AMERICAN ACADEMIC PUBLISHER

INTERNATIONAL JOURNAL OF DATA SCIENCE AND MACHINE LEARNING (ISSN: 2692-5141)

Volume 05, Issue 02, 2025, pages 374-384
Published Date: - 12-12-2025

Integrated Machine Learning and Distributed Fault Diagnosis Frameworks for Resilient Cyber-Physical Systems

Rohit A. Mendes
Department of Systems Engineering, University of Lisbon, Portugal

ABSTRACT

Background: The increasing complexity and interconnectedness of cyber-physical systems (CPS)—including multi-phase electric drives, unmanned aerial vehicles, underwater systems, building automation networks, and high-performance computing devices—demand robust approaches for the detection, isolation, and mitigation of faults. Traditional centralized diagnosis and single-method strategies face scalability, adaptability, and reliability challenges when applied to heterogeneous deployments that combine mechanical, electrical, and software components. This work synthesizes theoretical perspectives and applied developments from decentralized estimation, machine learning, sliding-mode observers, and health-aware control to propose an integrated, publication-ready framework for resilient fault diagnosis and fault-tolerant control across CPS domains.

Objectives: The goals are threefold: (1) to review and unify multiple methodologies for fault detection, isolation, and tolerant control drawn from recent literature; (2) to propose a structured methodology that harmonizes decentralized/distributed estimation with modern data-driven classifiers and domain adaptation techniques; and (3) to demonstrate, through descriptive analysis, how the hybrid framework addresses practical constraints—such as limited observability, actuator heterogeneity, and online transferability—highlighting performance tradeoffs and deployment pathways.

Methods: The proposed methodology builds on modular elements: multiple-model estimation for decentralized and distributed settings, residual generation and multiple-valued evaluation of residuals, sliding-mode observers for fault identification, and a spectrum of supervised and deep learning approaches for classification and localization of faults. Emphasis is placed on domain adaptation for online detection in nonstationary environments, actuator reliability modelling for health-aware control, and rule-based auto-correction mechanisms used in building management systems. The methodology describes algorithmic interfaces, data-flow patterns among modules, and practical considerations for implementation.

Results: Through descriptive analysis grounded in the literature, the framework demonstrates theoretical robustness in isolating both abrupt and incipient faults across representative applications: multi-phase power electronics, electric motors with trapezoidal back-EMF, multi-rotor UAV actuators, underwater thruster systems, bearings and rotating machinery, and building HVAC networks. The hybrid approach yields complementary advantages: decentralized multiple-model estimators provide scalability and local fault isolation (Straka & Punčochář, 2020); sliding-mode observers offer model-based precision for actuator faults (Zuev et al., 2020); data-driven classifiers and deep transfer methods provide adaptability to unknown operating regimes and noisy measurements (Mao et al., 2021; Li et al., 2021).

Conclusions: A synthesis of model-based and data-driven fault diagnosis with distributed estimation and health-aware control offers a pragmatic route to enhanced resilience in CPS. Challenges remain in standardized benchmarking, interpretability, and safe automatic correction, but the integrated pathway provides clear practical steps for engineering adoption across industries. The work concludes with a detailed discussion of limitations,

deployment recommendations, and prioritized future research directions.

KEYWORDS

Fault diagnosis; fault-tolerant control; distributed estimation; machine learning; health-aware control; sliding-mode observer

INTRODUCTION

Fault diagnosis and fault-tolerant control have matured into a multidisciplinary field that spans classical control theory, signal processing, and modern machine learning. The impetus for continued innovation arises from two concurrent trends. First, the pervasiveness of complex cyber-physical systems (CPS) in critical infrastructure, transportation, manufacturing, and consumer devices has vastly increased the number of potential fault modes and the cost of failures. Second, the availability of rich sensor data, networked communication, and computational resources enables more sophisticated diagnosis algorithms but also introduces new challenges in scalability, privacy, and cross-domain transferability. Collectively, these trends require rethinking diagnosis architectures to combine decentralization, adaptable data-driven models, and rigorous model-based observers to ensure dependable operation.

