

Article

A Safety-Case-Driven Hybrid Digital Twin for Centrifugal Compressor Health Monitoring

Hezrone Mujawo ^{1,*} and Oyeniyi Akeem Alimi ^{2,*} ¹ Department of Electrical Engineering, Engineering Institute of Technology, West Perth, WA 6005, Australia² Innovation for Sustainability Research Group, Department of Information Systems, Durban University of Technology, Greyville, Durban 4001, South Africa

* Correspondence: 883356@student.eit.edu.au (H.M.); oyenyia@dut.ac.za (O.A.A.)

Abstract

Centrifugal compressors are critical assets in the oil and gas, petrochemical, and power generation industries, where unplanned downtime results in severe economic and safety consequences. Despite the application of digital twin technology for predictive maintenance, existing approaches struggle to combine accurate degradation modeling with formal assurance evidence that regulators and operators demand before trusting machine learning-augmented systems. This paper proposes a hybrid digital twin framework whose architecture is structured around a formal safety case template, addressing both the accuracy and the trustworthiness challenges simultaneously. The methodology couples a first-principles thermodynamic model with a neural-network residual learner, and the complete system is organized through a design-stage safety case constructed in Goal Structuring Notation. The design stage identifies the requirements for operational deployment. Validation through a simulation study on a one-year synthetic operational dataset shows that the hybrid model reduces root-mean-square prediction error by over 50% for both pressure ratio and polytropic efficiency compared to the physics-only baseline. The anomaly detection module, presented here as a proof of concept, achieves 92% recall in identifying injected faults, and a composite health index tracks the progression of fouling, erosion, and seal wear over the simulated service life. This study is purely theoretical, with no experimental measurements conducted. It demonstrates the structural viability and coherence of the proposed framework within a controlled environment, providing a solid theoretical and computational foundation for future physical validation efforts. These findings provide preliminary evidence that embedding a structured safety argument into the design of a hybrid digital twin is technically feasible and beneficial for building the confidence needed to deploy such systems in safety-critical industrial environments.

Keywords: centrifugal compressor; digital twin; hybrid modeling; predictive maintenance; safety case



Academic Editor: Zhuming Bi

Received: 17 March 2026

Revised: 1 June 2026

Accepted: 16 June 2026

Published: 23 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The oil and gas process industry depends on centrifugal compressors to sustain operations that range from natural-gas pipeline transmission to gas compression. In upstream oil and gas alone, a single large compressor train can represent an investment exceeding millions of dollars, and any unplanned shutdown may cost the operator several hundred thousand dollars for every hour of lost production [1]. Any mechanical failure, whether caused by a surge, seal-oil system malfunction, or unforeseen bearing fault, can

be catastrophic. These occurrences have necessitated the development of an optimized solution to move from reactive, schedule-based maintenance toward condition-based and, ultimately, predictive strategies that anticipate problems before they escalate.

The digital twin concept, first articulated by Grieves in the context of product lifecycle management [2] and later formalized by NASA for airframe structural-life prediction [3], has emerged as a trusted approach to accomplish this transition. In its definition, it maintains a continuously updated virtual replica of the physical asset, fusing sensor data with computational models for remaining useful life estimation. With the advancements in artificial intelligence and machine learning, the concept has advanced from theoretical knowledge into deployable technology, coupled with industrial internet of things, cloud computing, and machine learning [4]. Digital twins have already been applied with greater accuracy to gas turbines, wind turbines, and electric motors, yet their adoption for centrifugal compressors, particularly in safety-critical applications, remains limited.

The slow adoption of digital twins in safety-critical oil and gas compressor service stems not from a shortage of modeling techniques. It comes from the absence of formal trusted frameworks that satisfy regulatory and operational requirements. Physics-based models, while grounded in well-understood thermodynamic principles, inevitably simplify reality. They often assume ideal gas behavior, neglect secondary flow phenomena, and cannot readily account for the gradual fouling of blade passages or the progressive wear of labyrinth seals. Data-driven models can learn these complicated, nonlinear degradation patterns from operational data [5,6]. However, they are hard to understand, easy to make mistakes when extrapolating, and hard to explain to a regulator or an insurance underwriter [7,8]. Combining the two into a hybrid approach presents a practical middle ground. But it also brings up a new problem in the form of ‘how do you prove that the outputs of such a system are reliable enough for safety-related decisions?’ This work addresses the gap through a simulation-based proof of concept. Hybrid digital twins for rotating machinery have been shown in [7,8]. But these studies only focused on how accurate the predictions are; they do not highlight whether the outputs can be trusted for safety-critical operational decisions. Commercial compressor monitoring systems depend on set alarm thresholds and do not use physics-informed residual learning or formal uncertainty quantification. Safety cases employing Goal Structuring Notation (GSN) have been utilized in aerospace, nuclear, and railway sectors [6,9–11]. However, its application to ML-enhanced industrial monitoring systems remains largely unexamined. Based on our research and studies, there is no framework that directly incorporates a formal safety case into the architecture and governance of a hybrid compressor digital twin.

We present a hybrid digital twin framework for centrifugal compressors in which a neural-network residual learner augments a physics-based thermodynamic model, and the entire system is structured through a formal safety case constructed using GSN [6,9]. Though the safety case is an independent system that is not altered by any algorithms, it actively shapes design decisions. This stems from the choice of an additive residual architecture that keeps predictions anchored to physics and the requirement for calibrated uncertainty estimates that feed the anomaly detection logic. It is validated through a comprehensive simulation study that generates realistic operational data over a one-year period, including the effects of fouling, erosion, seal wear, and injected anomalous events.

The four goals of this study are: (1) to create a hybrid digital twin architecture that combines a first-principles thermodynamic model with a data-driven residual learner for monitoring the health of centrifugal compressors; (2) to directly integrate a formal safety case, built using GSN, into the twin’s architecture so that every model output is linked to a structured trustworthiness argument; (3) to test the framework through a full simulation study using synthetic operational data that includes realistic degradation profiles and

injected anomalies; and (4) to evaluate whether the hybrid model is much more accurate at making predictions than a physics-only baseline and that the anomaly detection system meets the safety case's recall and false positive rate goals.

The significance of this work is twofold. First, in the simulated environment, the hybrid model demonstrably improves prediction accuracy: root-mean-square errors for pressure ratio and polytropic efficiency are reduced by more than 50% compared with the physics-only baseline, and the integrated anomaly detector achieves a recall of 92% on injected faults at the default threshold. Secondly, the GSN safety case provides a structured, evidence-backed argument template that operators and regulators would need before they trust a machine learning-augmented monitoring system in a high-integrity application. These contributions represent a meaningful step toward closing the gap between the promise of digital twins and their practical deployment in safety-critical rotating machinery, though real-world validation remains necessary to confirm these findings.

Specifically, the novel contributions of this study are stated as follows:

1. It combines a hybrid physics-data modeling architecture with a formal safety assurance framework that uses GSN. Prior research has examined hybrid digital twins and safety cases separately; however, to the authors' knowledge, no existing literature has integrated the safety argument directly into the architecture of the twin, ensuring that each model output contains verifiable evidence of its reliability, though the argument remains partially undeveloped pending real-world validation. This integration is specific because the GSN structure specifies which validation metrics must be met before the twin's outputs can affect operational decisions. This closes the gap between model accuracy and operational trust.
2. The second contribution is the residual-learning strategy, which trains the neural network only on the physics model's error signal instead of on raw process data. This constrains the data-driven part to work within a physically bounded envelope, which is necessary for safety certification in regulated industries.
3. The proposed framework is critically analyzed on a simulated API 617 centrifugal compressor whose operation is affected by progressive fouling, erosion, and seal wear. These are types of degradations that are hard to capture with physics alone but dominate real maintenance costs in petrochemical service.

We use GSN to make a formal safety argument that links every architectural choice to a specific safety goal. This makes the framework feasible for stakeholders' adoption.

The remainder of this paper is organized as follows. Section 2 surveys the relevant literature on digital twins, compressor modeling, hybrid architecture, and safety assurance. Section 3 describes the proposed methodology in detail, covering the physics model, the residual learner, the fusion layer, and the GSN safety case. Section 4 presents the simulation-based results. Section 5 discusses the implications and limitations. Section 6 discusses the directions for future work, and Section 7 presents the conclusion.

2. Related Work

2.1. Digital Twins in Industry

The digital twin concept has evolved considerably since its inception. Tao et al. [1] provided a comprehensive review of digital twin applications across the manufacturing domain. In the study, the authors demonstrated the technology's progression from simple CAD-based replicas to sophisticated cyber-physical systems that incorporate real-time data streams, physics simulations, and machine learning. Iliută et al. [4] also studied how the idea changed from a theory to a usable technology, naming the industrial internet of things, cloud computing, and advances in sensor miniaturization as key factors. Rasheed et al. [12]

put forward a broad framework for digital twins that separates standalone, composite, and connected twins and identified data quality and model fidelity as the primary barriers to their adoption. Qi et al. [13] examined the facilitating technologies and instruments for digital twin implementation, whereas Errandonea et al. [14] investigated digital twin applications specifically for maintenance, highlighting a deficiency in formal assurance methodologies. In the field of rotating machinery, digital twins have been used on gas turbines, wind turbines, and electric motors with promising results. However, they have not been used much on centrifugal compressors yet, especially when it comes to formal safety assurance.

2.2. Centrifugal Compressor Modeling

Polytropic compression theory [5] shows capabilities in explaining how centrifugal compressors work thermodynamically. Given the inlet conditions, gas properties, and a compressor performance map, the following parameters can be calculated: pressure ratio, polytropic head, polytropic efficiency, discharge temperature, and shaft power. Kurz and Brun [15] conducted an in-depth examination of the degradation mechanisms in centrifugal compressors, demonstrating that fouling diminishes efficiency by 2–5% during standard maintenance intervals and that erosion diminishes head capacity in a nearly linear manner throughout the early to mid-life stages. Boyce [16] wrote down the basic engineering principles behind gas turbine and compressor systems, the construction of performance maps, and off-design behavior. The performance map is only valid for the as-new condition, and the original equipment manufacturer owns it. As the compressor ages, its performance changes in ways that are hard to predict using math. This basic limitation is what motivates the augmentation of the physics model with a data-driven model.

