilities for smart Industrial Control Systems (ICS) that don't need to be connected to the cloud all the time. Traditional Programmable Logic Controllers (PLCs) and Supervisory Control and Data Acquisition (SCADA) systems predominantly depend on static, rule-based logic and face challenges in adapting to dynamic, data-intensive industrial settings that necessitate predictive maintenance, anomaly detection, and real-time process optimization. This paper examines the implementation of highly efficient Tiny-ML models on edge devices, including ARM-based micro-controllers and single-board computers, to attain low-latency inference, bandwidth optimization via local processing, and enhanced energy efficiency and security for industrial Internet of Things (IIoT) sensors. The paper talks about hardware and software limits, how to connect to old PLC/SCADA systems using industrial protocols, and gives an overview of contemporary industrial projects that show how edge AI is useful in modern workplaces.**

***Keywords*: Edge computing, predictive maintenance, PLC, SCADA, Tiny-ML, Industry 4.0, anomaly detection.**

1. Introduction

Industrial Control Systems (ICS) are the main parts of modern manufacturing, energy, and process industries. They are usually constructed around PLCs and SCADA systems that are meant to work in a deterministic, rule-based way. These systems are strong and safe, but they can't learn from data or change in real time as operating conditions change, equipment breaks down, or processes drift slightly. The rapid proliferation of Internet of Things (IoT) sensors and connectivity has resulted in the generation of substantial time-series data. Because of this, more and more people are using machine learning to discover faults, do predictive maintenance, and make things work better. On the other hand, cloud-based solutions have a lot of drawbacks, like latency, high bandwidth prices, concerns about data privacy, and the necessity for a stable network connection. These limitations could make it harder for industrial circuits that need to work quickly. Tiny-ML can help with these issues by letting small models operate directly on low-power edge devices and micro-controllers. The swift expansion of Internet of Things (IoT) sensors and connectivity has led to the creation of significant time-series data. This trend indicates an increasing adoption of machine learning for fault detection, predictive maintenance, and optimization of performance. On the other hand, cloud-based solutions have a lot of drawbacks, like latency, high bandwidth prices, concerns about data privacy, and the necessity for a stable network connection. The identified limitations may impede the efficiency of industrial circuits that require rapid operation. Tiny-ML addresses these challenges by enabling small models to function directly on low-power edge devices and micro-controllers. The device can have anywhere from a few megabytes to a few kilobytes of memory, which makes it possible to make inferences in real time on the device.

This research investigates the implementation of Tiny-ML models on edge devices, specifically focusing on STM32-class micro-controllers and embedded Linux boards. This study aims to investigate the potential of these models in enhancing decision-making processes and optimizing performance within industrial control systems in real time. The goal is to show how Tiny-ML can be used for predictive maintenance, problem diagnosis, and process improvement, all while working well with existing industrial environments and established automation frameworks.

2. Fundamental Concepts of Tiny-ML an its Two Industrial Applications

A. Applications of Tiny-ML in Industry

Tiny-ML represents a specialized form of machine learning designed to operate on devices characterized by constraints in memory, processing capabilities, and battery longevity. The majority of these devices consist of micro-controllers and closely interconnected systems. TensorFlow Lite for Micro-controllers, along with comparable frameworks, facilitates the use of 8-bit integer quantized networks. The run-times are relatively small, typically in the range of tens of kilobytes. This facilitates the process of drawing conclusions regarding Cortex M and other central processing units (CPUs) that operate without an operating system.

Tiny-ML applications in factories frequently address issues such as time-series classification or regression. To identify initial indicators of failure or alterations in the process, these challenges analyze data derived from vibrations, currents, pressures, or temperatures. Techniques such as post-training quantization, quantization-aware training, and pruning are

*Corresponding author: achwakkadari@qq.com employed to reduce the model's size and enhance its efficiency. This is implemented to ensure that activities such as identifying anomalies and conducting predictive maintenance are executed with sufficient precision.

B. Identifying and Monitoring Issues for Proactive Maintenance

The objective of predictive maintenance is to forecast the future performance of equipment and to schedule repairs proactively to prevent breakdowns. This reduces the probability of unforeseen downtime and elevated maintenance expenses. A recent study indicates that Tiny-ML enables predictive maintenance on devices by executing compact models on micro-controllers, which analyze vibration and other sensor data in real time. Consider Neuton as a case study. Arm-based micro-controllers utilizing Tiny-ML models are capable of identifying issues in compressor pumps and motors. Alternative platforms provide guidance on executing this process. These models issue warnings prior to significant failures occurring in the process.