This article draws on a curated selection of contemporary research addressing diverse domains: distributed active fault diagnosis algorithms (Straka & Punčochář, 2020); fault-tolerant control for multi-phase and fault-tolerant permanent magnet motors (Chen et al., 2021); multiple-valued residual evaluation for diagnostic reasoning (Kościelny et al., 2021); neural network approaches for spacecraft actuator faults (Li et al., 2021); machine learning for inverter and distributed generator faults (Ali et al., 2021); multi-phase transmission line fault location via machine learning (Eboule et al., 2022); deep residual convolutional neural networks for robotic systems (Oh et al., 2021); adversarial and transfer techniques for bearing early-fault detection (Mao et al., 2021); GPU diagnostic automation for high-performance computing hardware (Lulla et al., 2025); sliding-mode observers for underwater thrusters (Zuev et al., 2020); health-aware control for multirotor UAVs (Salazar et al., 2020); and building systems diagnostics and automatic correction algorithms (Katipamula & Brambley, 2005; Kim & Katipamula, 2018; Lin et al., 2020a, 2020b, 2021). Each of these contributions provides domain-specific insights; the central objective of the present work is to create a coherent architecture that leverages the strengths of these approaches while acknowledging their limitations.

The problem statement is simple in statement yet complex in realization: how can one design a diagnosis and tolerant control architecture that reliably detects, isolates, and mitigates faults in heterogeneous CPS where constraints include partial observability, communication limitations, evolving operating conditions, and diverse fault modes? Traditional centralized schemes often fail when network bandwidth is constrained, when latency matters, or when single points of failure exist. Conversely, purely data-driven methods may generalize poorly across operating conditions and can lack the explainability necessary for safety-critical applications. A hybrid architecture that systematically integrates decentralized model-based estimators with learning-based classifiers and domain adaptation emerges as a promising compromise. This article argues for such an integration and provides a detailed methodological blueprint for its realization, informed throughout by the cited literature.

The remainder of the introduction motivates each architectural choice by summarizing key contributions from the literature and explaining the rationale for their inclusion in the integrated framework. Decentralized and distributed active fault diagnosis approaches emphasize scalability and local decision-making, reducing communication needs and enhancing robustness to node failures (Straka & Punčochář, 2020). Multiple-model estimation algorithms afford explicit representations of hypothesized fault modes, enabling Bayesian or switching logic for residual interpretation and diagnosis. In parallel, residual evaluation strategies that employ multiple-valued logic allow nuanced reasoning beyond binary thresholds; this is particularly valuable for complex sensors and correlated residual patterns (Kościelny et al., 2021).

Model-based observers such as sliding-mode observers provide robustness to uncertainties and are well suited for actuator fault identification in dynamic systems such as underwater thrusters and motors (Zuev et al., 2020). Health-aware and fault-tolerant control methods offer systematic ways to reallocate control authority and schedule mitigation measures in platforms like octorotor UAVs (Salazar et al., 2020). These methods tie diagnosis to control, enabling systems to remain operational under degraded conditions.

However, model-based approaches alone are insufficient where models are incomplete or where there are numerous configuration variations—as in distributed generators with cascade H-bridge inverters or multi-phase transmission systems—thus creating room for machine learning algorithms that can classify open-switch faults, locate faults on transmission lines, or adapt feature extractors to new operating regimes (Ali et al., 2021; Eboule et al., 2022; Mao et al., 2021). The interplay between model and data approaches is central: models can provide structure and constraints that regularize learning and enhance interpretability, while data-driven methods can capture residual patterns and complex signatures that models may not represent.

Finally, the literature on building systems and analytics demonstrates the practical importance and business case for diagnostic automation, as well as the challenges in implementing automatic correction algorithms in operational settings (Holmberg, 2003; Lin et al., 2020a; Lin et al., 2020b; Lin et al., 2021). These works underscore the pragmatic constraints—sensor placement, upkeep, and human approval processes—that any integrated framework must consider for real-world adoption.

In sum, the introduction establishes the need for a structured hybrid framework, grounded in multiple threads of advanced research, to meet the diagnosis and control requirements of modern CPS. The sections that follow detail the methodology, describe anticipated results through integrative analysis, discuss limitations and research directions, and provide prescriptive recommendations for practitioners.