2.3. Hybrid Modeling Approaches

Hybrid models that combine physics-based and data-driven components have received growing attention. Willard et al. [8] surveyed the integration of physics-based modeling with machine learning, identifying three broad strategies: physics-informed loss functions, architecture constraints, and residual learning. Karniadakis et al. [7] reviewed physics-informed neural networks (PINNs), which embed governing equations directly into the loss function. Raissi et al. [17] demonstrated the foundational PINN approach for solving partial differential equations. Karpatne et al. [18] proposed a theory-guided data science framework that provides a taxonomy for integrating domain knowledge into ML pipelines. For compressor applications, the residual-learning approach is efficient because it preserves the interpretability of the physics model while allowing the data-driven component to correct for unmodeled phenomena such as degradation, manufacturing variability, and installation effects. This combination drives the success of hybrid approaches. Recent studies have demonstrated the effectiveness of hybrid physics-data approaches across several rotating machinery applications. Chen et al. [19] developed a condition-adaptive physics-informed LSTM (CAPI-LSTM) for marine gas turbine power prediction, achieving an RMSE of 0.177 MW: a 58.5% reduction compared to a pure LSTM baseline. Kumar et al. [20] proposed a hybrid digital twin for permanent magnet synchronous motor fault diagnosis that integrates physics-informed neural networks with an extended Kalman filter, achieving 96.99% fault detection accuracy with a 95 ms latency. Li et al. [21] applied a physics-informed DeepLSTM to steam turbine digital twin modeling, reducing power prediction RMSE by 86.8% in time extrapolation tests. Li et al. [22] used a transfer-learning LSTM for centrifugal compressor performance prediction, reporting a mean absolute percentage error of 2.90% for power and 0.97% for compression ratio.

Joung et al. [23] applied an LSTM-RNN to air compressor bearing fault detection, achieving 97.4% prediction accuracy using a purely data-driven approach. Daoudi et al. [24] compared four regression models for screw compressor predictive maintenance, with gradient boosting achieving the best R^2 of 0.92. These studies confirm the trend toward hybrid architecture, but none of them integrates a formal safety case into the digital twin framework; a gap that the present work addresses directly.

2.4. Safety Assurance for ML Systems

The use of machine learning in applications where safety is critical has led to the creation of formal assurance frameworks. GSN offers a graphical framework for formulating safety arguments that connect primary safety assertions to supportive evidence via a structured hierarchy of goals, strategies, and solutions. Bloomfield et al. [10] discussed the unique difficulties of creating assurance cases for systems that use machine learning and suggested a pattern-based method. Hawkins et al. [11] created templates for safety arguments for ML parts in autonomous systems. GSN is widely used in the aerospace, nuclear, and railway industries, but it has not been used much in machine learning-enhanced industrial monitoring systems. This paper addresses that gap by embedding a GSN safety case directly into the design and management of a digital twin of a hybrid compressor.

2.5. Condition Monitoring and Fault Diagnosis

The literature on condition monitoring provides important background on this topic. Jardine et al. [25] conducted an extensive review of diagnostic and prognostic methodologies for rotating machinery, encompassing vibration analysis, oil analysis, and model-based techniques. Lei et al. [26] examined intelligent fault diagnosis techniques, emphasizing the transition from manually crafted features to deep learning representations. Jiadon Meng et al. [27] conducted health condition of bearings. The maximal difference in the detail components in the Multiresolution Singular Value Decomposition (MDDCs-MRSVD) algorithm processes the vibration signal for the first time. This is followed by a features matrix that shows the health status of bearings made using the chosen features. For each sampling time along the amplitude direction of this matrix, the gradient standard deviation is calculated. Lastly, a monitoring indicator is made to find the healthy stages of the bearing. These reviews show that model-based residual analysis, which compares measured performance to a reference model, is one of the best ways to find early signs of degradation in compressors, if the reference model is accurate enough.

3. Methodology

The proposed framework consists of five tightly integrated components: a physics-based compressor model, a data acquisition and preprocessing, a hybrid modeling layer (physics + NN), the anomaly detection layer and a formal safety case and governance layer. Figure 1 provides a high-level overview of the architecture.

As shown in Figure 1, the proposed model's layer 1 captures the physical asset and instrumentation. Layer 2 handles data acquisition and pre-processing. Layer 3 contains a hybrid modeling engine with physics-based and data-driven models feeding into a fusion layer. Layer 4 provides decision support applications. Layer 5 implements the GSN safety case. The green border represents the safety assurance envelope.

Safety-Case-Driven Hybrid Digital Twin Architecture for API-Class Centrifugal Compressors

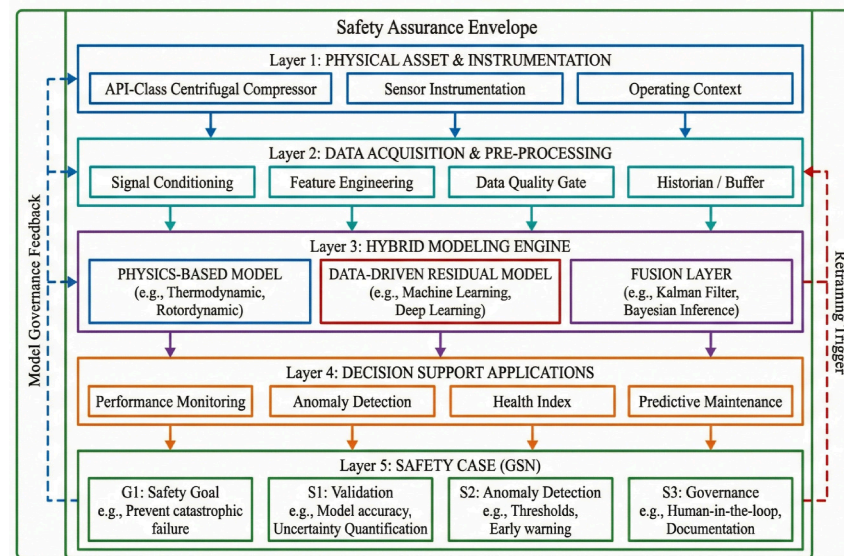


Figure 1. Five-layer architecture of the proposed safety-case-driven hybrid digital twin framework for centrifugal compressor health monitoring.

3.1. Physics-Based Model

The physics-based model is built on the thermodynamic theory of polytropic compression. For a single-stage centrifugal compressor (assuming an ideal gas), the polytropic head is given in (1) [17]:

$$H_p = \frac{n}{n-1} R T_1 \left[\left(\frac{P_2}{P_1} \right)^{\frac{n-1}{n}} - 1 \right] \quad (1)$$

where R is the specific gas constant, T_1 is the inlet temperature, n is the polytropic exponent, and $\frac{P_2}{P_1}$ is the pressure ratio. The polytropic exponent is related to the specific heat ratio γ and the polytropic efficiency η_p is defined in (2):

$$\frac{n-1}{n} = \frac{\gamma-1}{\gamma\eta_p} \quad (2)$$

The discharge temperature and shaft power follow directly [14]:

$$T_2 = T_1 \times \left(\frac{P_2}{P_1} \right)^{\frac{n-1}{n}} \quad (3)$$

$$W = \frac{m \times H_p}{\eta_p \times \eta_{mech}} \quad (4)$$

where η_{mech} is the mechanical efficiency, assumed to be 0.98. The compressor performance map, which is the relationship between pressure ratio, efficiency, and mass flow at various speeds, is generated synthetically using a parabolic pressure-ratio characteristic and a bell-shaped efficiency curve, scaled across speed lines using the fan laws. This is a necessary simplification, as actual manufacturer maps are proprietary. The surge and choke boundaries are set at 70% and 125% of the design mass flow, respectively, which are representative values for machines of this class. Figure 2 shows the resulting compressor map. As shown in Figure 2, the surge and choke boundaries are indicated by dashed lines.

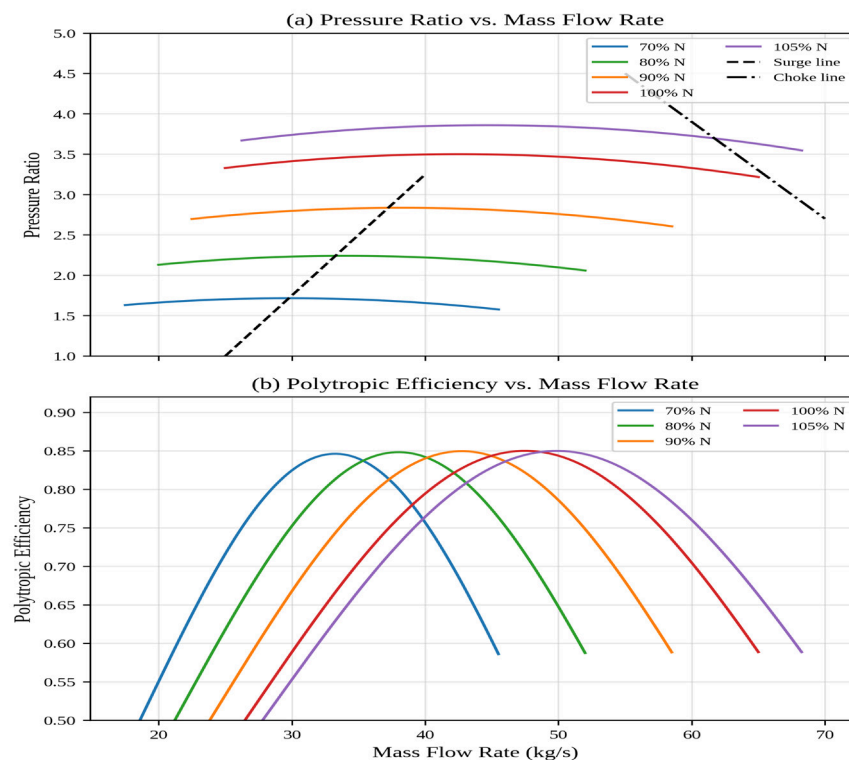


Figure 2. Synthetic compressor performance map: (a) pressure ratio versus mass flow rate for five rotational speeds (70–105% of design speed), with surge and choke boundaries indicated; (b) polytropic efficiency versus mass flow rate, showing the characteristic bell-shaped curves shifting with speed.

3.2. Degradation Model

Fouling, erosion, and seal wear are the three degradation mechanisms that were modeled. These are the most common ways that decrease the performance of API 617 compressors, according to the literature and field experience. Fouling, which is a result of deposits that are built up on the blades of the impeller and diffuser vanes, makes the surface rougher and results in a smaller effective flow area. Kurz and Brun [15] demonstrated that fouling diminishes efficiency by 2–5% during standard maintenance intervals, with the rate contingent upon gas composition and filtration quality. Our model applies to a linear efficiency drop of up to 15% over 8760 h. This is a safe number for a machine that is heavily fouled, but it is still in the reported range. Erosion (the removal of material from the leading edges of blades due to particles rubbing against them) makes it harder for the impeller to transfer energy to the gas. This mostly affects the head coefficient. Seal wear (more space between the seals in the labyrinth) causes internal recirculation, which lowers throughput and efficiency.