The implementation of predictive maintenance through Tiny-ML demonstrates a reduction in latency and bandwidth by eliminating the need for cloud round-trips. This has been confirmed through systematic literature reviews. This facilitates the monitoring of assets that were previously deemed too costly for inclusion in centralized analytic. Case studies within the industry demonstrate that the application of edge AI for detecting defect signals in HVAC systems and rotating equipment has led to improvements in both the mean time between failures (MTBF) and the operational uptime of the systems.

3. An Examination of PLC/SCADA Systems and their Limitations

The components that constitute an Industrial Control System (ICS) hold significant importance. Field devices, Programmable Logic Controllers (PLCs), Supervisory Control and Data Acquisition (SCADA), and Distributed Control Systems (DCS) constitute integral components of these systems. The components communicate with one another through industrial Ethernet and fieldbus protocols. Languages such as ladder logic, structured text, and function block diagrams serve as the foundation for developing cyclic control programs for programmable logic controllers (PLCs). Often, these systems possess limited random-access memory (RAM) and nonvolatile memory, necessitating real-time operation. Due to the discussed challenges, the direct application of traditional, large-scale machine learning models on PLC hardware is often unfeasible.

Typically, the integration of Tiny-ML into Industrial Control Systems necessitates the utilization of specific edge devices that interface with Programmable Logic Controllers and field sensors. The devices collect sensor data, perform Tiny-ML inference, and subsequently transmit the results to the control system through protocols such as OPC UA, MQTT, Modbus/TCP, or proprietary interfaces including Siemens S7 or EtherNet/IP.

A. Protocols and Middleware for Industrial Applications

The OPC Unified Architecture (OPC UA) has emerged as a critical standard for ensuring interoperability within the framework of Industry 4.0. It enables secure data sharing among edge devices, PLCs, SCADA, and various advanced systems across any platform without issues. MQTT is a lightweight publish-subscribe protocol frequently utilized for communication within the framework of the Industrial Internet of Things (IIoT). This is particularly beneficial for edge nodes equipped with Tiny-ML, as they can seamlessly transmit health indicators or anomaly scores without requiring additional effort.

An increasing number of businesses within the industrial sector are providing edge platforms that integrate these protocols with environments designed for executing applications or containers. Artificial intelligence models can be integrated into the production line using Siemens Industrial Edge, which operates in conjunction with TIA Portal. These models analyze sensor data independently at a local level to generate maintenance alerts and provide recommendations for optimizing operational efficiency. Rockwell Automation's FactoryTalk Analytics solutions come with analytics at the edge, which makes it easier to connect to PLC and SCADA systems that are already in place.

4. Four Technical Challenges Arise when Implementing Tiny-ML on Edge Devices within Industrial Control Systems

A. Constraints of Real-Time Systems and Hardware

Many PLCs and edge devices utilizing micro-controllers typically possess under one megabyte of random-access memory (RAM) and minimal flash storage capacity. There is a requirement for models that are compact in size and for memory management that is highly accurate due to this necessity. Tiny-ML techniques facilitate this process through methods such as static memory allocation, operator fusion, and quantized operators. Designers must carefully analyze and refine models to adhere to stringent real-time deadlines, frequently limited to just a few milliseconds for each control cycle.

For maintaining control stability, it is essential that real-time anomaly detection systems operate effectively within a time frame of ten milliseconds or less. Neural networks exhibit a level of simplicity and shallowness, facilitating the application of efficient architectures such as shallow convolutional neural networks (CNNs), compact recurrent networks, or conventional methods that utilize lightweight feature extraction. The absence of dynamic memory allocation and operating system services in numerous micro-controller targets introduces considerable complexity to the configuration and troubleshooting processes.

B. Adaptability in Response to Environmental Challenges, Protection, and Stability

In industrial environments, factors such as electrical interference, mechanical oscillation, and extreme temperatures can significantly compromise the accuracy of sensors and the reliability of gears. For Tiny-ML models to demonstrate robustness, it is essential that they undergo training and testing using real-world datasets gathered from actual scenarios. The datasets should include a range of operational modes, environmental conditions, and instances of sensor degradation.

