

Mechanical Edge Intelligence: From Physics-Informed Condition Monitoring to Autonomous Closed-Loop Control

Junhao Wang

*Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham
Ningbo China, Ningbo, China
ssyiw28@nottingham.edu.cn*

Abstract. With the rapid development of Industry 4.0, many more IoT sensors are being applied to monitor the condition of mechanical equipment continuously. To detect early-stage defects, such as bearing micro-cracks or cavitation, high-frequency sampling is often necessary; thus, the resulting data streams are large and non-stationary. Transmission of this data to the cloud platform will introduce bandwidth constraints and delay, thus failing to meet the requirements for real-time closed-loop control. Mechanical Edge Intelligence is a new way to solve the above problem by reducing the distance for computing and letting sensors perform local analysis for more prompt responses. Due to memory and computational constraints on many industrial microcontrollers, diagnostic algorithms cannot be used. Tiny Machine Learning reduces model size and computational requirements for edge deployment in this paper. Physics-Informed Edge AI is another direction of research that incorporates physical knowledge into the training process of learning models to enhance their robustness and interpretability in the presence of noise and variations. Applications in rotating and fluid machinery demonstrate how edge intelligence can support the transition from condition monitoring to autonomous closed-loop control.

Keywords: Edge intelligence, physics-informed AI, condition monitoring, closed-loop control, smart sensors

1. Introduction

With the emergence of Industry 4.0, Internet of Things (IoT) and intelligent manufacturing have begun to introduce sensor networks in many areas of industrial equipment. Now that their purpose is no longer restricted to basic data collection, they also need to be used for continuous monitoring of machine status and timely alert of maintenance or control needs. A wider application of sensors has led to an increase in the quantity of operating data produced by mechanical systems; vibration, acoustic emission, and other high-frequency signals are now among them.

Mechanical equipment may have only developed a problem that is difficult to spot at first. Fault-related features may be transient, low in amplitude, and easily masked by load changes or background noise. For example, early detection of failures in a planetary gear system of a rotating machine may require a sample rate as high as 48 kHz or more [1]. Although the above high-frequency measurements can be used to detect small-scale deterioration, they also have relatively

large requirements for data transmission, storage and processing. All the raw signals are sent to the same cloud platform and there will be a network bottleneck and transmission delays. The delay will be very difficult to handle in a high-precision machine that needs to work in real time [2].