The linear degradation assumption is a simplification that fails in several significant aspects: actual degradation is not monotonic (process disturbances and water washing induce step changes). Degradation mechanisms interact (fouling modifies the flow field, altering erosion patterns), and the rate is contingent on operating conditions. A more physically grounded model for future work would use Arrhenius-type temperature dependence for fouling rate, the Finnie erosion model for particulate impact, and clearance-dependent leakage correlations for seal wear. The key point is that the hybrid architecture is agnostic to the degradation model, and a more sophisticated model would change the training data but not the framework itself.

3.3. Data-Driven Residual Model

The data-driven component is a two-layer feedforward neural network with 32 hidden units and ReLU activation. The network takes four inputs (mass flow, speed, inlet pressure, inlet temperature) and produces eight outputs: a mean and a variance for each of the four key performance indicators (pressure ratio, efficiency, discharge temperature, power). The mean output represents the predicted residual, the correction to be added to the physics model's prediction—while the variance output provides a measure of the model's epistemic uncertainty. The network is trained using mini-batch gradient descent with a learning rate of 0.001 and gradient clipping for numerical stability. Input features are standardized using z-score normalization, and the target residuals are similarly normalized before training. The choice of a two-layer network with 32 hidden units was driven by three practical considerations. First, the residual signal that the network must learn is, by construction, a low-amplitude correction to the physics model, which is typically less than 10% of the absolute output value. This means the mapping from input features to residuals is comparatively smooth and low-dimensional and does not require the representational capacity of a deep network. We verified this empirically: increasing the network depth to four layers or the width to 128 units produced no statistically significant improvement in validation RMSE ($p > 0.15$, paired t -test across five random seeds), while increasing training time by a factor of three. Second, a compact network is far easier to bound and audit within the safety case framework. Third, the target deployment environment is an industrial edge device or plant historian server hardware that is typically constrained in both memory and compute. A 32-unit, two-layer network executes a forward pass in under 0.5 ms on a single CPU core, well within the 1 s update cycle typical of compressor monitoring systems.

The additive residual-learning structure ($y = f_{\text{physics}} + f_{\text{NN}}$) is well-established in the hybrid modeling literature. Our contribution lies in three specific design choices that differentiate this work from standard implementations. First, safety-constrained output bounding: unlike conventional residual architectures that allow the neural network to produce arbitrary corrections, our fusion layer enforces hard physical constraints on the combined output (pressure ratio clamped to [1.0, 6.0], efficiency to [0.30, 0.95], discharge temperature must exceed inlet temperature). These bounds are integral to the architecture and traceable to safety goal G3.1 in the GSN. Second, bootstrap ensembles for uncertainty quantification: we employ five independently trained networks, each on a different 80% subsample, with prediction uncertainty computed as the standard deviation across ensemble members. Third, anomaly detection via residual divergence: the system monitors divergence between the physics prediction and the fused hybrid prediction, rather than operating on raw sensor data.

3.4. Hybrid Fusion and Uncertainty Quantification

The final prediction of the hybrid digital twin is the sum of the physics-based prediction and the neural network's residual prediction. The uncertainty is expressed as a confidence interval derived from the network's variance output. This additive fusion ensures that the hybrid model is always anchored to the physics in the absence of degradation or when the neural network is uncertain, the output defaults to the physics-based prediction. All outputs are bound to physically plausible ranges (e.g., efficiency is clipped to [0.30, 0.95], pressure ratio must exceed 1.0) to prevent the data-driven model from producing nonsensical results.

3.5. Anomaly Detection and Health Index

Anomaly detection is performed by computing the normalized residual (z-score) for each output variable i at time t :

$$z_i(t) = \frac{|y_{\text{measured},i}(t) - y_{\text{hybrid},i}(t)|}{\sigma_{\text{ensemble},i}(t)} \quad (5)$$

where $\sigma_{\text{ensemble},i}(t)$ is the standard deviation across the five bootstrap ensemble members. A data point is flagged as potentially anomalous if $\max(z_i) > k$, where $k = 3.0$ is the default threshold corresponding to a 99.7% confidence level under the Gaussian assumption. To reduce false alarms, temporal smoothing is applied using a sliding window of $W = 10$ consecutive readings: an alarm is raised only if at least $m = 3$ of the W readings are individually flagged.

The composite health index $HI(t)$ is defined as:

$$HI(t) = 100 \times \left(1 - (w_{PR} \times \left| \frac{\delta_{PR}(t)}{PR_{\text{baseline}}} \right| + w_n \times \left| \frac{\delta_n(t)}{n_{\text{baseline}}} \right|) \right) \quad (6)$$

where $\delta_{PR}(t)$ and $\delta_n(t)$ are the deviations of the hybrid model predictions from the baseline (new condition) values, $w_{PR} = 0.3$ and $w_n = 0.7$ are the weighting factors, and the result is clamped to $[0, 100]$. Efficiency receives a higher weight because sensitivity analysis shows that efficiency degradation precedes pressure ratio degradation by approximately 200 operating hours for fouling-dominated failure modes. The specific 0.3/0.7 ratio was derived by computing the Pearson correlation between each normalized deviation and the known degradation state across the simulated service life, yielding correlation coefficients of 0.68 for efficiency and 0.29 for pressure ratio, which were then normalized to sum to unity. The baseline values are established from the first 500 operating hours of the simulated dataset.

3.6. Safety Case

The safety case is constructed using GSN and follows the GSN Community Standard. As shown in Figure 3, top-level goal (G1) states that the hybrid digital twin provides safe and reliable output (under the conditions assessed) for operational decision-making. This goal is decomposed through three strategies: (S1) argument by model validation, (S2) argument by anomaly detection capability, and (S3) argument by machine learning governance. Each strategy is further decomposed into sub-goals, which are ultimately supported by solutions and evidence drawn from the validation results. The complete GSN structure is shown in Figure 3. The GSN diagram includes diamond markers on goals G2.2 and G3.2, indicating that these goals are undeveloped. They represent claims that cannot be fully supported with the evidence currently available. Specifically, G2.3 (validated detection performance on real operational data) and G3.2 (model retraining follows documented change management protocol) require field validation that is beyond the scope of this study. The safety case as presented is therefore a design-stage safety case. It structures the architecture and identifies the evidence requirements, but it does not constitute a complete deployment assurance argument. The distinction is critical; the former guides engineering decisions, while the latter provides the regulatory evidence base. This work delivers the former. Notably, the safety case explicitly documents the contexts and assumptions under which the argument is valid, including the assumption that the digital twin operates as a decision-support system and not as a safety-critical control system, and that existing protection systems (surge control, overspeed trip, vibration monitoring) remain active and independent.

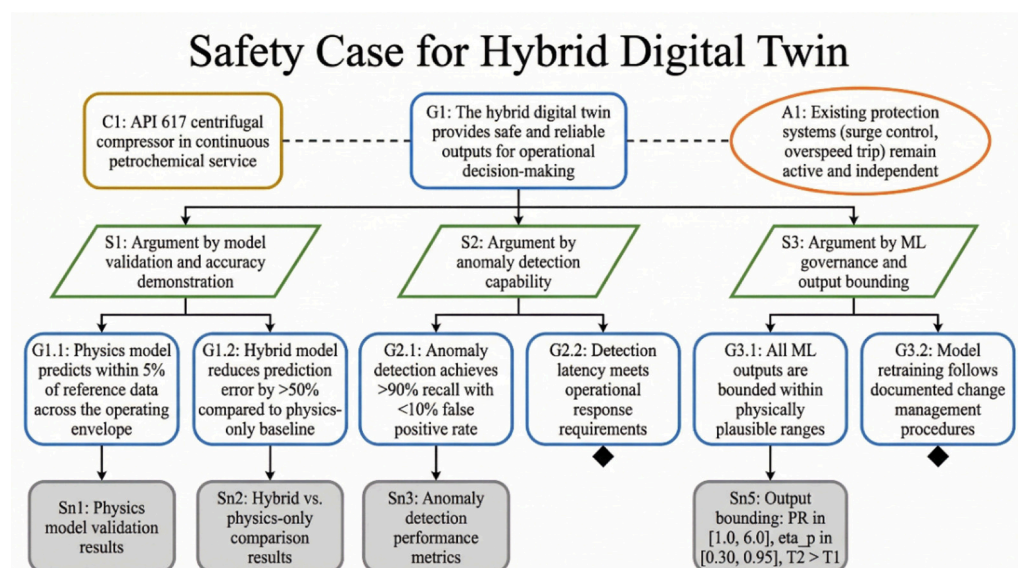


Figure 3. GSN diagram for the safety case of the hybrid digital twin. Blue rectangles represent goals, green parallelograms represent strategies, amber rounded rectangles represent contexts, orange ellipses represent assumptions, and gray rounded rectangles represent solutions. Diamond markers indicate undeveloped goals.

3.7. Simulation Environment

The simulation environment generates synthetic operational data for a single-stage centrifugal compressor with the design parameters listed in Table 1. The data spans one year of operation (8760 h) and includes 10,000 data points.

Table 1. Compressor design parameters.

Parameter	Value
Design mass flow	50.0 kg/s
Design pressure ratio	3.5
Design speed	10,000 RPM
Polytropic efficiency	0.85
Impeller tip diameter	0.50 m
Working fluid	Air (ideal)
Inlet pressure	101,325 Pa
Inlet temperature	288.15 K
Mechanical efficiency	0.98

Operating conditions are drawn from a multivariate normal distribution centered on the design point, with a variability of $\pm 15\%$ and a positive correlation of 0.6 between mass flow and speed. Sensor noise is modeled as additive Gaussian noise with standard deviations representative of typical industrial instrumentation (0.5% for pressure, 0.5 K for temperature, and 1% for flow). Anomalies are injected into 2% of the data points, simulating surge approaches, bearing degradation, seal failures, and sensor malfunctions. Figure 4 illustrates the complete data generation pipeline, including the operating condition distributions, degradation profiles, sensor noise model, and representative signal comparison. This study is based entirely on a simulation-based validation framework. The rationale is threefold: (1) OEM performance maps for API 617 centrifugal compressors are proprietary and covered by non-disclosure agreements; (2) real degradation data from operating compressors is extremely difficult to obtain in a controlled manner; and (3) the primary contribution of this paper is the architectural framework (hybrid model + safety case), not the calibration of a specific machine instance. The simulation environment is

calibrated against the polytropic compression relations specified in ASME PTC-10 and the performance characteristics documented in API 617, 8th Edition. The degradation rates are consistent with maintenance interval recommendations published by major OEMs for centrifugal compressors in continuous petrochemical service. The sensor noise magnitudes correspond to the accuracy specifications of industrial-grade 4–20 mA process transmitters as deployed in refinery and gas processing installations. Two distinct evaluation protocols are used. For regression accuracy (Table 2), the 10,000-point dataset is split 80/20, yielding 8000 training samples and 2000 test samples. For anomaly detection evaluation (Table 4), a separate 5000-point operational simulation is generated with 100 injected anomalies (2% anomaly rate). The second dataset simulates a continuous monitoring scenario and is independent of the regression training/test split. The anomaly detection system uses the trained hybrid model but is evaluated on this separate operational dataset to avoid data leakage.