METHODOLOGY

The methodology presented here is intentionally modular, reflecting the heterogeneous nature of CPS and the need for adaptable deployments. It is organized into constituent modules with defined inputs, outputs, and interaction protocols: (1) Local Model-Based Estimation and Residual Generation, (2) Residual Evaluation and Symptom Reasoning, (3) Data-Driven Classification and Localization, (4) Distributed Multiple-Model Estimation and Consensus, (5) Health-Aware Reconfiguration and Fault-Tolerant Control, and (6) Operational Decision and Auto-Correction Interfaces. Each module is described at length, with emphasis on algorithmic principles, implementation considerations, and integration patterns grounded in the literature.

Local Model-Based Estimation and Residual Generation

At the base layer of the architecture are local estimators that produce residuals—quantified deviations between observed and predicted behavior—which serve as primary indicators of faults. The rationale for deploying local observers is threefold: they reduce communication overhead by processing raw sensory streams near their source, they can be tailored to the physical dynamics of subsystems (e.g., motor phases, thrusters, HVAC components), and they provide interpretable signals rooted in physics.

Model selection for local observers depends on the subsystem dynamics and available measurements. For electrical drives, models incorporate electrical circuit dynamics and back-EMF characteristics; for mechanical actuators, rigid-body dynamics and actuator models are used. Sliding-mode observers are highlighted because of their inherent robustness properties—finite-time convergence and insensitivity to matched uncertainties make them well suited to actuator fault identification in noisy environments (Zuev et al., 2020). Residuals are generated by comparing measured outputs to observer outputs; additional processing includes smoothing, detrending, and adaptive baseline estimation to mitigate false positives due to operational transients.

Residual thresholds are commonly used to trigger alarms, but binary thresholding is simplistic for complex systems. Instead, the methodology emphasizes the extraction of multiple residual features—statistical moments, spectral descriptors, and time-domain signatures—that feed downstream reasoning modules. For rotating machinery, features may include kurtosis and spectral sidebands; for power electronics, harmonic content and instantaneous phase discrepancies are informative. Crucially, the model-based residuals act as structured features that constrain the interpretation space for learning algorithms, grounding them in domain knowledge.

Residual Evaluation and Symptom Reasoning

Rather than relying solely on threshold exceedance, the methodology employs a multiple-valued evaluation of residuals and symptom sequences to enable nuanced diagnostic reasoning (Kościelny et al., 2021). Multiple-valued logic extends binary decisions to graded interpretations—e.g., nominal, suspicious, degraded, and critical states—reflecting uncertainty and gradual degradation. The residual evaluation module maintains a sliding window of residual vectors and computes a multi-valued score based on magnitude, persistence, and correlation across channels.

Elementary symptoms are derived as temporal patterns within residual sequences: single-channel spikes, correlated multi-channel drifts, periodic modulations, or step changes. An elementary symptom sequence is the ordered occurrence of these patterns and serves as a fingerprint for certain fault classes. The diagnostic reasoning engine matches observed symptom sequences against a repository of known symptom patterns, using similarity metrics that account for temporal alignment and partial matches. This approach allows early detection of incipient faults that manifest as subtle symptom sequences before catastrophic failure.

Decision logic in this module also incorporates confidence measures and back-propagates uncertainty to upstream observers—allowing observers to adjust sensitivity or request additional measurements when ambiguity is high. The multiple-valued framework also facilitates human-readable explanations and graded alerts for operators, improving trust and actionability.

Data-Driven Classification and Localization

While model-based residuals and symptom reasoning capture structured deviations, many real-world faults manifest in complex, nonlinear ways that are more readily classified using machine learning. The methodology provides a supervised classification path that ingests residual features, raw sensor streams, and hand-crafted descriptors. A taxonomy of classifiers is recommended: lightweight tree-based algorithms (e.g., gradient boosting) for explainability and speed in resource-constrained nodes; deep convolutional residual networks for high-dimensional sensor data where temporal or spatial patterns are prominent (Oh et al., 2021); and ensemble techniques for robustness.

