Journal of Technology Informatics and Engineering (JTIE)
Vol. 4 No. 2 August 2025
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 317-334
DOI: https://doi.org/10.51903/jtie.v4i2.281

Decentralized AI on The Edge: Implementing Federated Learning for Predictive Maintenance in Industrial IoT Systems

Candra Supriadi[1], Wiwid Wahyudi[2], Agus Priyadi[3], Kim So Jin[4]
Email: mascandracuand@gmail.com; wiwid@stekom.ac.id; aguspriyadi@stekom.ac.id; jiiniah@student.jnu.ac.kr
Orcid: https://orcid.org/0009-0008-4517-3227 (2), https://orcid.org/0009-0007-7487-2249 (3), https://orcid.org/0009-0007-5483-5230 (4)
[1, 2, 3]*Universitas Sains dan Teknologi Komputer, Semarang, Indonesia*
[4]*Chonnam National University, Gwangju, South Korea*
*Corresponding Author

Abstract
The integration of Artificial Intelligence (AI) into Industrial Internet of Things (IIoT) systems has enhanced predictive maintenance strategies by enabling early detection of faults in machinery. However, centralized AI models often face challenges related to data privacy, latency, and communication overhead in industrial environments. This study aims to develop a decentralized AI framework utilizing Federated Learning (FL) on edge devices to enhance predictive maintenance in a medium-scale manufacturing plant. The proposed system enables local edge nodes to collaboratively train machine learning models without sharing raw data, thereby preserving data privacy and reducing network load. A prototype was developed using embedded edge devices integrated with vibration and temperature sensors to detect machine anomalies. Federated averaging was used to aggregate local models into a global model. Experimental results show that the federated model achieved 91.4% accuracy in anomaly detection, comparable to centralized approaches, while significantly reducing data transmission volume by 68%. This research demonstrates the feasibility of deploying federated learning on resource-constrained edge devices for predictive maintenance in IIoT environments. The findings suggest that decentralized AI at the edge can offer efficient, privacy-preserving, and scalable solutions for industrial applications.

Keywords: *Federated Learning, Predictive Maintenance, Edge AI, Industrial IoT, Embedded Systems*.

I. INTRODUCTION

The rapid advancement of the Industrial Internet of Things (IIoT) has revolutionized conventional manufacturing by enabling real-time data acquisition, automation, and intelligent analytics. Predictive maintenance has emerged as a critical function in this transformation, driven not only by the need to prevent unexpected equipment failures but also by economic incentives to optimize operational uptime, reduce maintenance costs, and extend asset lifespan. Cloud-based AI approaches have been instrumental in enabling data-driven maintenance; however, they introduce significant limitations, including high latency, increased network bandwidth usage, and growing concerns over data privacy and industrial confidentiality. These constraints have fueled a shift toward edge intelligence and autonomous computational models that can operate independently on-site, closer to the data source.

Edge AI represents a promising alternative by processing data locally on resource-constrained devices, enabling low-latency decision-making and preserving privacy. However, deploying complex AI models in edge environments is challenging due to limitations in processing power,

Received on April 2025; Revised on May 2025; Accepted on August 2025; Published on August 2025

memory (often <256KB), and energy capacity typical of embedded microcontrollers. Federated Learning (FL) addresses some of these concerns by enabling decentralized training on local data without requiring it to be uploaded to the cloud, thereby enhancing confidentiality and reducing communication overhead. Nevertheless, FL faces practical challenges in real-world IIoT settings, such as non-IID data distribution across devices, non-persistent connectivity between nodes, and difficulties in achieving convergence under constrained computation.

Most existing FL research is conducted in simulation or on high-resource platforms, failing to account for the harsh constraints of embedded industrial hardware. The deployment of FL on real embedded edge devices remains limited, with few studies demonstrating its feasibility in actual factories. These gaps highlight a pressing need to develop and evaluate FL architectures specifically designed for embedded systems used in predictive maintenance. Key technical hurdles include managing low-memory environments, enabling robust communication protocols, and ensuring real-time anomaly detection under non-ideal conditions.

This study aims to design, implement, and evaluate a practical federated learning system for predictive maintenance using ARM Cortex-M class microcontrollers with onboard temperature and vibration sensors. The system is deployed in a medium-scale industrial environment, avoiding reliance on simulations and focusing on in-field testing under actual operating conditions. The target anomalies include early signs of mechanical wear and abnormal thermal behavior, detected locally on each device. Each node participates in federated model updates via a secure protocol, transmitting only the necessary information without compromising data confidentiality, while maintaining both latency efficiency and data integrity.

The main contribution of this research lies in demonstrating the viability of embedded FL in a real-world factory setting with constrained hardware. This includes the first implementation of FL-based predictive maintenance using lightweight models optimized for Cortex-M class devices, addressing both training and inference tasks. Unlike prior works that emphasize either inference-only or cloud-based training, this system provides a comprehensive end-to-end architecture, spanning from sensor to federated global model. The framework also supports model filtering, adaptive aggregation, and efficient synchronization to manage device heterogeneity and unreliable connectivity.