The integration of artificial intelligence into edge devices facilitates easier access for hackers into control networks. This may result in the emergence of new vulnerabilities. Establishing standards such as OPC UA is crucial for safeguarding essential processes and minimizing the potential for detrimental control actions. It is essential for these standards to encompass encryption and authentication, network isolation, and comprehensive testing of a model's performance under adverse conditions.

C. Five Significant Changes that have Occurred in the Sector and their Implications

Many companies are seeking to integrate industrial control systems with localized intelligence, leading to the incorporation of edge AI into their products. Siemens is developing various predictive maintenance solutions that leverage artificial intelligence (AI) alongside the TIA Portal and Industrial Edge. These solutions examine machine data in real time on the factory floor. This results in a significant reduction in failures and downtime within the automotive industry and various other industrial sectors. Arm and its ecosystem partners demonstrate the application of Tiny-ML in illustrating the functionality of predictive maintenance on Arm micro-controllers. The responses illustrate the application of compact models in identifying issues related to pumps, motors, and compressors within the equipment.

Recent studies in academic and business environments indicate that Tiny-ML serves as a significant instrument for predictive maintenance and load optimization within the industrial sector. This indicates that lightweight models can achieve high accuracy despite limited resources. The integration of TinyML and ICS appears to facilitate the development of self-optimizing AI functionalities. The implementation of these features has the potential to reduce costs while simultaneously facilitating the objectives of Industry 4.0, such as increased uptime, reduced energy consumption, and enhanced production flexibility. TinyML enables edge devices with minimal power consumption to execute machine learning models effectively. This means that it is possible to make industrial assets and control systems better by using intelligence without having to always be connected to the cloud. Businesses can leverage TinyML for predictive maintenance, identifying issues, and optimizing processes. The outcome will be characterized by low-latency inference, reduced bandwidth and energy consumption, and enhanced security achieved through local data processing.

In order to actualize this promise, it is essential to address the constraints of the technology, the stringent real-time demands, the challenges associated with interfacing legacy PLC/SCADA systems, and the concerns surrounding safety and cyber security. Ongoing initiatives in both the industrial sector and academic research indicate that tinyML-enabled industrial control systems are poised to play a significant role in the evolution of Industry 4.0 and the development of future smart factories. These initiatives encompass the endeavors of leading companies in developing edge artificial intelligence platforms, alongside comprehensive analyses of Tiny-ML for predictive maintenance applications.

References

[1] Arm Editorial Team, “How predictive maintenance saves time, costs and improves reliability,” *Arm Newsroom*, Jun. 20, 2025. [Online]. Available: https://newsroom.arm.com/blog/predictive-maintenance-smart-factories Arm Newsroom

[2] Arm Editorial Team, “Siemens reinvents factory reliability with edge AI-driven predictive maintenance,” *Arm Newsroom*, Aug. 28, 2025. [Online]. Available: https://newsroom.arm.com/blog/siemens-arm-edge-ai-driven-predictive-maintenance Arm Newsroom

[3] S. O. Ooko and S. M. Karume, “Application of tiny machine learning in predictive maintenance in industrial settings,” *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 131–150, Aug. 02, 2024.

[4] M. Muthusamy *et al*., “TinyML with CTGAN-based smart industry power load monitoring and optimization,” *Sci. Rep.*, 2025

[5] “Real-time anomaly detection for predictive maintenance using TinyML,” *RuntimeRec*, 2025. [Online]. Available: https://runtimerec.com/real-time-anomaly-detection-for-predictive-maintenance-using-tinyml/

[6] S. Shitut, “AI-powered predictive maintenance: The future of industrial operations with Siemens TIA Portal & Industrial Edge,” 2025. [Online]. Available: https://www.linkedin.com/pulse/ai-powered-predictive-maintenance-future-industrial-siemens-shitut-qz0lf

[7] Silicon Labs, “TensorFlow Lite for Microcontrollers – Developer docs,” 2017. [Online]. Available: https://docs.silabs.com/gecko-platform/4.2/machine-learning/tensorflow/overview

[8] TensorFlow, “TensorFlow Lite for Microcontrollers – Overview,” 2025. [Online]. Available: https://www.tensorflow.org/lite/microcontrollers/overview