A critical innovation is the integration of domain transfer and joint adversarial training when deploying models across varying operating regimes or hardware variants (Mao et al., 2021). For example, bearings monitored in a laboratory may display different noise characteristics in the field; domain adaptation techniques adjust feature representations to align source and target distributions without retraining from scratch. Joint adversarial methods minimize domain discrepancy while preserving fault discriminatory features, enabling online detection of bearing early faults and improving generalization.

The methodology advocates for staged training: offline supervised learning on historical labeled faults where available, augmented with synthetic fault injections when safe and practicable, followed by online fine-tuning with semi-supervised learning. For systems where labeled faults are scarce—common in safety-critical systems where failures are rare—one-class classification and anomaly detection methods complement supervised classifiers by modeling nominal behavior and flagging outliers.

Fault localization—identifying the physical component or spatial location of the fault—is treated as a multi-label classification problem when multiple components may degrade concurrently. In electrical networks and transmission lines, localization algorithms exploit spatially distributed measurements in combination with physics-aware features to infer fault location (Eboule et al., 2022). Where communications permit, rapid sharing of location-relevant features among nodes aids cooperative localization.

Distributed Multiple-Model Estimation and Consensus

Decentralization is a central design principle to scale to large CPS with many nodes and to reduce single points of failure. Multiple-model estimation algorithms are naturally suited to decentralized active diagnosis: each node maintains a bank of candidate models—nominal and various fault hypotheses—and computes local likelihoods based on observations (Straka & Punčochář, 2020). Nodes periodically exchange compact belief messages with neighbors and perform consensus updates to refine global fault hypotheses.

The methodology delineates communication protocols and belief-fusion strategies that are robust to packet loss and asynchrony. Two complementary consensus schemes are proposed: weighted averaging of likelihoods with confidence weights based on measurement quality, and message-passing with loopy belief propagation when the network topology resembles a factor graph. The design ensures that nodes with richer information or higher confidence exert more influence, yet the consensus mechanism remains resilient to Byzantine failures by incorporating outlier suppression and trust metrics.

An important practical component is active fault diagnosis: nodes may direct experiments or probing actions (e.g., applying specific control inputs, initiating test sequences) to disambiguate competing hypotheses. Active strategies increase diagnostic speed at the expense of temporary perturbations to normal operation; the methodology describes tradeoffs and safety constraints, recommending conservative probing levels in safety-critical systems.

Health-Aware Reconfiguration and Fault-Tolerant Control

Diagnosis is only valuable insofar as it informs effective mitigation. The methodology integrates health modeling and reconfiguration strategies that translate diagnostic outcomes into control adjustments. Health-aware control involves quantifying actuator reliability and using that information to reallocate control commands, adjust trajectory plans, or engage contingency modes (Salazar et al., 2020). For example, in multi-rotor UAVs, actuator reliability maps can be used to compute control allocation matrices that redistribute forces among functioning motors to maintain stability with degraded performance.

In power electronics and motor drives, fault-tolerant control often entails topology adjustments (e.g., switching to redundant phases or reconfiguring inverter bridging strategies) to maintain torque production despite open-switch faults or phase losses (Chen et al., 2021; Ali et al., 2021). The methodology prescribes layered reconfiguration: immediate low-level control adjustments to maintain safe operation, followed by higher-level performance recovery where possible.

When automatic correction is feasible, building upon prior work in fault auto-correction, the methodology delineates safe auto-correction policies that consider human oversight, business rules, and energy impacts (Lin et al., 2020a; Lin et al., 2020b; Lin et al., 2021). Auto-correction policies are conservative, requiring diagnostic confidence and impact assessment before direct actuator changes are made; where possible, auto-corrections are suggested to operators rather than autonomously executed, particularly in early deployments.

Operational Decision and Auto-Correction Interfaces

The final module addresses human-machine interfaces, logging, and continuous improvement. Diagnostic outputs include graded alarms, suggested corrective actions with associated confidence and expected effects, and provenance metadata for auditability. Training datasets and labeled fault instances are stored for periodic retraining. The architecture supports gradual automation: operator-in-the-loop for critical corrections initially, moving toward greater automation as confidence and proven reliability increase.

