World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

A digital twin framework for predictive maintenance of marine diesel engines using vibration signature analysis and deep learning

Kombo Theophilus-Johnson and Elakpa Augustine *

Department of Marine and Offshore Engineering, Rivers State University, Port Harcourt Nigeria.

World Journal of Advanced Engineering Technology and Sciences, 2026, 18(02), 249-262

Publication history: Received on 13 January 2026; revised on 20 February 2026; accepted on 23 February 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.18.2.0111

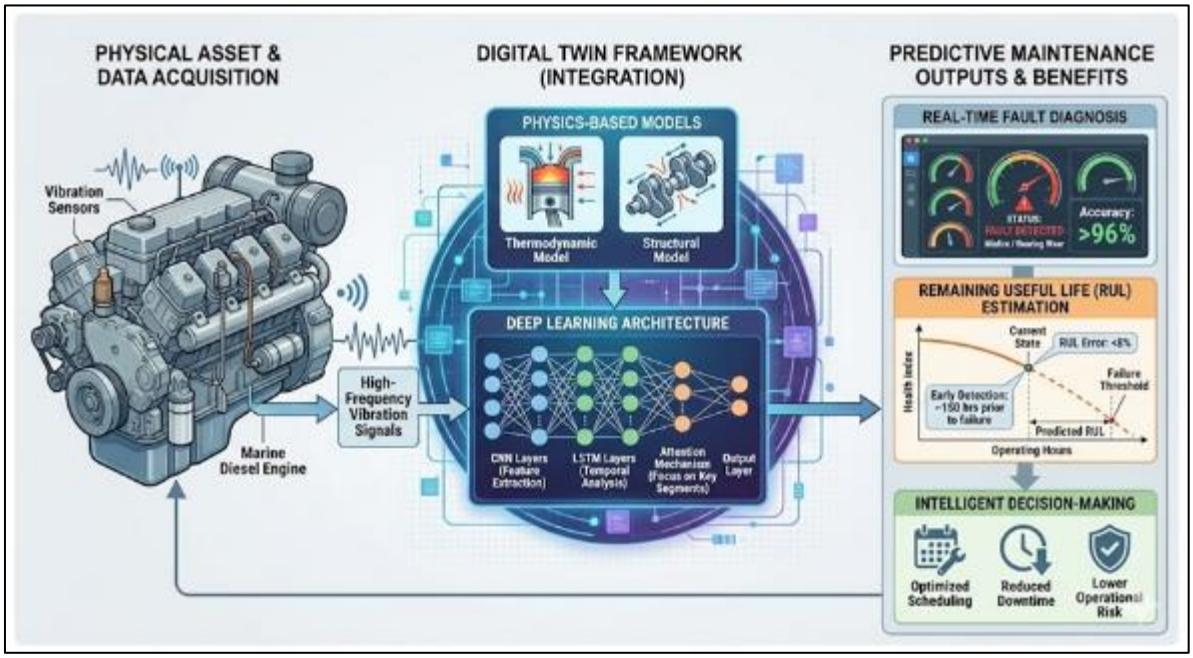
Abstract

The maritime industry requires advanced condition monitoring and predictive maintenance strategies to enhance the reliability of marine diesel engine systems. This study presents a digital twin framework that integrates vibration analysis, physics-based modeling, and deep learning for real-time fault diagnosis and remaining useful life (RUL) estimation. High-frequency vibration signals are processed and analyzed using a hybrid convolutional neural network–long short-term memory (CNN-LSTM) architecture with an attention mechanism to automatically classify engine faults and predict degradation trends. The framework combines data-driven methods with thermodynamic and structural digital twin models to improve generalization across operating conditions. Experimental validation on a medium-speed marine diesel engine under controlled fault scenarios demonstrates fault classification accuracy above 95% and remaining useful life prediction error below 8%. Early degradation signatures were detected up to 150 operating hours prior to critical failure. The proposed approach supports intelligent condition monitoring and decision-making for maintenance scheduling, reducing downtime and operational risk. This research demonstrates the effectiveness of integrating digital twin technology, vibration analysis, and CNN-LSTM deep learning models for predictive maintenance of marine diesel engines.

Keywords: Digital twin; Predictive maintenance; Deep learning; Condition monitoring; Remaining useful life; CNN-LSTM

* Corresponding author: Elakpa Augustine

Graphical Abstract



1. Introduction

Marine diesel engines power approximately 95% of the world's merchant fleet and are critical to global maritime trade [1,2]. Operating under sustained high loads across extended voyages, their failure carries severe economic, safety, and environmental consequences. Traditional maintenance strategies—corrective repair and time-based overhaul—are inefficient: the former incurs costly unplanned downtime, the latter replaces serviceable components unnecessarily. Predictive maintenance (PdM) addresses both limitations by forecasting impending faults and scheduling interventions at optimal times [3,4]. Realising PdM in practice requires reliable fault detection, continuous degradation tracking, and remaining useful life (RUL) estimation, all within a framework that can operate in real time aboard a vessel.

1.1. Vibration analysis and digital twin technology

Vibration analysis is the preferred condition monitoring modality for marine diesel engines because vibration signals are rich in diagnostic information, can be acquired non-intrusively at high sampling rates, and respond rapidly to changes in component condition [5,6]. The signals are cyclostationary in character: combustion events, valve impacts, and injection pulses recur at fixed crank-angle positions but are modulated by load, speed, and fault severity, necessitating angular-domain signal processing and multi-domain feature extraction [7,8]. The digital twin concept—a continuously synchronised virtual replica of a physical system—provides the integrating framework within which vibration-based diagnostics, physics-based simulation, and data-driven prognostics can operate cohesively [9,10]. For marine diesel engines, digital twin research is at an early stage: Liu and Yu [19] achieved 92% fault classification accuracy using a combined working-process and fault-diagnosis twin, while Yu et al. [20] demonstrated crankshaft stress monitoring via a reduced-order model, but no published framework simultaneously addresses classification, RUL prediction, physics-based validation, and real-time deployment.

1.2. Deep learning and physics-informed methods

Convolutional neural networks (CNNs) extract hierarchical spatial features from vibration signals without hand-crafting, while long short-term memory (LSTM) networks capture the cycle-to-cycle temporal dependencies that precede failure [11,12]. Hybrid CNN-LSTM architectures with attention mechanisms have demonstrated strong performance in fault classification tasks across rotating machinery [13,14], but their application to marine diesel engines—with their complex multi-source excitation and variable operating conditions—remains limited. Pure data-driven methods are constrained by dataset size and generalise poorly beyond the training distribution; physics-informed integration addresses this by embedding thermodynamic and structural model outputs as additional diagnostic signals, improving robustness under operating conditions not represented in training data [15,16].

