Article

Multi-Source Data Fusion Diagnosis of Rotating Machinery Based on Cloud-Edge Collaborative Architecture

Tingting Zhou[1,*], Xiaowei Xu[2] and Nan Ma[3]

[1] Taiyuan University of Science and Technology, Taiyuan, China
[2] Shenyang University of Technology, Shenyang, China
[3] Liaoning Technical University, Huludao, China
* Correspondence: Tingting Zhou, Taiyuan University of Science and Technology, Taiyuan, China

Abstract: Rotating machinery forms the operational backbone of industrial systems, where early and accurate fault diagnosis directly impacts safety, efficiency, and lifecycle management. Traditional diagnostic approaches often rely on single-source sensor data or centralized processing, leading to latency, bandwidth bottlenecks, and limited adaptability in dynamic operating environments. This research introduces a cloud-edge collaborative architecture specifically engineered for multi-source data fusion in rotating machinery diagnosis. The framework distributes computational responsibilities across hierarchical layers: edge devices perform real-time signal preprocessing, feature extraction, and lightweight anomaly detection using time-frequency domain representations; the cloud layer orchestrates heterogeneous data integration, model retraining, cross-machine knowledge transfer, and long-term health trend analysis. Data fusion occurs at two critical stages—sensor-level synchronization and semantic-level alignment—ensuring consistency across vibration, acoustic emission, temperature, and current signatures despite varying sampling rates and physical units. A unified normalization protocol and adaptive weighting mechanism enable robustness against missing modalities and sensor degradation. Validation was conducted across three industrial-grade testbeds simulating bearing faults, gear mesh defects, and rotor imbalance under variable load and speed conditions. Results demonstrate a 94.7% average classification accuracy with sub-second diagnostic latency at the edge and a 32% reduction in false alarm rate compared to monolithic cloud-only baselines. The architecture maintains diagnostic fidelity during network intermittency through local state retention and incremental model updates. This work establishes a scalable, modular, and operationally resilient paradigm for intelligent condition monitoring in distributed industrial ecosystems.

Keywords: data fusion; rotating machinery; cloud-edge computing; fault diagnosis

Received: 28 July 2026
Revised: 10 September 2026
Accepted: 21 September 2026
Published: 24 September 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Industrial Significance of Rotating Machinery Health Monitoring

Rotating machinery—including electric motors, centrifugal pumps, steam turbines, and mechanical gearboxes—constitutes the kinetic backbone of energy generation, process manufacturing, transportation, and infrastructure systems. Their operational integrity directly governs system availability, energy efficiency, and personnel safety. Undetected incipient faults—such as bearing spalls, rotor imbalance, misalignment, or gear tooth pitting—can rapidly propagate into catastrophic failures, triggering unplanned shutdowns, collateral damage to auxiliary systems, and hazardous release events [1]. Traditional time-based maintenance schedules prove increasingly inadequate in high-variability operational environments, while reactive repair strategies incur escalating lifecycle costs and compromise production continuity [2, 3]. Consequently, industry-wide imperatives now prioritize condition-based and predictive health monitoring frameworks capable of early fault detection, root-cause localization, and remaining useful life estimation. This paradigm shift demands robust diagnostic capability under real-world

constraints: heterogeneous sensor modalities, variable signal-to-noise ratios, non-stationary loading conditions, and stringent latency requirements [4]. Cloud-edge collaborative architectures emerge as a structurally coherent response, enabling distributed feature extraction, adaptive model updating, and scalable inference without compromising diagnostic fidelity or operational responsiveness [5, 6].

1.2. Limitations of Conventional Diagnostic Architectures

Conventional diagnostic architectures for rotating machinery exhibit several critical limitations that impede robust, real-time health assessment. First, they typically process sensor data in isolation, neglecting cross-modal correlations among vibration, acoustic emission, thermal, and current signatures. Second, centralized cloud-only computation introduces substantial latency, rendering them unsuitable for time-critical fault mitigation. Third, these systems lack native support for asynchronous, heterogeneous sampling rates and irregular temporal alignment across modalities, leading to information loss during resampling or interpolation. Fourth, scalability deteriorates markedly as device density increases, due to unbounded bandwidth demands and computational bottlenecks at the central node. Fifth, end-to-end dependency on continuous network connectivity renders them highly vulnerable to packet loss, intermittent links, or edge disconnection events—conditions frequently encountered in industrial wireless environments [3, 5]. Collectively, these constraints undermine diagnostic accuracy, responsiveness, and operational resilience, particularly under dynamic load conditions or evolving fault morphologies. Consequently, a paradigm shift toward distributed intelligence with coordinated cloud-edge orchestration is imperative to enable adaptive, low-latency, and communication-resilient multi-source fusion diagnostics.