Key implementation considerations include computational budgeting for edge nodes, data retention and privacy policies for shared data, secure communication channels to prevent adversarial interference, and fail-safe modes that default to safe halting of operations when ambiguity is excessive.

Integration Patterns and Practical Considerations

Integration of modules follows a publish-subscribe pattern with lightweight message schemas to minimize bandwidth: residuals and symptom scores are published locally, classifiers subscribe and return labels and confidence, and the distributed estimator subscribes to local belief updates. The methodology prescribes periodic heartbeat and watchdog messages to ensure liveness. Emphasis is placed on modularity to permit incremental deployment—starting with local observers and centralized analysis before migrating to full distributed consensus as infrastructure matures.

Real-world deployments must anticipate sensor faults, time synchronization issues, and missing data. Robust preprocessing, time-alignment procedures, and imputation strategies are therefore included as standard elements. The methodology also suggests continuous calibration routines for observers, periodic model validation using controlled test inputs, and use of redundancy where critical sensors are single points of failure.

RESULTS

Because this work synthesizes methods rather than reporting a single experimental dataset, the results section presents a comprehensive descriptive analysis of expected performance, comparative strengths of methods, and illustrative case syntheses across representative domains. Results are framed as reasoned outcomes supported by cited literature rather than empirical claims from new experiments.

Performance of Local Observers and Residual Features

Model-based observers, particularly sliding-mode observers, achieve high sensitivity to actuator faults in dynamic systems while being robust to matched uncertainties (Zuev et al., 2020). The literature indicates that properly tuned sliding-mode observers can detect abrupt faults such as thruster degradation with rapid convergence, enabling timely mitigation. For electrical drives, observers that model back-EMF and phase currents provide discriminative residuals that expose open-switch faults and commutation anomalies (Chen et al., 2021). Residual features enriched with temporal persistence and cross-channel correlation significantly reduce false alarms compared to naive thresholding (Kościelny et al., 2021).

Efficacy of Multiple-Valued Residual Reasoning

The adoption of multiple-valued residual evaluation permits graded detection that captures incipient faults manifesting as subtle deviations. Kościelny et al. (2021) demonstrate that multi-valued evaluation increases diagnostic nuance and supports procedural reasoning about sequences of elementary symptoms. Applied to HVAC and building systems, graded alerts enable prioritization of maintenance actions and correlate well with the observed operational impacts (Kim & Katipamula, 2018). The result is a diagnostic system that is more informative and actionable for operators.

Machine Learning for Classification and Localization

Supervised machine learning models demonstrate strong accuracy in classification tasks when sufficient labeled data exists. Ali et al. (2021) showed that classifiers can identify open-switch faults in cascade H-bridge inverters with high accuracy. Similarly, Eboule et al. (2022) illustrate that machine learning applied to multi-phase transmission data can accurately classify fault types and estimate locations. Deep learning architectures, particularly residual convolutional networks, enable robust detection across complex sensor modalities but require careful regularization and domain adaptation for deployment outside training distributions (Oh et al., 2021; Mao et al., 2021).

Domain adaptation methods reduce the performance degradation that arises from distributional shifts between training and deployment environments, a frequent occurrence in bearing fault detection and other rotating machinery contexts (Mao et al., 2021). Joint adversarial training aligns feature spaces while retaining discriminative capability, yielding more reliable online detection.

Distributed Multiple-Model Estimation and Consensus Outcomes

Distributed multiple-model estimation offers clear scalability benefits. Straka and Punčochář (2020) demonstrate algorithms that maintain diagnostic performance while distributing computation and thereby reducing single-point vulnerabilities. Consensus mechanisms ensure that compact belief messages suffice for maintaining a coherent global view, and active diagnosis strategies accelerate hypothesis disambiguation. When network constraints and asynchrony are considered, consensus with weighted confidence and outlier suppression maintains robustness.

