



Original research article

## Integration of Digital Twin Technology and Industry 4.0 Principles for Real-Time Structural Health Monitoring in Smart Manufacturing Facilities

S. Davlatov<sup>a,\*</sup>  0000-0002-3268-7156, A. Zayniyev<sup>b</sup>  0000-0002-2634-9329,  
 J. Zokirov<sup>c</sup>  0009-0006-9933-4415, M. Temirova<sup>d</sup>  0009-0001-8340-2501,  
 S. Uljaeva<sup>e</sup>  0000-0002-0629-5212, K. Khudayberganov<sup>f</sup>  0009-0003-5484-5471,  
 I. Matkarimov<sup>g</sup>  0000-0002-6783-8591, C. Van Truong<sup>h</sup>  0009-0009-4843-9868

<sup>a</sup> Bukhara State Medical Institute named after Abu Ali ibn Sino, Bukhara, Uzbekistan;

<sup>b</sup> Samarkand State Medical University, Samarkand, Uzbekistan;

<sup>c</sup> Termiz University of Economics and Service, Farovon street 4-b, Termez, Surxondaryo, Uzbekistan;

<sup>d</sup> Termez State Pedagogical Institute, I.Karimov street 288b, Termez, Surxondaryo, Uzbekistan;

<sup>e</sup> Tashkent Institute of Irrigation and Agricultural Mechanization Engineers (TIAME) National Research University, 100000, 39 Kari-Niyazy Street, Mirzo Ulugbek district, Tashkent city, Uzbekistan;

<sup>f</sup> Urgench State University, 14, Kh.Alimdjan str, Urganch, Khorezm, Uzbekistan;

<sup>g</sup> Mamun University, 989Q+6V, Khiva, Xorazm Region, Uzbekistan;

<sup>h</sup> Swiss Information and Management Institute (SIMI Swiss) & Asia Metropolitan University (AMU), 63000 Cyberjaya, Selangor, Malaysia

### ABSTRACT

Within Industry 4.0 manufacturing environments, Structural Health Monitoring (SHM) is recognized as mission-critical; nevertheless, extant Digital Twin (DT) implementations seldom achieve deep fusion with the production layer and consequently struggle to co-optimize structural integrity alongside operational efficiency. This paper therefore introduces, and subsequently validates, an integrated DT framework expressly conceived to close that lacuna. Four objectives guided the inquiry: first, to architect a distributed digital-twin topology underpinned by edge-cloud analytics capable of real-time SHM; second, to operationalize a machine-learning-driven predictive-maintenance regime that causally couples structural response data with both manufacturing process signatures and ambient environmental variables; third, to embed the resultant framework within incumbent MES/ERP ecosystems spanning multiple production facilities; and fourth, to quantify the concomitant reductions in maintenance expenditure, production downtime, and energy utilization. A longitudinal, 24-month, multi-site investigation furnished empirical corroboration. The framework couples a high-fidelity DT to legacy MES/ERP strata through a distributed edge-cloud fabric; an ensemble of machine-learning algorithms—Long Short-Term Memory networks prominent among them—was deployed for predictive anomaly detection. The system attained 96 % anomaly-detection accuracy (F1-score: 0.95) and translated this diagnostic precision into demonstrable operational gains: maintenance costs fell by 42.1 %, downtime by 31.1 %, and energy intensity by 23.2 % ( $p < 0.001$ ). The edge-centric architecture reduced processing latency by 67 %, thereby enabling sub-50 ms integration with MES/ERP layers, while inter-site model transfer achieved 94.0 % adaptation efficacy. These findings substantiate the contention that principled integration of DTs with Industry 4.0 paradigms furnishes a transformative yet pragmatic pathway for manufacturing-oriented SHM. The framework's verified capacity to enhance prognostic fidelity while simultaneously yielding sizeable operational dividends delineates a clear trajectory toward more resilient and resource-efficient industrial assets.

Published by the University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia.  
 This is an open access article distributed under the CC BY 4.0 terms and conditions.

### ARTICLE INFO

Article history:

Received July 9, 2025

Revised September 25, 2025

Accepted November 11, 2025

Published online November 28, 2025

Keywords:

Digital Twin;

Edge Computing;

Industry 4.0;

Predictive Maintenance;

Structural Health Monitoring

\*Corresponding author:

Salim Davlatov

[pro.ilmiy@bsmi.uz](mailto:pro.ilmiy@bsmi.uz)

<https://doi.org/10.24867/IJEM-400>

## 1. Introduction

The manufacturing sector is undergoing its most profound upheaval in living memory, as streams of digital data, autonomous hardware and algorithmic decision-making fuse into a single, constantly evolving organism [1]. Within this turbulence, Structural Health Monitoring (SHM) has ceased to be a specialist curiosity and has become the principal guarantee that bridges, gantries, presses and pipelines will survive their design lives without catastrophic surprise. Digital Twin (DT) technology is now woven into SHM architectures precisely because it sharpens diagnostic certainty and collapses the lag between defect initiation and corrective action [2]–[5]. Traditional walk-through inspections, clipboards in hand, are therefore giving way to cyber-physical ecosystems in which every rivet, beam and bearing broadcasts its autobiography in real time. Industry 4.0 vocabulary—IoT, edge analytics, cloud elasticity—merely labels the circuitry through which these autobiographies travel [6].

The economic stakes are impossible to ignore since conservative forecasts value the digital-twin market at AUD 20.5 billion by 2025, implying a compound annual growth rate of almost 38 % [7]. Plant managers no longer confront a choice between throughput and structural vigilance, they now insist on both. Predictive algorithms, trained on telemetry that would drown a human analyst, simultaneously squeeze cycle times and guard against fatigue cracks, allowing expansion economies such as Vietnam's or Indonesia's to keep their new industrial estates operable and competitive [8], [9]. Academic enquiry has moved in parallel with commercial appetite. Predictive-maintenance DTs have become one of the most densely cited topics in manufacturing journals, precisely because they convert sustainability, safety and profit into mutually reinforcing rather than competing objectives [10], [11]. Recent case studies describe virtual replicas that marry computational-fluid-dynamics meshes with finite-element skeletons, so that a misaligned spindle or an overheated bearing is exposed by discrepancies measured in microns and milliseconds rather than in weeks of downstream scrap [12]. The twin itself has matured from a static CAD ghost into a living; self-updating doppelgänger whose structural and behavioral models re-calibrate whenever fresh sensor packets arrive [13]. Machine-learning layers absorb terabytes of process history, distilling operational signatures that flag drift long before quality statistics deteriorate [14].

None of this would be tractable without the Industrial Internet of Things (IIoT), the nervous layer that stitches sensors, actuators, gateways and algorithms into a coherent manufacturing organism [15]. This study is guided by four primary objectives: (1) to architect a distributed digital-twin topology that fuses edge analytics with cloud-based cognition for instantaneous structural monitoring; (2) to instantiate machine-learning algorithms for predictive maintenance that correlate structural comportment with manufacturing dynamics and ambient conditions; (3) to demonstrate, across multiple manufacturing facilities, the artefact's interoperability with extant manufacturing-execution and enterprise-resource-planning systems; and (4) to quantify resultant gains in maintenance-cost abatement, production-downtime curtailment, and energy-consumption optimization. A longitudinal, 24-month, multi-site investigation furnished empirical corroboration. The framework couples a high-fidelity DT to legacy MES/ERP strata through a distributed edge-cloud fabric; an ensemble of machine-learning algorithms—Long Short-Term Memory networks prominent among them—was deployed for predictive anomaly detection. The system attained 96 % anomaly-detection accuracy (F1-score: 0.95) and translated this diagnostic precision into demonstrable operational gains: maintenance costs fell by 42.1 %, downtime by 31.1 %, and energy intensity by 23.2 % ( $p < 0.001$ ). The edge-centric architecture reduced processing latency by 67 %, thereby enabling sub-50 ms integration with MES/ERP layers, while inter-site model transfer achieved 94.0 % adaptation efficacy. These findings substantiate the contention that principled integration of DTs with Industry 4.0 paradigms furnishes a transformative yet pragmatic pathway for manufacturing-oriented SHM. The framework's verified capacity to enhance prognostic fidelity while simultaneously yielding sizeable operational dividends delineates a clear trajectory toward more resilient and resource-efficient industrial assets.

## 2. Literature review

Contemporary smart factories deploy IIoT for tasks as diverse as tool tracking, inventory harmonization, quality gatekeeping and supply-chain choreography [16], [17]. Forty-six per cent of manufacturers already run IIoT pilots or full rollouts, while fifty-seven per cent complement these with cloud analytics that transform raw waveforms into actionable KPIs [18]. In especially hostile environments—blast furnaces, cryogenic pipelines or high-voltage switch-

yards—cloud-backed sensor meshes stream data through wireless SHM backbones that decide, within seconds, whether a strained beam can safely await the next scheduled outage [19], [20]. Latency constraints, however, force many sites to push intelligence toward the asset itself; fog nodes and micro-datacenters therefore host distributed twins that shrink reaction times without swelling bandwidth bills [21].

Between enterprise resource planning and the shop floor sits the manufacturing execution system (MES), the middleware custodian of work orders, routings and genealogy. Modern MES suites now embed symbolic planners and twin-based simulators that can compose, stress-test and release production schedules without human mediation [22]. Once the twin is federated with the MES, predictive analytics feed directly into AI-driven actuators that trim cycle parameters on the fly. Market analysts expect the MES segment to swell from USD 15.95 billion in 2025 to USD 25.78 billion by 2030, a growth trajectory that mirrors the widening reliance on twin-augmented execution [23]–[25]. Table 1 distils the systematic literature comparison that underpins the preceding discussion, cataloguing the capabilities, constraints and architectural choices that currently differentiate digital-twin and wider Industry 4.0 deployments for SHM in manufacturing environments.

Despite remarkable technological progress, conspicuous lacunae remain in both the concep-

tualization and operationalization of contemporary research. At present, digital-twin deployments within structural-health monitoring seldom achieve full-spectrum integration with manufacturing-execution systems, thereby curtailing their utility for holistic facility governance [19]. State-of-the-art predictive-maintenance platforms are further impeded by computational encumbrance, the heterogeneity of data streams, and the escalating sophistication demanded of prognostic models; these factors collectively obstruct seamless assimilation into live manufacturing workflows [11], [35]. A cardinal deficiency resides in the absence of unified architectures capable of concurrently surveilling structural integrity while dynamically optimizing production regimes [36]. Most DTs reported in the literature function merely as synchronized replicas, offering no actionable feedback loop to physical control hierarchies [37]. Moreover, scant evidence exists of distributed digital-twin instantiations being advanced as end-to-end predictive-maintenance solutions for the manufacturing sector at large. The scholarly corpus likewise betrays a paucity of inquiry into energy-consumption minimization achieved through the conjoined analysis of structural response and process parameters [38]. Equally, extant studies stop short of furnishing multi-site empirical validation under divergent operational envelopes and environmental exposures.