1.3. Core Contribution and Structural Overview

This paper introduces a novel cloud-edge collaborative architecture for multi-source data fusion diagnosis of rotating machinery, designed to overcome latency, bandwidth, and heterogeneity constraints inherent in centralized approaches [5]. The framework enables synchronized ingestion of vibration, acoustic, thermal, and current signatures across distributed edge nodes, where lightweight feature extraction and anomaly pre-screening occur in real time [1]. Edge-local intelligence reduces upstream data volume while preserving discriminative information through adaptive quantization and semantic compression. At the cloud layer, hierarchical fusion logic integrates edge-processed streams with contextual metadata, applying dynamic weighting based on sensor reliability, operational mode, and fault evolution stage. A closed-loop feedback mechanism refines edge models using cloud-derived global patterns, ensuring continual adaptation to shifting degradation dynamics. The contribution lies in the co-design of distributed inference protocols, context-aware fusion operators, and resource-aware orchestration policies—unifying theoretical rigor with deployable engineering pragmatism. Subsequent sections formalize the architectural primitives, derive convergence guarantees for distributed fusion, validate performance against industrial benchmarks, and analyze scalability, robustness, and deployment readiness under realistic operational constraints.

2. Literature Review

2.1. Evolution of Multi-Source Sensing in Mechanical Diagnostics

The evolution of multi-source sensing in mechanical diagnostics reflects a paradigm shift from rudimentary single-sensor thresholding toward sophisticated multi-channel correlation analysis. Early approaches relied on isolated vibration or temperature measurements, applying fixed amplitude thresholds to trigger alarms—a strategy inherently limited by sensor-specific noise and contextual ambiguity. Subsequent advances emphasized synchronized acquisition across heterogeneous modalities, enabling temporal alignment of acoustic emission, current signature, and thermal imaging streams [5]. This synchronization unlocked modality complementarity: for instance, high-

frequency vibration features exposed incipient bearing faults, while low-frequency current harmonics revealed load-induced torque anomalies. Crucially, modern fusion frameworks prioritize physical interpretability—transforming raw signal combinations into physically grounded diagnostic indicators such as fault-related energy ratios $E_{\text{fault}}/E_{\text{total}}$ or phase-coherence metrics $\gamma_{\text{cross}}(f)$. These developments collectively support robust fault isolation under variable operating conditions, moving beyond statistical anomaly detection toward mechanistic reasoning [2, 4]. The integration of domain knowledge—such as gear mesh frequency harmonics or rotor dynamic constraints—further anchors fused representations in first-principles physics, enhancing generalizability across machine families and operational regimes.

2.2. Computational Paradigms for Real-Time Diagnosis

Centralized computing paradigms offer high model complexity and inference accuracy but suffer from latency bottlenecks, bandwidth saturation, and inflexible update cycles—particularly problematic in time-critical rotating machinery diagnostics. Distributed edge-only architectures reduce latency and enhance data privacy yet constrain model capacity due to limited memory, compute, and energy budgets, often compromising diagnostic fidelity under multi-source heterogeneity. Hybrid cloud-edge collaborative models reconcile these tensions by partitioning inference tasks: lightweight anomaly detection and feature extraction occur at the edge for sub-100 ms responsiveness, while computationally intensive fusion, retraining, and cross-machine knowledge distillation are delegated to the cloud. This division enables dynamic model updates, adaptive threshold calibration, and federated learning across geographically dispersed assets [5]. Crucially, the architecture supports hierarchical temporal modeling—where edge nodes process high-frequency vibration streams with $\tau_{\text{edge}} \leq 5\,\text{ms}$ resolution, while cloud layers integrate long-horizon spectral, thermal, and acoustic signatures [1]. Memory footprint remains bounded at the edge through quantized neural modules and sparse attention mechanisms, whereas cloud-side elasticity accommodates evolving model topologies. Update agility improves by two orders of magnitude relative to monolithic deployments, enabling rapid adaptation to new fault patterns without system-wide downtime [2].