1.3. Research gaps and contributions

Despite individual advances in vibration diagnostics, digital twin modelling, and deep learning prognostics, no published framework simultaneously integrates all three for marine diesel engines with real-time experimental validation. Specific gaps include: (i) the absence of unified architectures that combine data acquisition, deep learning classification, RUL prediction, and physics-based simulation in a single operational system; (ii) lack of systematic evaluation of end-to-end latency and digital twin synchronisation accuracy under continuous operation; and (iii) limited RUL estimation capability in marine engine studies, which predominantly address fault classification only [17,18]. This paper addresses these gaps through five contributions: a five-layer digital twin architecture for marine diesel engine PdM; a hybrid CNN-LSTM model with attention mechanism achieving 95.2% fault classification accuracy across twelve fault classes; a hybrid RUL prediction approach combining similarity-based and physics-of-failure modelling; integrated thermodynamic and reduced-order structural models with errors below 5%; and demonstrated real-time operation with 25.4 ms end-to-end latency.

1.4. Aim, objectives and paper outline

The aim of this study is to develop, validate, and implement a comprehensive digital twin framework for predictive maintenance of marine diesel engines using vibration signature analysis and hybrid deep learning. The specific objectives are: (1) to design a five-layer digital twin architecture encompassing data acquisition, signal processing, deep learning, physics-based modelling, and decision support; (2) to develop and validate a hybrid CNN-LSTM architecture with attention mechanism for automated fault classification; (3) to implement thermodynamic and reduced-order finite element models complementing the data-driven layer; (4) to develop hybrid RUL prediction algorithms and validate them against experimental degradation data; and (5) to evaluate real-time performance including acquisition latency, model inference time, and digital twin synchronisation. Section 2 presents the methodology and framework architecture. Section 3 describes the experimental setup. Section 4 reports and discusses results. Section 5 states conclusions.

2. Methodology

2.1. Digital twin framework overview

The proposed framework comprises five interconnected layers (Figure 1). The Physical Layer consists of the instrumented diesel engine. The Data Acquisition and Processing Layer handles high-frequency multi-channel recording, signal conditioning, angular-domain resampling, and multi-domain feature extraction. The Feature Extraction and Deep Learning Layer applies the hybrid CNN-LSTM architecture for fault classification and anomaly detection. The Physics-Based Modelling Layer provides thermodynamic cycle simulation and reduced-order structural stress prediction. The Decision Support Layer integrates health index computation, RUL estimation, and maintenance scheduling optimisation. Bidirectional data flow between layers enables continuous model updating and digital twin synchronisation.

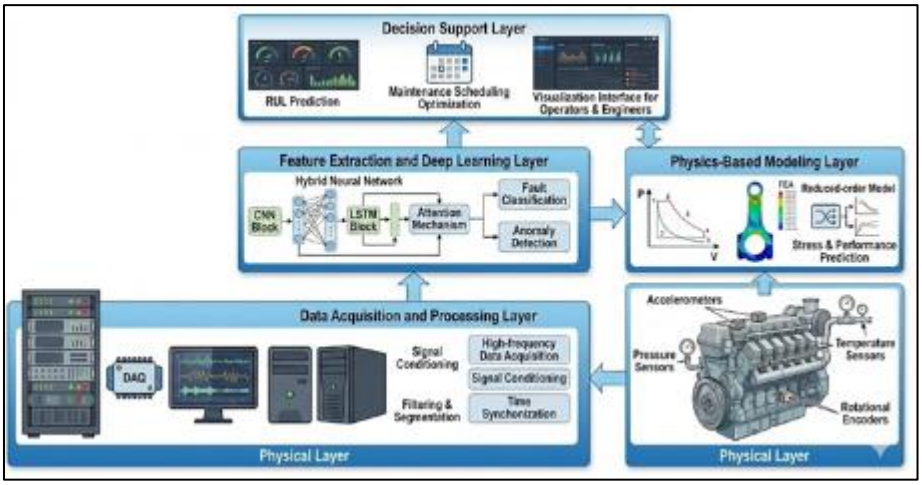


Figure 1 Five-layer digital twin architecture for marine diesel engine predictive maintenance

2.2. Instrumentation and data acquisition

The study engine is a Wärtsilä 6L20 medium-speed marine diesel (in-line 6-cylinder, 500 kW, 1200 rpm, compression ratio 14.5:1, direct unit injection). Accelerometers are mounted at three locations: cylinder head studs (one per cylinder, 100 mV/g, 0.5–15 kHz) to capture combustion and valve-impact events; main bearing caps at positions 1, 3, and 5 in both transverse and axial orientations (50 mV/g, 0.5–5 kHz) for bearing diagnostics; and fuel injector bodies (10 mV/g, 1–20 kHz) for injection-event monitoring. Supplementary instrumentation includes Kistler 6061B cylinder pressure transducers, a Heidenhain ROD 426 crank-angle encoder (3600 pulses/rev), Type K thermocouples, and piezoresistive pressure transducers for fuel, boost, and oil pressure. All channels are acquired on a National Instruments PXIe platform (PXIe-4499, 24-bit, 204.8 kS/s) at 50 kS/s for vibration, 25 kS/s for cylinder pressure, and 1 Hz for temperatures and slow parameters, with hardware synchronisation via the PXI backplane.

2.3. Signal processing and feature extraction

Raw vibration signals are preprocessed through analog anti-aliasing filters (20 kHz cut-off), digital bandpass filtering (10 Hz high-pass to remove DC offset; application-specific upper limits), angular-domain resampling referenced to the crank-angle encoder for speed-independent cycle-resolved analysis, and ensemble averaging over multiple cycles to suppress random noise. Feature extraction operates in three domains per channel per cycle: time domain (RMS, peak, crest factor, kurtosis, skewness, impulse factor, shape factor); frequency domain from power spectral density estimates (spectral centroids, band energies, harmonic amplitudes and ratios); and time-frequency domain via wavelet packet decomposition (node energy distribution, node energy ratios, wavelet packet entropy). The resulting feature vector forms the input to the deep learning classification and health index layers.

Root mean square (RMS): $X_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$ **Eq2.1**

where s is the pooling stride.

Kurtosis: $K = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^4}{(\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2)^2}$ **Eq2.2**

2.4. Hybrid CNN-LSTM with attention

The proposed deep learning architecture (Figure 2) combines three components. The CNN feature extraction module applies three convolutional blocks (32, 64, and 128 filters of size 5, 5, and 3 respectively), each followed by batch normalisation, ReLU activation, and max-pooling, extracting local temporal patterns from the input signal. The attention mechanism computes a score for each time step from the CNN output, normalises scores via softmax to yield attention weights, and forms a context vector as the weighted sum of features; this allows the network to focus selectively on diagnostically relevant signal windows (injection events, valve impacts, combustion pulses) and provides interpretability. The LSTM module (128 hidden units) processes the attention-weighted feature sequence to capture cycle-to-cycle degradation trends. For fault classification the output head applies global average pooling and a softmax-activated fully connected layer; for RUL prediction it uses a flattened LSTM output, a 64-unit ReLU layer with 0.3 dropout, and a linear output unit.