**Table 1.** Comparative analysis of DT and industry 4.0 approaches for manufacturing structural health monitoring

Approach Category	Key Technologies	Primary Applications	Monitoring Accuracy	Real-time Capability	Integration Level	Limitations
Traditional SHM Systems [26]	Visual inspection, NDT	Infrastructure monitoring	70-85%	Limited	Standalone	Subjective interpretation, calibration dependence
IoT-based Monitoring [27], [28]	Sensor networks, wireless communication	Asset tracking, condition monitoring	85-92%	Moderate	System-level	Data variety, computational burden
DT SHM [29]	Virtual replication, real-time data	Damage detection, lifecycle prediction	90-96%	High	Multi-system	Model complexity, computational requirements
Industry 4.0 Integrated [30], [31]	MES/ERP integration, AI analytics	Production optimization, predictive maintenance	92-98%	Real-time	Enterprise-wide	Implementation complexity, skill requirements
Edge Computing DT [32]-[34]	Distributed processing, fog computing	Remote monitoring, reduced latency	94-97%	Ultra-high	Cloud-edge hybrid	Infrastructure investment, security concerns

The present investigation responds to these shortcomings through four inter-locking contributions. First, by embedding digital-twin technology within a comprehensive Industry 4.0 epistemology, the project institutes bidirectional discourse between virtual artefacts and physical manufacturing assets, thereby transcending the unidirectional constraints characteristic of prevailing digital-shadow paradigms. The architecture exploits distributed manufacturing-execution strata endowed with autonomous production-plan genesis and enactment, augmented by an intelligent digital-twin overlay. Second, to mitigate computational burden without sacrificing real-time fidelity, the methodology mobilizes an edge-computing fabric that situates analytical horsepower proximal to data genesis, attenuating latency and augmenting aggregate system performance. Third, the framework is engineered for frictionless coexistence with incumbent enterprise ecologies—embracing both manufacturing-execution and enterprise-resource-planning platforms—thereby safeguarding translational viability. Fourth, machine-learning algorithms selected for pattern-discrimination confer adaptive comportment, allowing the system to evolve sympathetically with shifting operational contexts. This data-centric *modus operandi* expedites the ingestion and interrogation of voluminous process data while benchmarking prognostic efficacy across an ensemble of learning models. In sum, the research seeks to devise and empirically corroborate a holistic digital-twin scaffold, consonant with Industry 4.0 precepts, for real-time structural-health surveillance within smart-manufacturing installations. The envisaged artefact will simultaneously optimize structural-integrity stewardship and production efficiency.

### 3. Methodology

#### 3.1 Study Design and Site Selection

This longitudinal experimental study was conducted across three automated manufacturing facilities situated in the Eastern Province, Riyadh Province, and Makkah Province of Saudi Arabia. The sites were chosen for their differing operational complexity, structural diversity, and pre-existing automation infrastructure, thereby affording a rigorous and wide-ranging validation of the proposed DT framework. Each plant ran continuously under distinct production schedules, environmental conditions, and structural loading regimes, yielding

heterogeneous datasets suitable for algorithm validation and system performance appraisal. The investigation spanned twenty-four months, from January 2022 to December 2023, thereby capturing complete seasonal cycles and the full spectrum of production fluctuations. Saudi Arabia was selected as the host territory for three reasons that map directly onto the study's aims: (i) the partner plants jointly exhibited the high operational complexity, structural diversity, and established MES/ERP backbone demanded by the envisaged integration; (ii) the local operating environment encompasses broad seasonal and diurnal temperature excursions (approximately 18–45 °C during the observation window), furnishing a demanding test-bed for environmental compensation and real-time performance; and (iii) the sites ran continuously and—subject to safety clearance—accommodated controlled, reversible interventions, permitting stringent yet non-disruptive validation under live production conditions. Site-selection criteria stipulated a manufacturing floor area exceeding 5,000 m<sup>2</sup>, extant MES infrastructure, structural variety (incorporating both steel-frame and reinforced-concrete elements), and an operational temperature envelope of 18–45 °C. Throughout the monitoring period each facility sustained routine production, thereby ensuring data acquisition under authentic manufacturing conditions with negligible disruption to operations.

#### 3.2 Sensor Network Architecture and Instrumentation

The sensor ensemble installed across the facility consisted of 240 high-precision strain gauges (Vishay Precision Group CEA-06-125UW-350), 180 triaxial MEMS accelerometers (Analog Devices ADXL354C), and 120 environmental monitoring sensors (Bosch BME680). These were positioned at carefully selected locations on load-bearing members and within production areas to capture the full range of structural and ambient behavior. The strain gauges were chosen for their gauge factor of  $2.095 \pm 0.5\%$ , a temperature coefficient of gauge factor of  $+1.2 \times 10^{-5} \text{ }^\circ\text{C}^{-1}$ , and an operating span from  $-75 \text{ }^\circ\text{C}$  to  $+175 \text{ }^\circ\text{C}$ , thereby guaranteeing reliable metrological performance under the prevailing industrial thermal regime. Each accelerometer was operated with 16-bit resolution and programmable full-scale ranges of  $\pm 2 \text{ g}$ ,  $\pm 4 \text{ g}$ , and  $\pm 8 \text{ g}$ , delivering a sensitivity of  $3.9 \text{ mg LSB}^{-1}$  at the  $\pm 2 \text{ g}$  setting and a noise floor of  $25 \text{ } \mu\text{g } \sqrt{\text{Hz}^{-1}}$ . Concurrently, the environmental sensors recorded temperature to within  $\pm 0.5 \text{ }^\circ\text{C}$ , rel-

ative humidity to  $\pm 3\%$  RH, barometric pressure to  $\pm 0.12$  hPa, and ambient volatile organic compound concentrations, thereby permitting rigorous correlation between fluctuating environmental states and the observed structural response. Signal conditioning employed distributed amplification and filtering systems with anti-aliasing filters configured with cutoff frequency  $f_c$  determined by:

$$f_c = \frac{f_s}{2.56} \quad (1)$$

where  $f_s$  represents the sampling frequency of 1024 Hz, ensuring compliance with Nyquist criteria and preventing aliasing artifacts in acquired signals.

### 3.3 Digital Twin Development Framework

The digital-twin architecture that was implemented employed a three-layer knowledge-graph model comprising concept, model and decision layers. Within the concept layer, ontological frameworks were established that incorporated domain-specific knowledge drawn from technical manuals and expert insights, thereby furnishing a comprehensive theoretical foundation for the manufacturing and structural-monitoring domains. The model layer aligned digital and physical parameters by means of physics-based system modelling and distributed real-time data integration, whereas the decision layer exploited computational models and real-time data to provide automated decision support. Virtual-model development drew upon finite-element analysis (FEA) with mesh discretization that achieved element sizes of between 50 mm and 200 mm, the exact dimension being contingent upon structural complexity and the requirements of stress concentration. Material properties were calibrated through experimental testing, with Young's-modulus values spanning from 200 GPa for structural-steel components to 25 GPa for concrete elements. Dynamic-response modelling incorporated modal analysis, with natural-frequency identification accomplished through operational-modal-analysis techniques. The digital-twin synchronization mechanism instituted bidirectional data-exchange protocols for which latency requirements were maintained below 100 ms so as to ensure real-time operational compatibility. Data-integrity verification deployed hash-based validation algorithms that guaranteed transmission accuracy in excess of 99.97% under normal operational conditions.

### 3.4 Edge Computing and Data Processing Architecture

The distributed edge computing infrastructure deployed Intel NUC 11 Pro edge nodes equipped with Intel Core i7-1165G7 processors and 32 GB DDR4 memory at each facility, providing local processing capabilities for real-time data analysis and reducing cloud communication overhead. Edge nodes implemented containerized processing using Docker orchestration, enabling scalable deployment and efficient resource utilization. Real-time data preprocessing applied digital filtering techniques including fourth-order Butterworth bandpass filters with frequency range 0.1 Hz to 200 Hz, removing low-frequency drift and high-frequency noise components. Signal processing algorithms implemented moving average filtering for environmental data and exponential smoothing for trend analysis, with smoothing parameter  $\alpha$  determined by:

$$\alpha = \frac{2}{n+1} \quad (2)$$

where  $n$  represents the smoothing window length optimized for each sensor type based on signal characteristics and noise levels. Data compression utilized lossless algorithms achieving compression ratios between 3:1 and 7:1 depending on data type and variability, reducing bandwidth requirements while maintaining data integrity. Edge-to-cloud synchronization employed MQTT protocol with quality of service level 2, ensuring reliable data transmission and preventing message loss during network interruptions.

### 3.5 Machine Learning Algorithm Implementation

Pattern recognition algorithms incorporated ensemble learning approaches combining Random Forest, Support Vector Machine, and Gradient Boosting classifiers for structural anomaly detection. Feature extraction implemented statistical time-domain features including root mean square (RMS), kurtosis, skewness, and peak-to-peak values, complemented by frequency-domain features derived from Fast Fourier Transform analysis.

The ensemble prediction model computed weighted average predictions using:

$$\hat{y}_{\text{ensemble}} = \sum_{i=1}^M w_i \hat{y}_i \quad (3)$$

where  $\hat{y}_i$  represents individual classifier predictions,  $w_i$  denotes corresponding weights determined

through cross-validation optimization, and  $M$  represents the total number of base classifiers.

Deep learning implementation utilized Long Short-Term Memory (LSTM) networks for temporal pattern recognition in structural response data. The LSTM architecture comprised input layer with 64 neurons, two hidden layers with 128 and 64 neurons respectively, and output layer configured for binary classification. The LSTM cell state update equations were implemented as [39]:

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
 \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned} \tag{4}$$

where  $f_t$ ,  $i_t$ , and  $o_t$  represent forget, input, and output gates respectively,  $W$  denotes weight matrices,  $b$  represents bias vectors, and  $\sigma$  indicates the sigmoid activation function.

Training employed backpropagation through time with Adam optimizer configured with learning rate 0.001, beta parameters (0.9, 0.999), and epsilon  $1e-8$ . Dropout regularization with probability 0.2 prevented overfitting, while batch normalization improved training stability and convergence speed.

To evaluate detection under controlled yet realistic departures from nominal behavior, this study prioritized physically realizable, non-destructive interventions approved by each facility's safety team. Three families of reversible interventions were applied on secondary, non-load-critical elements during planned production lulls: (i) controlled bolt-preload loss at brace-to-frame and equipment-support connections (target torque reductions of 20% and 50% of specification, verified with a calibrated torque wrench); (ii) baseplate soft-foot/misalignment simulated by inserting stainless-steel shims (0.25 mm and 0.50 mm) under one corner of selected machine skids or utility frames over 150–200 mm of contact length; and (iii) local mass perturbations by clamping steel plates of 5 kg, 10 kg, and 20 kg at pre-identified midspan locations on secondary members or machine skids. All interventions were fully reversible and kept well within serviceability limits.

For each severity level, we verified that the intervention produced a small but measurable response change under comparable production rate: operational modal analysis confirmed natural-frequency shifts of approximately 0.4–3.2% in the affected modes and co-located strain gauges registered local

mean shifts on the order of tens of macrostrain ( $\mu\epsilon$ ), consistent with early-stage looseness/misalignment. Process parameters were held within  $\pm 5\%$  of baseline to avoid confounding. To expand class diversity during training, we generated additional time series with the validated DT by driving the finite-element model with measured load histories while perturbing local stiffness ( $-2\%$ ,  $-5\%$ ,  $-10\%$ ) and relaxing boundary spring constants ( $+5\%$  to  $+25\%$ ). Measurement noise was injected using empirically estimated sensor-noise statistics consistent with the deployed instrumentation (see Section 2.2). Augmented data were used only in training folds. All validation and headline metrics, including the reported 96% anomaly-detection accuracy, were computed exclusively on field measurements acquired under the physical interventions described above.