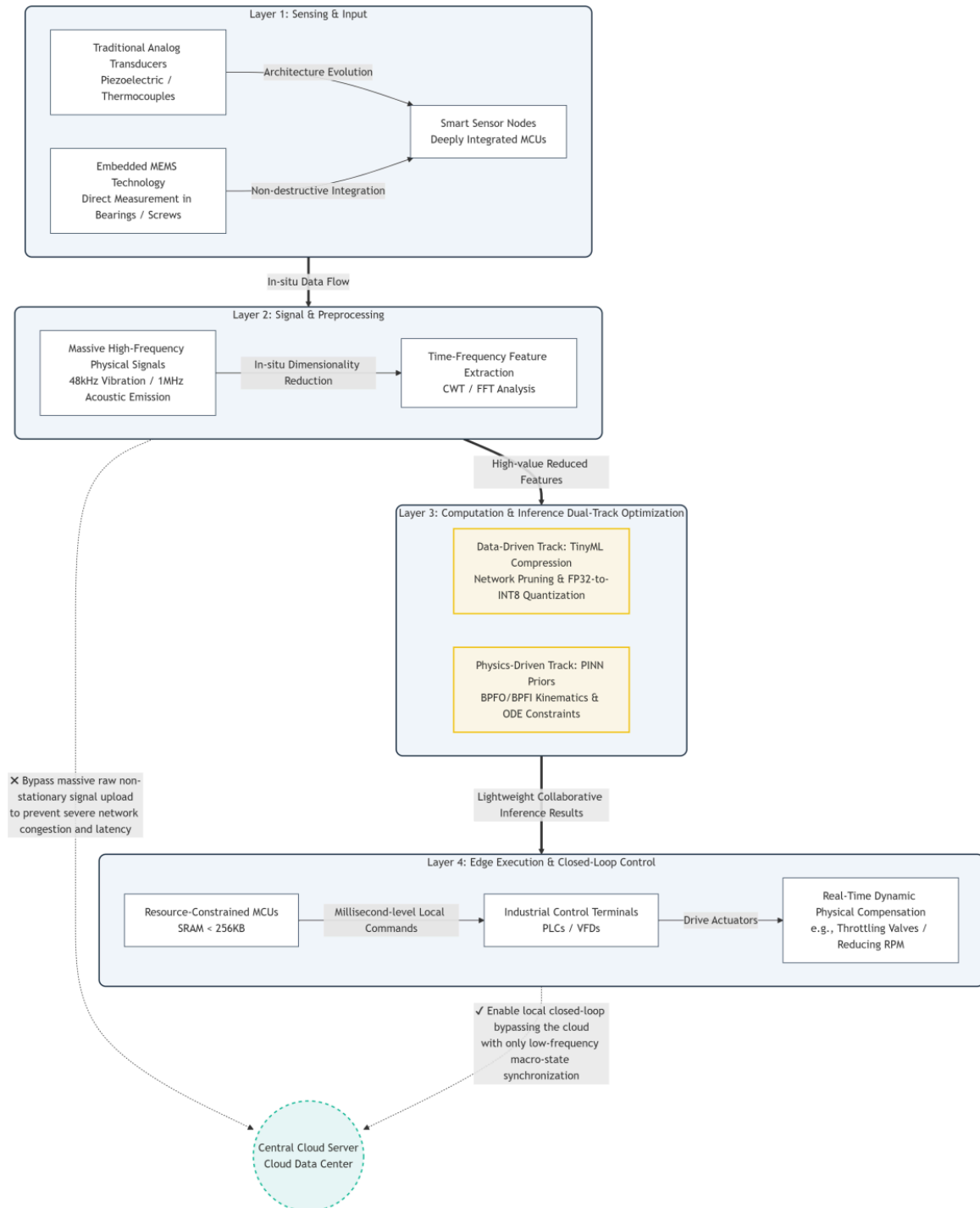


Figure 1. Holistic architecture of mechanical edge intelligence: from in-situ physical sensing to localized closed-loop control

Given the above reasons, Edge Intelligence is a better computation environment for mechanical monitoring and control. Rather than transmitting all sensor data remotely for processing, some of it

can be analyzed near the place where the sensors are located. Only the extracted diagnostic features, warning information and control-related results need to be sent to the controllers, actuators or cloud platforms. It will be less data transmission and thus have a faster response speed to abnormal operation conditions.

Therefore, this paper will refer to edge AI as a computing architecture and also as an expansion of the sensor. An intelligent sensor should be able to sense and process local information in the mechanical system to support local decision-making in this view. The review focuses on how this idea can be used to solve the problem of high-frequency physical diagnosis for rotating machinery and fluid machinery. Figure 1 shows the general architecture and data flow of mechanical edge intelligence for in-situ sensing and localized closed-loop control.

2. Evolution of sensors in mechanical engineering

2.1. The paradigm shift from passive perception to active in-situ computation

The Sensing Devices of older mechanical condition monitoring systems were generally passive. Piezoelectric accelerometers and thermocouples are typical examples, and both are commonly used to convert vibrations and temperature variations into electric signals. These signals were then sent through long cables to the main PC-based data acquisition unit for saving and analysis. The above arrangement was feasible in many traditional industrial environments, but it is relatively inconvenient in modern factories that have been compacted and spread out due to the arrangement of equipment and sensing points. A long wiring path will have higher installation and maintenance costs, and it may also be more prone to electromagnetic interference in plants with many electrical and mechanical devices [3].

Due to the above deficiency, a practical constraint has arisen in mechanical instrumentation. Smart sensors are not traditional transducers; they integrate a low-power microcomputer with the sensing module, and some basic signal processing and feature extraction can be performed locally at the measurement site [3, 4]. Recently, there have been further extensions of this idea in structural integration. MEMS sensors can be embedded in ball screws and spindle bearings now, for example, and conditions inside the mechanical structure of a machine can be observed. If done properly, the added integration will not reduce the performance or service life of the components and will still provide real-time monitoring [4]. Therefore, at present, the sensor has transcended its original function as only a signal emitter to become an integrated microcomputer for the whole machine system.

2.2. The complexity of mechanical signals and high-frequency demands

Mechanical signals measured during real operation are rarely clean or steady. They are often affected by load changes, speed fluctuations, assembly errors, and background vibration. Compared with many common audio or image signals, these data usually show stronger non-linearity and non-stationarity. In induction motors, for instance, defects such as broken rotor bars or air-gap eccentricity can disturb the electromagnetic field inside the machine. The resulting motor current may contain transient harmonic components, which are not always easy to identify through a simple frequency-domain analysis [5]. Methods such as the Fast Fourier Transform (FFT) remain useful, but they can miss time-varying details when the signal changes rapidly.

