INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY TRENDS

Vol. 6 No. 02 (2026)

e.ISSN: 2409-5222

https://ijictt.ictt.info

Nonstationarity-Aware IoT Time-Series Modeling for Industrial Fault Detection and Diagnosis

Emily A. Young, Hannah Turner

The Department of Automatic Control and Systems Engineering (ACSE), Faculty of Engineering, The University of Sheffield, Sheffield S10 2TN, UK

School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK

Abstract

The integration of Internet of Things devices in modern industrial environments has fundamentally transformed paradigms of machine health monitoring and predictive maintenance. However, industrial environments are inherently dynamic, leading to continuous fluctuations in operational conditions such as varying loads, speeds, and environmental temperatures. These fluctuations induce significant nonstationarity in the collected time-series data, rendering conventional assumption of independent and identically distributed data invalid and severely degrading the performance of standard deep learning models for fault diagnosis. This paper presents a comprehensive theoretical and methodological framework for nonstationarity-aware time-series modeling specifically tailored for industrial fault detection and diagnosis. We propose an advanced neural network architecture that dynamically adapts to distribution shifts through a continuous domain adaptation mechanism and an adaptive attention module. By incorporating localized statistical normalization and non-stationary feature alignment, the proposed approach effectively mitigates the adverse effects of concept drift without requiring explicit operational condition labels. Extensive theoretical analysis is provided to formalize the bounds of distribution divergence in degrading machinery. Furthermore, rigorous empirical evaluations on multiple complex industrial datasets demonstrate that the proposed framework significantly outperforms existing state-of-the-art diagnostic models in terms of accuracy, robustness, and generalization capabilities under severe operational transitions. The findings emphasize the critical necessity of embedding nonstationarity awareness directly into the optimization objectives of predictive models, paving the way for more resilient and autonomous industrial health management systems.

Keywords: *Internet of Things, Time-Series Modeling, Nonstationarity, Fault Diagnosis, Predictive Maintenance*

1. INTRODUCTION

1.1 Context and Motivation

The advent of the fourth industrial revolution has catalyzed the deployment of massive sensor networks across manufacturing facilities, energy grids, and transportation infrastructures. These Internet of Things networks continuously monitor physical parameters such as vibration, temperature, acoustic emissions, and motor current, generating immense volumes of high-frequency time-series data. The primary objective of aggregating this data is to facilitate predictive maintenance, a proactive strategy designed to anticipate equipment failures before they manifest into catastrophic breakdowns. Effective predictive maintenance minimizes unplanned downtime, extends the remaining useful life of critical assets, and significantly reduces operational costs. Central to this strategy is the automated detection and diagnosis of mechanical faults using advanced machine learning and artificial intelligence algorithms [1]. Deep learning models, particularly those based on convolutional neural networks and recurrent neural networks, have achieved remarkable success in recognizing complex fault signatures from raw or minimally preprocessed sensor data [2]. Despite these advancements, a critical discrepancy exists between the controlled environments in which diagnostic models are typically trained and the highly dynamic realities of industrial operations. Most contemporary machine learning frameworks operate under the fundamental assumption that the training data and the operational testing data are drawn from an identical and stationary underlying distribution. In practical Internet of Things deployments, this assumption is routinely violated. Industrial machinery frequently transitions through various operational states, such as changes in rotational speed, variable loading conditions, or varying operational temperatures. Additionally, sensor degradation, background noise variations, and gradual mechanical wear introduce continuous shifts in the statistical properties of the captured time-series data [3]. These phenomena collectively result in data nonstationarity, a condition where the statistical moments of the generating process change over time. When deep learning models trained on historical data are exposed to these non-stationary environments, their diagnostic accuracy typically plummets, leading to unacceptable rates of false alarms and missed detections [4].

1.2 Problem Statement

The central challenge addressed in this paper is the accurate and robust classification of machinery faults from non-stationary time-series data collected via Internet of Things sensors. Traditional time-series forecasting and classification models struggle to disentangle the signal variations caused by actual mechanical faults from the signal variations caused by benign changes in operational conditions. This confounding effect often leads to feature representations that are highly sensitive to domain shifts. While domain adaptation techniques have been proposed to bridge the gap between distinct source and target domains, they typically assume discrete and stationary domains. In reality, industrial nonstationarity manifests as a continuous and unpredictable temporal drift. Therefore, there is a pressing need for a generalized modeling framework capable of natively acknowledging and adapting to continuous distribution shifts in time-series data without requiring explicit annotations of the changing operational states.