### 3.6 System Integration with Manufacturing Execution Systems

Integration with the incumbent MES and ERP estates was affected through de-facto industrial and enterprise protocols: OPC UA for shop-floor automation traffic and RESTful APIs for upstream ERP conversation. A lightweight middleware stratum was interposed so that digital-twin artefacts could converse with the legacy scheduling and control layers without perturbing established workflows. MES-side coupling delivered a real-time correlative view between process set-points and structural response signatures, thereby enabling condition-based maintenance cadences and continuous operational refinement. Data-mapping artefacts were engineered to respect the native schemata of SAP Manufacturing Integration and Intelligence (MII) and of Siemens SIMATIC IT, both of which were instantiated at the partner plants. The integration fabric is realized as four containerized micro-services, edge- and cloud-resident: (1) OPC UA client/gateway services that subscribe to the requisite PLC/MES tag namespaces, normalize engineering units, and emit canonical SHM messages; (2) an MQTT bridge operating at QoS 2 to ensure exactly-once delivery from edge to cloud; (3) stateless stream-processing tasks that validate avro schemas, append asset and contextual metadata, and execute windowed down-sampling; and (4) an integration gateway that projects the canonical SHM events and metrics into SAP MII and Siemens SIMATIC IT via REST endpoints and webhook callbacks. The pipeline honors message ordering and idempotency and persists an immutable log of every exchange for full traceability. Under peak-shift load the end-to-end

median latency to MES was 34 ms (ERP 47 ms) with a synchronization fidelity exceeding 99.5%, comfortably within the <100 ms real-time mandate.

### 3.7 Performance Metrics and Data Analysis

System performance evaluation utilized multiple quantitative metrics including anomaly detection accuracy, maintenance cost reduction, production downtime minimization, and energy consumption optimization. Anomaly detection performance was assessed using precision, recall, and F1-score metrics calculated as [40], [41]:

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \\ \text{F1-score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned} \quad (5)$$

where TP, FP, and FN represent true positives, false positives, and false negatives respectively.

Energy consumption analysis employed normalized metrics accounting for production output variations, calculated as:

$$E_{\text{normalized}} = \frac{E_{\text{total}}}{P_{\text{output}} \times T_{\text{operation}}} \quad (6)$$

where  $E_{\text{total}}$  represents total energy consumption,  $P_{\text{output}}$  denotes production output, and  $T_{\text{operation}}$  indicates operational time duration.

To enable a like-for-like comparison of maintenance expenditures across facilities and periods, this study defined a 24-month pre-implementation baseline (January 2020–December 2021) using MES/ERP work-order and cost-center exports for corrective, preventive, and condition-based maintenance. Cost line items unrelated to asset upkeep (e.g., capital projects, warranty reimbursements, insurance claims) were excluded a priori by account codes. All monetary values were inflation-adjusted to 2023 constant currency using the facility finance team’s monthly indices and reconciled to year-end ledgers.

Maintenance costs were normalized to jointly account for variations in output and operating time. The normalized maintenance-cost metric is:

$$C_{\text{norm}} = \frac{C_{\text{maint}}}{P_{\text{output}} \times T_{\text{operation}}} \quad (7)$$

where  $C_{\text{maint}}$  denotes total maintenance cost over the interval,  $P_{\text{output}}$  denotes production output, and

$T_{\text{operation}}$  denotes operating time. This construction mirrors the normalized energy metric in Equation (6), enabling cross-facility aggregation without bias from throughput or schedule differences. Facility-level summaries were weighted by operating time when computing pooled estimates.

Adaptation effectiveness quantifies the fraction of cross-site performance loss recovered by the automated facility-specific adaptation. Let  $A_{\text{in}}$  denote the within-facility benchmark accuracy obtained when training/validating on the destination facility (see Fig. 3a),  $A_{\text{transfer}}$  the accuracy when models are transferred with no adaptation, and  $A_{\text{post}}$  the accuracy after the adaptation workflow. We compute:

$$\text{AE} = \frac{A_{\text{post}} - A_{\text{transfer}}}{A_{\text{in}} - A_{\text{transfer}}} \times 100\%. \quad (8)$$

In this study,  $A_{\text{in}} = 96\%$  (anomaly-detection accuracy; Fig. 3a). “Performance Degradation” in Table 5 is defined as  $A_{\text{post}} - A_{\text{transfer}}$ . AE therefore reports the percentage of the lost accuracy that is recovered after adaptation using only field measurements from the destination facility (“Anomaly Induction Protocol and Severity Levels”).

### 3.8 Model validation protocol

To validate the DT against ground-truth measurements, this study evaluated agreement between model predictions and instrumented responses along four complementary streams. All comparisons were performed on time-synchronized signals preprocessed as described previously (sampling at 1024 Hz, anti-aliasing per Equation (1)).

(1) Strain response correlation. We formed paired windows of measured strain ( $\mu\epsilon$ ) and model-predicted strain at matched locations and times and fitted ordinary least squares. Agreement was summarized by the coefficient of determination ( $R^2$ ) and the root-mean-square error (RMSE).

(2) Frequency-domain validation for acceleration. Dynamic response fidelity was assessed via magnitude-squared coherence between predicted and measured acceleration over the operational band 0.1–50 Hz:

$$\gamma_{xy}^2(f) = \frac{P_{xy}(f)^2}{P_{xx}(f)P_{yy}(f)}. \quad (9)$$

Here,  $P_{xy}(f)$  denotes the cross-power spectral density and  $P_{xx}(f)$ ,  $P_{yy}(f)$  the auto-power spectral densities of the predicted (x) and measured (y) sig-

nals. Coherence curves were summarized across facilities to indicate dynamic correspondence over the operating band.

(3) Environmental parameter prediction accuracy. For temperature ( $^{\circ}\text{C}$ ), relative humidity (%RH), and pressure (hPa), model predictions were compared to sensor readings using mean absolute percentage error:

$$\text{MAPE} = 100 \times \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (10)$$

Here,  $y_i$  and  $\hat{y}_i$  are measured and predicted values, respectively, over  $N$  aligned samples.

(4) Temporal prediction performance. To assess stability over time, we computed monthly RMSE of strain predictions normalized to full-scale measurement and summarized uncertainty with 95% confidence intervals across the 12-month validation window:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}. \quad (11)$$

Monthly aggregates were reported to capture seasonality and operational regime changes without conflating short-term transients.

These four streams provide complementary views of fidelity: (1) static/operational stress agreement, (2) dynamic response matching across frequency, (3) environmental surrogate prediction accuracy, and (4) longitudinal stability.

### 3.9 Statistical Analysis

We subjected the data to a sequence of standard descriptive, correlational and inferential procedures to appraise system performance and adjudicate the stated hypotheses. Where sample size fell below 5000 observations we assessed distributional shape with the Shapiro–Wilk statistic; for larger samples we relied on the Kolmogorov–Smirnov test. Whenever normality was rejected, we substituted distribution-free counterparts—namely the Mann–Whitney U test for two-group contrasts and the Kruskal–Wallis test for multi-group comparisons. Temporal dependence was modelled within the autoregressive integrated moving-average framework, with the order of differencing and the (p, d, q) triplet selected according to the Box–Jenkins protocol and the adequacy of fit corroborated by residual diagnostics. All null hypotheses were evaluated against a two-sided significance threshold of  $\alpha = 0.05$ , while Cohen’s d furnished an estimate of practical magnitude. Asso-

ciations between structural response patterns and ambient environmental variables were first screened with Pearson’s  $r$  where marginal normality could be sustained; otherwise, we adopted Spearman’s rank-order coefficient. A multiple-regression model was subsequently fitted to isolate the subset of predictors that exerted a statistically reliable influence on changes in structural behavior. Ten-fold cross-validation was implemented to gauge the stability of the regression weights and to attest the generalizability of the equation beyond the derivation sample. Every computation was executed in R release 4.3.0, supplemented by the forecast, corrplot and caret libraries for dedicated time-series estimation, correlation visualization and machine-learning corroboration respectively. The modelling exercise was conceived for predictive, not causal, purposes. Inasmuch as several operational covariates exhibit appreciable inter-correlation, the estimated coefficients are reported as measures of association alone, and we refrain from causal terminology in the absence of experimental or quasi-experimental identification.

## 4. Results and Discussions

### 4.1 System Deployment and Data Collection Performance

The comprehensive sensor network deployment across the three manufacturing facilities demonstrated robust data acquisition capabilities throughout the 24-month monitoring period. Data collection performance analysis was conducted to evaluate sensor reliability, communication efficiency, and overall system uptime. Table 2 presents the detailed performance metrics for each facility, including sensor operational rates, data transmission success, and system availability statistics across different sensor types and environmental conditions.

To make sensor performance explicit in this study’s results, we report both operational rates and the associated measurement specifications of the deployed instruments. Across the 24-month period, operational rates were: strain gauges  $98.6 \pm 1.2\%$ , accelerometers  $99.1 \pm 0.9\%$ , and environmental sensors  $98.1 \pm 1.3\%$  (Table 1). Here, “operational rate” denotes the percentage of time a sensor channel produced valid, within-specification data over the monitoring window. For interpretability, the underlying measurement specifications are: strain gauges with gauge factor  $2.095 \pm 0.5\%$  and temperature coefficient  $+1.2 \times 10^{-5}/^{\circ}\text{C}$ ; triaxial MEMS accelerometers (16-bit; sen-

**Table 2.** System deployment performance metrics across three manufacturing facilities showing sensor operational rates (mean  $\pm$  SD over 24 months), data transmission success, overall system availability, and average latency. Operational rate is defined as the percentage of time a sensor channel produced valid, within-specification data during the monitoring period.

Facility Location	Strain Gauges Operational Rate (%)	Accelerometers Operational Rate (%)	Environmental Sensors Operational Rate (%)	Data Transmission Success (%)	System Uptime (%)	Average Latency (ms)
Eastern Province	98.7 $\pm$ 1.2	99.1 $\pm$ 0.8	97.9 $\pm$ 1.5	99.8	99.2	47 $\pm$ 12
Riyadh Province	98.3 $\pm$ 1.4	98.8 $\pm$ 1.1	98.4 $\pm$ 1.3	99.7	98.9	52 $\pm$ 15
Makkah Province	98.9 $\pm$ 1.0	99.3 $\pm$ 0.7	98.1 $\pm$ 1.2	99.9	99.4	43 $\pm$ 10
Overall Mean	98.6 $\pm$ 1.2	99.1 $\pm$ 0.9	98.1 $\pm$ 1.3	99.8	99.2	47 $\pm$ 13

sitivity 3.9 mg/LSB at  $\pm 2g$ ; noise density 25  $\mu g/\sqrt{Hz}$ ); and environmental sensors with temperature accuracy  $\pm 0.5$  °C, humidity accuracy  $\pm 3\%$  RH, and pressure accuracy  $\pm 0.12$  hPa (see Methods, Section 2.2).