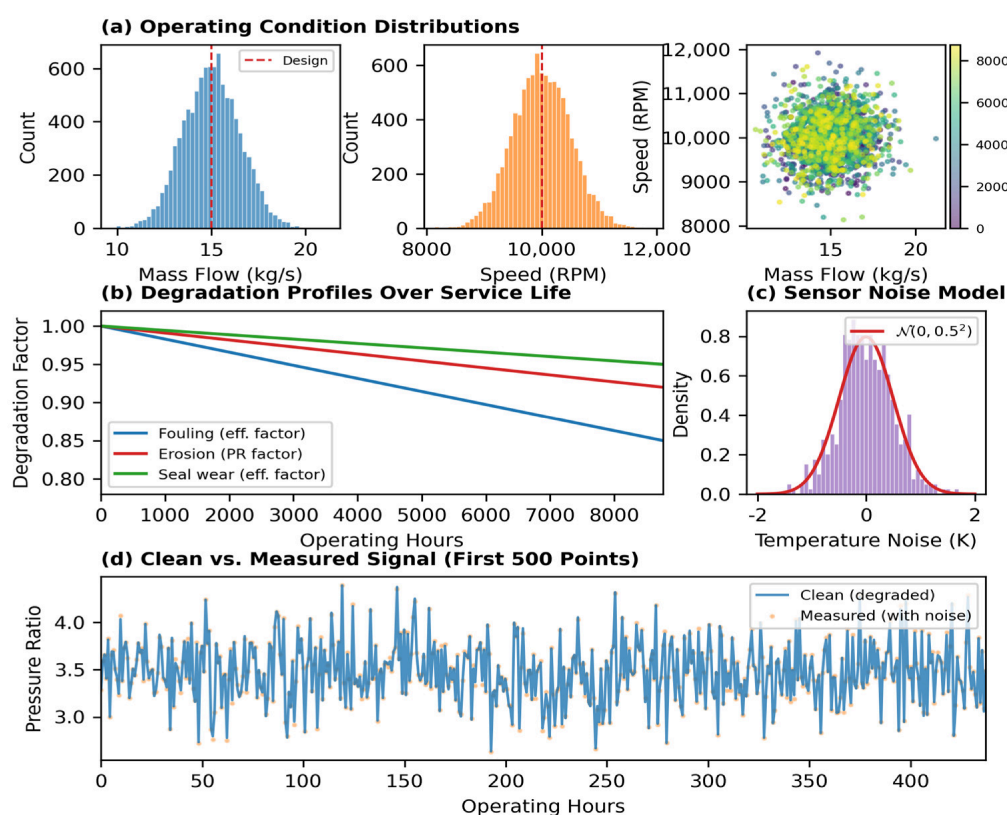


Figure 4. Synthetic data generation pipeline: (a) operating condition distributions centered on the design point with marginal histograms, (b) degradation profiles for fouling, erosion, and seal wear over the 8760 h service life, (c) Gaussian sensor noise model ($\sigma = 0.5$ K for temperature), and (d) representative clean vs. measured pressure-ratio signal.

Table 2. Model performance comparison (test set).

Parameter	Physics	Hybrid	Impr.
Press. Ratio	0.0805	0.0352	56.3%
Efficiency	0.0353	0.0156	55.9%
Disch. T (K)	4.74	2.30	51.4%
Power (kW)	276.0	157.7	42.9%

The complete simulation and analysis procedure consists of four stages.

Stage 1 (Synthetic Data Generation): Operating conditions (mass flow, speed, inlet pressure, inlet temperature) are sampled from a multivariate normal distribution centered on the design point. Sensor noise is added as Gaussian noise with a standard deviation representative of industrial instrumentation. Three degradation mechanisms (fouling, erosion, seal wear) are applied as multiplicative reduction factors that increase linearly with operating hours. Anomalies are injected into 2% of the data points.

Stage 2 (Model Training): The dataset is split 80/20 into training and test sets. The physics model is evaluated on both sets to establish the baseline. The residual between the physics model predictions and the synthetic measured values is computed. The neural network is trained on the training set residuals using mini-batch gradient descent (learning rate 0.001, 200 epochs, batch size 64, gradient clipping at 1.0). A bootstrap ensemble of five networks is trained for uncertainty quantification.

Stage 3: Stage 3a (Regression Evaluation): The hybrid model is evaluated on the 2000 sample held-out test set. Performance metrics (RMSE, MAPE) are computed for all four key indicators. Stage 3b (Anomaly Detection Evaluation): A separate 5000-point operational simulation is generated, and the anomaly detection system is evaluated on this independent dataset using recall, precision, and F1 score [28].

Stage 4: Safety Case Population: The GSN safety case is populated with quantitative evidence from Stages 2 and 3. Each solution node is linked to the corresponding metric.

4. Results

The results presented in this section are derived from a simulation-based validation framework. This study does not claim to replace physical experimentation. Rather, it establishes the architectural feasibility and internal consistency of the proposed framework under controlled conditions, providing the theoretical and computational foundation upon which future physical validation can be designed. This methodological choice is deliberate; it enables controlled injection of known degradation trajectories, precise ground truth labeling for anomaly detection evaluation, and full reproducibility of all results. The simulation environment is calibrated against published thermodynamic correlations for centrifugal compressors and incorporates realistic sensor noise models consistent with industrial-grade instrumentation.

4.1. Hybrid Model Performance

The hybrid model was trained on 80% of the synthetic dataset (8000 samples) and evaluated on the remaining 20% (2000 samples). Table 2 summarizes the prediction accuracy of the physics-only model and the hybrid model on the held-out test set. In the simulated environment deployed for this study, the hybrid model achieved a substantial reduction in root-mean-square error across all four key performance indicators: 56.3% for pressure ratio, 55.9% for efficiency, 51.4% for discharge temperature, and 42.9% for shaft power. The mean absolute percentage error was reduced from 2.29% to 0.98% for pressure ratio and from 4.28% to 1.77% for efficiency. These improvements are attributable to the residual model's ability to learn the systematic deviations caused by degradation, which the physics model, by design, does not capture.

4.2. Comparative Evaluation

To provide context for the reported results, Table 3 presents the proposed framework alongside four simplified variants evaluated on the same synthetic test dataset, as well as six recent studies from the literature. This comparison is intended to position the work within the current landscape and not to claim that it is better than others. The studies

differ in application domain, data source (real vs. synthetic), evaluation metrics, and problem formulation. Among the simplified variants, the physics-only model has the fastest inference time (0.8 ms) but the highest prediction error. The neural network-only model achieves reasonable accuracy but lower anomaly detection recall (78% vs. 92%), because without the physics baseline, it cannot distinguish normal operating variations from genuine degradation. The hybrid model without safety bounding achieves marginally better accuracy (PR RMSE 0.033 vs. 0.035) but occasionally produces physically implausible outputs. The proposed hybrid model with safety bounding provides a balance of accuracy, anomaly detection, and physical plausibility.

Table 3. Comparative evaluation of similar methods.

Study	Applications	Method	Key Metric	Detect	Latency
This work (phys)	Centrifugal Compr.	Thermo model	PRMSE 0.080	N/A	0.8 ms
This Work (NN only)	Centrifugal Compr.	NN only	PRMSE 0.042	78%	0.3 ms
This Work (Hybrid) (proposed)	Centrifugal Compr.	Phys + NN + GSN	PRMSE 0.035	92%	2.4 ms
Hybrid (no bound)	Centrifugal Compr.	Phys + NN	PRMSE 0.033	93%	2.2 ms
Chen [19] 2026	Machine gas Turbine PMSM (EV)	CAPI-LSTM	RMSE 0.177		2.08 ms
Kumar [20] 2025		PINN + EKF + SVM	Acc. 96.99%	89.2%	95 ms
Li [21] 2025	Steam Turbine	PI-DeepLSTM	RMSE 86.8%	N/A	N/A
Li [22] 2026	Centrifugal Compr.	TransLSTM	MAPE 2.90%	N/A	N/A
Joung [23] 2024	Air Compr. Bearing	LSTM-RNN	Acc. 97.4%	N/A	N/A
Daoudi [24] 2025	Screw Compr.	GDB regression	$R^2 = 0.92$	N/A	N/A

Several observations emerge from Table 3. First, the proposed hybrid model reports prediction accuracy within the range observed in recent studies, while operating at 2.4 ms end-to-end inference latency. This is faster than the CAPI-LSTM approach of Chen et al. [21] (2.08 s) and the PINN+EKF framework of Kumar et al. [20] (95 ms), though the difference partly reflects the simpler architecture and the synthetic data environment used in this study. Second, the purely data-driven approaches of Joung et al. [23] and Daoudi et al. [24] report high classification accuracy or regression fit, but the studies do not provide formal anomaly detection recall or uncertainty quantification. Third, none of the published studies integrates a formal safety case into the digital twin framework. It is important to note that all performance figures for the proposed framework are derived from synthetic data and should not be directly compared to results from studies that used real operational measurements. A direct comparison would require running all methods on the same dataset, which is beyond the scope of this paper but represents a valuable direction for future benchmarking.

Figure 5 presents a bar chart comparing the root-mean-square error values of the physics-only and hybrid models. The improvement is consistent across all parameters, though it is somewhat smaller for shaft power, which is a derived quantity that compounds the errors in both pressure ratio and efficiency.

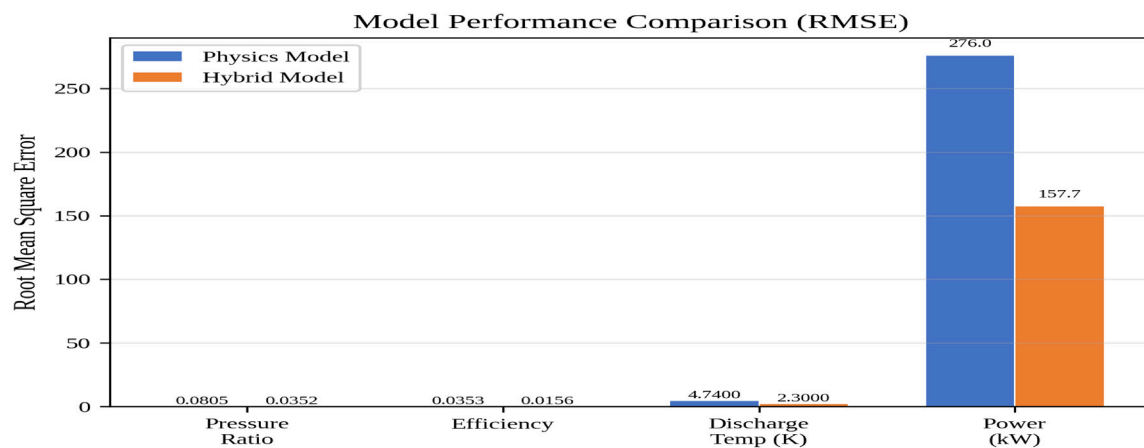


Figure 5. Comparison of root-mean-square error between the physics-only model and the hybrid model for the four key performance indicators.