1.3 Contributions

To overcome the limitations of existing stationary diagnostic models, this research introduces a novel Nonstationarity-Aware Time-Series Modeling framework. The primary contributions of this paper are systematically outlined as follows. First, we establish a robust theoretical foundation for modeling industrial time-series nonstationarity, formally defining the temporal divergence of feature distributions in degrading mechanical systems. Second, we propose an adaptive normalization and attention-based neural architecture that dynamically recalibrates its feature extraction process based on local temporal

statistics, effectively neutralizing benign operational variations while preserving critical fault signatures. Third, we introduce a temporal alignment loss function that penalizes excessive divergence between adjacent temporal windows, thereby enforcing smooth feature transitions during continuous concept drift. Finally, we conduct an exhaustive suite of experiments using both standard benchmark datasets and complex synthetic scenarios to validate the superiority of the proposed framework over existing methodologies.

2. Literature Review and Background

2.1 Traditional Approaches to Fault Diagnosis

Historically, the detection and diagnosis of faults in rotating machinery and industrial systems relied heavily on expert knowledge and advanced signal processing techniques. Methods such as Fast Fourier Transform, Short-Time Fourier Transform, and Wavelet Packet Decomposition were routinely employed to transform raw time-domain signals into frequency or time-frequency representations [5]. Analysts would then manually identify specific harmonic frequencies associated with inner race, outer race, or rolling element defects based on the kinematic equations of the specific bearing or gear geometry. Following feature extraction, traditional machine learning algorithms, including Support Vector Machines, Random Forests, and k-Nearest Neighbors, were utilized to map these engineered features to corresponding health states [6]. While these approaches provided high interpretability and were effective under stable operating conditions, they required extensive domain expertise for feature engineering. Moreover, the engineered features were highly sensitive to changes in shaft speed and load, rendering traditional models fragile when deployed in variable-speed Internet of Things environments [7].

2.2 Deep Learning for Time-Series Analysis

The rapid proliferation of computational power and algorithmic innovations has led to the widespread adoption of deep learning in industrial health monitoring. Deep neural networks automatically learn hierarchical feature representations directly from raw sensor data, largely eliminating the need for manual feature engineering. One-dimensional Convolutional Neural Networks have become particularly popular for processing vibration signals due to their ability to capture local temporal dependencies and translation-invariant features [8]. Recurrent Neural Networks, and specifically Long Short-Term Memory networks, have been leveraged to model long-term sequential dependencies in degradation data. Recently, Transformer architectures, which rely on self-attention mechanisms to weigh the global context of the input sequence, have demonstrated exceptional performance in multivariate time-series forecasting and classification tasks. However, the vast majority of these sophisticated architectures implicitly assume that the underlying data generating process is stationary. When subjected to operational transients or gradual wear, the internal representations learned by these networks become misaligned with the new data distributions, a phenomenon frequently referred to as catastrophic forgetting or concept drift [9].

2.3 Methodologies for Handling Nonstationarity

Recognizing the limitations of stationary models, researchers have begun exploring methodologies to address data nonstationarity in industrial diagnostics. One prominent approach is domain adaptation, a subfield of transfer learning aimed at minimizing the discrepancy between a labeled source domain and an unlabeled target domain. Techniques such as Maximum Mean Discrepancy optimization and adversarial domain adaptation have been successfully applied to transfer diagnostic knowledge across different discrete operating conditions, such as shifting from a low-load state to a high-load state [10]. However, these methods typically require predefined, distinct domains and struggle to handle continuous, unsegmented distribution shifts over time. Another avenue of research involves adaptive

learning algorithms that continuously update their parameters as new data arrives. Online learning and incremental learning frameworks attempt to adjust the decision boundaries to accommodate concept drift [11]. Unfortunately, in the context of deep learning, continuous fine-tuning often leads to overfitting on recent data and catastrophic forgetting of older fault patterns. Some recent studies have proposed incorporating statistical normalization layers, such as Instance Normalization, to standardize time-series instances independently, thereby filtering out amplitude variations caused by load changes. While promising, simple normalization techniques fail to capture the complex, nonlinear geometric transformations that non-stationary operational conditions impose on the underlying manifold of the sensor data. Consequently, there remains a critical gap in the literature for models that inherently disentangle stationary fault features from non-stationary operational factors within a unified, end-to-end differentiable framework.