The problem becomes more demanding when the target is an early-stage fault. Small cracks, weak impacts, or cavitation-related events may only appear in a high-frequency range. Conventional

vibration monitoring is often concerned with signals below 20 kHz, whereas micro-cracks in low-speed rolling bearings or cavitation in fluid machinery may require acoustic emission or ultrasonic measurement methods [6]. In these cases, sampling rates can rise from tens of kilohertz to nearly 1 MHz. The benefit is that subtle fault information becomes more visible; the drawback is the sudden growth of data volume. Sending all of these raw signals to a remote server would place heavy pressure on bandwidth and would also introduce delay. For systems that rely on fast closed-loop responses, local processing is therefore not just a matter of efficiency, but part of the control requirement itself.

2.3. Severe resource constraints in extreme industrial environments

Relocation of computation to the sensor end does not solve all problems. Industrial edge devices need to work in different environments than laboratory computing facilities. Most of the smart sensors are based on very low-power microcontrollers, such as the ARM Cortex-M series of chips; therefore, they must meet strict requirements for size, cost, power consumption and durability. At the same time, there may be problems such as high temperature, dust and mechanical vibration or electromagnetic interference in extended use [3]. Therefore, it cannot be simply ported from a desktop computer to an edge device. In practice, there is often a trade-off among the accuracy of the model, speed and hardware cost [7].

Memory is generally a relatively limited resource. Recently, many TinyML studies have shown that many industrial microcontrollers have less than 256KB of static random-access memory (SRAM), and the flash memory is also relatively small [8]. This is much lower than the resource levels provided by a smartphone, desktop processor or cloud GPU. Therefore, many traditional data-driven AI models cannot be deployed directly on these devices, even if they perform well during offline testing. Therefore, the models for edge diagnosis need to be relatively small, simple and closely aligned with the actual task. Now, to make intelligent diagnosis more feasible in real industrial environments, methods of model compression, lightweight feature extraction and physics-informed optimisation have been introduced.

3. Edge AI as a mechanical tool

3.1. Traditional signal processing meets lightweight edge ML

Many applications of engineering do not need to run large deep learning models on raw time-series data because they do not provide new diagnostic information. A more efficient way is to use the old mechanical signal processing method and a light classifier together. Continuously Wavelet Transform can be used to obtain features in the time-frequency domain. The above characteristics are suitable for classification by a light-weight machine learning model, such as Random Forest, and thus can achieve low-cost fault pattern recognition [9]. FFT can be used to transform a high-frequency sound signal into a frequency-domain representation and then extract harmonic features. The above attributes can serve as the input features for a classifier, and for example, in a Support Vector Machine (SVM) algorithm, reliable fault detection in low-memory devices has been achieved [5, 9].

3.2. Engineering dimensionality reduction via TinyML

To realise in-depth fault diagnosis by deep neural networks, it is necessary to first compress the neural network for deployment. TinyML is a system of many techniques to fit machine learning models on small-scale embedded devices and is used to reduce dimensionality effectively. Pruning can also be used to reduce the number of connections and neurons in a network to make it smaller without losing the fault detection ability. Pruning is generally understood as the reduction of the structure by removing non-functional sections and keeping only the necessary, operational parts [10]. Quantization is also a reduction in memory; by using low-precision numbers, such as 8-bit integers, for the weights of the neurons, a considerable improvement in efficiency can be achieved [11]. Recent studies have shown that TinyML compression can deploy neural networks on very small edge devices, such as Cortex-M4 microcontrollers with 192kB of RAM, and meet the latency requirements [11]. Thus, TinyML can be used to realise real-time diagnosis with hardware constraints for deep learning.

3.3. Physics-informed edge AI: embedding mechanical priors

An important trend in mechanical edge intelligence is the shift from pure data-driven to physics-informed learning. Data-driven models generally have good predictive accuracy but can fail to work well with changing operating conditions; in particular, industrial environments are often characterized by changing rotation speed, load conditions, and noise levels, all of which can affect the performance of statistical models [12]. By embedding mechanical priors into machine learning, physics-informed approaches solve this issue.

There are two major ways of doing this: feature weighting and integration of physical principles into the model architecture. Feature weighting can focus on degradation-related features invariant to changing speeds and loads [12]. On the other hand, one can explicitly use mechanical concepts in the neural network architecture itself. For example, ball pass frequencies BPFO and BPFI can be considered physical references for the spectral attention mechanism, making sure the attention focuses on physically meaningful harmonics [12].

More generally, one can introduce classical mechanical models to the learning process. The Jeffcott rotor model is one such classic; its introduction as latent Ordinary Differential Equations constrains the learning process to follow Newtonian physics and thus disfavors unreasonable outcomes [13].