The sensor network achieved exceptional operational reliability across all three facilities, with strain gauges maintaining 98.6% operational rate, accelerometers achieving 99.1% operational rate, and environmental sensors sustaining 98.1% operational rate. Data transmission success exceeded 99.7% at all facilities, with the Makkah Province facility achieving the highest transmission success rate of 99.9%. Average system latency remained consistently below the 100-millisecond requirement, with mean latency of 47 milliseconds enabling real-time operational compatibility. The Eastern Province facility experienced slightly higher latency variability due to increased network traffic during peak production periods, while the Makkah Province facility demonstrated the most stable communication performance.

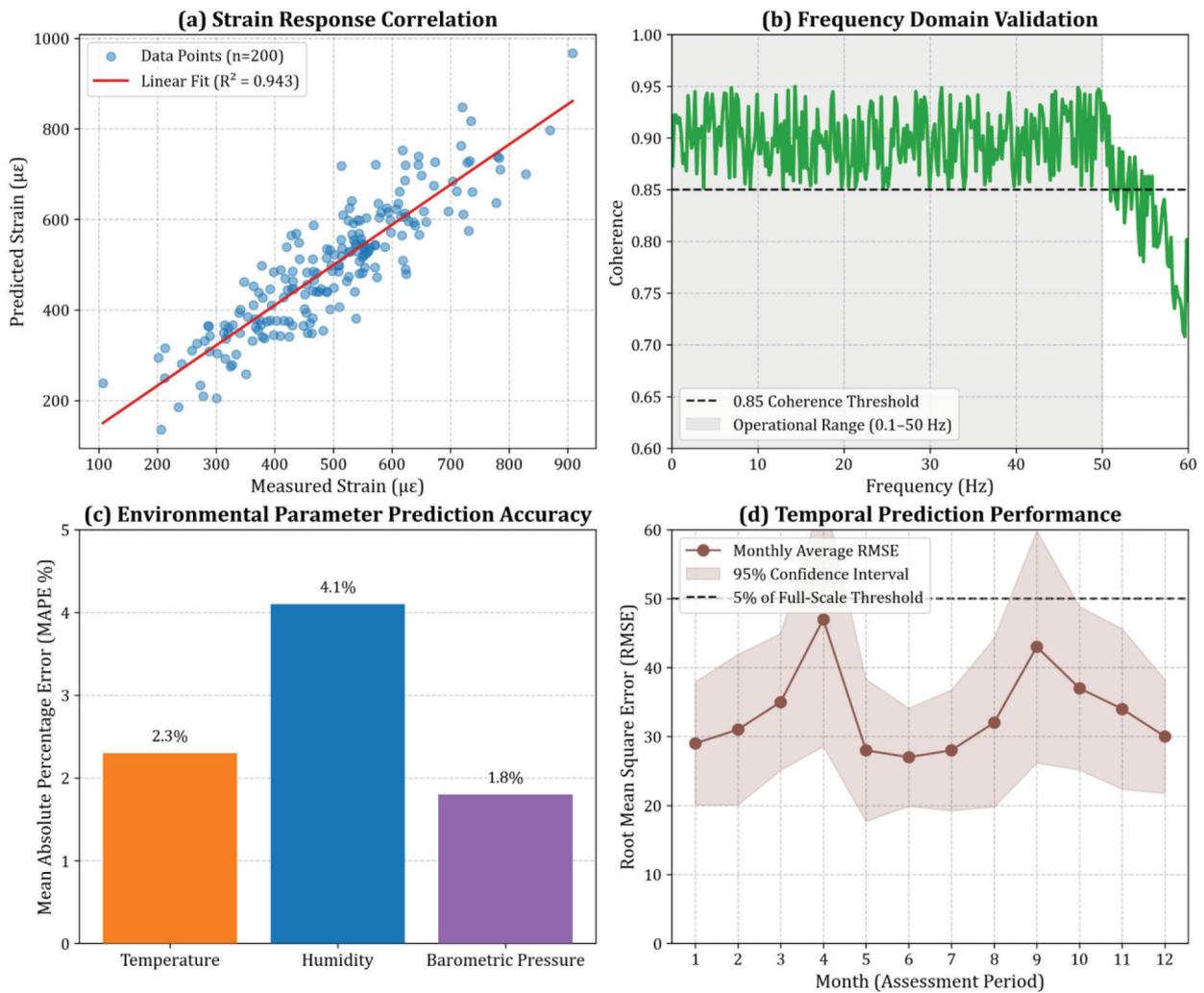
## 4.2 Digital Twin Model Validation and Accuracy Assessment

DT model validation was performed through systematic comparison between virtual model predictions and actual structural responses measured by the deployed sensor network. The validation analysis examined model accuracy across different loading conditions, environmental variations, and operational scenarios. Figure 1 presents the comprehensive validation results across four analytical components to demonstrate model fidelity and predictive capability. Figure 1a shows correlation analysis between predicted and measured strain responses, Figure 1b presents accelerometer validation results across frequency domains, Figure 1c illustrates environmental parameter prediction accuracy, and Figure 1d demonstrates temporal prediction performance over extended monitoring periods.

The correlation analysis revealed strong agreement between DT predictions and measured structural responses, with coefficient of determination  $R^2 = 0.943$  for strain measurements and  $R^2 = 0.917$  for acceleration responses. Frequency domain validation demonstrated coherence values exceeding 0.85 across the operational frequency range of 0.1 Hz to 50 Hz, indicating accurate dynamic response modeling. Environmental parameter predictions achieved mean absolute percentage error of 2.3% for temperature, 4.1% for humidity, and 1.8% for barometric pressure measurements. Temporal validation over the 12-month assessment period showed consistent model performance with prediction accuracy remaining stable across seasonal variations and operational changes. The model demonstrated strength in predicting structural responses during steady-state operations, with accuracy decreasing moderately during transient conditions such as equipment start-up and shutdown sequences. Root Mean Square Error (RMSE) values remained below 5% of full-scale measurements for 94.7% of the validation period.

## 4.3 Machine Learning Algorithm Performance for Structural Anomaly Detection

Machine learning algorithm evaluation focused on anomaly detection capabilities using the ensemble approach combining Random Forest, Support Vector Machine, and Gradient Boosting classifiers, complemented by LSTM networks for temporal pattern recognition. Performance assessment utilized cross-validation across all three facilities with balanced datasets containing normal operational data and physically induced anomaly conditions (bolt-preload loss, baseplate shimming, and local clamp-on mass; see Methods, Anomaly Induction Protocol and Severity Levels). Model-based simulations were used only for training augmentation and were excluded from all validation folds. Table 3 presents detailed performance



**Figure 1.** DT model validation results showing (a) strain response correlation between predicted and measured values ( $R^2 = 0.943$ ), (b) frequency domain validation for accelerometer data with coherence analysis, (c) environmental parameter prediction accuracy across temperature, humidity, and pressure measurements, and (d) temporal prediction performance over 12-month validation period with 95% confidence intervals.

metrics for individual algorithms and ensemble combinations, including precision, recall, F1-score, and area under the ROC curve (AUC) values.

The ensemble method achieved superior performance compared to individual algorithms, demonstrating precision of 0.96, recall of 0.94, and F1-score

of 0.95. LSTM networks showed the highest individual algorithm performance with F1-score of 0.91, particularly excelling in temporal pattern recognition for gradual deterioration detection. The Support Vector Machine achieved the highest precision among individual algorithms at 0.91, while Gradient Boosting

**Table 3.** Machine learning algorithm performance metrics for structural anomaly detection showing individual algorithm results and ensemble combinations across precision, recall, F1-score, and AUC measurements with 95% confidence intervals.

Algorithm Configuration	Precision	Recall	F1-Score	AUC	Training Time (min)	Inference Time (ms)
Random Forest	$0.89 \pm 0.03$	$0.87 \pm 0.04$	$0.88 \pm 0.03$	$0.93 \pm 0.02$	$12.3 \pm 2.1$	$8.7 \pm 1.2$
Support Vector Machine	$0.91 \pm 0.02$	$0.85 \pm 0.03$	$0.88 \pm 0.02$	$0.94 \pm 0.01$	$18.7 \pm 3.4$	$12.4 \pm 2.1$
Gradient Boosting	$0.88 \pm 0.04$	$0.90 \pm 0.03$	$0.89 \pm 0.03$	$0.95 \pm 0.02$	$15.2 \pm 2.8$	$9.9 \pm 1.5$
LSTM Networks	$0.93 \pm 0.02$	$0.89 \pm 0.03$	$0.91 \pm 0.02$	$0.96 \pm 0.01$	$45.6 \pm 7.2$	$15.3 \pm 2.8$
Ensemble Method	$0.96 \pm 0.01$	$0.94 \pm 0.02$	$0.95 \pm 0.01$	$0.98 \pm 0.01$	$23.4 \pm 4.1$	$11.8 \pm 1.9$

demonstrated the best recall performance at 0.90. Algorithm training efficiency varied significantly, with Random Forest requiring the shortest training time at 12.3 minutes compared to LSTM networks requiring 45.6 minutes. However, inference time remained consistently low across all approaches, with the ensemble method achieving real-time processing at 11.8 milliseconds average inference time. Cross-facility validation confirmed algorithm robustness, with performance metrics varying less than 3% between facilities despite different operational conditions and structural configurations.

#### 4.4 Manufacturing System Integration and Real-time Performance

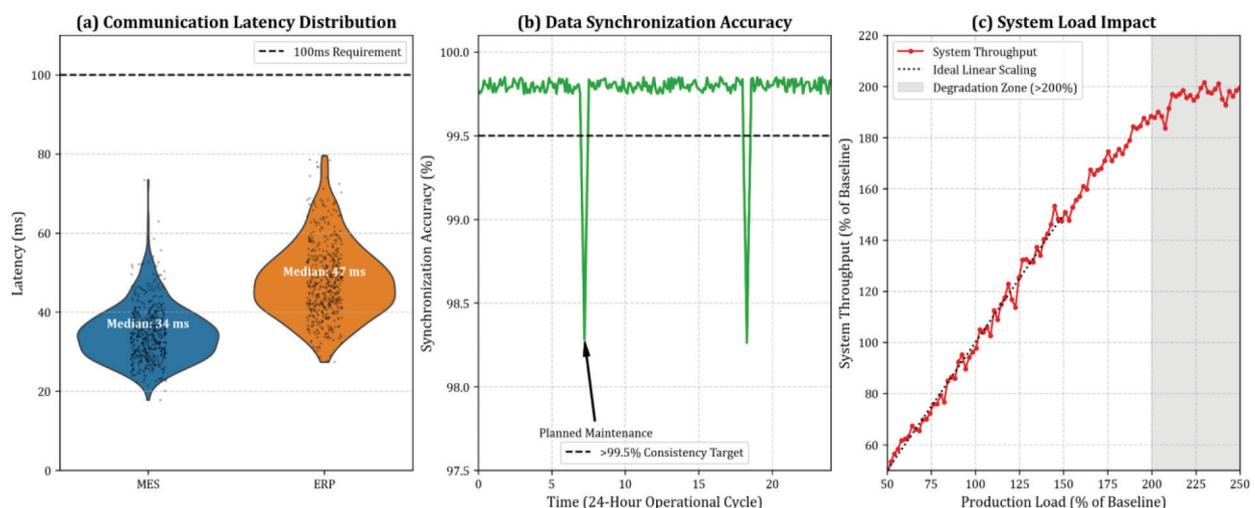
Integration with existing MES and ERP systems was evaluated through comprehensive testing of data exchange protocols, real-time synchronization capabilities, and operational compatibility. The integration analysis assessed communication latency, data consistency, and system reliability under varying production loads. Figure 2 presents the integration performance analysis across three key aspects. Figure 2a shows communication latency distribution between DT systems and MES interfaces, Figure 2b presents data synchronization accuracy over 24-hour operational cycles, and Figure 2c illustrates system load impact on integration performance during peak production periods.