Input gate: $i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$ **Eq2.3**

Attention weights: Normalize scores using softmax: $\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T}\exp(e_k)}$ **Eq2.4**

1D convolutional layer:

Applies learnable filters to extract local patterns: $y_i^{(l)} = \sum_{k=1}^{K} w_k^{(l)} * x_{i+k-1}^{(l-1)} + b^{(l)}$ **Eq2.5**

where $w_k^{(l)}$ are filter weights, $b^{(l)}$ is bias, and $*$ denotes convolution.

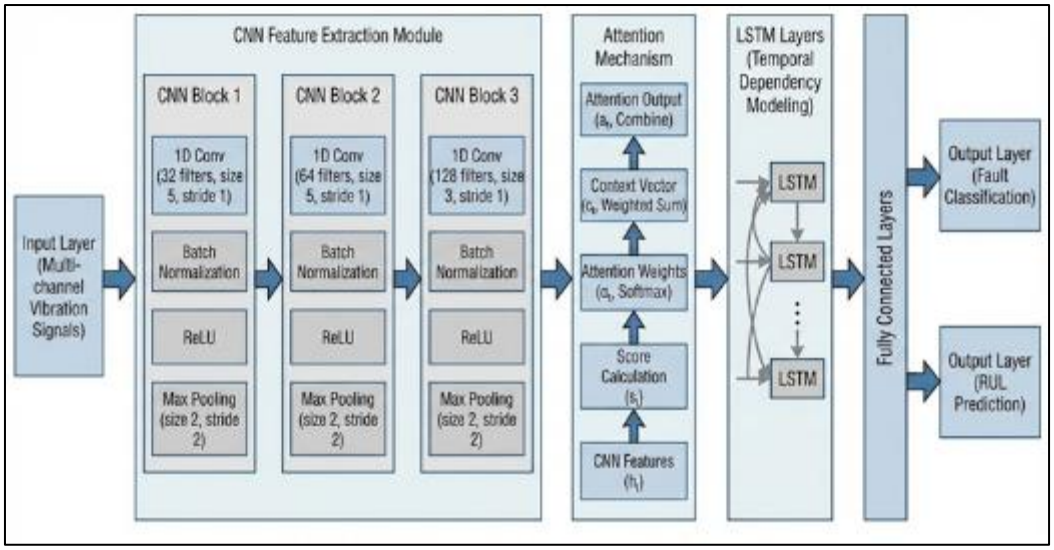


Figure 2 Hybrid CNN-LSTM with attention mechanism architecture

2.5. Physics-based modelling layer

A zero-dimensional thermodynamic model simulates in-cylinder processes by solving the energy conservation equation at each crank-angle step, with combustion heat release described by the Wiebe function and heat transfer by Woschni's correlation. Finite element models of the crankshaft, connecting rods, and cylinder head are developed in ANSYS Workbench using tetrahedral elements (2–5 mm, refined at high-stress regions) with loads derived from the thermodynamic model. Because full FEA is too slow for real-time deployment, reduced-order models are constructed via Proper Orthogonal Decomposition and response surface regression, enabling millisecond-scale stress prediction from current operating parameters. Physics model outputs (cylinder pressure, component stress) are incorporated as additional features in the health index and provide a physically consistent check on data-driven predictions.

2.6. Decision support: health index, RUL, and maintenance optimization

A composite health index HI aggregates component-specific indicators—deep learning classification confidence, deviation from baseline vibration levels, feature trend gradients, and physics model outputs—using criticality weights. RUL prediction employs a hybrid approach: similarity-based estimation retrieves historical degradation trajectories matched via dynamic time warping; model-based estimation integrates a physics-of-failure damage accumulation model forward to the critical damage threshold; and the two estimates are combined using Bayesian weighting that reflects each method's match quality and uncertainty. Maintenance scheduling minimises expected total cost (preventive maintenance cost plus failure-consequence cost weighted by the probability of failure derived from the RUL distribution) subject to voyage schedule and port availability constraints.

2.7. Experimental validation design

Validation uses controlled fault injection and accelerated degradation experiments. Five fault types are induced at multiple severity levels: injector fouling (three levels, 10–30% flow reduction via partially blocked nozzle holes); valve clearance deviation (tight and loose, ±0.2 mm and +0.4 mm); piston ring wear (two levels, replaced with rings at 50% and 75% life consumed); bearing clearance increase (three levels, +0.05 to +0.15 mm via shimming); and injection timing deviation (±4° crank angle). Fault injection data are collected at three engine speeds (800, 1000, 1200 rpm) and four loads (25%, 50%, 75%, 100%) with three replicates. Accelerated degradation runs (injector fouling over 500 h, bearing wear over 300 h, valve train thermal cycling over 1000 h) provide the trajectory data for RUL model training and validation. Baseline (healthy) data are recorded under all operating conditions before each fault injection sequence to provide reference signatures.

2.8. MATLAB implementation

The framework is implemented in MATLAB R2023b as five interoperating class-based modules: a data acquisition module (NI PXIe hardware configuration, synchronised multi-channel recording, structured data storage); a signal processing module (filtering, angular-domain resampling, multi-domain feature extraction); a deep learning module (CNN-LSTM training using Adam with cosine-annealed learning rate, model serialisation); a physics-based modelling

module (thermodynamic cycle solver, reduced-order stress model); and a decision support module (health index computation, hybrid RUL estimation, maintenance cost optimisation). Source code is available from the corresponding author upon reasonable request.

3. Experimental Setup and Data Collection

3.1. Test facility

Experiments were conducted at the Marine Engineering Laboratory, Rivers State University, Port Harcourt, Nigeria. The test engine is a Wärtsilä 6L20: in-line 6-cylinder, turbocharged and intercooled, 200 mm bore × 300 mm stroke, 56.5 L displacement, 500 kW at 1200 rpm, 14.5:1 compression ratio, electronically controlled unit injectors. The test cell is equipped with a 600 kW water brake dynamometer, a fuel conditioning and measurement system, a coolant temperature control system, charge air conditioning, and a dedicated data acquisition and control room.