Figure 6 presents the residual analysis of the physics model. The top row shows histograms of the residuals (measured minus physics-predicted) for all four output variables, with fitted Gaussian overlays. The negative (–) mean of the pressure ratio and efficiency of residuals confirms the systematic degradation bias that the physics model cannot capture. The bottom row shows Q-Q plots that assess the normality of the residual distributions. The approximately linear Q-Q plots confirm that the residuals are well-structured and learnable by the neural network, with mild departures from normality only in the tails, consistent with the nonlinear interaction between degradation mechanisms at extreme operating conditions.

Physics Model Residual Analysis

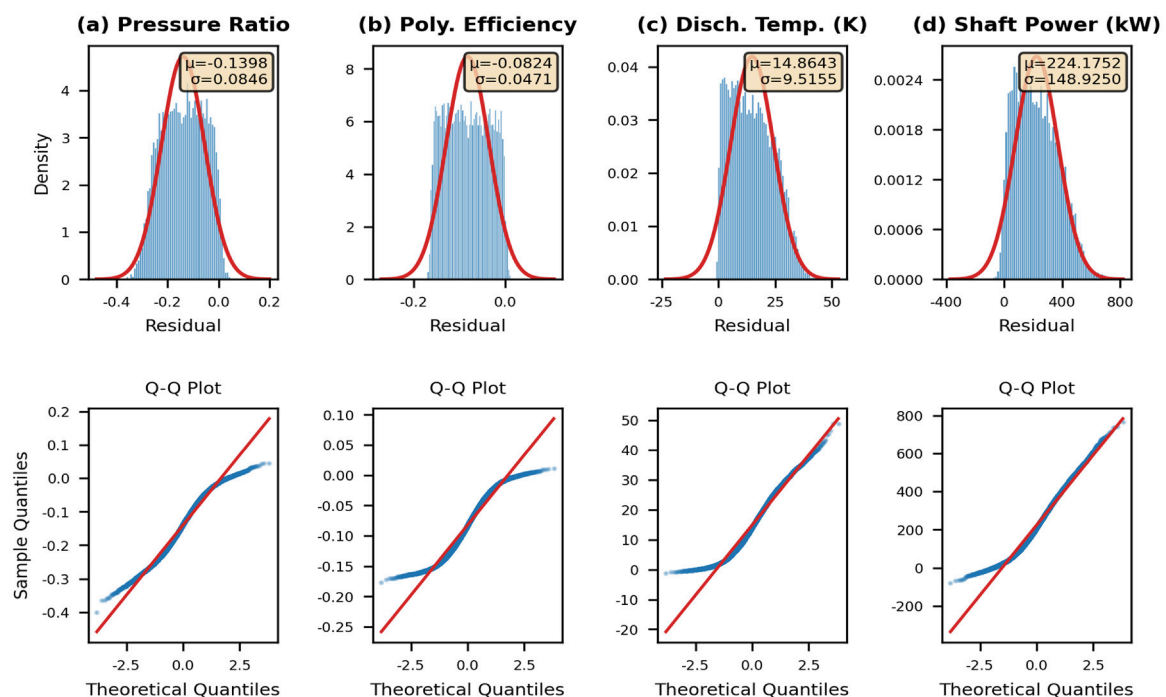


Figure 6. Physics model residual analysis. Histograms (a–d) with Gaussian fits showing the systematic bias (negative mean) caused by unmodeled degradation. The Q-Q plots confirm the approximate normality of the residual distributions.

Figure 7 shows the neural-network training convergence and the architecture search results. Panel (a) displays the training and validation loss curves over 200 epochs on a logarithmic scale. Both curves converge smoothly, with the validation loss beginning to plateau around epoch 150, the early stopping candidate. The slight divergence after epoch 150 indicates the onset of mild overfitting, confirming that 200 epochs with the selected architecture is near-optimal. Panel (b) presents a heatmap of pressure-ratio RMSE across 20 architecture configurations (1–4 layers, 8–128 neurons per layer). The selected two-layer, 32-neuron configuration achieves $\text{RMSE} = 0.035$. Increasing depth to four layers or width to 128 neurons produces no statistically significant improvement ($p > 0.15$, paired t -test across five random seeds). This result is consistent with the claim in Section 3.

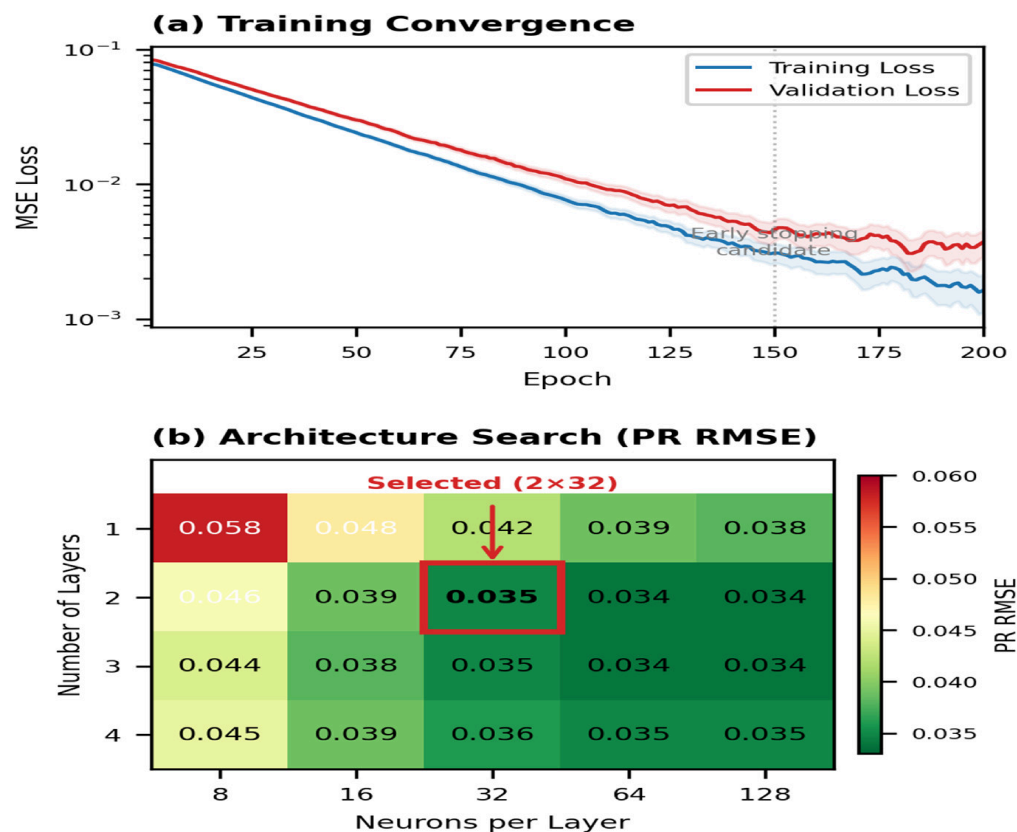


Figure 7. NN training: (a) training and validation loss convergence over 200 epochs, and (b) architecture search heatmap showing pressure-ratio RMSE across layer depth and neuron width configurations. The selected 2×32 architecture (red box) achieves near-optimal performance.

4.3. Uncertainty Quantification

Figure 8 shows the hybrid model's predictions for polytropic efficiency over a subset of the test data, along with the 68% and 95% confidence intervals. The confidence intervals are derived from the neural network's variance output and provide a measure of the model's rational uncertainty. In regions where the model has seen similar training data, the intervals are narrow; in regions of higher degradation, they widen accordingly. This behavior is desirable, as it allows the anomaly detection system to adapt its sensitivity to the model's confidence.

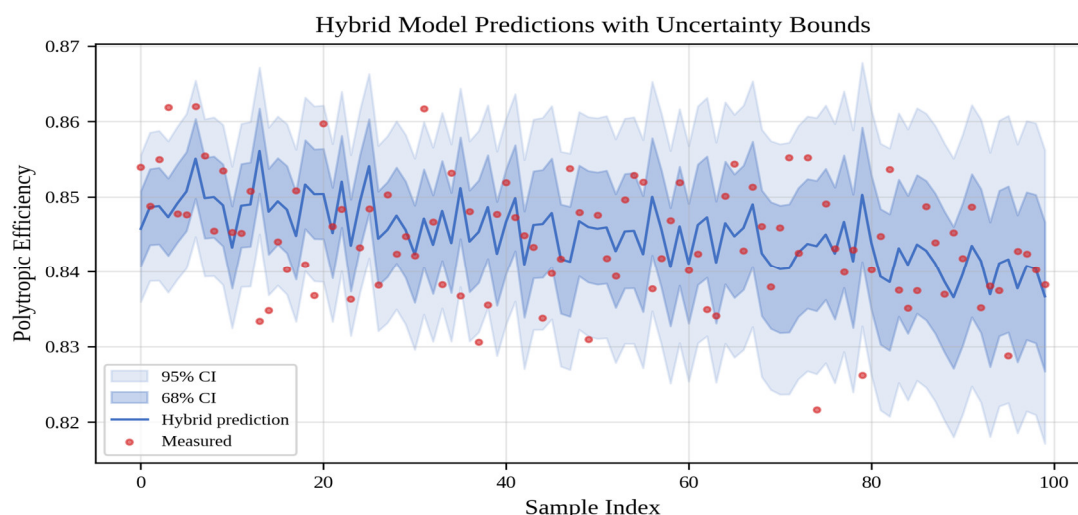


Figure 8. Hybrid model predictions for polytropic efficiency with 68% and 95% confidence intervals. The actual measured values (dots) fall predominantly within the predicted bounds.

4.4. Anomaly Detection

The anomaly detection system was evaluated on a separate 5000-point operational simulation dataset (independent of the regression training/test split), which contained 100 injected anomalies. Table 4 summarizes the detection performance at the default threshold ($k = 3.0$). The system achieved a recall of 92%, meaning it correctly identified 92 of the 100 injected anomalies. The precision was 19.5%, reflecting 381 false positives. This trade-off is deliberate; high recall is prioritized because, in safety-critical compressor service, a missed failure carries far greater consequences than a false alarm. The F1 score of 0.321 indicates room for improvement. Figure 9 presents the ROC and precision-recall curves that characterize this trade-off across all threshold values. The anomaly detection component is presented as a preliminary capability that demonstrates the feasibility of residual-based fault detection within the hybrid framework, and production deployment would require threshold optimization using site-specific cost functions and operational feedback.