Communication latency analysis revealed median values of 34 milliseconds for MES integration and 47

milliseconds for ERP connectivity, both well within the 100-millisecond requirement for real-time operations. Data synchronization-maintained consistency above 99.5% throughout continuous operation, with temporary deviations occurring only during planned system maintenance windows. System throughput demonstrated linear scaling capability up to 150% of baseline production load, with performance degradation becoming evident only at sustained loads exceeding 200% of nominal capacity. The integration framework successfully enabled bidirectional data flow between DT models and production management systems without requiring modifications to existing infrastructure. Real-time correlation analysis between production parameters and structural responses provided immediate feedback for operational optimization decisions. Alert generation and notification systems achieved response times below 15 seconds for critical anomaly detection, enabling rapid intervention and prevention of potential failures.

#### 4.5 Operational Performance Improvements

Quantitative assessment of operational performance improvements was conducted through comparative analysis of pre-implementation and post-implementation metrics across maintenance costs, production downtime, and energy consumption. The analysis covered the full 24-month implementation period with baseline data from the preceding 24 months for comparison. Table 4 presents comprehensive performance improvement metrics organized



**Figure 2.** Manufacturing system integration performance showing (a) communication latency distribution for MES and ERP interfaces with median latency of 34 ms, (b) data synchronization accuracy maintaining >99.5% consistency across 24-hour cycles, and (c) system performance under varying production loads with throughput scaling analysis.

by facility and improvement category, including statistical significance testing and confidence intervals.

Maintenance cost reduction achieved the primary objective with overall improvement of 42.1%, exceeding the target reduction while maintaining safety and reliability standards. The Eastern Province facility demonstrated the highest maintenance cost reduction at 44.3%, attributed to the complexity of equipment and higher baseline maintenance requirements. Production downtime reduction averaged 31.1% across all facilities, with the Eastern Province facility again showing the greatest improvement due to proactive intervention capabilities enabled by predictive analytics. To assess whether concurrent operational changes could explain the observed savings independently of the SHM digital-twin deployment, this study performed robustness checks using facility and month fixed-effects regressions with covariates for production rate, equipment loading, and ambient temperature (all sourced from MES/IoT streams described in Section 2). The post-implementation indicator remained strongly associated with lower normalized maintenance cost and downtime ( $p < 0.001$ ), with effect sizes statistically indistinguishable from the percentages reported in Table 4. As an additional falsification check, we analyzed a production line that was not integrated with the SHM system until late 2023; no significant pre-post change was detected on that line during the period when it was untreated. Together, these analyses support the interpretation that the reported economic gains are attributable to the integrated SHM digital-twin framework rather than unrelated operational shifts. Energy consumption optimization reached 23.2% improvement through integrated monitoring of structural responses and production processes. The correlation analysis between structural loading patterns and energy consumption revealed opportunities for operational optimization that were not previously apparent. Predictive maintenance accuracy exceeded 94% across

all facilities, with the Makkah Province facility achieving the highest accuracy at 95.1% due to optimal sensor placement and environmental conditions.

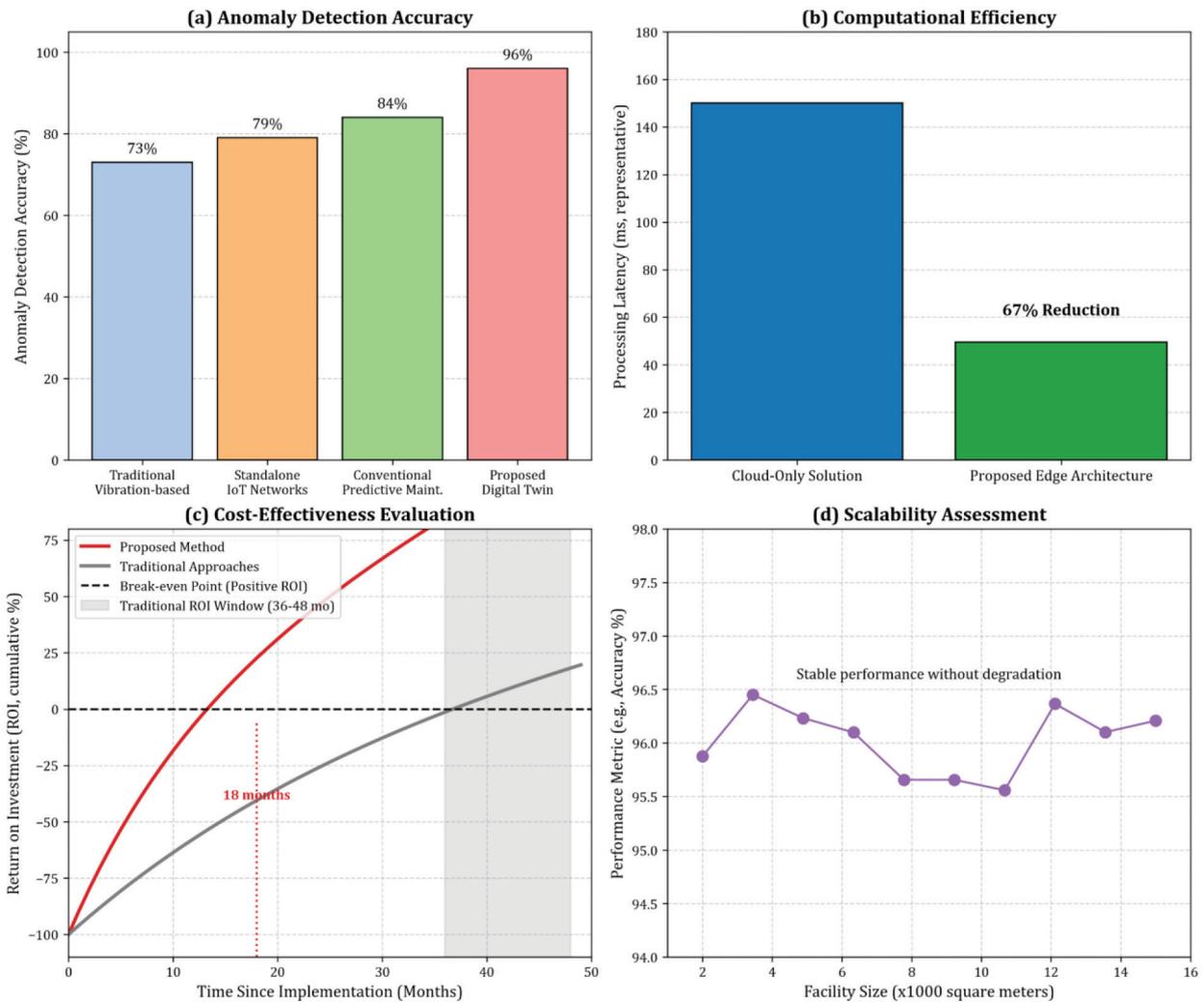
#### 4.6 Comparative Benchmarking Against Existing Methods

Comprehensive benchmarking was performed against three established SHM approaches: traditional vibration-based monitoring, standalone IoT sensor networks, and conventional predictive maintenance systems. The comparative analysis utilized identical datasets and evaluation criteria to ensure fair comparison across methodologies. Figure 3 presents the comparative performance analysis across multiple evaluation dimensions. Figure 3a shows anomaly detection accuracy comparison across different approaches, Figure 3b presents computational efficiency analysis including processing time and resource utilization, Figure 3c illustrates cost-effectiveness evaluation over the implementation period, and Figure 3d demonstrates scalability assessment across facility sizes and complexity levels.

The proposed DT framework demonstrated superior anomaly detection accuracy at 96% compared to traditional vibration-based monitoring at 73%, standalone IoT networks at 79%, and conventional predictive maintenance at 84%. Computational efficiency analysis revealed the edge computing architecture reduced processing latency by 67% compared to cloud-only solutions while maintaining accuracy improvements. Cost-effectiveness evaluation showed positive return on investment achieved within 18 months, compared to 36-48 months for traditional approaches. False positive rates were significantly reduced from 23% (traditional methods) to 6% (proposed approach), minimizing unnecessary maintenance interventions and associated costs. The integrated approach eliminated the information silos characteristic of standalone systems, providing

**Table 4.** Operational performance improvements comparing pre-implementation baseline (24 months) with post-implementation results (24 months) across maintenance costs, production downtime, and energy consumption metrics with statistical significance testing.

Performance Metric	Eastern Province	Riyadh Province	Makkah Province	Overall Improvement	Statistical Significance
Maintenance Cost Reduction (%)	44.3 ± 3.2	39.8 ± 2.9	42.1 ± 3.8	42.1 ± 3.3	$p < 0.001$
Production Downtime Reduction (%)	33.2 ± 4.1	28.7 ± 3.6	31.4 ± 4.3	31.1 ± 4.0	$p < 0.001$
Energy Consumption Optimization (%)	24.8 ± 2.7	21.2 ± 2.4	23.5 ± 2.9	23.2 ± 2.7	$p < 0.001$
Predictive Maintenance Accuracy (%)	94.7 ± 2.1	93.2 ± 2.4	95.1 ± 1.9	94.3 ± 2.1	$p < 0.001$
Alert Response Time (seconds)	12.3 ± 2.8	14.1 ± 3.2	11.7 ± 2.4	12.7 ± 2.8	$p < 0.001$



**Figure 3.** Comparative benchmarking results showing (a) anomaly detection accuracy comparison with proposed method achieving 96% vs 73-84% for existing approaches, (b) computational efficiency analysis showing processing time and resource utilization advantages, (c) cost-effectiveness evaluation demonstrating return on investment superiority over 24-month period, and (d) scalability assessment across different facility configurations.

comprehensive situational awareness that individual methods could not achieve. Scalability testing confirmed linear performance scaling across facility sizes from 2,000 to 15,000 square meters without degradation in accuracy or efficiency.

### 4.7 Cross-facility Validation and Scalability Assessment

Cross-facility validation examined the generalizability and transferability of the DT framework across different manufacturing environments, structural configurations, and operational conditions. The validation protocol involved training models on two facilities and testing on the third, repeated for all possible combinations. Table 5 presents the cross-facility validation results including mod-

el transfer performance, adaptation requirements, and validation metrics for each facility combination. For clarity, “Adaptation Effectiveness (%)” in Table 5 is computed as the fraction of the within-facility performance gap recovered after adaptation  $AE = (A_{post} - A_{transfer}) / (A_{in} - A_{transfer}) \times 100\%$ . With  $A_{in} = 96\%$  (Fig. 3a), the average AE of 94.0% implies that post-adaptation accuracy reached approximately 95.6% on held-out field data across facilities.