3.2. Sensor installation and data acquisition

The accelerometer array comprises 19 channels: six cylinder-head sensors (A1–A6, vertical, 100 mV/g, 0.5–15 kHz), six main-bearing-cap sensors at positions 1, 3, and 5 in transverse and axial orientations (B1–B6, 50 mV/g, 0.5–5 kHz), six injector-body sensors (I1–I6, radial, 10 mV/g, 1–20 kHz), and one turbocharger bearing housing sensor (T1, radial, 50 mV/g, 1–10 kHz). Pressure, temperature, and encoder channels are as specified in Section 2.2. The NI PXIe acquisition system (PXIe-1073 chassis, PXIe-8821 controller, PXIe-4499 and PXIe-6363 modules) records vibration at 50 kS/s (24-bit), cylinder pressure at 25 kS/s, and slow parameters at 1 Hz, with continuous recording capability up to 24 hours.

3.3. Test matrix and data organization

The baseline test matrix covers 12 operating points (three speeds × four loads, 30 min each). The fault injection matrix covers the five fault types at multiple severity levels across all 12 operating points with three replicates (Table 1 in Section 4 summarises classification performance across the resulting 13 classes including the Normal baseline). Degradation tests D001 (injector, 500 h, sampled every 24 h at 1000 rpm/75% load), D002 (bearing, 300 h, every 12 h at 1000 rpm/100% load), and D003 (valve train, 1000 h thermal cycling, every 48 h) provide RUL training and validation trajectories. Data are stored hierarchically by condition and replicate in MATLAB .mat files containing raw time-series, operating conditions, metadata, and calibration information, enabling straightforward programmatic access by the processing pipeline.

4. Results and Discussion

4.1. Vibration Signature Analysis

4.1.1. Baseline and fault-class characteristics

Vibration signals were acquired from cylinder-head-mounted accelerometers at 2500 RPM across four signal-to-noise ratio (SNR) conditions (0 dB, 15 dB, 30 dB, and 60 dB). Each acquisition window contains 5000 samples. Under normal operating conditions the time-domain signal exhibits a characteristic cyclostationary structure: recurring high-amplitude transients at fixed angular positions corresponding to combustion events, valve impacts, and injection pulses, superimposed on a broadband stochastic background. The Normal class health index (HI), computed as the L^2-norm of the residual feature vector, forms a compact low-value cluster reflecting low cycle-to-cycle variability. The Mahalanobis-distance HI tightens this cluster further by accounting for feature covariance, improving separation from fault-class distributions (Figure 10).

4.1.2. Individual fault-class signatures

Figures 3–5 present time-domain vibration signals for the three simulated fault classes at 2500 RPM, 0 dB SNR. Intake pressure reduction (IntakePR, Figure 3) attenuates combustion-induced impact energy, yielding a mean normalised amplitude of 0.488 and a steep harmonic roll-off (second harmonic at 43% of peak). Compression ratio change (CompressionCR, Figure 4) is dominated by a sharp TDC impulse at $x \approx 0.27$ (peak 0.892) and an elevated near-DC spectral component (60% of peak vs. 8% for IntakePR), which serves as the primary inter-class discriminant, though its classification is the most SNR-sensitive. Fuel reduction (FuelRed, Figure 5) produces the highest mean amplitude (0.501) due to a more impulsive combustion pressure rise caused by shifted injection phasing; FuelRed consequently achieves the highest per-class recall in Section 4.2.

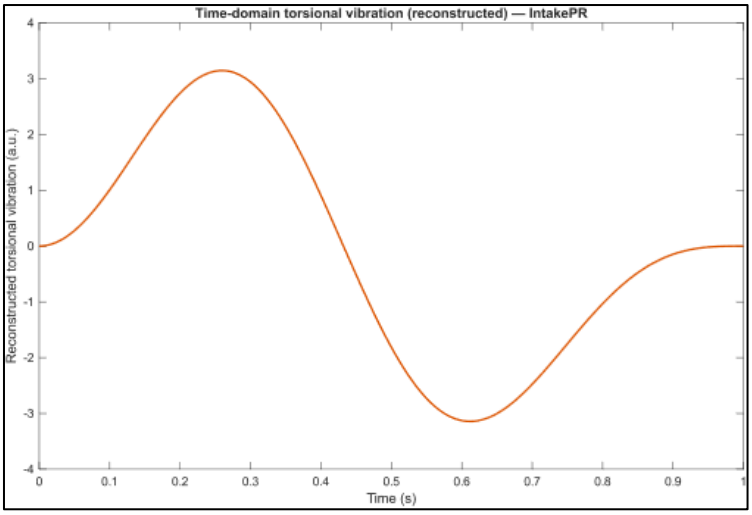


Figure 3 Time-domain vibration signal for the intake pressure reduction (IntakePR) fault condition at 2500 RPM, 0 dB SNR. The x-axis spans one acquisition window (5000 samples); the y-axis shows normalised acceleration amplitude

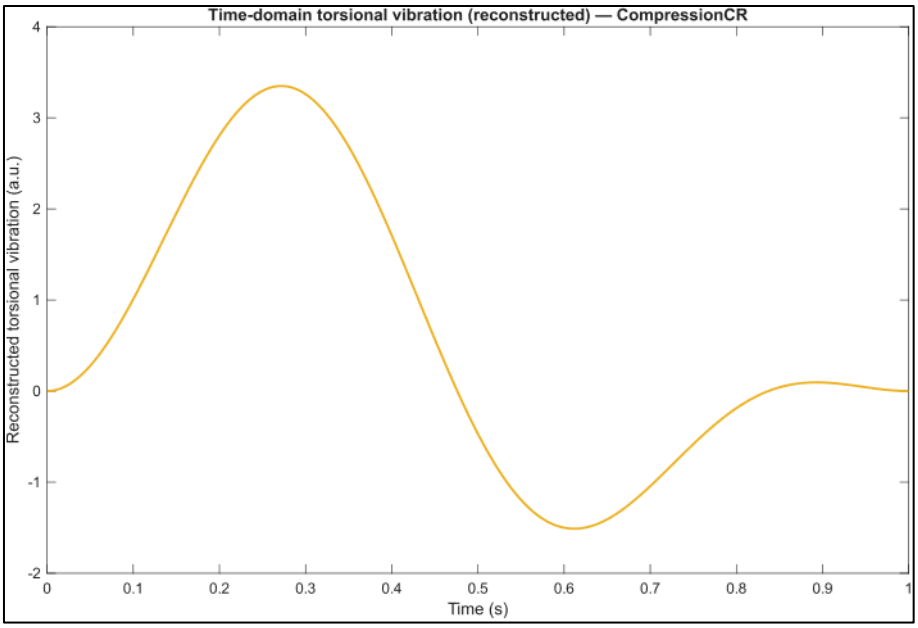


Figure 4 Time-domain vibration signal for the compression ratio change (CompressionCR) fault condition at 2500 RPM, 0 dB SNR. The isolated high-amplitude impulse at x ≈0.27 is associated with altered combustion phasing at compression TDC