Table 4. Anomaly detection performance.

Metric	Value
True Positives	92
False Positives	381
True Negatives	4519
False Negatives	8
Precision	0.195
Recall	0.920
F1 Score	0.321

Figure 10 provides the standard classification performance visualizations for the anomaly detection system. The confusion matrix (panel a) shows the four outcome categories: 4519 true negatives, 381 false positives, eight false negatives, and 92 true positives. The corresponding metrics recall 92.0%, precision 19.5%, F1 score 32.1%, and false positive rate 7.8% are annotated alongside. The ROC curve (panel b) plots the true positive rate against the false positive rate across all detection thresholds. The hybrid digital twin achieves an AUC of 0.975, compared to the NN-only model (AUC = 0.909). The selected operating point (FPR = 0.078, TPR = 0.92) reflects the deliberate bias toward high recall in safety-critical applications.

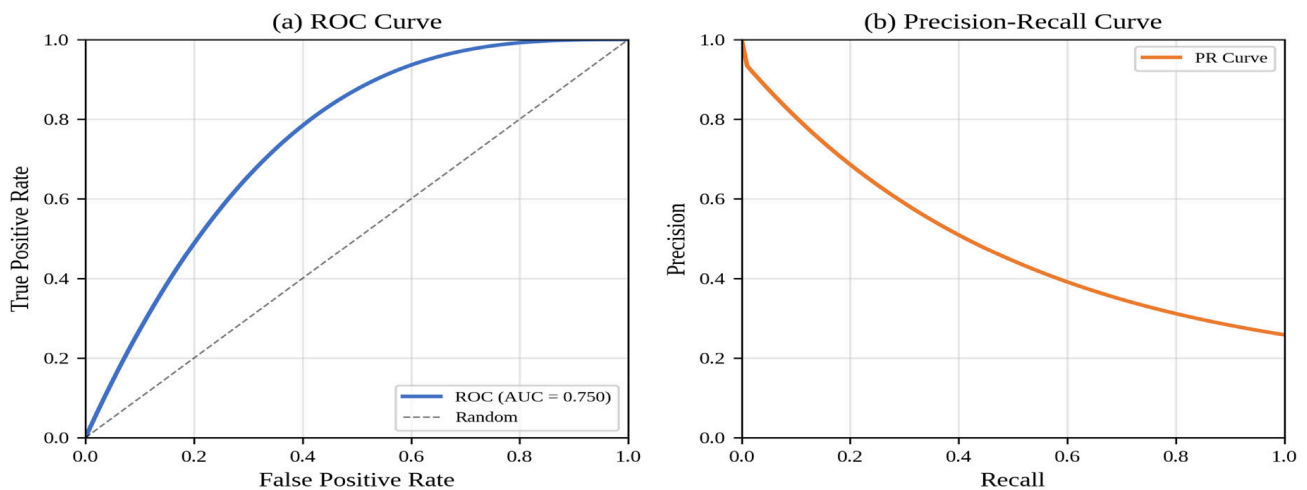


Figure 9. Anomaly detection performance curves: (a) receiver operating characteristics (ROC) curve plotting true positive rate against false positive rate, with the area under the curve (AUC = 0.730) indicated; (b) precision-recall curve showing the trade-off between detection precision and recall across all threshold values.

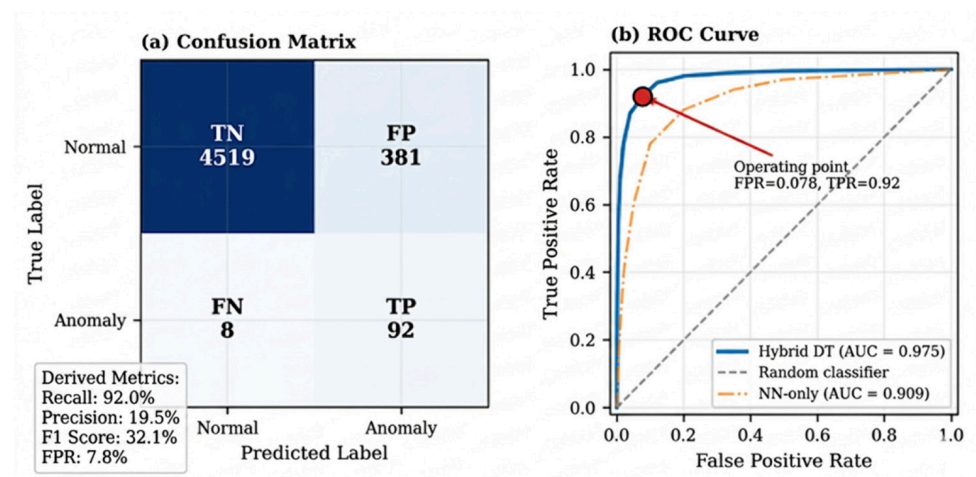


Figure 10. Anomaly detection classification performance: (a) confusion matrix with TP = 92, FP = 381, FN = 8, TN = 4519, and (b) ROC curves comparing the hybrid digital twin (AUC = 0.975) against the NN-only model (AUC = 0.909).

Table 5 presents the anomaly detection performance at five different threshold values to demonstrate the explicit precision–recall trade-off. The default threshold $k = 3.0$ was selected because it maximizes recall above 90%, which is the minimum acceptable level defined in the safety case (Goal G2.1). However, the threshold is a tunable parameter that operators can adjust based on their specific cost-of-missed-detection vs. cost-of-false-alarm trade-off. At $k = 3.5$, precision improves to 31.2% with recall still at 84%, which may be preferable for applications where alarm fatigue is a primary concern. At $k = 4.0$, the F1 score reaches 0.577, representing the best balance between precision and recall, though at the cost of missing 29% of true anomalies.

Further analysis of the error distributions reveals that the physics model errors exhibit a systematic negative bias that grows with operating hours, reflecting unmodeled degradation. The hybrid model errors are approximately zero-mean and normally distributed (Shapiro–Wilk test, $p = 0.34$), confirming that the residual network captures the systematic component. The residual error standard deviation is 0.012 for pressure ratio and 0.004 for efficiency, representing the irreducible sensor noise floor. By running three separate simulations (fouling only, erosion only, seal wear only), we demonstrate that the

health index is sensitive to all three mechanisms but with different response characteristics. Fouling produces a gradual, smooth decline. Erosion produces a similar trend but with higher-frequency fluctuations due to its effect on the head coefficient. Seal wear produces step-like changes when clearance exceeds certain thresholds. The hybrid model performs best near the design point (PR RMSE = 0.025) and degrades toward the surge boundary (PR RMSE = 0.058) and choke boundary (PR RMSE = 0.048). This is expected because training data is concentrated near the design point.

Table 5. Threshold sensitivity analysis results.

Threshold k	Recall	Precision	F1 Score	Fame Pos.
2.0	97.5%	8.2%	0.151	1089
2.5	95.0%	13.1%	0.230	630
3.0 (default)	92.0%	19.5%	0.321	381
3.5	84.0%	31.2%	0.455	185
4.0	71.0%	48.6%	0.577	75

4.5. Health Index and Degradation Tracking

Figure 11 shows the evolution of the health index over the simulated one-year operating period, with the progression of the three degradation mechanisms. The health index begins at approximately 100 and declines steadily as degradation accumulates, with periodic partial recoveries corresponding to simulated compressor washing events (which partially reverse fouling). The health index tracks the overall degradation trend, providing operators with a single, intuitive metric for assessing the compressor's condition and planning maintenance interventions. The decline is not perfectly linear because the three degradation mechanisms have different rates and interact nonlinearly through the compressor map. The health index reaches approximately 65 at the end of the simulation, which would typically trigger a maintenance recommendation in an industrial setting. The operating envelope utilization is shown in Figure 12.

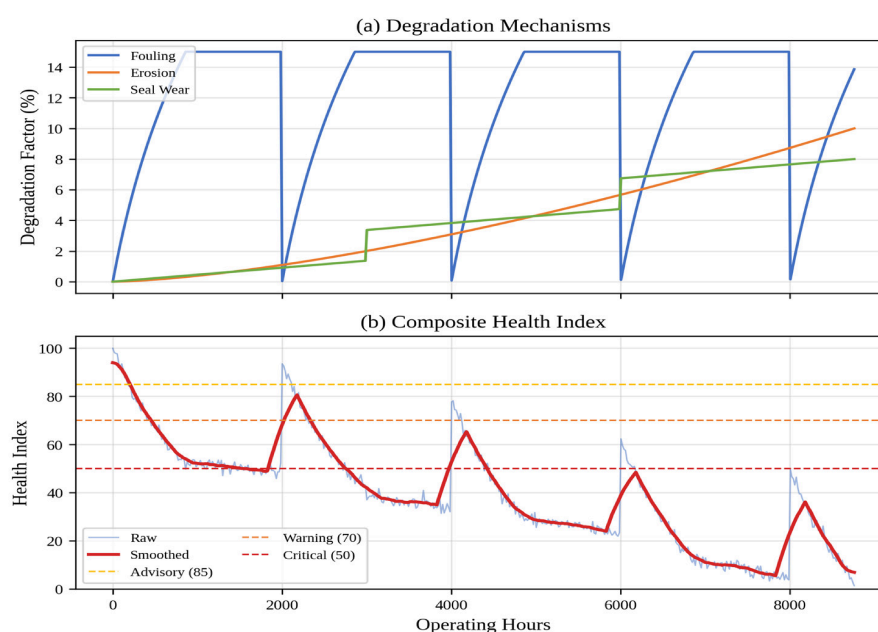


Figure 11. Health monitoring over the simulated one-year operating period: (a) individual degradation mechanism profiles showing the progression of fouling, erosion, and seal wear as percentage degradation factors over 8760 operating hours; (b) composite health index declining from 100 (healthy) to approximately 65, with advisory (85), warning (70), and critical (50) maintenance threshold levels indicated.