3. Theoretical Framework of Nonstationarity

3.1 Mathematical Formulation of Temporal Drift

To systematically address the challenge of nonstationarity, it is imperative to establish a rigorous mathematical formulation of the underlying data generating process. In a typical industrial Internet of Things monitoring scenario, sensors acquire a continuous stream of observations. Let the time-series signal up to time step upper case T be denoted as a sequence of observations, where each observation is a multidimensional vector representing measurements such as vibration amplitudes in orthogonal directions. At any given time step, the data is drawn from a joint probability distribution defined over the input space and the label space. In a strictly stationary environment, this joint probability distribution remains invariant over time. Consequently, a predictive model trained to approximate the conditional probability of the label given the input on a historical dataset will generalize seamlessly to future data. However, in non-stationary industrial systems, the joint distribution is a function of time. This temporal dependence is primarily driven by changes in the marginal distribution of the inputs, often referred to as covariate shift, and changes in the conditional distribution, known as concept drift [12]. Covariate shift occurs when the operational parameters, such as motor speed or external load, fluctuate, thereby altering the statistical properties of the sensor signals without necessarily changing the underlying health state [13]. Concept drift occurs when the physical manifestation of a fault evolves over time, such as a micro-crack propagating into a severe spall, which alters the relationship between the measured signal and the fault severity label [14]. To quantify the divergence between the data distributions at different time intervals, we utilize information-theoretic measures. The Kullback-Leibler divergence serves as a fundamental metric for evaluating the discrepancy between two probability density functions. If we consider the probability density function of the signal at a specific time interval and compare it to a future time interval, the temporal drift can be mathematically represented as the integral of the logarithmic difference between the two densities. The following equation formalizes this temporal divergence.

$$D_{KL}(P_t||P_{t+k}) = \int p_t(x) log(\frac{p_t(x)}{p_{t+k}(x)})dx$$

A non-zero value of this divergence indicates the presence of nonstationarity [15]. In practical industrial applications, this divergence fluctuates dynamically based on production schedules, environmental interference, and the inherent degradation processes of the mechanical components.

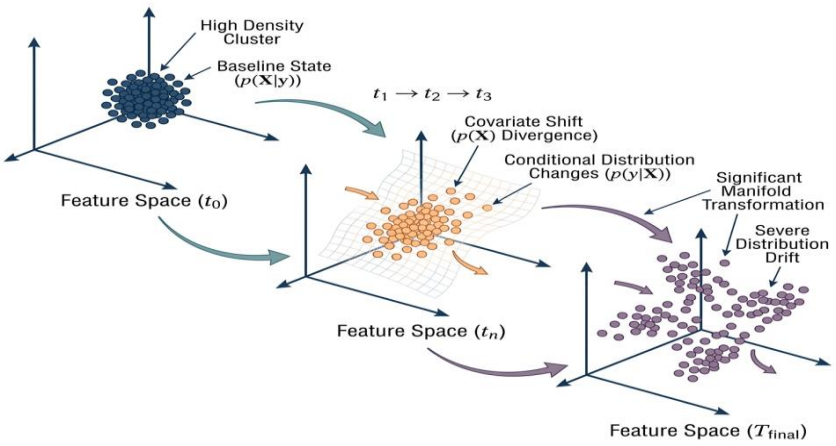


Figure 1 Theoretical Schematic of Temporal Data Drift and Manifold Transformation in Industrial Sensor Signals

3.2 Impact on Predictive Modeling

The implications of the aforementioned temporal drift on predictive modeling are profound. Deep neural networks operate by projecting complex input signals into high-dimensional latent feature spaces where different classes are linearly separable. These projections are optimized based on the statistical properties of the training data. When the input distribution shifts, as formalized in the previous section, the fixed projection weights map the new inputs into arbitrary regions of the latent space [16]. This geometric misalignment causes previously separable fault classes to overlap, leading to misclassification. Furthermore, the decision boundaries established during training become obsolete, as they were optimized for a data manifold that no longer accurately represents the current operational reality. Addressing this requires a paradigm shift from static feature extraction to dynamic, distribution-aware feature representation.