4. Application scenarios in mechanical systems

4.1. Rotating machinery and induction motors

As the main power drivers of industrial machinery, AC induction motors and rotating equipment are particularly sensitive to mechanical degradation, thus requiring multi-modal on-site data acquisitions (e.g., vibration, electrical current, and temperature) as a premise of predictive maintenance [14]. Although traditional monitoring solutions depend greatly on distant analysis, the latest innovations highlight the advantages of performing Motor Current Signature Analysis (MCSA) and high-frequency vibration data processing right on the extremely resource-limited edge devices. By performing local extraction of the statistical features and dimensionality reduction, lightweight classifier models can be run on edge nodes with even extremely limited capacity of SRAM (as little as 264 KB) to effectively distinguish rotor broken bars and bearing degradation problems [5].

Moreover, by utilizing 1D Convolutional Neural Networks (1D-CNNs), it becomes possible to perform multi-class fault recognition directly from time-series vibration signals on edge nodes with millisecond-level delay [12]. Such localized machine learning models effectively eliminate the cloud communication delay, achieving instantaneous fault detection of mechanical unbalance, misalignment, and complex structure-related damages.

4.2. Fluid machinery and complex flow systems

Fluid dynamics of mechatronic systems (e.g., centrifugal pumps, gas and liquid compressors, and hydraulic valves) brings much higher complexity than that of the rigid-body dynamics. Local pressure drop and bubble collapse lead to destructive nonlinear high-frequency pressure fluctuations and acoustic emission (AE) noise. Such complicated flow dynamics can significantly corrupt data-driven AI algorithms. Nevertheless, the latest industrial practices show that by applying localized feature sets informed by physics knowledge and running lightweight autoencoder algorithms on the microcontroller itself, one can isolate these fluid-dynamic anomalies in spite of the interference [15]. Thus, edge-based smart monitoring instantaneously detects early-stage fluid cavitation or leakage in hydraulic valves, while avoiding unnecessary data upload to the cloud server, hence ensuring that diagnostic results always conform to fluid dynamics.

4.3. From condition monitoring to intelligent closed-loop control

The ultimate goal of edge intelligence in the field of mechatronics goes beyond the open-loop condition monitoring (e.g., issuing warning messages) to intelligent closed-loop control. From the perspective of classical control theory, the delays associated with data upload and cloud computation are manifested in phase lag, leading to decreased stability margins of fast mechanical systems. Latest studies emphasize the intrinsic ability of extreme edge intelligence in this regard. An adaptive framework of smart sensing that performs on-site data analytics can help to balance the latency vs. accuracy trade-off and ensure that processed diagnostic results are delivered quickly enough for dynamic systems to preserve stability [7]. For further clarification of the control paradigm shift caused by extreme edge computing, Figure 2 shows the schematic difference between the conventional cloud-centric control loop and the proposed intelligent edge-control loop. The key advantage of the latter is its ability to skip the cloud stage and directly interface with the Programmable Logic Controller (PLC) or the Variable Frequency Drive (VFD). Hence, once an anomaly is detected by an in-sensor computation module, the edge node immediately triggers control commands sent to actuators, such as throttling of a hydraulic valve or reducing the motor revolutions per minute (RPM) [2].

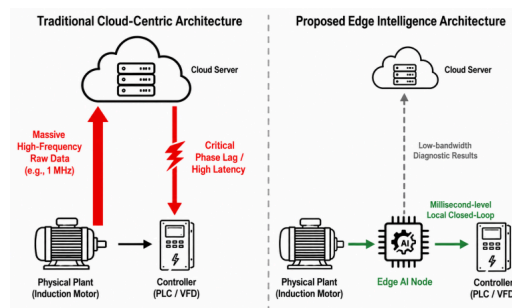


Figure 2. Topological comparison of control feedback loops between traditional cloud-centric and the proposed edge intelligence architectures

5. Engineering challenges and future outlook

Despite the promising potential for mechatronic application of edge intelligence, its practical implementation faces major challenges in large-scale engineering.

First, it is the hardware's physical survival threshold under the conditions of extreme operation. Besides the extremely limited Static Random-Access Memory (SRAM) capacity, edge nodes in the field of industry are also required to maintain reliable operation in a harsh environment of high temperature, dense dust, mechanical vibration, and high-frequency EM interference. This requires extremely tough design, material selection, and heat dissipation measures, which go beyond those of general-purpose computing devices.