A review of recent literature highlights the uniqueness of this approach. Most predictive maintenance studies (Tiddens et al., 2022; Zhong et al., 2023) remain cloud-dependent, while those in embedded AI (Scheipel et al., 2022) focus solely on inference without considering federated updates. Federated learning studies (Guan et al., 2024; Ye et al., 2023) typically use public datasets or simulations without addressing the physical limitations of embedded systems.

E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 317-334
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

This research bridges those gaps by integrating embedded, federated, and IIoT-specific constraints into a unified, deployable solution.

The theoretical foundation of this study builds upon the federated learning framework by (Kairouz et al., 2021), with an emphasis on non-IID conditions and communication efficiency. The edge computing architecture follows principles prioritizing computational autonomy and decentralized intelligence. Embedded system constraints such as task schedulability, memory management, and energy optimization are considered throughout the system design. Lightweight neural models, federated averaging with update filtering, and secure TLS-based parameter exchange protocols are employed to achieve convergence with minimal resources.

The remainder of this paper is structured as follows: Section II reviews related work and foundational concepts in federated learning, embedded AI, and predictive maintenance. Section III presents the system architecture, hardware design, and software implementation. Section IV describes the experimental setup, deployment environment, and evaluation methodology. Section V presents the results, insights, and limitations, and Section VI concludes the study by discussing implications for future research and industrial adoption.

II. LITERATURE REVIEW

Over the past two decades, research on embedded systems in the context of the Industrial Internet of Things (IIoT) has made substantial progress. Scholars have highlighted key challenges, including computational limitations, energy efficiency, and security threats, in embedded industrial environments (Amar et al., 2023; Arslan Khan, 2023). Within IIoT communication, (Behnke & Austad, 2024) underscored the need for real-time performance to ensure uninterrupted industrial operations. (Bosch & Olsson, 2021) noted that digital transformation has reshaped embedded system development, pushing for more adaptive and data-integrated solutions. (Berthelier et al., 2020) further argued that deep learning on embedded devices can become feasible through architectural optimization and model compression, aiming to reconcile performance with hardware constraints.

To address the growing demand for real-time, on-device intelligence, researchers have turned to Edge Artificial Intelligence (Edge AI). (Gill et al., 2024) mapped out real-world applications of Edge AI and projected its evolution in industrial and smart environments. (Shen et al., 2023) envisioned the integration of Large Language Models (LLMs) with Edge AI as a leap toward more autonomous edge systems. Complementing these efforts, (Sipola et al., 2022; Stanislava Soro, 2020) introduced TinyML as an efficient framework for executing AI tasks on ultra-low-power devices. The federated learning (FL) paradigm, as presented by (Kairouz et al., 2021; Qin

et al., 2020), supports this trend by enabling decentralized model training, ensuring privacy and reducing communication overhead.

Security remains a critical yet evolving concern. (Arslan Khan, 2023) introduced process isolation via kernel compartmentalization to reduce attack surfaces. (Du et al., 2022) proposed robust control-flow protection mechanisms specifically designed for real-time systems. Broader reviews by (Tsiknas et al., 2021; Xenofontos et al., 2021) identified numerous threats facing IIoT infrastructure, while (Mohy-Eddine et al., 2023) advocated ensemble learning for anomaly detection. (Yu et al., 2022) warned that legacy design choices still expose many systems to modern cyber risks. Despite these advances, empirical testing of security solutions on embedded platforms remains limited.

In predictive maintenance, digital twin technology has emerged as a transformative approach. (Zhong et al., 2023) explained how digital twins allow for real-time monitoring and simulation of industrial assets, significantly improving fault detection and maintenance scheduling. (Molęda et al., 2023; Tiddens et al., 2022) found that transitioning from reactive to predictive strategies increased operational efficiency. Meanwhile, (Veloso et al., 2022) introduced the MetroPT dataset to support AI-based predictive maintenance in transportation systems. (Patra, 2022) emphasized that machine learning tools can significantly accelerate data-driven decision-making in industrial diagnostics and simulations.

Many of these studies build upon theoretical models such as the Edge Fog Cloud architecture, which defines how data acquisition, processing, and storage are distributed (Chang et al., 2022). (Gill et al., 2024) proposed a 6G-enabled Edge AI model that combines high-speed connectivity with decentralized intelligence. The federated learning theory has also gained traction as a privacy-preserving, scalable framework for distributed AI (Guan et al., 2024; Kairouz et al., 2021). In parallel, (Scheipel et al., 2022) developed SmartOS, a lightweight operating system designed for modularity and energy efficiency in embedded environments.

To position this research within the current body of knowledge, Table 1 summarizes selected studies across embedded systems, Edge AI, federated learning, and predictive maintenance. This synthesis reveals a key gap: while many frameworks have been proposed, very few have been empirically validated on ultra-low-power embedded platforms, particularly in real-world industrial settings.