Health-Aware Control and Fault Mitigation

Health-aware control methods allow systems to remain operational under degraded conditions. Salazar et al. (2020) demonstrate that octorotor UAVs, when equipped with actuator reliability models, can reallocate thrust and preserve safe flight. In motor drives, topology reconfiguration and torque reallocation sustain performance under certain open-switch faults (Chen et al., 2021). The literature suggests performance tradeoffs: the system may maintain safety and stability but at reduced efficiency, increased energy consumption, or diminished precision.

Practical Deployment Case Syntheses

Combining modules, the integrated framework yields the following practical outcomes for selected domains:

● Multi-phase Power Electronics and Motors: Local observers detect phase abnormalities; classifiers identify open-switch signatures; distributed consensus helps localize faults in systems with multiple inverters (Ali et al., 2021; Chen et al., 2021). Health-aware reconfiguration reduces torque ripple and preserves operation during fault correction.

● UAVs and Multirotor Systems: Residual reasoning and sliding-mode observers identify actuator loss and degradation; health-aware control reassigns control allocation to sustain flight, while machine learning supports classification of partial faults that originate from mechanical wear or sensor degradation (Salazar et al., 2020; Oh et al., 2021).

● Underwater Thrusters and Marine Systems: Sliding-mode observers detect thruster faults and asymmetric thrust; distributed estimation supports cooperative unmanned vehicle fleets in isolating failing actuators (Zuev et al., 2020).

● Rotating Machinery and Bearings: Domain-adapted deep learning methods detect early bearing faults in noisy environments; model-based residuals provide structural constraints that reduce false positives (Mao et al., 2021).

● Building Automation and HVAC: Integration of residual reasoning with business rules and auto-correction policies can reduce energy waste and prevent recurring faults; conservative auto-correction procedures combined with operator oversight ensure safety and acceptance (Lin et al., 2020a; Lin et al., 2020b; Lin et al., 2021).

● High-Performance Computing Hardware Diagnostics: Factory-grade diagnostic automation for GPUs benefits from automated symptom extraction and classifiers trained on large operational traces, enabling faster root-cause isolation and maintenance scheduling (Lulla et al., 2025).

These syntheses indicate that the integrated approach achieves complementary strengths: model-based observers deliver interpretability and precise timing for fault onset; machine learning enhances pattern recognition for complex signatures; distributed estimation enables scalable, robust decision making; and health-aware control ties diagnosis to practical mitigation.

DISCUSSION

This section offers a deep interpretation of the synthesized results, explores limitations and counter-arguments, and outlines prioritized future directions. The analysis emphasizes theoretical implications and provides practical guidance for deployment.

Interpretation and Theoretical Implications

The integrated framework reflects a broader theoretical shift from monolithic, centralized diagnosis systems toward hybrid, layered architectures that balance model structure and data adaptivity. Several theoretical themes emerge:

1. Complementarity of Model and Data: Model-based and data-driven methods are not mutually exclusive; instead, they form a symbiotic pairing where models impose physics-based constraints and priors, and machine learning captures residual complexities that models omit. This duality improves generalization and interpretability compared to either approach alone.

2. Scalability via Decentralization: Distributed multiple-model estimation and consensus algorithms respond to the scaling challenge inherent in modern CPS. The theoretical underpinning is that compact belief messages carry sufficient statistical information when accompanied by local processing, reducing the information bottleneck inherent in centralized architectures (Straka & Punčochář, 2020).

3. Tradeoffs between Speed, Accuracy, and Safety: Active probing accelerates diagnosis but may interfere with nominal operation. Conservative auto-correction preserves safety but slows mitigation. The theoretical design space requires formal tradeoff analysis—e.g., cost functions balancing diagnostic delay, energy consumption, and risk exposure—to guide policy choices.

4. Uncertainty Management: Multiple-valued logic for residual evaluation and probabilistic belief fusion explicitly incorporate uncertainty, aligning with decision-theoretic perspectives on fault diagnosis. Formalizing uncertainty propagation and its effect on control decisions remains a theoretical necessity.

Limitations and Counter-Arguments

Several limitations temper the framework's practical application:

● Data Scarcity and Labeling: High-quality labeled fault data is scarce for rare, catastrophic failures. The methodology mitigates this through synthetic fault injection and one-class modeling but cannot fully replicate the breadth of real events (Ali et al., 2021). Counter-arguments propose reliance on physics-based simulations to augment datasets; however, simulation fidelity limits generalizability.