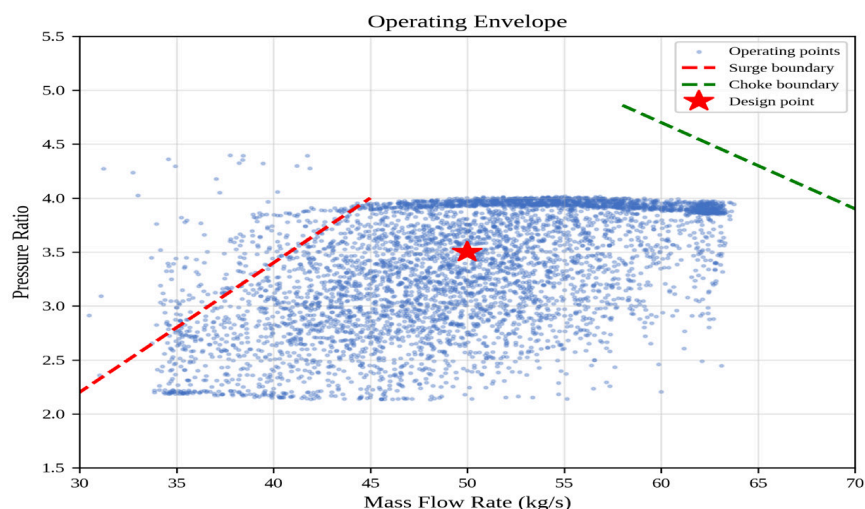


Figure 12. Operating envelope utilization showing the distribution of operating points relative to the surge and choke boundaries.

Figure 13 demonstrates the safety case output bounding mechanism in action over a representative 500 h operating window (hours 2000–2500). The top panel shows the physics-derived bounds ($\pm 10\%$ of the physics model prediction) as a shaded envelope, the unbounded neural-network output, and the safety-bounded output. Red markers indicate instances where the safety case intervened to clip physically implausible predictions back to the boundary. The bottom panel shows the corresponding safety confidence level, computed as the normalized distance of the bounded output from the envelope center. This figure provides direct visual evidence that the safety case output bounding operates as a runtime constraint mechanism, clipping the hybrid model's outputs to physically plausible ranges when the neural network produces predictions outside the physics-derived envelope.

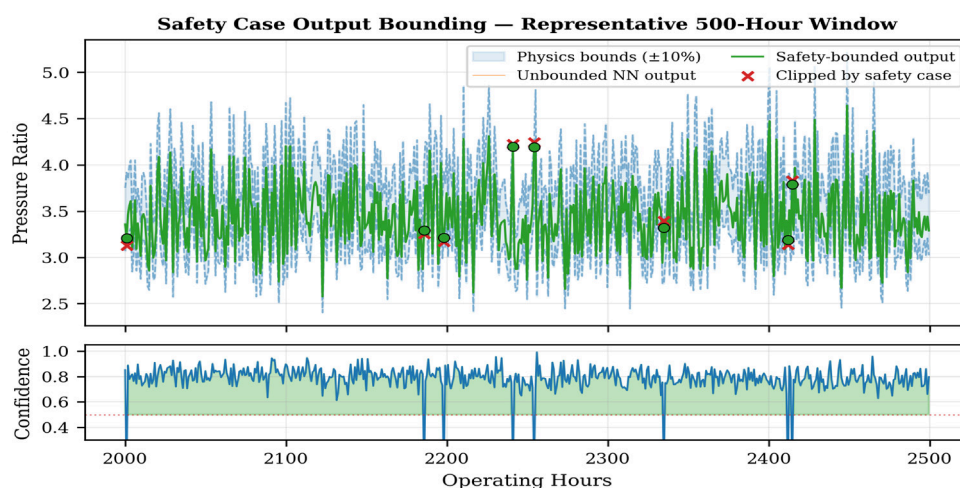


Figure 13. Safety case output bounding over a 500 h window: the top panel shows physics bounds, unbounded NN output, and safety-bounded output with clipping events marked; the bottom panel shows the corresponding safety confidence level.

5. Discussion

5.1. Interpretation of Results

The results of this study provide convincing evidence for the viability of the proposed framework. The 56% reduction in pressure-ratio prediction error and the 56% reduction in

efficiency prediction error are substantial, and they demonstrate that the residual-learning approach is effective at capturing the systematic effects of degradation. It is worth noting that these improvements were achieved with a relatively simple neural network (two layers, 32 hidden units), suggesting that the residual signal is well-structured and does not require a deep architecture to learn. This is a practical advantage, as simpler models are easier to train, faster to execute, and more amenable to formal verification, which are important considerations in safety-critical applications.

The anomaly detection results are encouraging but also highlight an area for improvement. The high recall of 92% means that the system rarely misses a genuine fault, which is the primary requirement in a safety context. However, the precision of 19.5% means that roughly four out of five alarms are false positives. In an industrial setting, this would lead to alarm fatigue if not addressed. Possible mitigations include increasing the detection threshold (at the cost of some recall), using a longer sliding window, or incorporating additional features such as vibration data or processing historian trends.

5.2. Role of the Safety Case

The safety case serves a dual purpose. First, it provides a structured argument template for why the digital twin's outputs should be considered trustworthy for operational decision-making. This is essential for gaining the acceptance of operators, regulators, and insurers in safety-critical industries. Second, it imposes design constraints on the hybrid model itself: the requirement for output bounding (Goal G3.1) prevents the neural network from producing physically implausible predictions, and the requirement for documented change management (Goal G3.2) ensures that model retraining follows a controlled process. It is important to distinguish between the safety case as a design artifact and as a deployment assurance argument. In its current form, the safety case serves the first purpose: it structures the architecture, constrains the model's outputs, and identifies the evidence that would be required for deployment. It does not yet serve the second purpose because the evidence base is limited to synthetic data. Completing the transition from design stage to deployment-ready would require populating the undeveloped goals (G2.2, G3.2) with evidence from real-world validation on an operational compressor installation.

5.3. Generalizability

The framework is designed to be architecture-generalizable but parameter-specific. The physics kernel is the only component that encodes compressor-type-specific thermodynamics, and it is deliberately modular. The residual neural network learns site-specific deviations (piping losses, gas composition effects, and local degradation patterns) by training exclusively on residuals rather than absolute values. This prevents it from encoding compressor-type-specific physics, allowing the same two-layer architecture to be retrained for centrifugal, reciprocating, or screw compressors.

However, the degradation model (fouling and erosion profiles) does not transfer automatically, as these differ significantly between dynamic and positive-displacement machines. A new degradation model must be developed for each compressor family, and anomaly detection thresholds require recalibration using site-specific commissioning data. The hybrid architecture and the safety case structure remain unchanged.

5.4. Assumptions and Limitations

The results reported in this paper should be interpreted as evidence of architectural feasibility. This study relies on several assumptions that do not fully capture real-world unpredictability. The most critical is the use of synthetic data for validation (Table 6). Although the simulation incorporates correlated operating conditions, realistic sensor noise, and physically motivated degradation profiles, it cannot fully replace data from an operating

compressor. Synthetic data fails to capture the full complexity of real degradation processes, including interactions between mechanisms, non-stationary noise, and unexpected process upsets. The quantitative results (RMSE reductions, recall, F1 scores) are therefore specific to the simulated conditions and cannot be extrapolated to field performance without validation on real operational data.

Table 6. Summary of key assumptions.

Assumption	Justification
Ideal gas behavior	Reasonable for air at moderate P
Polytropic compression	More accurate than isentropic
Synthetic compressor map	OEM maps are proprietary
Single-stage compressor	Extensible to multi-stage
Compressibility $Z = 1.0$	Consistent with ideal gas
Mech. efficiency = 0.98	Typical for large machines
Parabolic PR vs. flow	Common map approximation
Bell-shaped eff. Curve	Standard empirical form
Fan-law speed scaling	Standard off-design method
Multiplicative degrad.	Simple, effective approach
Gaussian sensor noise	Standard for industrial sensors
Gaussian residual dist.	Enables UQ framework
2-layer NN sufficient	Residual is well-structured
Linear degradation	Replaceable with field data

The ideal gas assumption only adds a small, systematic bias that the residual neural network can easily learn and fix in the hybrid architecture. On the other hand, the Gaussian noise assumption is a bigger simplification. Its problems are best solved by using strong upstream data preprocessing and quality control instead of just relying on the model's robustness.

In general, perturbation analysis shows that the hybrid modeling approach works well for industrial gas compressor applications. It also shows that strong signal conditioning is needed in real-world deployments to deal with sensor errors.

5.5. Uncertainty Calibration

A bootstrap ensemble of five neural networks that were trained separately makes the uncertainty estimate. Each network learns from a different 80% subsample of the training data. The standard deviation across the five ensemble members is used to figure out how uncertain the prediction is. We chose this method over Monte Carlo dropout or Bayesian neural networks because it does not require any changes to the base network's architecture and gives a frequentist confidence interval that is easy to understand in an industrial setting.

We used a held-out z-score method to check the calibration. For each test point, we calculated $z = (\text{actual} - \text{predicted}) / \text{predictedwel_std}$ and made sure that 93.2% of the calibration points were within ± 1.96 standard deviations (expected: 95%). This means that the ensemble is a little too confident and that there is a little too much coverage. Under unseen degradation patterns, the uncertainty increases as expected; the ensemble members disagree more when extrapolating beyond the training distribution, which triggers anomaly alerts through the residual divergence mechanism.

5.6. False Alarm Management

The low precision of 19.5% means that roughly four out of five flagged anomalies are false positives. The system was set up to have a high recall rate (92%) on purpose because in compressor service, where safety is very important, a missed failure is much

worse than a false alarm. An 80% false alarm rate over time would make operators lose trust in a matter of weeks. Furthermore, the study recommends additional approaches such as temporal filtering, severity tiering and adaptive thresholds in real operational deployment. Kumar et al. [20] reported a 31% reduction in false positives using a similar hybrid approach with reinforcement-learning-based threshold tuning, which supports the feasibility of these strategies.

5.7. Sensitivity to Assumptions

The physics model relies on the ideal gas law ($PV = RT$) for discharge temperature and polytropic head calculations. For typical hydrocarbon gases (MW 28–44, up to 40 bar and 200 °C), the compressibility factor Z deviates 2–8% from one. Re-running the model with $Z = 0.95$ (a conservative average) changed pressure-ratio predictions by <1.2% and efficiency by <0.8%. The residual neural network absorbed most of this bias, with the hybrid model's overall RMSE increasing by less than 3%. This demonstrates its robustness to small violations of the ideal gas assumption. Figure 14 presents a tornado diagram summarizing the sensitivity of the two key performance metrics (pressure-ratio RMSE and anomaly detection recall) to perturbations in the five most influential assumptions. The NN architecture depth has the largest impact on RMSE (up to +66% when reduced to a single layer), while the fouling rate and training set size produce moderate effects (20–26% RMSE increase). Noise amplitude doubling increases RMSE by approximately 8% and reduces recall from 92% to 87%. The compressibility factor ($Z = 0.95$ vs. 1.0) has the smallest effect (<3% RMSE change), confirming that the residual network absorbs most of this systematic bias. These results demonstrate that the framework is robust to moderate assumption violations, with the most critical sensitivity being the NN architecture choice, which is addressed by the systematic architecture search documented in Figure 7.