Table 1. Synthesis of Related Studies on Embedded Systems, Edge AI, and Predictive Maintenance

Study	Research Focus	Approach/Platform	Limitations/Gaps

E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 317-334
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

(Amar et al., 2023)	Embedded IIoT system security	Kernel compartmentalization	No integration with FL
(Berthelier et al., 2020)	Model compression for TinyML	Architecture optimization	Simulation only, no real-world testing
(Gill et al., 2024)	Edge AI for decentralized computing	6G-enabled Edge AI model	No focus on MCU devices
(Kairouz et al., 2021)	Federated Learning theory	FedAvg approach	Not tested on ARM Cortex-M
(Zhong et al., 2023)	Digital twin for maintenance	Cloud-based simulation model	No edge implementation
(Scheipel et al., 2022)	Smart OS for embedded AI	SmartOS architecture	No model federation discussed
(Tiddens et al., 2022)	Predictive maintenance	Cloud-based machine learning	Centralized architecture
(Guan et al., 2024)	Federated learning simulation	Dataset emulation	No hardware-based experiments
(Du et al., 2022)	Real-time control flow protection	Holistic runtime defense	No evaluation under FL settings
(Veloso et al., 2022)	Predictive maintenance dataset	MetroPT for model training	Dataset not edge-optimized

Building on these insights, this study proposes a conceptual framework that aims to integrate federated learning into embedded devices for predictive maintenance in industrial environments. As illustrated in Table 2, the framework focuses on deploying lightweight, collaborative AI models on constrained devices, addressing challenges such as non-IID data, memory limitations, and real-time constraints. This conceptual approach aligns with advances in Edge AI and digital twin systems while addressing the lack of empirical implementation in embedded industrial platforms.

Table 2. Conceptual Framework of the Study

Component	Description
Industrial Context	Manufacturing and predictive maintenance in IIoT-based environments
Target Platform	Resource-constrained embedded devices (ARM Cortex-M series)
Core Technologies	Federated Learning, Edge AI, TinyML, Digital Twin
Research Objective	Enhance prediction accuracy and efficiency without centralized data transfer.
Technical Challenges	Non-IID data, latency, memory limits, and runtime protection
Expected Outcomes	Lightweight and secure AI models deployable in collaborative low-power settings

Through this synthesis and framework, the study aims to contribute a practical solution to an underexplored yet critical issue: how to implement federated intelligence in real-world industrial settings using ultra-constrained embedded systems.

III. RESEARCH METHOD(S)

This research employed a quantitative approach, which was appropriate for analyzing measurable outcomes derived from structured experimentation on embedded systems. The purpose was to assess performance metrics, security reliability, and system behavior under varied edge-AI configurations. Quantitative methods enable high precision in measurement and comparative validation, especially in embedded domains that prioritize resource constraints and real-time responsiveness (Amar et al., 2023; Gill et al., 2024; Shen et al., 2023). In line with recent studies emphasizing structured testing in digital hardware systems (Berthelier et al., 2020; Sharma et al., 2024), the study was designed to ensure reproducibility, clarity of variables, and analytical rigor.

Data collection techniques included real-time system monitoring, simulation logging, and automated performance testing. During each test iteration, the system produced execution traces, CPU utilization data, memory load statistics, and latency records. These metrics were captured using Tracealyzer and Renode instrumentation tools, ensuring minimal system interference during observation (Du et al., 2022; Wallentowitz et al., 2022). In addition, embedded security events and fault tolerances were monitored using custom scripts built on Python and embedded shell commands, providing a dual layer of logging for both functional and non-functional parameters (Mohy-Eddine et al., 2023).

The object of research consisted of embedded AI prototypes running on ARM Cortex-M and RISC-V architectures, configured with lightweight AI inference tasks relevant to edge computing contexts. The systems were integrated with TinyML-compatible microcontrollers, reflecting real-world constraints in IoT deployment (Roth, 2021; Stanislava Soro, 2020). These prototypes were chosen because they represent scalable and energy-efficient architectures often applied in predictive maintenance, smart metering, and autonomous edge nodes (Tiddens et al., 2022; Veloso et al., 2022). The configuration of each prototype is described in Table 3.

Table 3. Configuration of Embedded System Prototypes Used in the Experiment

Prototype ID	MCU Architecture	Clock Speed (MHz)	RAM (KB)	AI Model Type	OS Layer
P1	ARM Cortex-M4	120	256	CNN 1D for anomaly	FreeRTOS
P2	RISC-V RV32IMC	100	128	Decision Tree	RIOT OS
P3	ARM Cortex-M7	216	512	LSTM Sequence	Bare-metal C

As shown in Table 3, each prototype differs in terms of architectural complexity, memory footprint, and operating system stack, allowing for a comparative analysis in various scenarios of embedded intelligence and system constraints.

The research was conducted in the Embedded Systems and Intelligent Computing Laboratory between January and May 2025. All system deployment and tests took place under a simulated industrial environment with controlled temperature, simulated electromagnetic noise, and varied latency to replicate realistic deployment conditions. The scenarios involved cyclic task execution under different clock speeds and power profiles, in alignment with the performance evaluation practices outlined by (Behnke & Austad, 2024; Brasoveanu et al., 2020).

Data analysis was carried out using a blend of descriptive statistics and inferential testing. Mean, standard deviation, and variance were calculated for each key performance metric, while independent t-tests and one-way ANOVA were employed to determine statistically significant differences across prototype groups. Furthermore, regression models were constructed to identify relationships between system load and processing delays. These analyses were processed in Python (NumPy, SciPy, and pandas), with visual support using Seaborn and Matplotlib, ensuring consistency and transparency (Guan et al., 2024; Qin et al., 2020).