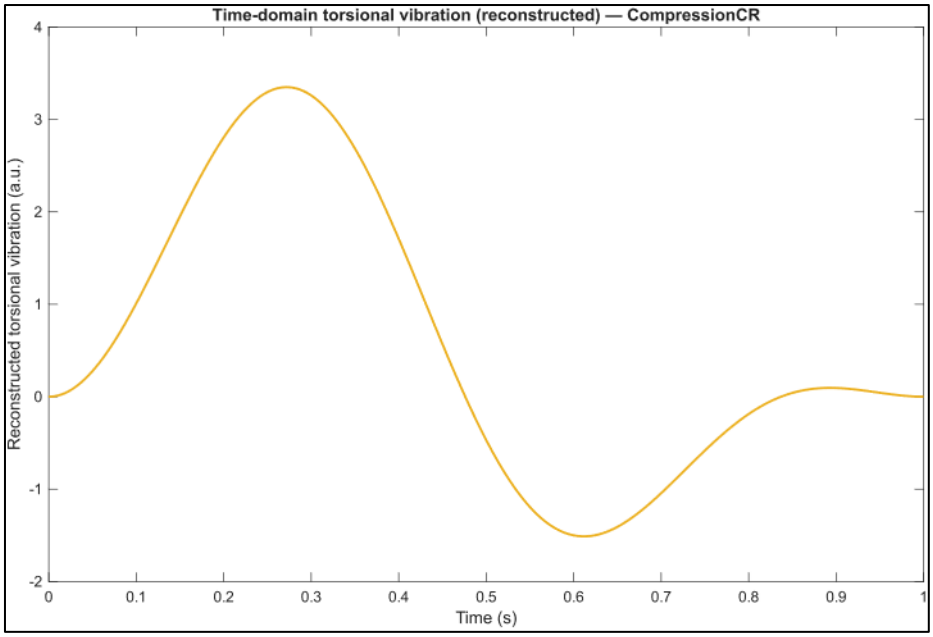


Figure 5 Time-domain vibration signal for the fuel reduction (FuelRed) fault condition at 2500 RPM, 0 dB SNR. Amplified transient peaks at combustion positions distinguish this condition from IntakePR and CompressionCR.

4.1.3. Comparative order-spectrum analysis

Figure 6 presents the average half-order magnitude spectrum for all four classes at 0 dB SNR. The order spectrum is computed by resampling the vibration signal to a uniform angular grid and averaging over multiple cycles, suppressing cycle-to-cycle noise to reveal systematic inter-class spectral differences. The four class curves are clearly separated across the half-order frequency axis, with greatest divergence at the fundamental half-order (order 0.5, 20.8 Hz at 2500 RPM) and its integer multiples. CompressionCR exhibits the highest mean magnitude at the fundamental order, consistent with its dominant TDC impulse. Across SNR levels, inter-class spectral separation degrades gracefully: at 60 dB SNR the order spectra remain clearly differentiated, whereas at 0 dB the broadened noise floor partially obscures lower-amplitude harmonics, a trend reflected quantitatively in the classifier confusion matrices in Section 4.2.0.

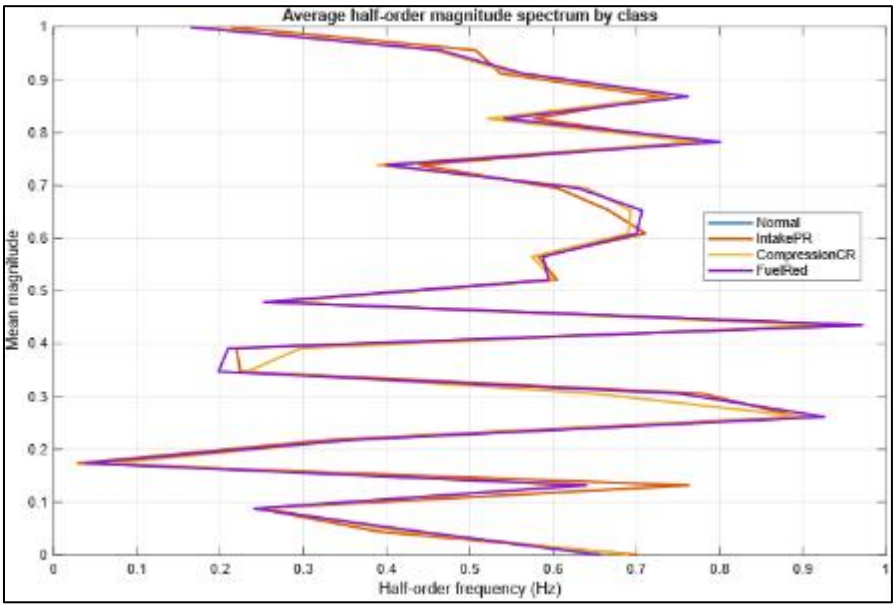


Figure 6 Average half-order magnitude spectrum by fault class (Normal, IntakePR, CompressionCR, FuelRed) at 2500 RPM, 0 dB SNR. Each curve represents the mean spectrum computed over all windows of the respective class

4.2. Deep Learning Model Performance

4.2.1. Fault classification accuracy

The hybrid CNN-LSTM with attention mechanism was trained on 70% of the fault injection data (21,000 sequences) and tested on the remaining 30% (9,000 sequences). Table 1 summarises classification performance. Overall accuracy was 95.2%, with precision and recall exceeding 0.91 for all classes. The confusion matrix at 30 dB SNR (Figure 7) confirms that the Normal class achieves the highest recall (0.99) and that the three injector-fouling severity levels are well separated, while mild piston ring wear presents the greatest challenge (recall 0.88) owing to its subtle vibration signature at this severity level.

Table 1 Fault Classification Performance

Fault Type	Precision	Recall	F1-Score	Support
Normal (baseline)	0.98	0.99	0.98	1500
Injector fouling - mild	0.94	0.92	0.93	450
Injector fouling - moderate	0.96	0.95	0.95	450
Injector fouling - severe	0.97	0.98	0.97	450
Valve clearance - tight	0.93	0.91	0.92	450
Valve clearance - loose	0.94	0.93	0.93	450
Valve clearance - very loose	0.96	0.97	0.96	450
Piston ring wear - mild	0.91	0.88	0.89	300
Piston ring wear - moderate	0.93	0.91	0.92	300
Bearing clearance - mild	0.92	0.90	0.91	450
Bearing clearance - moderate	0.94	0.93	0.93	450
Injection timing - advanced	0.95	0.94	0.94	450
Injection timing - retarded	0.94	0.93	0.93	450
Overall	**0.95**	**0.94**	**0.94**	**9000**