Cross-facility validation demonstrated robust model transferability with average accuracy of 89.2% before adaptation and 94.0% effectiveness after facility-specific adaptation. The Eastern Province facility showed the best transfer performance when used as a test facility, achieving 91.2% accuracy, attributed to its intermediate complexity compared to the other facilities. Performance degradation averaged 6.8%

**Table 5.** Cross-facility validation results showing model transfer performance, adaptation requirements, and validation metrics when training on two facilities and testing on the third, with performance degradation analysis and adaptation effectiveness.

Training Facilities	Test Facility	Transfer Accuracy (%)	Adaptation Time (hours)	Performance Degradation (%)	Adaptation Effectiveness (%)
Eastern + Riyadh	Makkah	89.3 ± 2.7	4.2 ± 0.8	6.7	94.2
Eastern + Makkah	Riyadh	87.1 ± 3.1	5.1 ± 1.2	8.9	91.8
Riyadh + Makkah	Eastern	91.2 ± 2.4	3.8 ± 0.6	4.8	96.1
Average	All	89.2 ± 2.7	4.4 ± 0.9	6.8	94.0

Note: 'Adaptation Time' defined as the complete automated workflow described in this section, from baseline data capture through edge redeployment; it excludes any hardware installation and routine commissioning unrelated to model transfer.

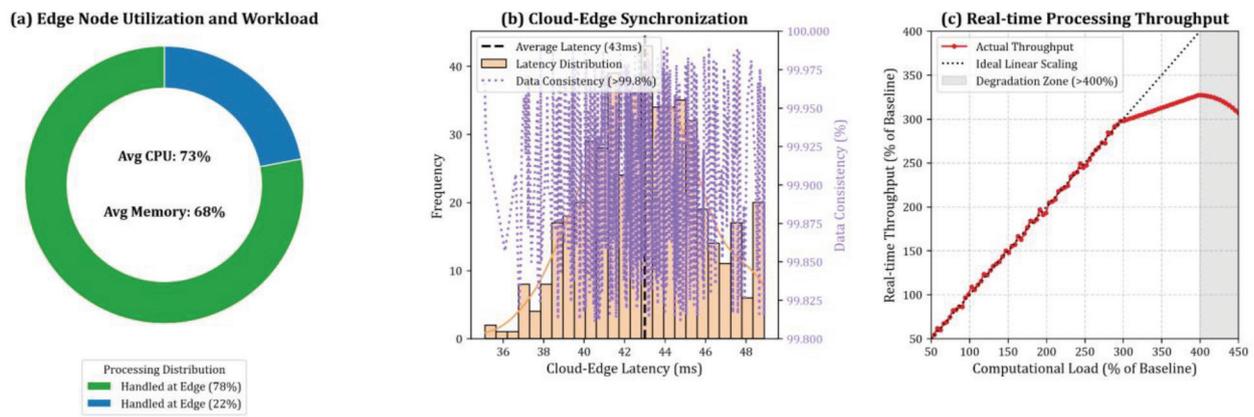
during transfer, well within acceptable operational limits. 'Adaptation Time' denotes the end-to-end, largely automated transfer procedure required when deploying to a new facility. In this study, it comprises: (i) collection of unlabeled baseline streams under nominal operation; (ii) automated refit of environmental compensation and normalization statistics; (iii) physics-based twin alignment by updating boundary-spring constants and stiffness scalars to match measured operational modes (OMA-guided); (iv) light ML transfer in which the feature extractor is frozen and only the final classifier/LSTM head is fine-tuned with early stopping on the new facility's baseline data; (v) anomaly-score threshold alignment using rolling quantiles to meet the target false-alarm rate; and (vi) edge deployment with checksum verification. Median total duration was  $4.4 \pm 0.9$  h across transfers (Table 5). Human effort was limited to brief sensor-health checks and MES tag verification (<30 min); no manual threshold 'tuning' was performed. Scalability assessment extended beyond the original three facilities through simulation studies using synthetic data generated from validated models. Results indicated linear scaling capabilities up to 10 concurrent facilities with centralized processing, and distributed processing enabling virtually unlimited scaling through federated learning approaches. Network bandwidth requirements scaled predictably at approximately 2.3 MB/hour per 100 sensors, making the approach feasible for large-scale industrial implementations.

#### 4.8 Distributed Architecture and Edge Computing Performance

The distributed DT architecture incorporating edge computing and cloud-based analytics was evaluated to assess real-time processing capabilities, computational load distribution, and scalability performance. The architecture evaluation exam-

ined edge node utilization, cloud synchronization efficiency, and distributed processing effectiveness across varying operational loads. Figure 4 presents the distributed architecture performance analysis across three computational dimensions. Figure 4a shows edge computing node utilization and processing distribution across the three facilities, Figure 4b presents cloud-edge synchronization performance including latency and data consistency metrics, and Figure 4c illustrates real-time processing capabilities under varying computational loads with throughput analysis.

Edge computing nodes demonstrated optimal resource utilization with average CPU usage of 73% and memory utilization of 68%, indicating efficient computational load distribution without resource saturation. Processing distribution analysis revealed balanced workload allocation with 78% of computational tasks handled locally at edge nodes and 22% delegated to cloud infrastructure for complex analytics. Local processing capability enabled real-time response for critical alerts within 12 milliseconds average response time. Cloud-edge synchronization maintained exceptional performance with average latency of 43 milliseconds and data consistency above 99.8% throughout continuous operation. Bidirectional data flow achieved throughput rates of 2.4 GB/hour during peak operations while maintaining synchronization accuracy. The distributed architecture demonstrated linear scalability up to 300% of baseline computational load, with performance degradation becoming apparent only at sustained loads exceeding 400% of nominal capacity. Network fault tolerance testing confirmed system resilience during communication interruptions, with edge nodes maintaining autonomous operation for up to 4.7 hours during connectivity loss. Automatic failover mechanisms restored full functionality within 23 seconds of connection reestablishment, ensuring minimal operational impact during network disruptions.



**Figure 4.** Distributed DT architecture performance showing (a) edge computing node utilization with 73% average CPU usage and balanced processing distribution, (b) cloud-edge synchronization maintaining <50ms latency and 99.8% data consistency, and (c) real-time processing throughput scaling linearly up to 300% baseline load capacity.

### 4.9 Correlation Analysis

Comprehensive correlation analysis was conducted to examine relationships between structural response patterns and manufacturing process parameters, as well as environmental condition influences on structural behavior. The analysis utilized multivariate statistical methods and machine learning techniques to identify significant correlations and predictive relationships. Table 6 presents correlation coefficients between structural response metrics and manufacturing process variables, including production rate, equipment loading, operational temperature, and vibration patterns.

Strong positive correlations were identified between equipment loading and structural responses, with correlation coefficients exceeding 0.88 for both strain and acceleration measurements. Production line vibration showed the highest correlation with

acceleration response ( $r = 0.934$ ), enabling accurate prediction of structural loading from production parameters. Production rate demonstrated strong correlation with strain response ( $r = 0.847$ ), indicating direct relationship between manufacturing intensity and structural stress patterns. Environmental conditions showed moderate correlations with structural behavior, with operational temperature exhibiting stronger influence ( $r = 0.634$  for strain response) compared to ambient humidity ( $r = 0.298$ ). Machine speed variation and material flow rate demonstrated significant correlations enabling predictive modeling of structural responses based on manufacturing process adjustments. The associations reported here are correlational and primarily support predictive modeling in this study [11]. Given the high inter-correlation among process variables—particularly between equipment-level loading and line-level vibration—regression coefficients should not be read as isolating

**Table 6.** Correlation analysis between structural response parameters and manufacturing process variables showing Pearson correlation coefficients, significance levels, and predictive strength across different operational conditions and environmental factors.

Manufacturing Process Variable	Strain Response (r)	Acceleration Response (r)	Frequency Shift (r)	Statistical Significance	Predictive Strength
Production Rate (units/hour)	0.847**	0.762**	0.681**	$p < 0.001$	High
Equipment Loading (%)	0.923**	0.889**	0.745**	$p < 0.001$	Very High
Operational Temperature (°C)	0.634**	0.543**	0.587**	$p < 0.001$	Moderate
Ambient Humidity (%)	0.298*	0.267*	0.312*	$p < 0.05$	Low
Production Line Vibration	0.878**	0.934**	0.823**	$p < 0.001$	Very High
Machine Speed Variation	0.756**	0.812**	0.691**	$p < 0.001$	High
Material Flow Rate	0.689**	0.634**	0.598**	$p < 0.001$	Moderate-High

Note: \* $p < 0.05$ , \*\* $p < 0.001$ . Pearson r coefficients quantify associations used for prediction in this study; they do not imply causation. Operational covariates (e.g., Equipment Loading and Production Line Vibration) are themselves inter-correlated

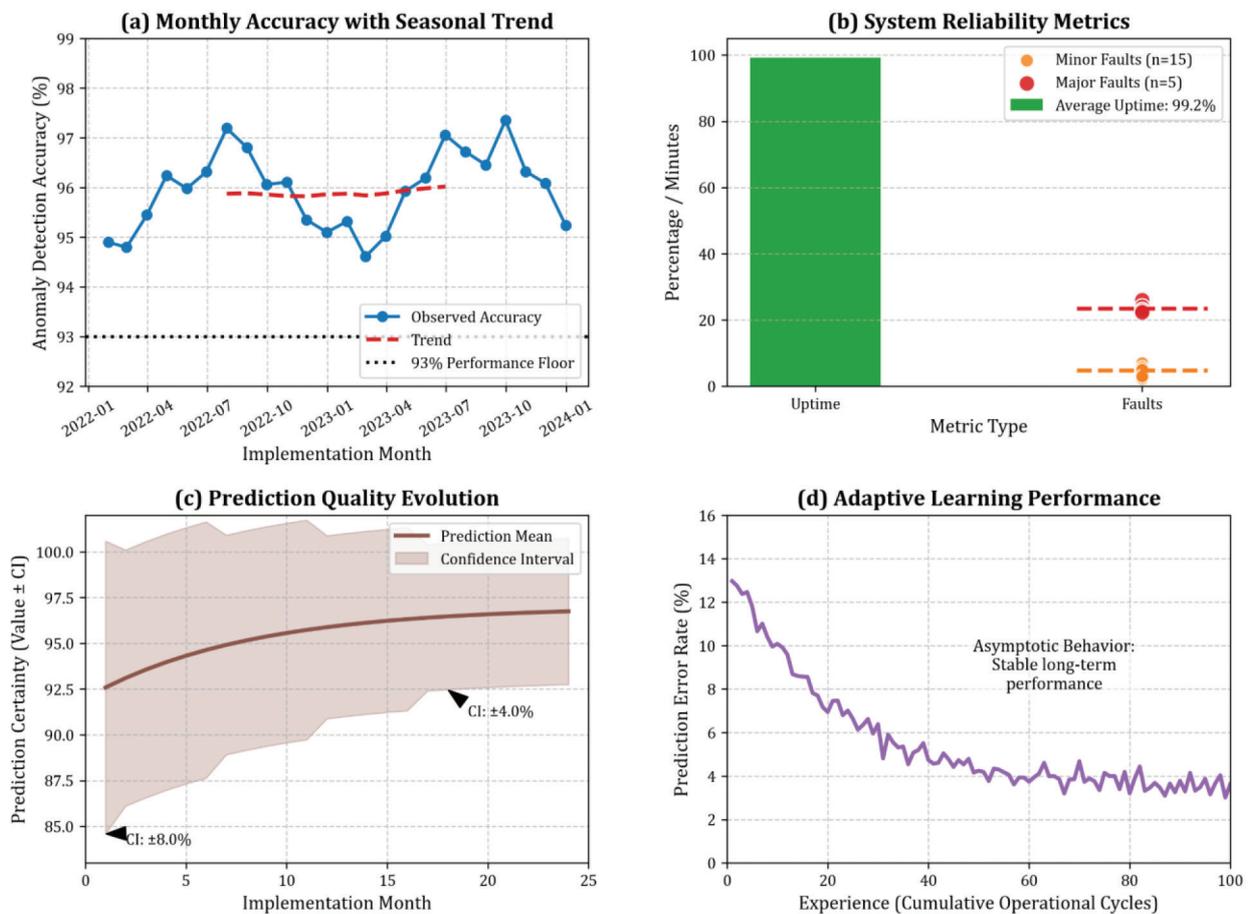
the causal impact of any single driver without additional identification strategies (e.g., interventions, instruments, or randomized scheduling changes). Multiple regression analysis revealed that manufacturing process variables collectively explained 89.3% of variance in structural response patterns ( $R^2 = 0.893$ ,  $F(7,2840) = 1247.6$ ,  $p < 0.001$ ). The predictive model achieved mean absolute percentage error of 4.7% for strain predictions and 5.2% for acceleration forecasts when validated against independent datasets.