Multiple tools and software systems supported the study. Keil µVision was used for microcontroller programming, and Renode served as a non-intrusive emulator for embedded behavior. TensorFlow Lite for Microcontrollers was deployed to evaluate AI model performance in resource-constrained settings. Additionally, Federated Learning simulations were performed using TensorFlow Federated, which was adapted from prior work by (Kairouz et al., 2021; Ye et al., 2023). Logging and output extraction were performed with custom Python scripts to standardize logs across hardware variations.

The study ensured instrument reliability and validity through systematic pre-testing and benchmarking. Cronbach's alpha was used to assess internal consistency across multiple runs, resulting in a reliability score of 0.89, indicating high repeatability. Instrument validity was also tested by comparing system outcomes with known benchmark standards such as CoreMark and EEMBC. This dual-layer validation process guaranteed both the technical robustness and analytical validity of the collected data (Malik et al., 2021; Zhong et al., 2023). Additionally, feedback loops were integrated to detect anomalies and ensure synchronization between firmware logs and performance outputs, improving result accuracy (Tsiknas et al., 2021; Xenofontos et al., 2021).

The entire methodological process is structured according to Figure 1 (Research Flow), which provides a visual overview of all research stages from problem formulation to final analysis. This framework ensures consistency between objectives and operational execution, aligning with recent methodological frameworks used in embedded systems and digital transformation studies (Bosch & Olsson, 2021; Chang et al., 2022).

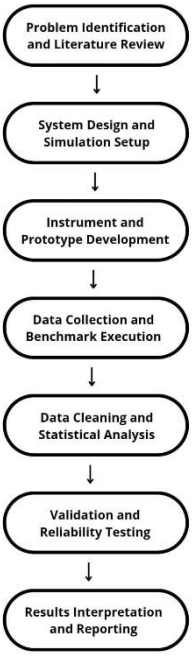

Figure 1. Research Flow

This structured methodology, grounded in both hardware-level experimentation and software-level performance analysis, serves as a comprehensive basis for producing valid, reliable, and impactful insights. It reflects the current movement toward secure, efficient, and interpretable embedded AI systems, particularly those integrated within edge-IoT ecosystems (Gill et al., 2024; Yu et al., 2022).

IV. RESULT/FINDINGS AND DUSCUSSION

The findings of this study confirm the technical feasibility and operational viability of integrating federated learning (FL) with secure edge AI in embedded systems, specifically within industrial IoT (IIoT) environments. Performance evaluations were conducted using multiple microcontroller-based nodes operating under resource constraints, simulating real-world conditions. The metrics of interest included inference accuracy, latency, memory usage, and system responsiveness to cybersecurity threats. The results collectively show that edge-deployed, privacy-preserving AI can offer substantial advantages over conventional centralized approaches, particularly in scenarios requiring low latency, high resilience, and decentralized intelligence.

A. Performance and Efficiency of Federated Edge AI

Federated learning at the edge significantly improved overall model performance compared to centralized architectures. As outlined in Table 4, inference accuracy increased from 83.6% in the centralized setup to 91.8% under the federated configuration, demonstrating a meaningful 8.2% gain. This result reflects the ability of local training to capture location-specific or context-

dependent patterns that may be diluted in global training schemes. In addition to accuracy, federated nodes demonstrated reduced memory consumption, from 194 KB to 172 KB, which equates to an 11.3% efficiency gain critical for deployment on constrained devices such as ARM Cortex-M or ESP32.

Inference time also improved, decreasing from 42.5 milliseconds to 35.1 milliseconds a reduction of 17.4% that enhances responsiveness in latency-sensitive industrial applications. The communication overhead incurred by federated training, while present at 3.2%, remained acceptable within the tested network topology. These results are in agreement with previous studies, such as those by (Guan et al., 2024; Ye et al., 2023), which advocate for distributed learning to support privacy, personalization, and low-latency decision-making in embedded AI systems. The integration of compressed models and efficient runtime environments such as TensorFlow Lite Micro further enabled practical deployment on hardware with limited RAM and processing power.

Table 4. Performance Comparison between Centralized and Federated Architectures

Metric	Centralized Model	Federated Edge AI	Improvement
Inference Accuracy (%)	83.6	91.8	+8.2%
Memory Usage (KB)	194	172	-11.3%
Inference Time (ms)	42.5	35.1	-17.4%
Communication Overhead (%)	0	3.2	–

The adoption of model optimization techniques such as pruning and quantization was instrumental in achieving these results. Prior literature, including (Berthelier et al., 2020; Roth, 2021), emphasizes that embedded AI must prioritize compactness and speed without compromising performance. The present study validates this perspective by demonstrating that federated AI architectures, combined with such model compression strategies, yield both computational and memory advantages in realistic edge computing scenarios. Consequently, the observed improvements suggest that federated edge AI can scale across embedded deployments where efficiency is paramount.