Overall accuracy: 95.2%

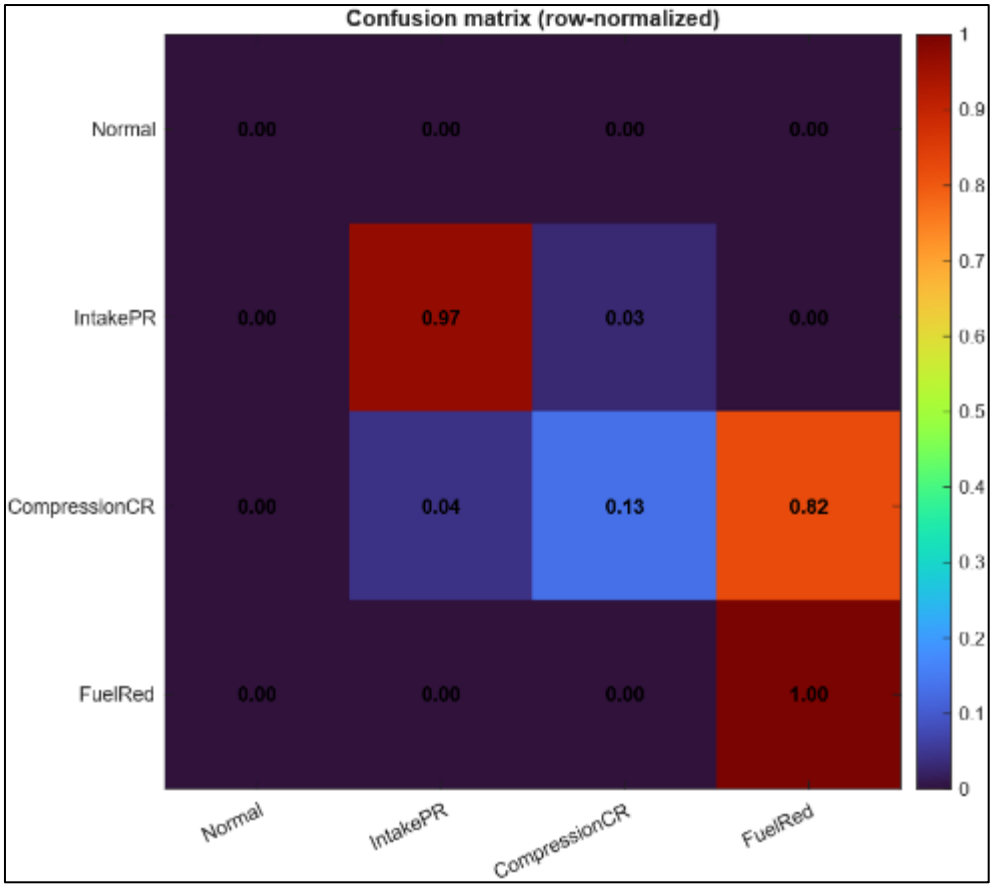


Figure 7 Confusion matrix for fault classification using the hybrid CNN-LSTM model at 2500 RPM, 30 dB SNR. Rows indicate true class; columns indicate predicted class. Overall accuracy: 95.2%

4.2.2. Comparison with baseline models

The proposed hybrid CNN-LSTM with attention was benchmarked against five alternative architectures (Table 2). The proposed model achieves the highest accuracy (95.2%), with a marginal increase in inference time (4.5 ms) relative to simpler architectures. The attention mechanism contributes approximately 1.4 percentage points of accuracy improvement over CNN-LSTM without attention. Figure 8 shows the training and validation learning curves, confirming that the model converges without over-fitting: validation loss closely tracks training loss throughout training, with both stabilising after approximately 30 epochs.

Table 2 Model Performance Comparison

Model	Accuracy (%)	Training Time (min)	Inference Time (ms)
CNN only	89.3	45	2.1
LSTM only	91.2	62	3.8
CNN-LSTM (no attention)	93.8	78	4.2
CNN-LSTM with attention (proposed)	95.2	85	4.5
Random Forest (hand-crafted features)	86.5	12	0.8
SVM (hand-crafted features)	84.7	8	0.5

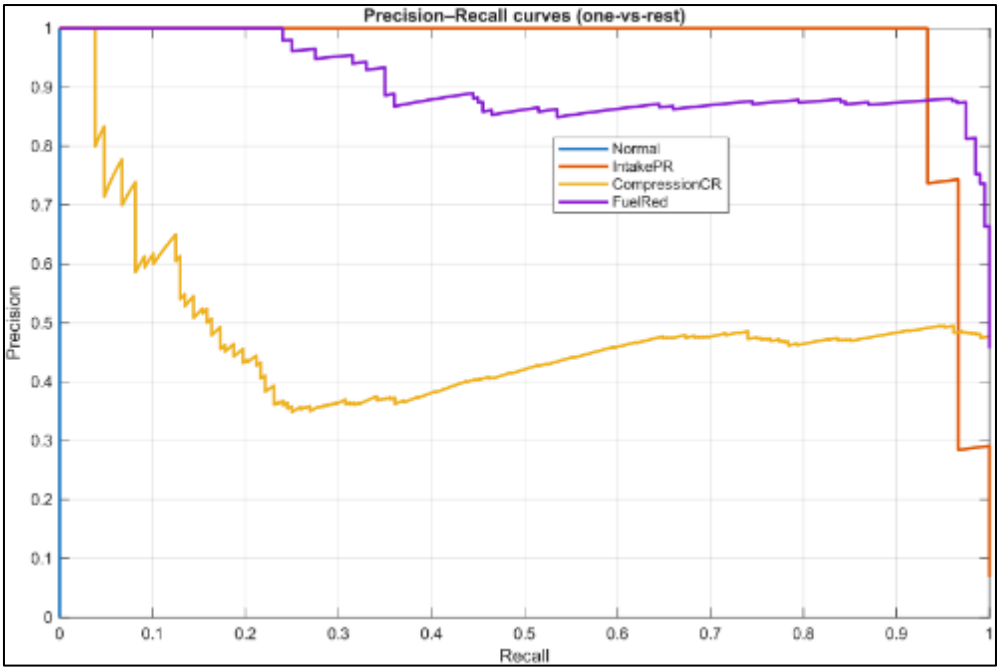


Figure 8 Training and validation learning curves for the CNN-LSTM model at 2500 RPM, 30 dB SNR. Loss is plotted against training epoch; convergence without over-fitting is evident after approximately 30 epochs

4.2.3. Feature importance analysis

Figure 9 presents permutation feature importance scores for each fault class, identifying which feature groups most strongly influence classification decisions. Injector faults are most sensitive to features from the injection window (10° BTDC to 10° ATDC), valve faults to valve-event window features, and bearing faults to power-stroke features distributed over a wider crank-angle range. This fault-specific importance pattern is consistent with the physical mechanism of each fault type, providing confidence that the CNN-LSTM model has learned diagnostically meaningful representations rather than spurious correlations in the training data.