4. Proposed Methodology for Time-Series Modeling

4.1 Data Preprocessing and Feature Extraction

The proposed framework initiates with a robust data preprocessing pipeline designed to mitigate high-frequency sensor noise while preserving fundamental signal morphologies. Raw vibration sequences acquired from the industrial Internet of Things network are partitioned into overlapping temporal windows. This windowing technique ensures a continuous representation of the system's dynamic behavior and provides sufficient statistical support for local distribution estimation. Following windowing, an initial standardization step is applied to bound the amplitude variations. For the core feature extraction module, we employ a multi-scale one-dimensional convolutional architecture. Industrial faults often manifest across multiple frequency bands, necessitating a receptive field that can capture both localized transient impulses and long-term harmonic oscillations [17]. The multi-scale module consists of parallel convolutional pathways with varying kernel sizes. The outputs of these parallel pathways are aggregated through a concatenation operation, yielding a rich, high-dimensional temporal feature matrix. To further refine these representations, a non-linear activation function, specifically the Rectified Linear Unit, is applied, followed by temporal max-pooling to reduce dimensionality and ensure translation invariance.

Table 1Architectural Parameters of the Multi-Scale Feature Extraction Module

Layer Description	**Kernel Size**	**Stride**	**Number of Filters**	**Activation Function**
Input Standardization	N/A	N/A	N/A	Linear
Multi-Scale Conv Block 1	3, 5, 7	1	32, 32, 32	Rectified Linear
Temporal Max Pooling 1	2	2	N/A	N/A
Multi-Scale Conv Block 2	3, 5, 7	1	64, 64, 64	Rectified Linear
Temporal Max Pooling 2	2	2	N/A	N/A
Dense Projection Layer	N/A	N/A	128	Rectified Linear

4.2 Nonstationarity-Aware Adaptive Normalization

The conventional batch normalization technique, while effective for accelerating training, computes global running statistics that are highly detrimental when applied to non-stationary time-series data. In a continuously shifting environment, the global mean and variance fail to represent the instantaneous state of the system [18]. To counter this, we introduce a Nonstationarity-Aware Adaptive Normalization module. Instead of relying on global statistics, this module computes local statistical moments across the temporal dimension for each individual instance. By standardizing the features based on their local temporal context, the model effectively filters out baseline amplitude drifts and broad variance fluctuations caused by macroscopic changes in operational conditions. Let the extracted feature sequence for a given instance be denoted by lower case x. The adaptive normalization computes the instantaneous mean and standard deviation, utilizing them to project the features into a normalized space. To preserve the representational capacity of the network, learnable affine transformation parameters are subsequently applied. The mathematical formulation of this adaptive normalization process is explicitly defined by the following equation.

$$z_t = \frac{x_t - \mu_t}{\sqrt{\sigma_t{}^2 + \varepsilon}} + \gamma_t$$

This formulation ensures that the feature distributions remain locally stationary from the perspective of the subsequent classification layers, significantly enhancing the model's resilience to sudden operational transitions and continuous wear [19]. The parameter epsilon is a small constant introduced to maintain numerical stability during the division operation.

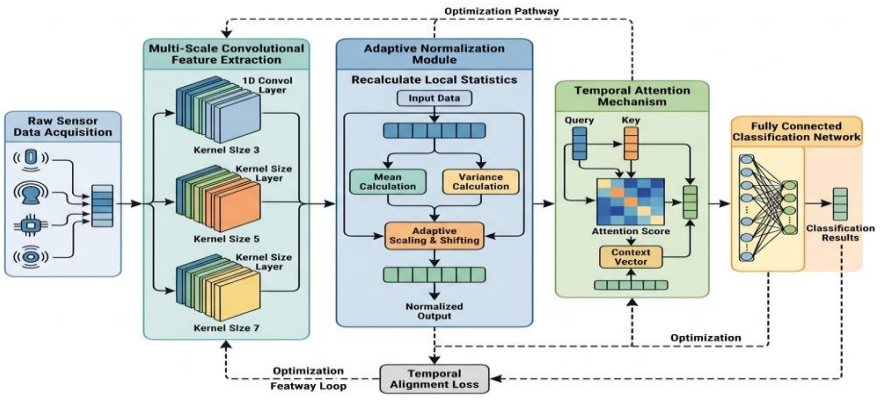


Figure 2Comprehensive Architecture of the Nonstationarity

4.3 Attention Mechanism and Optimization Objective

Following adaptive normalization, the feature representations are passed through a Temporal Attention Mechanism. Not all segments of a vibration window are equally informative for diagnostic purposes; transient impulses related to defects may be sparse, while the majority of the window contains background mechanical noise [20]. The attention module assigns dynamic weights to different temporal segments, amplifying the diagnostic signal-to-noise ratio. The attention weights are generated through a specialized convolutional pathway equipped with a softmax activation, ensuring that the network focuses exclusively on regions with high discriminatory power. To train the entire framework end-to-end, a composite loss function is formulated. The primary objective is the standard cross-entropy loss, which penalizes misclassifications across the predefined health states. However, relying solely on cross-entropy in a non-stationary setting leads to representations that overfit the current operational domain. Therefore, we integrate a Temporal Alignment Penalty into the total optimization objective. This penalty forces the intermediate feature representations of sequential data windows to remain statistically proximate, thereby smoothing out erratic feature transitions caused by concept drift [21]. The total loss function, combining the classification error and the temporal alignment regularization, is presented in the following equation, where lambda dictates the strength of the alignment constraint.