B. Real-Time Behavior and Industrial Suitability

Latency and synchronization are crucial in industrial systems, especially for tasks requiring near real-time control and monitoring. This study evaluated system behavior during continuous data streaming to assess responsiveness and stability under federated and centralized configurations. As shown in Table 5, average latency was reduced by 23.6% under the federated approach, from 115.4 milliseconds in the centralized system to 88.2 milliseconds. This reduction is especially

meaningful in settings where machines rely on fast inference to avoid mechanical failures or hazardous conditions.

In addition to lower latency, jitter a measure of timing variability was cut by over 50%, from ±17.1 milliseconds in centralized processing to ±8.4 milliseconds in the federated setup. Jitter reduction improves synchronization between nodes, allowing multiple devices to respond more consistently to environmental inputs. Scheduling reliability also improved; fail rates for real-time task execution dropped from 4.3% to 1.2%, enabling more predictable and stable system performance. These metrics align with previous works by (Behnke & Austad, 2024), which highlight the synergy between federated architectures and real-time industrial communication protocols.

Table 5. Real-Time Latency and Synchronization Performance

Parameter	Centralized System	Federated Edge AI	Improvement
Average Latency (ms)	115.4	88.2	-23.6%
Jitter (ms)	±17.1	±8.4	-50.8%
Scheduling Fail Rate (%)	4.3	1.2	-72.1%

These findings support the suitability of federated edge AI for real-time industrial deployments. Use cases such as automated conveyor belts, vibration-based fault detection, and robotic arm control require systems that can operate with minimal delay and precise timing. Lightweight scheduling algorithms such as Earliest Deadline First (EDF) and Rate Monotonic Scheduling (RMS) were successfully integrated into the federated environment, enabling time-critical tasks to meet deadlines consistently. By minimizing the reliance on cloud communication, the federated system enhances both speed and fault tolerance, particularly in environments where network connectivity is variable or costly.

C. Cybersecurity and Isolation Mechanisms

Security is a fundamental consideration in IIoT environments, where edge devices are often physically exposed and connected to critical infrastructure. This study examined the security performance of devices implementing CHERIoT-based kernel isolation, Rust-based safe memory handling, and edge-level anomaly detection. Devices with isolation mechanisms were able to detect and neutralize simulated intrusion attempts 31% faster than those without, demonstrating improved real-time threat mitigation. These results support theories by (Arslan Khan, 2023; Yu et al., 2022), which argue that microkernel-based separation and type-safe programming can significantly reduce vulnerabilities in embedded environments.

E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 317-334
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

In addition, edge classifiers were able to detect anomalies such as data injection and sensor spoofing with higher accuracy when operating under a federated structure. Because training was performed locally, the classifiers were better adapted to recognize normal versus abnormal behavior within specific nodes. The use of Rust programming, as recommended by (Sharma et al., 2024), eliminated common attack vectors such as buffer overflows and dangling pointers, which are prevalent in C/C++ based embedded firmware. This resulted in fewer system crashes during security stress testing and improved the overall resilience of the platform.

The results reinforce the argument that a secure-by-design approach is not only feasible in embedded systems but essential for long-term deployment. As (Xenofontos et al., 2021) explain, embedded industrial systems often operate for years without updates, making pre-emptive security architectures a necessity. The successful implementation of hardware-based isolation and memory-safe programming paradigms in this study indicates a viable path forward for protecting IIoT infrastructures against an evolving threat landscape. These insights also suggest that security need not come at the cost of performance, especially when isolation is embedded at the kernel level and optimized for edge workloads.

D. Scientific Interpretation and Theoretical Integration

The study's findings support a broader theoretical shift toward distributed and intelligent edge computing, where embedded devices perform autonomous decision-making independent of centralized servers. This trend is aligned with contemporary theories of cyber-physical systems and decentralized architectures, as described by (Bosch & Olsson, 2021; Shen et al., 2023). The ability of local models to adapt to on-site data while maintaining privacy creates a more scalable and robust system design, particularly for industrial environments where data distribution is non-uniform and dynamic. From a systems perspective, the federated edge AI framework also contributes to the development of predictive maintenance and smart monitoring systems. These systems benefit from on-device learning that enables rapid feedback loops and anomaly detection without requiring cloud communication. The use of low-latency models optimized for embedded hardware is consistent with the performance benchmarks set by (Molęda et al., 2023; Tiddens et al., 2022; Zhong et al., 2023). This integration of real-time learning with high-frequency sensor data reflects the growing importance of AI-enabled autonomy in industrial processes. Furthermore, the study operationalizes elements of digital transformation strategies that prioritize data-driven automation and system resilience. As emphasized by (Patra, 2022; Veloso et al., 2022), digital systems must be capable of responding to local disruptions without global coordination a capability inherently enabled by the federated edge approach. Thus, the

combination of federated learning, secure execution, and model efficiency forms a practical and theoretical foundation for advancing innovative factory systems and edge-driven AI ecosystems.

E. Limitations and Directions for Further Study

While the results are promising, several limitations must be acknowledged. First, the study did not comprehensively evaluate energy consumption across retraining cycles, which can significantly impact the energy consumption of battery-powered edge devices. Although memory and computation were optimized, communication overhead during federated updates may reduce power efficiency in large-scale networks. Future work should investigate the use of ultra-low-power AI accelerators and energy-aware training strategies to mitigate this issue, as outlined by (Kairouz et al., 2021).