2.3. Fusion Strategies Across Abstraction Levels

Data-level fusion integrates raw sensor measurements prior to any feature extraction, offering maximal information preservation but demanding strict temporal synchronization and high bandwidth—challenging under heterogeneous sampling rates and latency constraints inherent in rotating machinery monitoring. Feature-level fusion operates on extracted representations, enabling partial compensation for temporal misalignment and sensor heterogeneity through alignment-aware embedding spaces; however, it risks information loss if discriminative patterns are discarded during dimensionality reduction. Decision-level fusion aggregates outputs from independent diagnostic models, providing robustness against individual sensor failures and noise-induced degradation, particularly advantageous under low signal-to-noise ratios where early-stage features become unreliable. Each strategy exhibits distinct trade-offs: data-level methods require precise clock synchronization and suffer from redundant computation across modalities; feature-level approaches necessitate domain-invariant representation learning to harmonize disparate physical units and spectral characteristics; decision-level schemes depend critically on calibration consistency across classifiers and may obscure root-cause interpretability. Cloud-edge collaborative architectures modulate these trade-offs by distributing fusion stages—edge nodes perform lightweight feature or decision fusion under latency constraints, while cloud platforms execute computationally intensive data-level or hierarchical fusion with global context awareness [3]. The optimal configuration depends on real-time requirements, communication reliability, and the degree of sensor heterogeneity present in the deployment environment [4].

3. Materials and Methods

3.1. System Architecture and Layered Responsibility Allocation

The proposed cloud-edge collaborative architecture adopts a rigorously partitioned three-tier structure to optimize latency, bandwidth, and computational efficiency in multi-source data fusion for rotating machinery diagnosis. Edge nodes constitute the foundational layer, implemented via microcontroller-based acquisition units deployed proximal to sensors; they execute real-time signal sampling, hardware-triggered preprocessing, and lightweight feature extraction—such as time-domain statistical moments and spectral energy ratios—while enforcing strict temporal synchronization across heterogeneous sensor modalities. Fog gateways form the intermediate aggregation layer, responsible for low-latency data ingestion from multiple edge nodes, time-aligned buffering, and preliminary cross-modal fusion using adaptive weighted averaging and entropy-driven modality selection. Communication between edge and fog layers adheres to MQTT over TLS with QoS level 1, ensuring reliable delivery under intermittent connectivity. The cloud platform serves as the centralized intelligence layer, orchestrating deep ensemble models, executing long-term trend analysis on historical fused datasets, and enabling cross-site federated learning through encrypted model parameter exchange. Inter-layer communication employs RESTful APIs with JSON payloads for fog-to-cloud interactions, incorporating OAuth2.0 authentication and payload compression [7]. Functional boundaries are enforced via strict API contracts: edge nodes never perform classification, fog gateways never retain raw waveforms beyond $\tau_{\text{buffer}} = 30$ seconds, and cloud services never issue direct actuation commands. This layered responsibility allocation ensures deterministic real-time behavior at the edge while preserving scalability and analytical depth in the cloud.

3.2. Multi-Source Data Acquisition and Preprocessing Pipeline

Synchronized multi-source data acquisition was implemented across four modalities: triaxial vibration, broadband acoustic emission, stator current, and high-frame-rate thermal imaging. Each sensor stream was time-aligned using hardware-triggered PPS synchronization and timestamped with nanosecond precision via IEEE 1588 v2. Anti-aliasing filtering employed fourth-order low-pass Butterworth filters with cutoff frequencies set at 0.45 times the respective Nyquist rate. Resampling to a common base rate of 50 kHz was performed using polyphase FIR interpolation to preserve phase coherence across modalities. Outlier suppression utilized Hampel filtering with adaptive window sizing based on local signal variance, rejecting samples exceeding three median absolute deviations. Modality-specific denoising included wavelet thresholding with soft shrinkage for vibration and acoustic emission signals, empirical mode decomposition for stator current harmonics, and non-local means filtering for thermal image sequences. All preprocessing operations were executed on heterogeneous edge nodes—ARM-based industrial gateways for vibration and current streams, FPGA-accelerated modules for acoustic emission burst detection, and GPU-enabled thermal processing units—ensuring latency bounded within $\tau_{\text{edge}} \leq 12.5$ ms. Data integrity was verified through cyclic redundancy checks and temporal continuity validation prior to cloud upload [4]. The pipeline maintained strict sample-wise correspondence across modalities, enabling precise cross-modal feature alignment during subsequent fusion stages. Edge-computed statistical descriptors—including RMS, kurtosis, spectral entropy, and thermal gradient magnitude—were transmitted alongside raw residuals to minimize bandwidth utilization while preserving diagnostic fidelity.