● Computational and Communication Constraints: Edge devices in many CPS have limited computation and constrained communication links. While the architecture emphasizes lightweight message passing and local processing, certain deep learning components may require offloading to cloud resources, introducing latency and privacy concerns (Lulla et al., 2025).

● Model Mismatch and Nonstationarity: Model-based observers require accurate models, and machine learning models risk degradation under distributional shifts. Domain adaptation addresses this to an extent, but continual retraining and model validation are operationally demanding (Mao et al., 2021).

● Safety and Regulatory Acceptance: Automatic corrections in regulated or safety-critical contexts (e.g., aircraft systems) may be restricted by certification and human factors requirements. Thus, conservative operator-in-the-loop designs are necessary initially (Lin et al., 2020a).

● Security Concerns: Networked diagnosis systems may be subject to adversarial attacks that mimic faults or suppress alarms. Robust consensus mechanisms with trust metrics help, but security must be designed alongside diagnostic functions (Holmberg, 2003).

Future Research Directions

Prioritized research avenues include:

1. Formal Guarantees for Hybrid Architectures: Developing theoretical bounds on detection delay, false alarm rates, and mitigation effectiveness for hybrid model/data architectures under network constraints would strengthen confidence in deployments.

2. Simulation-To-Reality Transfer Techniques: Enhanced domain adaptation and physics-informed learning methods that combine simulated fault scenarios with limited real data can improve sample efficiency.

3. Human Factors and Explainability: Creating diagnostic outputs that are both actionable and explainable for operators is essential for adoption. Research into natural language summarization of symptom sequences and recommended actions could bridge this gap.

4. Benchmarking and Open Datasets: The field would benefit from standardized datasets and benchmarks across diverse domains to compare algorithms fairly and accelerate progress.

5. Secure and Resilient Consensus Protocols: Investigating consensus algorithms robust to adversarial nodes and network partitioning will be critical as diagnosis systems become more interconnected.

6. Integration with Lifecycle Management: Diagnostic systems should feed into maintenance scheduling and supply-chain planning. Research into closed-loop integration between diagnosis outputs and lifecycle decisions will yield operational benefits.

Deployment Recommendations

For practitioners, the following pragmatic steps are advised:

● Start with local observers and residual reasoning to establish baseline detection capability.

● Augment with supervised classifiers where labeled fault data exist; where data are scarce, prioritize model-based reasoning and anomaly detection.

● Adopt distributed multiple-model estimation only after establishing reliable local processing and secure communications; implement consensus protocols that tolerate packet loss.

● Use conservative auto-correction policies initially, escalating automation as confidence grows and human operators become comfortable.

● Invest in continuous monitoring, retraining pipelines, and data management practices to sustain model relevance.

● Ensure security and privacy by design, including encryption of belief messages and authentication of nodes.

CONCLUSION

This article synthesizes diverse strands of contemporary research to present a comprehensive, integrated framework for fault diagnosis and fault-tolerant control in cyber-physical systems. By combining local model-based observers, multiple-valued residual reasoning, machine learning classification with domain adaptation, distributed multiple-model estimation, and health-aware control, the architecture addresses core challenges of scalability, adaptability, and safety. The descriptive results highlight the complementary strengths of each component and illustrate their application across multiple domains—from electric drives and UAVs to underwater vehicles and building automation.

Limitations persist—data scarcity, model mismatch, computational constraints, and security threats—and future work should pursue formal guarantees, improved transfer learning methods, operator-centric explainability, benchmarking, and resilient consensus protocols. Practitioners are advised to adopt the architecture incrementally, emphasizing human oversight in early deployments and prioritizing secure communication and maintenance integration.

Ultimately, resilient diagnosis and tolerant control require systems to be not only technically adept at fault detection but also operationally aligned with safety procedures, human workflows, and business objectives. The integrated pathway proposed here moves the field toward that synthesis by charting concrete, literature-backed steps for the design, implementation, and continued evolution of dependable diagnostic systems.