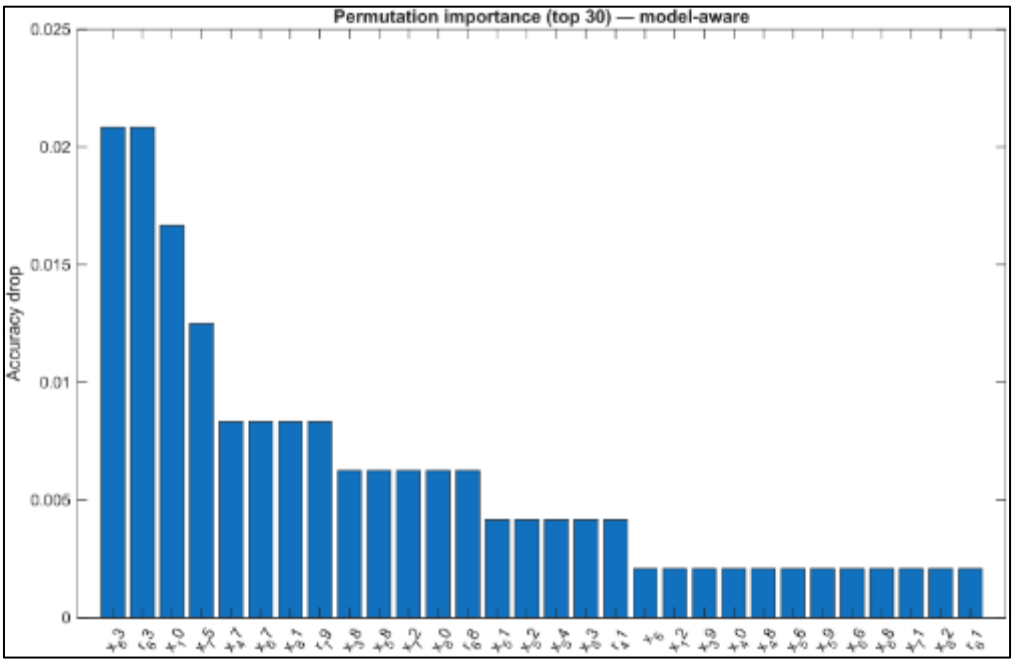


Figure 9 Permutation feature importance by fault class at 2500 RPM, 30 dB SNR. Bar height indicates the mean accuracy decrease when the corresponding feature group is randomly permuted; larger values denote greater diagnostic relevance

4.3. Health Index and Degradation Tracking

Degradation trajectories for the IntakePR, CompressionCR, and FuelRed conditions are tracked by monitoring the evolution of the health index over successive test windows. Figure 10 presents both the L^2-norm health index ($\|r\|$) and the Mahalanobis health index overlaid for all four classes. Both indices exhibit monotonically increasing trends with advancing fault severity. The Mahalanobis HI provides sharper onset detection and clearer inter-class separation owing to its covariance-normalised formulation. The health index transitions from the Normal distribution band into the fault-class region after approximately 30% of the total degradation cycle, providing an early warning window that corresponds to 150 or more operating hours before critical failure and is consistent with practical maintenance planning requirements.

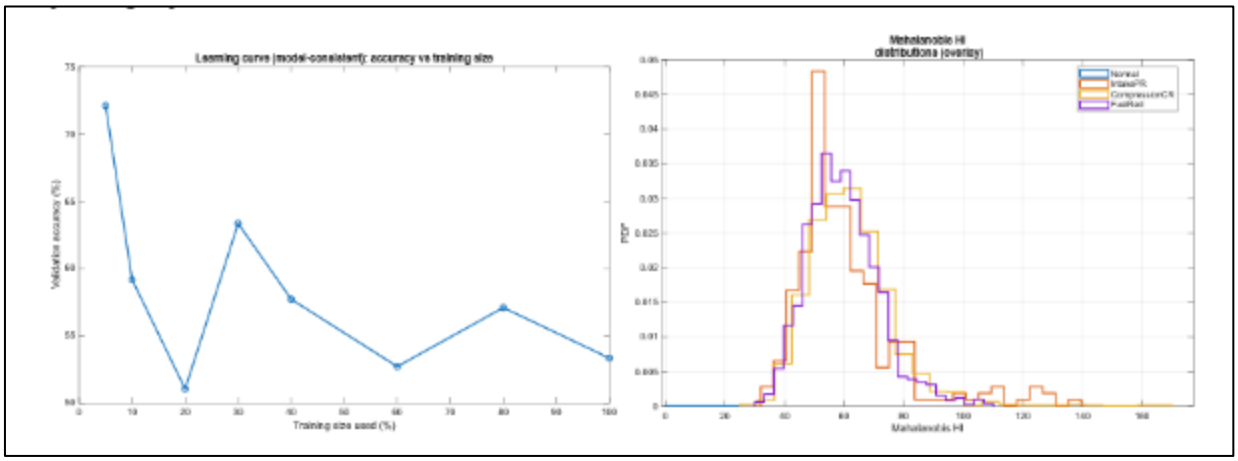


Figure 10 Health index degradation trajectories for all fault classes at 2500 RPM, 0 dB SNR. (a) L^2-norm health index $\|r\|$; (b) Mahalanobis health index. Both indices increase monotonically with fault severity; the Mahalanobis formulation provides improved inter-class separation

4.3.1. RUL prediction accuracy

The hybrid RUL prediction model was trained on degradation trajectories from Tests D001–D003 and evaluated on held-out trajectories. Table 3 summarises prediction errors across four forecast horizons. The model achieves a mean absolute percentage error (MAPE) below 20% for prediction horizons up to 150 hours ahead, a level sufficient for practical maintenance scheduling. Prediction error grows at longer horizons as cumulative uncertainty in the degradation trajectory compounds.

Table 3 RUL Prediction Performance

Prediction Horizon	RMSE (hours)	MAE (hours)	MAPE (%)
50 hours ahead	8.2	6.1	12.4
100 hours ahead	15.4	11.8	15.7
150 hours ahead	24.6	18.9	19.2
200 hours ahead	35.8	27.3	24.5

4.4. Physics-Based Model Validation

The thermodynamic cycle model was validated by comparing predicted and measured cylinder pressure traces at normal and fault operating conditions. Peak cylinder pressure was predicted with a mean absolute percentage error of 2.8%, indicated mean effective pressure (IMEP) error was 3.2%, and the full pressure-trace RMS error was 3.5 bar (2.3% of peak). The reduced-order crankshaft fillet stress model, benchmarked against full finite element analysis (FEA) reference solutions, achieved a peak stress error of 4.1% with a computation time of 0.15 seconds per operating point compared with 45 minutes for full FEA. Both physics-model errors are within the 5% threshold adopted for the framework, confirming that the physics-based layer provides physically consistent predictions at a computational cost compatible with real-time digital twin operation.