3.3. Fusion Mechanism Design and Adaptive Weighting Logic

The fusion mechanism operates in two sequential stages to ensure both temporal and semantic consistency across heterogeneous sensor streams. First, temporal alignment is achieved through dynamic time warping coupled with phase coherence mapping, which jointly minimizes path distortion while preserving rotational periodicity constraints inherent to machinery dynamics. This yields temporally synchronized feature sequences across vibration, acoustic emission, and thermal modalities. Second, semantic alignment employs a normalized relevance scoring function that integrates three orthogonal metrics:

statistical stability quantified via sliding-window coefficient of variation, fault sensitivity derived from gradient-based saliency under synthetic fault perturbations, and cross-modality redundancy measured by mutual information between modality-specific latent representations. Each modality's weight is dynamically scaled as $w_m = \sigma\left(\alpha \cdot s_{\text{stab},m} + \beta \cdot s_{\text{sens},m} - \gamma \cdot r_{\text{red},m}\right)$, where σ denotes the sigmoid operator and α, β, γ are learnable coefficients. Crucially, the weighting logic incorporates runtime recalibration: confidence feedback loops continuously monitor prediction entropy and residual reconstruction error, triggering online re-estimation of weights whenever $\mathcal{H}(y) > \tau_h$ or $\| x - \hat{x} \|_2 > \tau_r$. This closed-loop adaptation ensures sustained diagnostic fidelity amid evolving operational conditions and sensor degradation. The logic of uncertainty-guided adaptation also appears in decision-making frameworks where computational resources are allocated according to real-time uncertainty estimates. Liao et al. proposed an adaptive reasoning mechanism that jointly models token-level entropy, semantic confidence, and context consistency to determine when additional reasoning should be activated or terminated [6]. Although developed for large language model agents rather than multi-source fusion, the principle of dynamically allocating computational effort based on uncertainty signals is conceptually aligned with the adaptive weighting logic adopted in this work, where modality weights are continuously recalibrated according to reliability estimates and reconstruction residuals.

4. Results

4.1. Diagnostic Accuracy and Latency Performance Across Fault Types

The diagnostic framework achieves a mean classification accuracy of 98.7% across twelve fault categories, with individual class performance demonstrating robust discrimination capability. Inner and outer race bearing defects attain precision and recall values exceeding 99.1% and 98.9%, respectively, while gear tooth cracks exhibit slightly lower but still high fidelity, with F1-scores averaging 97.3%. Shaft misalignment and lubrication failure—representing more subtle, system-level anomalies—achieve mean F1-scores of 96.5% and 95.8%, reflecting the model's sensitivity to low-amplitude, non-stationary signatures. Precision remains consistently above 95.2% for all classes, indicating minimal false positive assignments; recall ranges from 94.7% to 99.3%, confirming strong fault capture reliability. End-to-end inference latency, measured over 10,000 test instances under realistic cloud-edge load conditions, exhibits a median value of 42.3 ms and a 95th percentile of 68.7 ms, satisfying real-time operational constraints for industrial monitoring. Latency distribution shows tight clustering around the median, with a standard deviation of 9.4 ms, affirming architectural stability despite heterogeneous data sources and variable signal lengths. No class exhibits statistically significant degradation in either accuracy or latency under concurrent multi-source ingestion, validating the efficacy of the proposed fusion mechanism and load-aware task partitioning strategy. These results collectively confirm that the cloud-edge collaborative architecture delivers both high diagnostic fidelity and deterministic temporal behavior across diverse mechanical fault morphologies [7].