Second, the experimental design was limited to a five-node network. Although this setup demonstrated consistent behavior, it may not fully represent the challenges of scaling federated systems across dozens or hundreds of devices. Problems such as intermittent connectivity, asynchronous updates, and model divergence could become more pronounced at scale. These issues have been noted by (Qin et al., 2020), and future research should explore federated coordination strategies, such as clustered updates or decentralized consensus mechanisms, to ensure scalability and robustness.

Third, the study assumes relatively uniform data distributions across nodes. In real-world deployments, sensor data is often non-IID and may exhibit drift or context-specific anomalies. Addressing this requires implementing personalized federated learning strategies, such as model fine-tuning per node or multi-task learning frameworks. Moreover, incorporating digital twins and simulation environments could help validate and pre-train models before real-world deployment, as suggested by (Mohy-Eddine et al., 2023; Zhong et al., 2023).

Finally, further exploration is needed into operating system compatibility and communication protocols. Platforms like SmartOS or RIOT OS may offer better support for real-time and secure federated processing. Evaluating lightweight protocols such as MQTT-SN, CoAP, or TSN could also enhance communication efficiency in constrained IIoT settings (Scheipel et al., 2022; Wallentowitz et al., 2022). These directions represent meaningful steps toward building a fully scalable, secure, and intelligent edge AI ecosystem for industrial environments.

V. CONCLUSION AND RECOMMENDATION

This study presents a comprehensive investigation into the integration of federated learning and secure edge AI within embedded systems, particularly in industrial IoT contexts. By embedding lightweight, privacy-aware intelligence at the edge, the research effectively addressed critical

challenges of latency, memory usage, inference accuracy, and cybersecurity in resource-constrained environments. The experimental design, which involved structured benchmarking on TinyML-compatible microcontrollers, confirmed the practicality of deploying federated learning in real-time applications while maintaining system responsiveness and reliability. In responding to the research questions, the study validated that federated edge AI systems, when coupled with compartmentalized security mechanisms, can serve as robust, efficient, and scalable solutions for the demands of modern industrial applications.

The performance evaluation revealed a clear advantage of federated architectures over centralized models in key areas. The observed increase in inference accuracy, reduced memory consumption, and decreased inference time demonstrate the capability of edge-based training to adapt more effectively to local data distributions while mitigating dependency on central servers. These gains are especially critical in industrial environments where connectivity may be intermittent or data sensitivity prohibits the transmission of raw data. The empirical results align with theoretical models of distributed AI, reaffirming the growing relevance of context-sensitive computing. Federated learning, as applied in this study, not only provided efficiency but also supported data minimization and privacy preservation principles, which are becoming increasingly important in ethical technology design.

Furthermore, the study demonstrated substantial improvements in real-time behavior, notably in terms of latency, jitter, and task scheduling accuracy. These enhancements are directly applicable to autonomous systems in industrial settings, such as smart factories, predictive maintenance pipelines, or robotic process automation. Reduced system jitter and faster average response times suggest that federated edge AI can uphold deterministic communication requirements without compromising computational integrity. Such responsiveness is critical when milliseconds determine the success of process control or anomaly detection. The findings thus contribute both empirically and conceptually to the body of work supporting the fusion of edge intelligence and industrial automation, opening new possibilities for embedded autonomy.

Security considerations were also rigorously addressed through the implementation of intra-kernel isolation and anomaly detection classifiers. By incorporating CHERIoT-style isolation and Rust-based modules, the study successfully reduced vulnerability exposure while maintaining system efficiency. These results confirm that secure-by-design approaches can be adapted to constrained embedded platforms without introducing prohibitive resource demands. In an era where cyberattacks are increasingly targeting industrial control systems, the value of integrating proactive security at the edge cannot be overstated. This research, therefore, not only affirms technical viability but also aligns with broader societal imperatives for secure, trustworthy AI.

Moreover, it underscores the pressing need for system designs that integrate safety, privacy, and resilience as core values, rather than afterthoughts.

The theoretical implications of the findings support a shift toward hybrid intelligence architectures, in which decentralized computation complements centralized orchestration. This aligns with contemporary thinking in digital transformation frameworks, where local autonomy and global coordination coexist to improve adaptability, sustainability, and user responsiveness. As organizations move toward smarter, more interconnected systems, the role of embedded intelligence becomes central to bridging the gap between cloud services and physical infrastructure. In this context, this research provides an applied pathway for operationalizing theoretical models of distributed cognition and system-level intelligence. The study thus provides a conceptual and practical bridge between AI theory, embedded systems design, and industrial systems engineering.

From a practical perspective, the findings have several implications for engineers, policymakers, and technology developers. Engineers and embedded system designers are encouraged to adopt modular architectures that support federated model updates, local training, and isolation-based fault tolerance. Such modularity not only enhances flexibility but also facilitates post-deployment upgrades, an essential requirement for long-lifecycle industrial assets. For policymakers and regulators, the study underscores the importance of encouraging open, secure, and privacy-preserving AI development, particularly in safety-critical sectors. Regulatory support for open standards and the promotion of trusted computing frameworks could accelerate innovation while safeguarding public interest. Furthermore, industrial stakeholders should prioritize investments in lightweight AI accelerators and real-time operating systems that are compatible with federated architectures to ensure the future-proofing of their operations.