REFERENCES

1. Straka, O.; Punčochář, I. Decentralized and distributed active fault diagnosis: Multiple model estimation algorithms. Int. J. Appl. Math. Comput. Sci. 2020, 30, 239–249.
2. Chen, Q.; Xia, Y.; Wang, J.; Zhao, W.; Liu, G. Short-circuit fault-tolerant control for five-phase fault-tolerant permanent magnet motors with trapezoidal back-EMF. Fundam. Res. 2021, in press.
3. Kościelny, J.M.; Syfert, M.; Wnuk, P. Diagnostic Row Reasoning Method Based on Multiple-Valued Evaluation of Residuals and Elementary Symptoms Sequence. Energies 2021, 14, 2476.
4. Li, Y.; Hu, Q.; Shao, X. Neural network-based fault diagnosis for spacecraft with single-gimbal control moment gyros. Chin. J. Aeronaut. 2021, 7, 261–273.
5. Ali, M.; Din, Z.; Solomin, E.; Cheema, K.M.; Milyani, A.H.; Che, Z. Open switch fault diagnosis of cascade H-bridge multi-level inverter in distributed power generators by machine learning algorithms. Energy Rep. 2021, 7, 8929–8942.
6. Eboule, P.S.P.; Pretorius, J.H.C.; Mbuli, N. Machine learning techniques applied on a nine-phase transmission line for faults classification and location. Energy Rep. 2022, 8, 801–810.
7. Oh, Y.; Kim, Y.; Na, K.; Youn, B.D. A deep transferable motion-adaptive fault detection method for industrial robots using a residual–convolutional neural network. ISA Trans. 2021.

8. Mao, W.; Ding, L.; Liu, Y.; Afshari, S.S.; Liang, X. A new deep domain adaptation method with joint adversarial training for online detection of bearing early fault. ISA Trans. 2021, 122, 444–458.
9. Lulla, K.; Chandra, R.; Ranjan, K. Factory-grade diagnostic automation for GeForce and data centre GPUs. International Journal of Engineering, Science and Information Technology 2025, 5(3), 537–544.
10. Zuev, A.; Zhirabok, A.; Filaretov, V. Fault identification in underwater vehicle thrusters via sliding mode observers. Int. J. Appl. Math. Comput. Sci. 2020, 30, 679–688.
11. Salazar, J.C.; Sanjuan, A.; Nejjari, F.; Sarrate, R. Health–Aware and Fault–Tolerant Control of an Octorotor UAV System Based on Actuator Reliability. Int. J. Appl. Math. Comput. Sci. 2020, 30, 47–59.
12. Holmberg, D. BACnet Wide Area Network Security Threat Assessment, NIST Interagency/Internal Report, National Institute of Standards and Technology, Gaithersburg, MD, 2003.
13. Johnson Controls Inc. Help files, 2020.
14. Kramer, H.; Lin, G.; Curtin, C.; Crowe, E.; Granderson, J. Proving the business case for building analytics. Lawrence Berkeley National Laboratory, 2020.
15. Katipamula, S.; Brambley, M.R. Methods for fault detection, diagnostics, and prognostics for building systems—A review, part I. HVAC and R Research 2005, 11(1), 3–25.
16. Kim, W.; Katipamula, S. A review of fault detection and diagnostics methods for building systems. Science and Technology for the Built Environment 2018, 24(1), 3–21.
17. Lin, G.; Pritoni, M.; Chen, Y.; Granderson, J. Development and implementation of fault-correction algorithms in fault detection and diagnostics tools. Energies 2020a, 13(10).
18. Lin, G.; Pritoni, M.; Chen, Y.; Moromisato, R.; Kozlen, S.; Granderson, J. Can we fix it automatically? Development of fault auto-correction algorithms for HVAC and lighting systems. 2020 ACEEE Summer Study on Energy Efficiency in Buildings.
19. Lin, G.; Pritoni, M.; Chen, Y.; Moromisato, R.; Kozlen, S.; Granderson, J. Fault “auto correction” for HVAC systems, a preliminary study. 6th International High Performance Buildings Conference, 2021.