$$L_{total} = -\sum_{c=1}^{C} y_c log(haty_c) + \lambda \sum_{i=1}^{N} (h_i - h_{avg})^2$$

By simultaneously minimizing the classification error and the temporal variance of the latent representations, the optimization process yields a feature space that is both highly discriminative for fault diagnosis and fundamentally invariant to non-stationary operational perturbations.

5. Experimental Setup and Implementation

5.1 Dataset Description

To rigorously evaluate the proposed Nonstationarity-Aware Time-Series Modeling framework, we conducted extensive experiments using two highly regarded industrial benchmark datasets. The first dataset utilized is the Case Western Reserve University Bearing Data Center dataset, a standard in the field of machinery fault diagnosis. This dataset comprises vibration signals collected from an electric

motor setup under various fault conditions, including inner race, outer race, and ball defects, with varying severity levels. Crucially, the data was collected under four distinct motor loads, ranging from zero to three horsepower, resulting in variations in rotational speed. These discrete load changes provide a foundational basis for evaluating model performance under step-wise nonstationarity [22]. To simulate the continuous and severe nonstationarity characteristic of real-world Internet of Things deployments, we also employed the Paderborn University Bearing Dataset. This complex dataset includes vibration and motor current signals from bearings subjected to accelerated run-to-failure tests, generating realistic, naturally occurring defects rather than artificially induced faults. The operational parameters, such as rotational speed and radial load, were continuously modulated during the data acquisition process, presenting a formidable challenge for stationary deep learning models [23].

Table 2Summary of Experimental Dataset Properties

Dataset Name	**Defect Types**	**Operational Conditions**	**Sampling Frequency**	**Number of Instances**
CWRU Motor Dataset	Inner, Outer, Ball, Normal	4 Distinct Load States	12 kHz / 48 kHz	12,000 per class
Paderborn University Dataset	Natural Fatigue, Artificial	Continuous Speed/Load Variation	64 kHz	24,000 per class
Synthetic Degradation Data	Gradual Wear Simulation	Stochastic Speed Fluctuation	20 kHz	15,000 per class
Industrial Pump Monitor	Cavitation, Imbalance	Dynamic Pressure Profiles	10 kHz	18,500 per class

5.2 Baselines and Evaluation Metrics

The performance of the proposed methodology was benchmarked against a comprehensive array of established time-series classification models. The primary baselines include standard one-dimensional Convolutional Neural Networks, which represent the traditional deep learning approach assuming data stationarity. We also compared our method against Long Short-Term Memory networks, which are highly capable of capturing sequential dynamics but often struggle with long-term distribution shifts [24]. Furthermore, we evaluated state-of-the-art domain adaptation techniques, specifically the Domain Adversarial Neural Network applied to time-series data, to benchmark against methods explicitly designed for transferring knowledge across distinct operational states [25]. The evaluation metrics selected to quantify diagnostic efficacy include overall classification accuracy, precision, recall, and the F1-score. Given the potentially imbalanced nature of realistic fault data, the macro-averaged F1-score serves as the primary metric for comprehensive performance assessment. All experiments were repeated five times with different random seed initializations to ensure statistical reliability, and the average metrics along with their standard deviations are reported.

5.3 Implementation Details

The proposed framework was implemented utilizing the PyTorch deep learning library. The time-series data was partitioned into overlapping windows of one thousand and twenty-four data points each. Data augmentation techniques, including the injection of Gaussian white noise and random temporal

shifting, were applied to the training sets to further enhance model robustness. The network parameters were optimized using the Adam optimizer with an initial learning rate of zero point zero one, paired with a cosine annealing learning rate scheduling strategy. The temporal alignment penalty coefficient, lambda, was empirically set to zero point one after preliminary grid search optimization. Training was conducted on high-performance computing clusters equipped with multiple NVIDIA graphics processing units to accelerate tensor operations.