Second, there is an absolute demand of industrial closed-loop systems for safety and reliability, which drives the pressing need for Explainable Artificial Intelligence (XAI). In terms of servo control and automated actuators, it is necessary for the decision-making model to have sufficient theoretical support and traceability. Traditional data-driven black-box models find it difficult to earn the industry's trust at the engineering execution layer. In the future, algorithms for edge diagnostic purposes should incorporate deeper insight into mechanics and kinematic principles, guaranteeing a strict physical interpretation of every step in fault assessment and feedback trigger, just like PID parameters do in classical control theory.

Finally, there is the problem of interoperability and integration in heterogeneous systems. Modern active computing sensors must seamlessly interact with multiple legacy control modules and old programmable logic controllers (PLCs) of many years operation history in the factory. This involves not only solving the compatibility problem of underlying bus protocols, but also ensuring absolute time synchronization between new-generation edge nodes and legacy actuators at millisecond control cycle scale.

6. Conclusion

To conclude, extreme edge intelligence is not merely a migration of the underlying computing architecture, but rather a revolution of the control loop towards the front line of physical perception. Faced with the overwhelming volume of raw physical signal acquisitions, as well as the critical latency due to traditional cloud-based architectures, algorithmic extreme compression and physics injection emerge as inevitable ways out. By compressing massive models to tiny yet powerful analytical units under extremely limited memory capacity, and limiting their decisions within rigid constraints provided by rigid-body and fluid physics, edge nodes can realize millisecond-scale feedback on physical parameters of extreme industrial systems. Such radical transition from passive observation to active on-site computation establishes an indispensable bridge between the engineering stages of condition monitoring and adaptive closed-loop control.

References

- [1] He, C., Yassenjiang, J., Lv, L., et al. (2024). Gearbox fault diagnosis based on MSCNN-LSTM-CBAM-SE. *Sensors*, 24(14), 4682.
- [2] Li, X., Jing, T., Wang, Y., et al. (2026). Cloud-edge-end collaborative SC3 system in smart manufacturing: A survey. *Comput. Mater. Contin.*, 87(2), 1–10.
- [3] Peng, C.-C. (2021). Low-cost embedded real-time handheld vibration smart sensor for industrial equipment onsite defect detection. *IEEE Open J. Ind. Electron. Soc.*, 2, 469–478.
- [4] Beyer, V. (2023). Structural integration of smart sensors for the industrial internet of things. Oktober.
- [5] Ūselis, R., Serackis, A., & Pomarnacki, R. (2025). Signal processing optimization in resource-limited IoT for fault prediction in rotating machinery. *Electronics*, 14(18), 3670.

- [6] Ompusunggu, A. P., & Carcel, C. R. (2023). Low-cost vibration sensor with low frequency resonance for condition monitoring of low speed bearings: A feasibility study. *IET Conf. Proc.*, 2023(16), 158–164.
- [7] Ballotta, L., Peserico, G., Zanini, F., et al. (2024). To compute or not to compute? Adaptive smart sensing in resource-constrained edge computing. *IEEE Trans. Netw. Sci. Eng.*, 11(1), 736–749.
- [8] Lin, J., Zhu, L., Chen, W.-M., et al. (2023). Tiny machine learning: Progress and futures [Feature]. *IEEE Circuits Syst. Mag.*, 23(3), 8–34.
- [9] Le, T. M., Tran, H. M., Wang, K., et al. (2025). An internet-of-things-integrated deep learning model for fault diagnosis in industrial rotating machines. *IEEE Access*, 13, 57266–57286.
- [10] Nguyen, H.-A.-H., & Kim, C. H. (2025). Efficient bearing fault diagnosis for edge computing using grayscale spectrograms and hybrid neural model compression. *IEEE Access*, 13, 147494–147510.
- [11] Fu, L., et al. (2023). EdgeCog: A real-time bearing fault diagnosis system based on lightweight edge computing. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–11.
- [12] Lu, H., et al. (2023). A physics-informed feature weighting method for bearing fault diagnostics. *Mechanical Systems and Signal Processing*, 191, 110171.
- [13] Ullah, R., et al. (2026). Conformer-PhyFaultNet: Physics-informed spectral attention conformer for generalizable bearing fault diagnosis. *IEEE Internet of Things Journal*, 13(8), 17538–17552.
- [14] Yousuf, M., Alsuwian, T., Amin, A. A., et al. (2024). IoT-based health monitoring and fault detection of industrial AC induction motor for efficient predictive maintenance. *Measurement and Control*, 57(8), 1146–1160.
- [15] Nguyen, V. L., Chu, D. L., & Nguyen, V. H. (2026). From dataset to deployment: A full-stack autoencoder-based anomaly detection pipeline for gas turbine BLDC fuel pumps. *IEEE Access*, 14, 58103–58122.