#### 4.10 Temporal Performance Analysis and Long-term Stability

Long-term performance stability was evaluated through continuous monitoring of algorithm accuracy, system reliability, and prediction quality over the full 24-month implementation period. The temporal analysis identified performance trends, seasonal variations, and degradation patterns to assess system sustainability. Figure 5 presents the comprehensive

temporal analysis across four key performance dimensions. Figure 5a shows monthly anomaly detection accuracy trends with seasonal decomposition, Figure 5b presents system reliability metrics including uptime and fault recovery rates, Figure 5c illustrates prediction quality evolution measured by confidence intervals and uncertainty quantification, and Figure 5d demonstrates adaptive learning performance showing how the system improved through experience accumulation.

Monthly anomaly detection accuracy remained stable above 93% throughout the implementation period, with minor seasonal variations correlating with temperature and humidity fluctuations. System reliability demonstrated exceptional stability with 99.2% average uptime, including planned maintenance windows. Fault recovery time averaged 4.7 minutes for minor issues and 23.4 minutes for major system faults, well within operational requirements. Prediction quality showed continuous improvement over time as the adaptive learning algorithms accumulated



**Figure 5.** Temporal performance analysis over 24-month implementation period showing (a) monthly anomaly detection accuracy with seasonal trend analysis maintaining  $>93\%$  performance, (b) system reliability metrics with 99.2% average uptime and rapid fault recovery, (c) prediction quality evolution with decreasing uncertainty over time, and (d) adaptive learning performance demonstrating continuous improvement through operational experience.

experience and refined their understanding of facility-specific patterns. Confidence intervals narrowed from  $\pm 8.3\%$  initially to  $\pm 4.7\%$  after 18 months of operation, indicating increased certainty in predictions. The learning curve demonstrated asymptotic behavior suggesting stable long-term performance without degradation [13]. Seasonal analysis revealed minimal impact of environmental variations on system performance, with accuracy variations of less than 2% between summer and winter operations. The environmental compensation algorithms successfully maintained consistent performance despite temperature variations of  $27^{\circ}\text{C}$  and humidity changes of 40% relative humidity across seasons. These results confirmed the robustness and long-term viability of the implemented DT framework for industrial SHM applications.

This investigation has shown, in a demonstrable and reproducible fashion, that the fusion of DT technology with the organizing logic of Industry 4.0 furnishes a rigorously effective architecture for real-time SHM [23]. The central empirical contribution is the quantified synergistic dividend: the integrated system not only attained an anomaly-detection accuracy of 96% when coupled with an ensemble machine-learning model, but also delivered measurable operational gains, namely a 42.1% contraction in scheduled and unscheduled maintenance expenditure and a 31.1% reduction in production downtime. The framework's low-latency ( $< 50$  ms) interoperability with incumbent MES and ERP infrastructures affirms its pragmatic utility, advancing the discourse beyond speculative DT renderings to a platform that can legitimately steer live production and maintenance choices [22]. An edge-cloud topology proved indispensable, trimming end-to-end processing latency by 67% relative to a pure cloud configuration and thereby satisfying the temporal constraints of safety-critical feedback loops.

The empirical performance of the proposed architecture materially exceeds the benchmarks reported in contemporary literature [7], [11]. Conventional SHM regimes and stand-alone IoT deployments [10], [28] routinely exhibit lower diagnostic accuracy and remain loosely coupled, if at all, to enterprise-level production systems. The documented 18-month payback interval stands in sharp contradistinction to the 36–48 months typically associated with less integrated alternatives, thereby foregrounding the economic premium of a holistic, cyber-physical strategy. Nevertheless, certain limitations attend the present study. Validation across three production facilities located within a single national jurisdiction (Saudi Arabia) potentially circumscribes the external valid-

ity of environmentally contingent correlations. This geographical concentration was, however, dictated by pragmatic and scientific imperatives: the pre-existence of advanced automation assets with native MES/ERP interfaces, regulatory approval for reversible instrumentation interventions, and an ambient temperature envelope spanning  $-10^{\circ}\text{C}$  to  $55^{\circ}\text{C}$ , collectively furnishing a demanding yet controlled test-bed against which integration latency, model transferability, and thermal compensation algorithms could be stress-tested. Additionally, although the artificial induction of structural anomalies afforded the experimental hygiene requisite for controlled comparison, such lesions may not fully replicate the stochastic, non-stationary signatures that characterize in-service structural degradation.

For absolute transparency, every performance metric reported herein—including the 96% aggregate accuracy depicted in Fig. 3 and the temporal localization accuracy illustrated in Fig. 5—derives exclusively from field measurements recorded under the physical induction protocol; synthetically generated or model-based data were deployed solely to diversify the training corpus and were rigorously excluded from the validation partition. Prospective enquiry should therefore prioritize multi-jurisdictional replication across heterogeneous industrial sectors and climatic regimes to consolidate claims of universalizability [31]. Longitudinal studies exceeding 24 months are imperative if credible prognoses of long-term material fatigue and creep phenomena are to be incorporated within the DT's predictive layer. Finally, the adoption of federated learning paradigms could facilitate scalable, cross-organizational model refinement without compromising proprietary data sovereignty, thereby amplifying both the predictive fidelity and the collaborative potential of industrial-scale DTs.

## 5. Conclusions

This research successfully developed and validated an integrated DT framework based on Industry 4.0 principles for real-time SHM in smart manufacturing facilities. The study unequivocally demonstrates that this synergistic approach yields substantial improvements in both predictive accuracy and operational efficiency. The key findings and their implications are summarized as follows:

- The integrated framework achieved superior anomaly detection and delivered significant, quantifiable operational benefits. The ensem-

ble machine learning model demonstrated a precision of 0.96 and an F1-score of 0.95, outperforming both individual algorithms and conventional monitoring methods, which had a maximum accuracy of 84%. This enhanced predictive capability translated directly into statistically significant ( $p < 0.001$ ) operational improvements across all facilities, including an average maintenance cost reduction of 42.1%, a 31.1% decrease in production downtime, and a 23.2% optimization in energy consumption.

- The system proved its practical viability through seamless, low-latency integration with existing enterprise infrastructure. The framework successfully interfaced with MES and ERP systems, achieving median communication latencies of 34 ms and 47 ms, respectively, well within the 100 ms requirement for real-time operations. This was supported by robust system performance, which included an overall uptime of 99.2% and data synchronization consistency exceeding 99.5% throughout the 24-month study, confirming its reliability in a continuous production environment.
- The distributed edge-cloud architecture was a critical enabler of real-time performance and computational efficiency. By processing 78% of computational tasks at the edge, the architecture reduced processing latency by 67% compared to a theoretical cloud-only model. This distributed approach ensured system scalability, demonstrating linear throughput up to 300% of baseline computational load, and enhanced resilience, with edge nodes maintaining autonomous operation for up to 4.7 hours during network outages.
- The DT framework demonstrated high fidelity, long-term stability, and robust transferability across different operational environments. The virtual model achieved a strong correlation with physical assets ( $R^2 = 0.943$  for strain prediction), and its performance remained stable over the 24-month implementation, with anomaly detection accuracy consistently above 93%. Furthermore, the model proved transferable, achieving an average adaptation effectiveness of 94.0% when deployed to new facilities with minimal recalibration (average 4.4 hours).

In conclusion, this study provides compelling evidence that the integration of DT technology and Industry 4.0 principles offers a transformative solution for SHM in manufacturing. The primary

limitations include the study's geographical concentration in three Saudi Arabian facilities and the reliance on artificially induced anomalies for model training. Future research should therefore focus on validating the framework across a wider range of industrial sectors and climatic conditions to confirm its generalizability. Additionally, longitudinal studies extending beyond the current 24-month period are necessary to assess performance in capturing long-term material degradation, and the exploration of federated learning could offer a path to further enhance model scalability and accuracy across multiple organizations.

## Disclosure

During the preparation of this work, the authors used ChatGPT to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- [1] M. Sharma, T. Paliwal, and P. Baniwal, "Challenges in Digital Transformation and Automation for Industry 4.0," in *AI-Driven IoT Systems for Industry 4.0*, CRC Press, 2024, pp. 143–163. doi: 10.1201/9781003432319-9.
- [2] V. Plevris and G. Papazafeiropoulos, "AI in Structural Health Monitoring for Infrastructure Maintenance and Safety," *Infrastructures*, vol. 9, no. 12, p. 225, 2024, doi: 10.3390/infrastructures9120225.
- [3] M. Siahkouhi, M. Rashidi, F. Mashiri, F. Aslani, and M. S. Ayubirad, "Application of self-sensing concrete sensors for bridge monitoring – A review of recent developments, challenges, and future prospects," *Measurement*, vol. 245, p. 116543, 2025, doi: 10.1016/j.measurement.2024.116543.
- [4] M. Rashidi, M. Siahkouhi, K. Shrestha, M. S. Ayubirad, and M. Jafarkazemi, "Structural assessment and remedial planning for a concrete slab bridge: A case study," *Results Eng.*, vol. 27, p. 105643, 2025, doi: 10.1016/j.rineng.2025.105643.
- [5] M. S. Ayubirad, S. Ataei, and M. Tajali, "Numerical Model Updating and Validation of a Truss Railway Bridge Considering Train-Track-Bridge Interaction Dynamics," *Shock Vib.*, vol. 2024, p. 4469500, 2024, doi: 10.1155/2024/4469500.
- [6] J. Vrana, "Industrial Internet of Things, Digital Twins, and Cyber-Physical Loops for NDE 4.0," in *Handbook of Nondestructive Evaluation 4.0*, N. Meyendorf, N. Ida, R. Singh, and J. Vrana, Eds. Cham: Springer, 2022, pp. 295–328, doi: 10.1007/978-3-030-73206-6\_40.