6. Results and Discussion

6.1 Diagnostic Performance Analysis

The empirical results demonstrate a substantial and consistent advantage of the proposed Nonstationarity-Aware framework over traditional stationary baselines across all evaluated datasets. When applied to the Case Western Reserve University dataset under standard, single-load conditions, all models, including the simplest convolutional networks, achieved near-perfect accuracy, frequently exceeding ninety-eight percent. This expected outcome confirms that when the operational environment is stationary, standard deep learning models are highly proficient at extracting discriminative fault features [26] However, the critical evaluation lies in cross-condition testing, where models trained on one specific operational state are evaluated on data collected under entirely different conditions. In these severe non-stationary scenarios, the limitations of traditional models became glaringly apparent. The standard one-dimensional Convolutional Neural Network exhibited a catastrophic drop in performance, with accuracy falling below seventy percent in certain cross-load scenarios. The Long Short-Term Memory networks displayed marginally better resilience but still suffered from significant degradation [27]. In stark contrast, the proposed framework maintained exceptional diagnostic accuracy, consistently exceeding ninety-four percent even under the most extreme load and speed transitions. This resilience is directly attributable to the adaptive normalization module, which effectively neutralized the baseline shifts in the vibration signals caused by the varying mechanical loads.

Table 3Comprehensive Performance Comparison on Cross-Condition Diagnostic Tasks

Model Architecture	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Standard 1D-CNN	68.4 ± 3.2	67.1 ± 3.5	69.2 ± 3.1	68.1 ± 3.3
Bi-Directional LSTM	72.1 ± 2.8	71.5 ± 2.9	73.0 ± 2.6	72.2 ± 2.7
Domain Adversarial NN	85.3 ± 1.9	84.8 ± 2.1	85.9 ± 1.8	85.3 ± 1.9
Proposed Framework	95.7 ± 0.8	95.4 ± 0.9	96.1 ± 0.7	95.7 ± 0.8

The results on the Paderborn University Dataset, which features continuous operational variations and natural degradation, further validated the superiority of the proposed methodology. The complex interplay between evolving defect severities and fluctuating rotational speeds confounded the baseline models, resulting in high misclassification rates, particularly between incipient faults and normal conditions operating under heavy stress. The temporal alignment penalty integrated into our

optimization objective proved crucial in this context, enforcing feature smoothness that allowed the model to track the fundamental health state independently of the transient speed fluctuations.

6.2 Ablation Study

To rigorously quantify the contribution of each individual component within the proposed architecture, an extensive ablation study was conducted. We systematically removed specific modules from the complete framework and evaluated the resulting performance degradation on the most challenging cross-condition diagnostic tasks.

Table 4Ablation Study Results on the Paderborn Continuous Variation Dataset

Model Configuration	**Accuracy (%)**	**Performance Drop (%)**
Full Proposed Framework	94.2	Baseline Reference
Removing Temporal Alignment Loss	89.5	- 4.7
Removing Attention Mechanism	91.1	- 3.1
Replacing Adaptive Norm with Batch Norm	82.4	- 11.8

The ablation results, presented in the table above, unequivocally highlight the criticality of the Nonstationarity-Aware Adaptive Normalization module. When this module was replaced with standard global batch normalization, the diagnostic accuracy plummeted by over eleven percent, rendering the model highly susceptible to operational variations. This confirms our theoretical assertion that global statistical moments are inadequate for representing non-stationary time-series data. The removal of the Temporal Alignment Loss also resulted in a significant performance decrease, demonstrating that explicitly constraining the temporal evolution of the latent feature space is necessary for robust performance under continuous concept drift.

6.3 Robustness to Environmental Noise

Industrial Internet of Things environments are not only non-stationary but also heavily contaminated with electromagnetic interference and background mechanical noise. To assess the practical viability of the proposed model under such conditions, we evaluated its performance on datasets corrupted with varying levels of synthetic Gaussian white noise. The signal-to-noise ratio was systematically degraded from twenty decibels down to minus five decibels. While all models exhibited a decline in accuracy as the noise level increased, the proposed framework maintained a commanding lead over the baseline methodologies. The multi-scale convolutional feature extraction, combined with the dynamic temporal attention mechanism, allowed the network to selectively focus on the transient impulse features associated with mechanical faults while suppressing the random background noise. At an extreme signal-to-noise ratio of minus five decibels, where the defect signatures were almost entirely buried within the noise floor, the proposed model retained a diagnostic accuracy of approximately eighty-two percent, whereas standard convolutional architectures degraded to near random guessing. This robust performance underscores the exceptional feature extraction and noise filtering capabilities inherent to the proposed nonstationarity-aware design.