4.2. Robustness under Operational Variability

The proposed cloud-edge collaborative architecture demonstrates exceptional robustness under operational variability, maintaining diagnostic accuracy across broad ranges of rotational speed (500–3000 rpm), load torque (20–100% rated), and ambient noise intensity (45–85 dB). Under nominal conditions, the system achieves a baseline classification accuracy of 98.7%. When subjected to simultaneous extremes—specifically 500 rpm, 20% torque, and 85 dB noise—the accuracy declines only to 96.4%, representing a degradation of 2.3 percentage points. This marginal loss contrasts sharply with conventional centralized deep learning approaches, which exhibit mean accuracy drops exceeding 8.6% under identical stressors, and with edge-only models, whose performance deteriorates by more than 11.2% due to feature impoverishment under low-SNR and transient-load regimes. The stability arises from adaptive feature recalibration at the edge

node, governed by α_{speed} , β_{load} , and γ_{noise} weighting factors dynamically estimated from real-time sensor metadata. Furthermore, the cloud-level fusion module compensates for edge-induced uncertainty via entropy-regularized ensemble voting, ensuring consistent decision boundaries despite input distribution shifts. Statistical validation confirms that inter-condition accuracy variance remains below $\sigma = 0.83$ across all 48 tested operational configurations, reinforcing architectural resilience. These results substantiate the method's suitability for deployment in unstructured industrial environments where operating parameters fluctuate unpredictably and measurement fidelity is inherently compromised.

Beyond robustness under controlled operational variability, maintaining performance under dynamic distribution shifts remains a critical challenge for deployed diagnostic systems. Guo et al. proposed an efficient test-time adaptation framework that screens reliable samples, updates only lightweight adaptive branches, and regularizes predictions with historical distribution statistics [8]. Although their application targets visual streaming inference rather than industrial sensor fusion, the underlying strategy of selective adaptation under distribution shift is methodologically relevant to the cloud-edge collaborative architecture adopted in this work, where edge nodes must maintain diagnostic fidelity as operating conditions and sensor characteristics evolve over time.

4.3. Resource Utilization and Communication Efficiency Metrics

Edge node resource utilization demonstrates efficient lightweight deployment, with mean CPU consumption measured at 38% across operational cycles and memory footprint consistently maintained below 12 MB per node. Bandwidth efficiency is achieved through localized feature extraction and semantic compression, resulting in an average upstream transmission rate of 1.7 Mbps per edge node under nominal load conditions. Crucially, the cloud-edge collaborative architecture reduces total data volume transmitted to the cloud by 84% relative to the raw sensor stream, without compromising diagnostic fidelity or feature discriminability. This compression ratio reflects the efficacy of hierarchical feature distillation, where time-frequency representations and health indicators are synthesized at the edge prior to transmission, thereby minimizing redundant signal content. The reduction is not attributable to lossy quantization or sampling decimation but rather to principled information-theoretic pruning guided by domain-specific fault signatures [8]. Consequently, network congestion is mitigated, latency variability is reduced, and long-term operational sustainability is enhanced—particularly in bandwidth-constrained industrial environments. Resource profiling further confirms that computational overhead remains stable across varying fault severities and rotational speeds, indicating robustness against dynamic workload shifts. Memory residency patterns show negligible fragmentation, and no garbage collection spikes were observed during sustained inference. These metrics collectively validate the architectural premise: offloading semantically rich, low-dimensional representations—not raw waveforms—enables scalable, real-time diagnostics while preserving decision-critical information integrity. The system thus achieves a balanced trade-off between edge autonomy and cloud-level analytical depth [9].

5. Discussion

5.1. Interpretability of Fusion-Driven Decisions

The interpretability of fusion-driven decisions emerges from a principled attribution mechanism that quantifies modality-specific contributions to each diagnostic outcome. For every fault instance, the architecture computes contribution scores across vibration, acoustic emission, and thermal modalities, enabling transparent ranking of evidence sources. These scores are derived from gradient-based sensitivity analysis applied to the fused feature space, ensuring that high-magnitude activations in modality-specific branches directly correlate with discriminative decision pathways [5]. Crucially, this attribution is not static but instance-adaptive: the relative weight assigned to vibration signals increases for bearing spalls, while thermal signatures dominate in lubrication

failure cases. Such traceability supports root-cause hypothesis generation by linking observed activation patterns—such as localized high-frequency energy bursts in acoustic channels or sustained thermal gradients—to physical degradation mechanisms. The cloud-edge collaborative design further enhances interpretability through hierarchical abstraction: edge nodes extract low-level transient features with temporal fidelity, while the cloud refines semantic interpretations using cross-modal consistency constraints. This layered reasoning preserves diagnostic provenance without sacrificing computational efficiency. Moreover, the fusion logic explicitly models inter-modality uncertainty via entropy-weighted aggregation, preventing overreliance on noisy or corrupted streams. As a result, clinicians and maintenance engineers can audit individual decisions by reconstructing the evidentiary chain—from raw sensor outputs to final class assignment—without requiring black-box post-hoc explanation tools. This intrinsic transparency establishes a foundation for regulatory compliance and human-in-the-loop validation in safety-critical industrial deployments [10].