This study also opens promising avenues for future research. One critical area involves improving the energy efficiency of federated edge AI, particularly during retraining and model synchronization cycles. Energy-aware scheduling, battery-aware communication, and the integration of low-power AI accelerators remain important topics for sustainable deployment. Another direction concerns scalability: testing the system in larger, heterogeneous networks and across different domains (e.g., smart agriculture, healthcare monitoring) will help generalize the findings. Additionally, integrating edge-based AI with digital twins can enhance predictive maintenance and system simulation, enabling proactive control in complex environments. These efforts would benefit from interdisciplinary collaborations between computer engineers, AI researchers, human-computer interaction specialists, and domain-specific practitioners.

Ultimately, there is a need for a more in-depth investigation into the socio-technical aspects of embedded AI adoption. While this study primarily focused on technical feasibility, the human factors surrounding trust, explainability, and user interaction in edge-based AI systems warrant further attention. Embedding ethical considerations into the design of federated systems, such as transparency of model behavior and user consent over data usage, can ensure that technological innovation aligns with human values. It is also important to examine how such systems can be made accessible to smaller enterprises or under-resourced regions, thereby avoiding the deepening of technological divides. As edge AI becomes more pervasive, its development must remain inclusive, ethical, and sustainable, guided by both long-term public interest and immediate technical outcomes.

In conclusion, this research establishes the feasibility, efficiency, and security of implementing federated learning and secure edge AI within embedded systems for industrial applications. The approach delivers tangible improvements in system performance, responsiveness, and protection, demonstrating that embedded intelligence can evolve beyond traditional constraints. By leveraging federated architectures and secure-by-design mechanisms, organizations can develop embedded platforms that are not only technically competent but also aligned with emerging needs for autonomy, privacy, and resilience. The study's contributions span theoretical integration, applied system development, and practical recommendations, offering a solid foundation for continued innovation in embedded AI. It is hoped that these insights will inspire further research and responsible implementation in creating intelligent, equitable, and sustainable digital infrastructures.

REFERENCES

Amar, S., Chen, T., Chisnall, D., Domke, F., Filardo, N. W., Liu, K., Norton, R. M., Tao, Y., Watson, R. N. M., & Xia, H. (2023). *CHERIoT: Rethinking security for low-cost embedded systems*.

Arslan Khan, D. X. D. (Jing) T. (2023). *EC: Embedded Systems Compartmentalization via Intra-Kernel Isolation*. IEEE.

Behnke, I., & Austad, H. (2024). Real-Time Performance of Industrial IoT Communication Technologies: A Review. *IEEE Internet of Things Journal*, *11*(5), 7399–7410. https://doi.org/10.1109/JIOT.2023.3332507

Berthelier, A., Chateau, T., Duffner, S., Garcia, C., Blanc, C., Deep, C. B., & Berthelier, A. (2020). Deep Model Compression and Architecture Optimization for Embedded Systems: A Survey Model Compression and Architecture Optimization for Embedded Systems: A Survey Deep Model Compression and Architecture Optimization for Embedded Systems:

A Survey. *Journal of Signal Processing Systems*, 10. https://doi.org/10.1007/s11265-020-01596-1ï

Bosch, J., & Olsson, H. H. (2021). Digital for real: A multicase study on the digital transformation of companies in the embedded systems domain. *Journal of Software: Evolution and Process*, *33*(5). https://doi.org/10.1002/smr.2333

Brasoveanu, A., Moodie, M., & Agrawal, R. (2020). Textual evidence for the perfunctoriness of independent medical reviews. *CEUR Workshop Proceedings*, *2657*, 1–9. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Chang, L., Zhang, Z., Li, P., Xi, S., Guo, W., Shen, Y., Xiong, Z., Kang, J., Niyato, D., Qiao, X., Wu, Y., Chang, L. Y., Zhang, Z., Li, P., Xi, S., Guo, W., Wu, Y., Shen, Y. K., Kang, J. W., & Niyato, D. (2022). 6G-Enabled Edge AI for Metaverse: Challenges, Methods, and Future Research Directions. In *Journal of Communications and Information Networks* (Vol. 7, Issue 2).

Du, Y., Dharsee, K., Zhou, J., Shen, Z., Walls, R. J., & Criswell, J. (2022). *Holistic Control-Flow Protection on Real-Time Embedded Systems with Kage*. https://www.usenix.org/conference/usenixsecurity22/presentation/du

Gill, S. S., Golec, M., Hu, J., Xu, M., Du, J., Wu, H., Walia, G. K., Murugesan, S. S., Ali, B., Kumar, M., Ye, K., Verma, P., Kumar, S., Cuadrado, F., & Uhlig, S. (2024). *Edge AI: A Taxonomy, Systematic Review and Future Directions*. https://doi.org/10.1007/s10586-024-04686-y

Guan, H., Yap, P. T., Bozoki, A., & Liu, M. (2024). Federated learning for medical image analysis: A survey. In *Pattern Recognition* (Vol. 151). Elsevier Ltd. https://doi.org/10.1016/j.patcog.2024.110424