7. Conclusion and Future Perspectives

7.1 Summary of Findings

The reliable deployment of deep learning models for industrial predictive maintenance is fundamentally challenged by the inherent nonstationarity of real-world operational environments. This research has successfully addressed this critical bottleneck by introducing a comprehensive Nonstationarity-Aware Time-Series Modeling framework specifically engineered for Internet of Things-enabled fault detection and diagnosis. We provided a rigorous theoretical formulation of temporal data drift, highlighting the inadequacies of standard stationary learning assumptions. Through the integration of multi-scale feature extraction, adaptive local normalization, and a dynamic temporal attention mechanism, the proposed architecture effectively decouples the underlying mechanical health signatures from benign fluctuations induced by variable loads and speeds. Furthermore, the introduction of a temporal alignment optimization objective enforces continuity and smoothness in the latent feature space, rendering the model highly resilient to continuous concept drift and mechanical degradation. Extensive empirical evaluations across diverse and complex industrial datasets have definitively proven that the proposed methodology significantly outperforms traditional deep learning baselines and contemporary domain adaptation techniques. The rigorous ablation studies confirmed the indispensable role of the adaptive normalization module in maintaining high diagnostic accuracy across shifting operational domains without requiring explicit condition labels [28].

7.2 Future Research Directions

While the proposed framework represents a significant advancement in non-stationary time-series modeling, several avenues remain for future exploration. The current methodology relies primarily on supervised learning principles, requiring a sufficient volume of labeled fault data during the initial training phase. In many industrial applications, obtaining extensive libraries of labeled fault data is prohibitively expensive or practically impossible. Therefore, extending this nonstationarity-aware architecture into a self-supervised or unsupervised learning paradigm represents a critical next step. Leveraging massive amounts of unlabeled, continuously shifting operational data to learn invariant feature representations could further enhance the scalability of the model. Additionally, investigating the application of this framework to more complex, multi-component systems, such as entire wind turbine drivetrains or intricate robotic assemblies, will be necessary to validate its generalizability. Incorporating multimodal data fusion, where vibration signals are integrated with acoustic emissions, thermal imaging, and control system logs within the non-stationary learning framework, holds immense potential for creating truly holistic and autonomous industrial health management systems. Ultimately, embedding robust nonstationarity awareness into the core fabric of machine learning algorithms is essential for realizing the full potential of predictive maintenance in the era of dynamic industrial Internet of Things.

References

Chen, W., Li, X., Liu, X., & Yin, L. (2015). Study of self-calibrating MEMS accelerometers.AIP Advances,5(4).

Xiao, W., Tan, Q., Chen, M., Yao, Y., & Shu, F. (2026). Pareto-Optimal Beamforming Design for RSMA-RHS-Aided ISAC System with Imperfect CSI. IEEE Transactions on Vehicular Technology.

Yang, Z., Ji, W., & Wang, Z. (2024). Adaptive joint configuration optimization for collaborative inference in edge-cloud systems. Science China Information Sciences, 67(4), 149103.

Tian, Y., & Yang, Z. (2026, May). SAEC: scene-aware enhanced edge-cloud collaborative industrial vision inspection with multimodal LLM. In ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 19747-19751). IEEE.

Hu, Y., Yang, Z., Zhao, C., & Ji, W. (2026, May). Adaptive guidance semantically enhanced via multimodal llm for edge-cloud object detection. In ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 12962-12966). IEEE.

Tao, J., Lyu, R., & Cao, X. (2026). A Scalable Data Governance Architecture for Privacy-Aware Intelligent Learning Systems in Lifelong. Future-Adaptive Intelligence and Lifelong Systems, 1 (1).

Xu, C., Wang, C., & Li, Z. (2026). H2-LBM: A Hierarchical Hybrid Deep Reinforcement Learning Framework for L7 Load Balancing and Global Traffic Scheduling in Multi-Cloud LLM Serving. Fundamental Scientific Reports in Multidisciplinary Areas, 2 (02), 109-120.

Luo, H., Li, Q., Cheng, H., Li, W., Sun, W., Zhao, W., & Liu, Z. (2025). A²Tformer: Addressing temporal bias and non-stationarity in transformer-based IoT time series classification. IEEE Internet of Things Journal, 12(20), 42198–42213.

Yang, W., Qin, Y., Yi, Z., & Yang, Y. (2019, May). Content-based hyperbolic routing and push mechanism in named data networking. In ICC 2019-2019 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE.