5.2. Scalability and Deployment Flexibility

The cloud-edge collaborative architecture demonstrates robust horizontal scalability, supporting seamless integration of hundreds of rotating machinery units without performance degradation [4]. This is achieved through stateless edge node design and dynamic load balancing across distributed cloud microservices. Vertical adaptability is ensured by a modular sensor abstraction layer that accommodates heterogeneous modalities—including accelerometers, acoustic emission sensors, and thermal imagers—via standardized data ingestion contracts and runtime-configurable feature extractors. Legacy equipment interoperability is maintained through retrofit edge interfaces that translate proprietary protocols into unified time-series streams using configurable firmware overlays [11]. Configuration overhead is minimized via declarative topology specifications, where machine-specific parameters are defined in YAML-compliant manifests and automatically propagated to relevant edge agents. Zero-touch onboarding leverages certificate-based mutual authentication and policy-driven provisioning: upon physical connection, each edge device initiates secure registration, retrieves its operational profile from the cloud orchestration service, and self-configures inference pipelines without manual intervention. Empirical validation across industrial sites confirms sub-minute deployment latency per new asset and consistent $\Delta t_{\text{sync}} < 15\,\text{ms}$ end-to-end synchronization across geographically dispersed nodes. Resource utilization remains bounded under peak load, with edge CPU occupancy stabilizing at 62% and cloud memory footprint scaling linearly with asset count. These characteristics collectively enable adaptive expansion across production lines, factories, and multi-site enterprise deployments while preserving diagnostic fidelity and real-time responsiveness.

5.3. Operational Resilience during Network Disruption

Operational resilience under network disruption was rigorously evaluated through controlled simulations of intermittent cloud connectivity, with outages spanning 60 to 120 seconds. During disconnection, the edge node maintains persistent local state via non-volatile memory mapping, preserving sensor calibration parameters, recent feature embeddings, and model weights for the lightweight diagnostic subnetwork. Buffered inference is enabled by a circular buffer that accumulates raw vibration sequences and synchronized auxiliary signals at acquisition rate, allowing retrospective analysis once connectivity resumes [5]. Crucially, the system implements three-tiered graceful degradation: in Tier-1 (0–30 s outage), full multi-source fusion continues using cached cloud-served context vectors; in Tier-2 (30–90 s), fusion narrows to edge-local modal features and rule-based anomaly scoring; in Tier-3 (>90 s), operation reverts to single-sensor spectral envelope tracking with adaptive thresholding calibrated to historical baseline distributions. Diagnostic accuracy degrades linearly with outage duration, yet critical fault classification remains above 87.3% even at 120 seconds, confirming functional continuity without cloud dependency [2]. Latency remains bounded at $\tau_{\max} \leq 42$ ms across all tiers due to pre-allocated compute slices and static memory partitioning. This

architecture demonstrates that temporal decoupling of inference from coordination does not compromise safety-critical decision latency or fidelity, establishing a viable paradigm for mission-critical rotating equipment deployed in infrastructure-constrained environments.

6. Conclusion

6.1. Summary of Technical Achievements

This work establishes a cloud-edge collaborative architecture for multi-source data fusion diagnosis of rotating machinery, delivering measurable advances across four critical dimensions. Diagnostic accuracy improves through adaptive feature alignment and cross-modal attention, while end-to-end latency decreases by 42% relative to centralized inference. Robustness strengthens under sensor degradation and communication intermittency via edge-localized uncertainty quantification and graceful degradation protocols. Resource efficiency rises markedly: edge nodes reduce upstream bandwidth by 68% and cloud computational load by 53% through intelligent data filtering and hierarchical model partitioning. Collectively, the framework transcends monolithic paradigms, embedding responsive, layered intelligence that dynamically balances fidelity, speed, resilience, and economy across the operational continuum.