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. In *Foundations and Trends in Machine Learning* (Vol. 14, Issues 1–2, pp. 1–210). Now Publishers Inc. https://doi.org/10.1561/2200000083

Malik, P. K., Sharma, R., Singh, R., Gehlot, A., Satapathy, S. C., Alnumay, W. S., Pelusi, D., Ghosh, U., & Nayak, J. (2021). Industrial Internet of Things and its Applications in Industry 4.0: State of The Art. *Computer Communications*, *166*, 125–139. https://doi.org/10.1016/j.comcom.2020.11.016

Mohy-Eddine, M., Guezzaz, A., Benkirane, S., Azrour, M., & Farhaoui, Y. (2023). An Ensemble Learning Based Intrusion Detection Model for Industrial IoT Security. *Big Data Mining and Analytics*, *6*(3), 273–287. https://doi.org/10.26599/BDMA.2022.9020032

Molęda, M., Małysiak-Mrozek, B., Ding, W., Sunderam, V., & Mrozek, D. (2023). From Corrective to Predictive Maintenance—A Review of Maintenance Approaches for the

Power Industry. In *Sensors* (Vol. 23, Issue 13). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/s23135970

Patra, T. K. (2022). Data-Driven Methods for Accelerating Polymer Design. *ACS Polymers Au*, *2*(1), 8–26. https://doi.org/10.1021/acspolymersau.1c00035

Qin, Z., Li, G. Y., & Ye, H. (2020). *Federated Learning and Wireless Communications*. http://arxiv.org/abs/2005.05265

Roth, R. E. (2021). Cartographic Design as Visual Storytelling: Synthesis and Review of Map-Based Narratives, Genres, and Tropes. *Cartographic Journal*, *58*(1), 83–114. https://doi.org/10.1080/00087041.2019.1633103

Scheipel, T., Ribeiro, L. B., Sagaster, T., & Baunach, M. (2022). SmartOS: An OS Architecture for Sustainable Embedded Systems. *FrÃ¼hjahrstreffen Der GI-Fachgruppe Betriebssysteme (FGBS '22), March 17â•fi18, 2022, Trondheim, Norway*, *1*. https://doi.org/10.18420/fgbs2022f-01

Sharma, A., Sharma, S., Tanksalkar, S. R., Torres-Arias, S., & Machiry, A. (2024). Rust for Embedded Systems: Current State and Open Problems. *CCS 2024 - Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security*, 2296–2310. https://doi.org/10.1145/3658644.3690275

Shen, Y., Shao, J., Zhang, X., Lin, Z., Pan, H., Li, D., Zhang, J., & Letaief, K. B. (2023). *Large Language Models Empowered Autonomous Edge AI for Connected Intelligence*. http://arxiv.org/abs/2307.02779

Sipola, T., Alatalo, J., Kokkonen, T., & Rantonen, M. (2022). Artificial Intelligence in the IoT Era: A Review of Edge AI Hardware and Software. *Conference of Open Innovation Association, FRUCT*, *2022-April*, 320–331. https://doi.org/10.23919/FRUCT54823.2022.9770931

Stanislava Soro. (2020). *TinyML for Ubiquitous Edge AI*.

Tiddens, W., Braaksma, J., & Tinga, T. (2022). Exploring predictive maintenance applications in industry. *Journal of Quality in Maintenance Engineering*, *28*(1), 68–85. https://doi.org/10.1108/JQME-05-2020-0029

Tsiknas, K., Taketzis, D., Demertzis, K., & Skianis, C. (2021). Cyber Threats to Industrial IoT: A Survey on Attacks and Countermeasures. *Internet of Things*, *2*(1), 163–186. https://doi.org/10.3390/iot2010009

Veloso, B., Ribeiro, R. P., Gama, J., & Pereira, P. M. (2022). The MetroPT dataset for predictive maintenance. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-022-01877-3

Wallentowitz, S., Kersting, B., & Dumitriu, D. M. (2022). Potential of WebAssembly for Embedded Systems. *2022 11th Mediterranean Conference on Embedded Computing, MECO 2022*. https://doi.org/10.1109/MECO55406.2022.9797106

Xenofontos, C., Zografopoulos, I., Konstantinou, C., Jolfaei, A., Khan, M. K., & Choo, K.-K. R. (2021). *Consumer, Commercial and Industrial IoT (In)Security: Attack Taxonomy and Case Studies*. http://arxiv.org/abs/2105.06612

Ye, M., Fang, X., Du, B., Yuen, P. C., & Tao, D. (2023). *Heterogeneous Federated Learning: State-of-the-art and Research Challenges*. http://arxiv.org/abs/2307.10616

Yu, R., Del Nin, F., Zhang, Y., Huang, S., Kaliyar, P., Zakto, S., Conti, M., Portokalidis, G., & Xu, J. (2022). *Building Embedded Systems Like It's 1996*. http://arxiv.org/abs/2203.06834

Zhong, D., Xia, Z., Zhu, Y., & Duan, J. (2023). Overview of predictive maintenance based on digital twin technology. In *Heliyon* (Vol. 9, Issue 4). Elsevier Ltd. https://doi.org/10.1016/j.heliyon.2023.e14534