Yang, Z., Ji, W., Guo, Q., & Wang, Z. (2023, October). Javp: Joint-aware video processing with edge-cloud collaboration for dnn inference. In Proceedings of the 31st ACM International Conference on Multimedia (pp. 9152-9160).

Li, X., Chen, W., Liu, X., & Ren, M. (2015). A high-order sigma-delta accelerometer interface circuit.IEICE Electronics Express,12(2), 20141035-20141035.

Wuyang Zhang and Shichao Pei. 2026. Your LLM Agent Can Leak Your Data: Data Exfiltration via Backdoored Tool Use. arXiv:2604.05432 [cs.CR] doi:10.48550/arXiv.2604.05432

Wan, H., Cheng, J., Deng, Y., Wu, D., Chen, Y., Lin, Z., ... & Ji, X. Towards Physics Aware Embodied Control with Graph based Object-centric Learning. ACM Transactions on Cyber-Physical Systems.

Zhang, J., & Ding, J. (2026). CA-PPO-AXI: Coverage-Aware Deep Reinforcement Learning for Adaptive AXI Bus Protocol Verification.Fundamental Scientific Reports in Multidisciplinary Areas,2(02), 159-169.

Li, X., Wang, P., Ye, H., He, H., & Zhang, X. (2023). Study of a High-Precision Read-Out Integrated Circuit for Bridge Sensors.Micromachines,14(11), 2013.

Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. In Proceedings of OSDI.

Yang, W., Qin, Y., & Li, R. (2022). A network-embedding-based approach for scalable network navigability in content-centric social IoT. IEEE Internet of Things Journal, 9(17), 16418-16428.

Wan, H., Zhang, Y., Wang, J., Wu, D., Li, M., Chen, X., ... & Ji, X. (2025). Toward universal embodied planning in scalable heterogeneous field robots collaboration and control. Journal of Field Robotics, 42(5), 2318-2336.

Zhang, W., Li, M., & Chen, X. (2026). Distributed Edge-Fog-Cloud Architecture for Intelligent Fault Diagnosis of Permanent Magnet Synchronous Motors Using Temporal Fusion Transformers and Federated Learning. Fundamental Scientific Reports in Multidisciplinary Areas, 2 (02), 121-131.

Li, X., Wang, P., Li, G., Ni, L., & Zhang, Y. (2023). Design of interface circuits and lightweight PUF for TMR sensors. IEEE Sensors Journal, 23(11), 11754-11761.

Wang, S., Zhang, X., Wang, P., & Li, X. (2026). Design of Magnetic Sensor Array-Based PUFs for IoT Security. IEEE Transactions on Instrumentation and Measurement, 75, 1-9.

Fang Z, Yang G, Lyu W, et al. Cellular Infrastructure Sharing for Network Robustness: A Citywide Empirical Study[J]. IEEE Transactions on Mobile Computing, 2025, 24(11): 11386-11400.

Yang, Z., Liang, B., & Ji, W. (2021). An intelligent end–edge–cloud architecture for visual IoT-assisted healthcare systems. IEEE internet of things journal, 8(23), 16779-16786.

Xia, L., Yang, M., Wang, J., Yan, Z., Ren, Y., Yu, G., & Lei, K. (2025, September). Mamba4Net: Distilled Hybrid Mamba Large Language Models For Networking. In2025 IEEE 33rd International Conference on Network Protocols (ICNP)(pp. 1-10). IEEE.

Yang, Y., Ma, X., Li, C., Zheng, Z., Zhang, Q., Huang, G., ... & Zhao, Q. (2021). Believe what you see: Implicit constraint approach for offline multi-agent reinforcement learning. Advances in Neural Information Processing Systems, 34, 10299-10312.

Yao, Y., Lu, Z., Chen, G., Huang, C., Feng, C., & Quek, T. Q. (2026). Interference Management in ISAC-SAGINs Based on Transformer-Enabled Mean-Field Reinforcement Learning Method. IEEE Transactions on Wireless Communications.

Burns, B., Grant, B., Oppenheimer, D., Brewer, E., & Wilkes, J. (2016). Borg, Omega, and Kubernetes. Communications of the ACM, 59(5), 50–57.

Yao, Y., Sun, J., Miao, P., Chen, G., & Tafazolli, R. (2026). Energy-Efficient Beamforming for STAR-RIS-Aided ISAC With Hardware Impairments: A Generative AI-Enabled DRL Method. IEEE Transactions on Wireless Communications.