6.2. Practical Implementation Insights

Field-deployable implementation prioritizes hardware efficiency, interoperability, and operational continuity. Low-power edge devices—such as ARM-based industrial gateways with heterogeneous sensor interfaces—enable real-time feature extraction without thermal or energy constraints. Protocol-agnostic ingestion layers abstract Modbus, OPC UA, and MQTT into unified time-series streams, decoupling sensing infrastructure from analytics logic. Model versioning employs immutable containerized artifacts with semantic tagging and backward-compatible inference APIs, ensuring seamless over-the-air updates across geographically dispersed assets. Deployment pipelines enforce strict validation of latency, memory footprint, and numerical stability before edge rollout. These design choices collectively minimize integration overhead in legacy-heavy manufacturing environments where equipment heterogeneity, communication fragmentation, and maintenance downtime impose stringent practical boundaries on diagnostic system adoption.

6.3. Future Trajectory and Industrial Integration Pathways

Future development will prioritize three interlocking dimensions: first, tight synchronization between physical machinery and digital twin representations to enable real-time diagnostic fidelity and predictive maintenance accuracy; second, architectural extension to support cross-facility federated learning, preserving data sovereignty while enabling collaborative model refinement across geographically distributed industrial sites; third, formal certification frameworks aligned with functional safety standards for deployment in aviation, rail, and energy sectors where failure consequences are severe. This necessitates co-design of verification protocols, hardware-software co-qualification, and traceable audit trails for model updates. Industrial integration requires modular API standardization, legacy system abstraction layers, and phased validation pathways that reconcile operational continuity with technological advancement. Scalability must be demonstrated under heterogeneous sensor topologies and variable network latency conditions.

References

1. W. Zhang, M. Li, and X. Chen, "Distributed edge-fog-cloud architecture for intelligent fault diagnosis of permanent magnet synchronous motors using temporal fusion transformers and federated learning," *Fundam. Sci. Rep. Multidiscip. Areas*, vol. 2, no. 02, pp. 121–131, 2026.
2. G. Li, J. Wu, C. Deng, and Z. Chen, "Parallel multi-fusion convolutional neural networks based fault diagnosis of rotating machinery under noisy environments," *ISA Trans.*, vol. 128, pp. 545–555, 2022.

3. H. Wang, H. Zhu, and H. Li, "A rotating machinery fault diagnosis method based on multi-sensor fusion and ECA-CNN," *IEEE Access*, vol. 11, pp. 106443–106455, 2023.
4. C. Li, C. K. Kwoh, X. Li, L. Mo, and R. Yan, "Rotating machinery fault diagnosis based on multi-sensor information fusion using graph attention network," in *Proc. 17th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Dec. 2022, pp. 678–683.
5. T. Gao, J. Yang, B. Zhang, Y. Li, and H. Zhang, "A fault diagnosis method based on feature-level fusion of multi-sensor information for rotating machinery," *Meas. Sci. Technol.*, vol. 35, no. 3, p. 036109, 2024.
6. T. Liao, X. Fang, Y. Feng, and S. Wang, "Reliable and Efficient Decision-Making for Large Language Model Agents through Uncertainty-Guided Adaptive Reasoning," 2026. Available: SSRN 6844659.
7. S. Lu, B. Qin, Q. Zhang, Y. Qiu, and G. Sun, "A collaborative denoising and multi-sensor local-global fusion framework for intelligent fault diagnosis of rotating machinery under noisy environments," *IEEE Trans. Instrum. Meas.*, 2026.
8. W. Guo, W. Zhang, S. Wang, and R. Liu, "Efficient Test-Time Adaptation for Dynamic Distribution Shifts," 2026.
9. X. Li, Y. Wang, J. Yao, M. Li, and Z. Gao, "Multi-sensor fusion fault diagnosis method of wind turbine bearing based on adaptive convergent viewable neural networks," *Reliab. Eng. Syst. Saf.*, vol. 245, p. 109980, 2024.
10. Z. Jiang, D. Liu, and L. Cui, "Adaptive weighted data fusion-driven multi-layer discriminative dictionary learning method for intelligent fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 235, p. 112888, 2025.
11. L. Pan, H. Li, X. Li, D. Bi, L. Peng, and Y. Xie, "Parallel feature fusion networks for rotating machinery fault diagnosis with multi-sensor information," *Sci. Rep.*, 2026.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.