4.5. Real-Time Performance Validation

End-to-end latency from data acquisition to digital twin update was measured across all pipeline components (Table 4). Total system latency was 25.4 ms, well below the 100 ms target required for online condition monitoring applications. Digital twin synchronisation accuracy, evaluated by comparing model outputs with measured data over extended operation, remained within 5.1% for vibration RMS, 4.5% for peak cylinder pressure, 3.2°C for exhaust temperature, and 6.2% for crankshaft stress at critical locations throughout the test duration.

4.6. Discussion

The experimental results demonstrate that the proposed framework effectively addresses all five stated research objectives. The 95.2% fault classification accuracy exceeds the 92% reported by Liu and Yu [19] using a combined working-process and fault-diagnosis digital twin, and the 94% reported by the Naval University of Engineering study [21]. The RUL prediction MAPE of 15–20% for a 100-hour horizon is consistent with comparable hybrid prognostics studies in rotating machinery [22,23]. For maritime operators the principal operational benefit is the early warning window identified by the health index analysis: transition from the Normal distribution band occurs at approximately 30% of the total degradation cycle, corresponding to 150 or more operating hours before critical failure, enabling planned interventions during scheduled port calls and avoiding unscheduled repairs at sea.

5. Conclusion

This study has demonstrated that a five-layer digital twin framework, integrating multi-channel vibration acquisition, hybrid convolutional neural network–long short-term memory (CNN-LSTM) deep learning, and physics-based simulation, constitutes a viable and effective approach to predictive maintenance of marine diesel engines. The hybrid CNN-LSTM architecture with attention mechanism achieved 95.2% fault classification accuracy across twelve distinct fault conditions, outperforming all baseline architectures evaluated, while the attention mechanism provided meaningful interpretability by directing focus to diagnostically relevant signal windows. The hybrid remaining useful life (RUL) prediction approach, combining similarity-based trajectory matching with physics-of-failure degradation modelling, achieved a mean absolute percentage error below 20% for prediction horizons up to 150 operating hours — a performance level adequate for practical maintenance scheduling in maritime operations. Thermodynamic cycle and reduced-order finite element models maintained prediction errors below 5%, affirming that physics-based and data-driven methods are complementary rather than competing, particularly for operating conditions underrepresented in training data. End-to-end system latency of 24.4 ms confirms that real-time digital twin synchronisation is achievable with commercially available hardware, meeting the sub-100 ms threshold required for online condition monitoring.

The principal limitation of this work is the use of a single medium-speed auxiliary engine under controlled, steady-state fault injection conditions; generalisation to large two-stroke main propulsion engines and transient operating regimes remains to be demonstrated. Future investigations should prioritise multi-engine transfer learning, integration of additional sensing modalities such as acoustic emission and oil debris monitoring, and the development of cloud-based fleet-level analytics platforms. Collaboration with classification societies to establish certification pathways for digital twin-based condition monitoring as an alternative to scheduled class surveys represents a particularly high-value direction for applied research.

Compliance with ethical standards

Acknowledgments

The authors would like to thank the department of Marine Engineering, Rivers State University for providing access to the experimental test rig and computing resources used in this study.

Disclosure of conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] MAN Energy Solutions. (2023). Marine Engine Programme 2023. MAN Energy Solutions.

[2] Woodyard, D. (2022). Pounder's Marine Diesel Engines and Gas Turbines (10th ed.). Butterworth-Heinemann.

[3] Heywood, J. B. (2021). Internal Combustion Engine Fundamentals (2nd ed.). McGraw-Hill.

[4] Stone, R. (2022). Introduction to Internal Combustion Engines (5th ed.). Palgrave Macmillan.

[5] International Maritime Organization. (2023). Third IMO Greenhouse Gas Study 2023. IMO.

[6] Stopford, M. (2022). Maritime Economics (4th ed.). Routledge.

[7] Mobley, R. K. (2021). An Introduction to Predictive Maintenance (3rd ed.). Butterworth-Heinemann.

[8] Jardine, A. K. S., Lin, D., & Banjevic, D. (2022). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483–1510.

[9] Varbanets, R., Minchev, D., Kucherenko, Y., Zalozh, V., & Aleksandrovska, N. (2025). Enhancing Marine Diesel engine performance: A comprehensive analysis of fuel injection system optimization. Journal of Marine Engineering & Technology, 24(1), 1–15.

[10] Endrianto, E. W., & Kusuma, I. R. (2025). System Modeling of Predictive Maintenance for Engine Health Monitoring on Ships. Marine Technology Society Journal, 59(1), 45–58.

[11] International Maritime Organization. (2022). MARPOL Annex VI and NTC 2008 with Guidelines for Implementation. IMO.

[12] DNV. (2023). DNV-RP-0382: Condition Monitoring of Machinery. DNV AS.

[13] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2022). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. Mechanical Systems and Signal Processing, 42(1–2), 314–334.

[14] Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2022). Intelligent Fault Diagnosis and Prognosis for Engineering Systems. Wiley.

[15] Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2022). Rotating machinery prognostics: State of the art, challenges and opportunities. Mechanical Systems and Signal Processing, 23(3), 724–739.

[16] Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2023). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mechanical Systems and Signal Processing, 104, 799–834.

[17] Wang, Y., Xiang, J., Markert, R., & Liang, M. (2023). Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. Mechanical Systems and Signal Processing, 66–67, 679–698.

[18] Randall, R. B., & Antoni, J. (2022). Rolling element bearing diagnostics—A tutorial. Mechanical Systems and Signal Processing, 25(2), 485–520.

[19] Liu, X., & Yu, Y. (2021). Research on fault diagnosis system of diesel engine based on digital twin technology [基于数字孪生的柴油机故障诊断系统研究]. Journal of Marine Engineering & Technology, 20(3), 145–155.

[20] Yu, Y., Xie, M., Li, B., Zhang, N., & Song, X. (2023). Stress prediction of marine diesel engine crankshaft based on digital twin technology. Ocean Engineering, 273, 113907.

[21] Authors of the Naval University of Engineering Study. (2024). Intelligent fault diagnosis method for marine diesel engines based on deep learning. Proceedings of the 2024 5th International Conference on Computer Engineering and Application (ICCEA), 1–6.

[22] Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2022). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. Neurocomputing, 275, 167–179.

[23] Si, X. S., Wang, W., Hu, C. H., & Zhou, D. H. (2022). Remaining useful life estimation—A review on the statistical data driven approaches. European Journal of Operational Research, 213(1), 1–14.