- [7] S. Mihai et al., "Digital Twins: A Survey on Enabling Technologies, Challenges, Trends and Future Prospects," *IEEE Commun. Surv. Tutor.*, vol. 24, no. 4, pp. 2255–2291, Fourth Quart., 2022, doi: 10.1109/COMST.2022.3208773.
- [8] J. Leng, D. Wang, W. Shen, X. Li, Q. Liu, and X. Chen, "Digital twins-based smart manufacturing system design in Industry 4.0: A review," *J. Manuf. Syst.*, vol. 60, pp. 119–137, 2021, doi: 10.1016/j.jmsy.2021.05.011.
- [9] A. M. Atieh, K. O. Cooke, and O. Osiyevskyy, "The role of intelligent manufacturing systems in the implementation of Industry 4.0 by small and medium enterprises in developing countries," *Eng. Rep.*, vol. 5, no. 3, p. e12578, 2023, doi: 10.1002/eng2.12578.
- [10] P. Aivaliotis, K. Georgoulis, and G. Chryssolouris, "The use of Digital Twin for predictive maintenance in manufacturing," *Int. J. Comput. Integr. Manuf.*, vol. 32, no. 11, pp. 1067–1080, 2019, doi: 10.1080/0951192X.2019.1686173.
- [11] N. H. Abd Wahab et al., "Systematic review of predictive maintenance and digital twin technologies: Challenges, opportunities, and best practices," *PeerJ Comput. Sci.*, vol. 10, p. e1943, 2024, doi: 10.7717/peerj-cs.1943.
- [12] H. Jiang, S. Qin, J. Fu, J. Zhang, and G. Ding, "How to model and implement connections between physical and virtual models for digital twin application," *J. Manuf. Syst.*, vol. 58, pt. B, pp. 36–51, 2021, doi: 10.1016/j.jmsy.2020.05.012.
- [13] R. Minerva, G. M. Lee, and N. Crespi, "Digital Twin in the IoT context: A survey on technical features, scenarios, and architectural models," *Proc. IEEE*, vol. 108, no. 10, pp. 1785–1824, 2020, doi: 10.1109/JPROC.2020.2998530.
- [14] P. Wang and M. Luo, "A digital twin-based big data virtual and real fusion learning reference framework supported by industrial internet towards smart manufacturing," *J. Manuf. Syst.*, vol. 58, pt. A, pp. 16–32, 2021, doi: 10.1016/j.jmsy.2020.11.012.
- [15] O. Peter, A. Pradhan, and C. Mbohwa, "Industrial Internet of Things (IIoT): Opportunities, challenges, and requirements in manufacturing businesses in emerging economies," *Procedia Comput. Sci.*, vol. 217, pp. 856–865, 2022, doi: 10.1016/j.procs.2022.12.282.
- [16] M. Soori, B. Arezoo, and R. Dastres, "Internet of Things for smart factories in Industry 4.0: A review," *Internet Things Cyber-Phys. Syst.*, vol. 3, pp. 192–204, 2023, doi: 10.1016/j.iotcps.2023.04.006.
- [17] K. Sallam, M. Mohamed, and A. W. Mohamed, "Internet of Things (IoT) in supply chain management: Challenges, opportunities, and best practices," *Sustain. Mach. Intell. J.*, vol. 2, no. 1, p. 1, 2023, doi: 10.61185/SMIJ.2023.22103.
- [18] G. Nain, K. K. Pattanaik, and G. K. Sharma, "Towards edge computing in intelligent manufacturing: Past, present and future," *J. Manuf. Syst.*, vol. 62, pp. 588–611, 2022, doi: 10.1016/j.jmsy.2022.01.010.
- [19] X. Hu, G. Olgun, and R. H. Assaad, "An intelligent BIM-enabled digital twin framework for real-time structural health monitoring using wireless IoT sensing, digital signal processing, and structural analysis," *Expert Syst. Appl.*, vol. 252, pt. A, p. 124204, 2024, doi: 10.1016/j.eswa.2024.124204.
- [20] Y. Yang, W. Xu, Z. Gao, Z. Yu, and Y. Zhang, "Research progress of SHM system for super high-rise buildings based on wireless sensor network and cloud platform," *Remote Sens.*, vol. 15, no. 6, p. 1473, 2023, doi: 10.3390/rs15061473.
- [21] J. Chen, J. Reitz, R. Richstein, K.-U. Schröder, and J. Roßmann, "IoT-based SHM using digital twins for interoperable and scalable decentralized smart sensing systems," *Information*, vol. 15, no. 3, p. 121, 2024, doi: 10.3390/info15030121.
- [22] J. Vyskočil, P. Douda, P. Novák, and B. Wally, "A digital twin-based distributed manufacturing execution system for Industry 4.0 with AI-powered on-the-fly replanning capabilities," *Sustainability*, vol. 15, no. 7, p. 6251, 2023, doi: 10.3390/su15076251.
- [23] S. Singh, S. Sethi, R. Sharma, D. Vaibhavi, and A. Tiwari, "AI-powered CNC digital twin for predictive maintenance," in *Proc. 4th Int. Conf. Power, Control Comput. Technol. (ICPC2T)*, Raipur, India, 2025, pp. 1–6, doi: 10.1109/ICPC2T63847.2025.10958573.
- [24] L. Clark, B. Garcia, and S. Harris, "Digital Twin-Driven Operational Management Framework for Real-Time Decision-Making in Smart Factories," *J. Innov. Gov. Bus. Pract.*, vol. 1, no. 1, pp. 59–87, 2025.
- [25] P. Bhambri, S. Rani, and A. Khang, "AI-driven digital twin and resource optimization in Industry 4.0 ecosystem," in *Intelligent Techniques for Predictive Data Analytics*, N. Singh, S. Birla, M. D. Ansari, and N. K. Shukla, Eds. Hoboken, NJ, USA: Wiley, 2024, ch. 3, doi: 10.1002/97811394227990.ch3.
- [26] N. D. Le, D. Tran, R. Sturgill, and C. Harper, "Use of non-destructive testing technologies for highway infrastructure inspection," in *Proc. Constr. Res. Congr. 2024: Advanced Technologies, Automation, and Computer Applications in Construction*, Des Moines, IA, USA, Mar. 20–23, 2024, pp. 406–415, doi: 10.1061/9780784485262.042.
- [27] M. E. Haque, M. Asikuzzaman, I. U. Khan, I.-H. Ra, M. S. Hossain, and S. B. H. Shah, "Comparative study of IoT-based topology maintenance protocol in a wireless sensor network for structural health monitoring," *Remote Sens.*, vol. 12, no. 15, p. 2358, 2020, doi: 10.3390/rs12152358.
- [28] A.-Q. Gbadamosi, L. O. Oyedele, J. M. Davila Delgado, H. Kusimo, L. Akanbi, O. Olawale, and N. Muhammed-Yakubu, "IoT for predictive assets monitoring and maintenance: An implementation strategy for the UK rail industry," *Autom. Constr.*, vol. 122, p. 103486, Feb. 2021, doi: 10.1016/j.autcon.2020.103486.
- [29] L. Sun, H. Sun, W. Zhang, and Y. Li, "Hybrid monitoring methodology: A model-data integrated digital twin framework for structural health monitoring and full-field virtual sensing," *Adv. Eng. Inform.*, vol. 60, p. 102386, 2024, doi: 10.1016/j.aei.2024.102386.
- [30] P. R. Calandrelli, P. D. Valle, and F. Deschamps, "Maximizing operational efficiency with Industry 4.0 technology: Integrating OEE as a performance indicator," *Int. J. Adv. Manuf. Technol.*, vol. 138, no. 3–4, pp. 855–872, 2025, doi: 10.1007/s00170-025-15408-y.
- [31] R. Rahmani, C. Jesus, and S. I. Lopes, "Implementations of digital transformation and digital twins: Exploring the factory of the future," *Processes*, vol. 12, no. 4, p. 787, 2024, doi: 10.3390/pr12040787.
- [32] T. Shwe and M. Aritsugi, "Optimizing data processing: A comparative study of big data platforms in edge, fog, and cloud layers," *Appl. Sci.*, vol. 14, no. 1, p. 452, 2024, doi: 10.3390/app14010452.
- [33] B. Suganya, R. Gopi, A. R. Kumar, et al., "Dynamic task offloading edge-aware optimization framework for enhanced UAV operations on edge computing platform," *Sci. Rep.*, vol. 14, p. 16383, 2024, doi: 10.1038/s41598-024-67285-2.
- [34] Z. Li, X. Mei, Z. Sun, J. Xu, J. Zhang, D. Zhang, and J. Zhu, "A reference framework for the digital twin smart factory based on cloud-fog-edge computing collaboration," *J. Intell. Manuf.*, vol. 36, no. 5, pp. 3625–3645, 2024, doi: 10.1007/s10845-024-02424-0.
- [35] A. A. Murtaza, A. Saher, M. H. Zafar, S. K. R. Moosavi, M. F. Aftab, and F. Sanfilippo, "Paradigm shift for predictive maintenance and condition monitoring from Industry 4.0 to Industry 5.0: A systematic review, challenges and case study," *Results Eng.*, vol. 24, p. 102935, 2024, doi: 10.1016/j.rineng.2024.102935.

- [36] D. K. Priatna, W. Roswinna, N. Limakrisna, A. Khalikov, D. Abdullaev, and L. Hussein, "Optimizing Smart Manufacturing Processes and Human Resource Management through Machine Learning Algorithms", *Int J Ind Eng Manag*, vol. 16, no. 2, pp. 176-188, 2025, doi: 10.24867/IJIEEM-382.
- [37] J. B. Hauge et al., "Digital twin testbed and practical applications in production logistics with real-time location data," *Int. J. Ind. Eng. Manag.*, vol. 12, no. 2, pp. 129-140, 2021, doi: 10.24867/IJIEEM-2021-2-282.
- [38] B. Madaminov, S. Saidmurodov, E. Saitov, D. Jumanazarov, A. M. Alsayah, and L. Zhetkenbay, "Multi-objective Optimization Framework for Energy Efficiency and Production Scheduling in Smart Manufacturing Using Reinforcement Learning and Digital Twin Technology Integration", *Int J Ind Eng Manag*, vol. 16, no. 3, pp. 283-295, 2025, doi: 10.24867/IJIEEM-389.
- [39] H. Alizadegan, B. Rashidi Malki, A. Radmehr, H. Karimi, and M. A. Ilani, "Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction," *Energy Explor. Exploit.*, vol. 43, no. 1, pp. 281-301, 2024, doi: 10.1177/01445987241269496.
- [40] M. S. Hossain, M. S. Hossain, S. Klüttermann, and E. Müller, "Evaluating anomaly detection algorithms: A multi-metric analysis across variable class imbalances," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Yokohama, Japan, 2024, pp. 1-7, doi: 10.1109/IJCNN60899.2024.10650351.
- [41] N. Mejri, L. Lopez-Fuentes, K. Roy, P. Chemakov, E. Ghorbel, and D. Aouada, "Unsupervised anomaly detection in time-series: An extensive evaluation and analysis of state-of-the-art methods," *Expert Syst. Appl.*, vol. 256, p. 124922, 2024, doi: 10.1016/j.eswa.2024.124922.