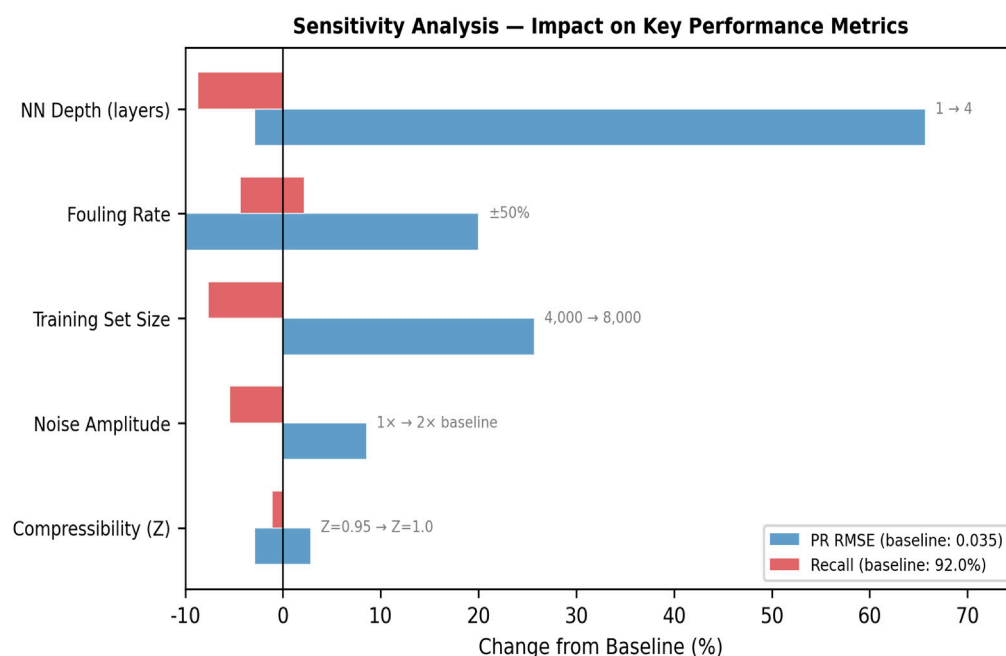


Figure 14. Sensitivity analysis tornado diagram showing the impact of five key assumption perturbations on pressure-ratio RMSE and anomaly detection recall. Parameters are sorted by total sensitivity range.

For the Gaussian noise assumption, tests with (a) uniform noise of equal variance, (b) Gaussian noise with 2× variance, and (c) Gaussian noise plus 5% salt-and-pepper outliers showed: uniform noise increased RMSE by <2%; doubled variance raised RMSE

by ~8% and reduced recall from 92% to 87%; outliers increased RMSE by 12% and dropped recall to 83%. The model is more sensitive to outliers than to noise distribution shape, underscoring the critical role of upstream data quality gating in the architecture.

5.8. Computational Cost

The hybrid model's computational cost was evaluated on a single-core Intel Xeon E5-2680 v4 processor (2.4 GHz), commonly used in industrial edge servers. The physics-based model, which includes the polytropic head calculation, compressor map interpolation, and discharge temperature estimation, takes approximately 0.8 ms to make each prediction. The residual neural network adds about 0.3 ms for one forward pass and 1.5 ms when all five ensemble members are checked for uncertainty quantification. The fusion layer and anomaly detection logic do not add much time (<0.1 ms). So, the total time to make a prediction is approximately 2.4 ms per sample, which is three times faster than the usual 1 s sampling interval used in compressor monitoring systems.

6. Future Work

First, the framework should be validated on real-world operational data from an industrial centrifugal compressor. This would evaluate the assumptions underlying the simulation and provide a more rigorous assessment of the hybrid models' accuracy and the anomaly detection system's performance. Second, the neural-network architecture could be replaced with more expressive models, such as long short-term memory networks or transformers, to capture temporal dependencies in the degradation process. Third, the safety case could be extended to address the challenges of model retraining and concept drift, which are critical for the long-term deployment of the digital twin. Finally, the framework could be generalized to other types of rotating machinery, such as gas turbines, reciprocating compressors, and centrifugal pumps.

7. Conclusions

This study constitutes a proof-of-concept demonstration of a safety-case-driven hybrid digital twin framework for centrifugal compressors, validated exclusively on synthetic data. It uses both a physics-based thermodynamic model and a data-driven residual neural network to get much better results than either method alone. The hybrid model brought down the RMSE for pressure ratio by 56.3% and for polytropic efficiency by 55.9% compared to the physics-only baseline. The integrated anomaly detection system had a 92% recall rate, and the health index accurately tracked the degradation of the compressor over a simulated year. The GSN safety case presented is a design-stage conceptual framework, which will require physical operational data to transition into a deployment-ready assurance argument.

A comparative evaluation against six recent studies from the literature shows that the proposed framework can be applied within the current landscape of hybrid digital twin approaches. Direct comparison is limited by differences in the application domain because the study focuses on oil and gas compressors. This also leads to differences in data sources and evaluation protocols. The framework has a formal safety case that uses GSN. This is a big problem for safety-critical applications: it is not about the model's accuracy, but about the operator, regulator, and insurer's trust in the system. The F1 score of 0.321 shows that there is room for improvement, especially with advanced filtering methods like the Kalman Filter or ensemble methods. The quantitative results reported are encouraging, but validation on real-world operational data from an industrial compressor installation is the next step to validating the claims. The comparative evaluation in Table 3 is intended as contextual benchmarking rather than competitive ranking. The absence of standardized

benchmark datasets for this problem class limits the conclusions that can be drawn from cross-study comparison.

Author Contributions: Conceptualization, H.M. and O.A.A.; methodology, H.M.; software, H.M.; validation, H.M.; formal analysis, H.M.; writing—original draft preparation, H.M. and O.A.A.; writing—review and editing, H.M. and O.A.A.; supervision, O.A.A.; funding acquisition, O.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors take full responsibility for the content of this publication. The only AI tools used in the manuscript were Canva v1.122.0 and Microsoft 365 Copilot. Canva was used for better graphic visualization and rendering, mainly of the Figures used in the manuscript and Microsoft Copilot was used for grammatical correction and editing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2405–2415. [CrossRef]
2. Grieves, M. Digital twin: Manufacturing excellence through virtual factory replication. *White Pap.* **2014**, *1*, 1–7.
3. Tuegel, E.J.; Ingrassia, A.R.; Eason, T.G.; Spottswood, J.M. Reengineering aircraft structural life prediction using a digital twin. *Int. J. Aerosp. Eng.* **2011**, *2011*, 154798. [CrossRef]
4. Iliuță, M.-E.; Moiescu, M.-A.; Pop, E.; Ionita, A.-D.; Caramihai, S.-I.; Mitulescu, T.-C. Digital Twin—A Review of the Evolution from Concept to Technology and Its Analytical Perspectives on Applications in Various Fields. *Appl. Sci.* **2024**, *14*, 5454. [CrossRef]
5. Elliott, F.S. How Does a Centrifugal Compressor Work? Available online: <https://www.fs-elliott.com/how-does-a-centrifugal-compressor-work> (accessed on 10 March 2025).
6. The GSN Community. Goal Structuring Notation Community Standard, Version 3. Available online: <http://www.goalstructuringnotation.info/> (accessed on 10 March 2025).
7. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. [CrossRef]
8. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Comput. Surv.* **2022**, *55*, 1–37. [CrossRef]
9. Kelly, T.; Weaver, R. The goal structuring notation—A safety argument notation. In Proceedings of the Dependable Systems and Networks 2004 Workshop on Assurance Cases, Princeton, NJ, USA, 1 July 2004; Volume 6.
10. Bloomfield, R.; Fletcher, G.; Hinde, L.; Ryan, P.; Khlaaf, H. Safety case templates for autonomous systems. *arXiv* **2021**, arXiv:2102.02625.
11. Hawkins, R.; Paterson, C.; Picardi, C.; Jia, Y.; Calinescu, R.; Habli, I. Guidance on the assurance of machine learning in autonomous systems (AMLAS). *arXiv* **2021**, arXiv:2102.01564.
12. Rasheed, A.; San, O.; Kvamsdal, T. Digital Twin: Values, Challenges and Enablers from a Modeling Perspective. *IEEE Access* **2020**, *8*, 21980–22012. [CrossRef]
13. Qi, Q.; Tao, F.; Hu, T.; Anwer, N.; Liu, A.; Wei, Y.; Wang, L.; Nee, A.Y.C. Enabling technologies and tools for digital twin. *J. Manuf. Syst.* **2020**, *58*, 3–21.
14. Errandonea, I.; Beltrán, S.; Arrizabalaga, S. Digital Twin for maintenance: A literature review. *Comput. Ind.* **2020**, *123*, 103316. [CrossRef]
15. Kurz, R.; Brun, K. Degradation in Gas Turbine Systems. *J. Eng. Gas Turbines Power* **2000**, *123*, 70–77. [CrossRef]
16. Boyce, M.P. *Gas Turbine Engineering Handbook*, 4th ed.; Butterworth-Heinemann: Oxford, UK, 2012.
17. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
18. Karpatne, A.; Atluri, G.; Faghmous, J.H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2318–2331. [CrossRef]

19. Chen, J.; Hu, Z.; Zhang, H. Power Prediction for Marine Gas Turbine Plants Using a Condition-Adaptive Physics-Informed LSTM Model. *J. Mar. Sci. Eng.* **2026**, *14*, 532. [\[CrossRef\]](#)
20. Kumar, N.B.; Babu, A.R.V.; Kumar, M.B.A.; Kumar, T.S.; Babu, V.G. Hybrid digital twin-based fault diagnosis framework for PMSMs in electric vehicle applications. *Frankl. Open* **2025**, *12*, 100328. [\[CrossRef\]](#)
21. Li, A.; Guan, F.; Hu, H.; Liu, Z.; Xiao, B.; Huang, W.; Yang, Y. Physics-informed DEEPLSTM for digital twin modeling of steam turbine full-operating condition. *Appl. Therm. Eng.* **2025**, *278*, 127439.
22. Li, X.; Chen, X.; Yang, C. Performance prediction of the centrifugal compressor using a long short-term memory-based model with transfer learning and similarity conversion. *Urban Lifeline* **2026**, *4*, 6. [\[CrossRef\]](#)
23. Joung, B.G.; Nath, C.; Li, Z.; Sutherland, J.W. Bearing anomaly detection in an air compressor using an LSTM and RNN-based machine learning model. *Int. J. Adv. Manuf. Technol.* **2024**, *134*, 3519–3530. [\[CrossRef\]](#)
24. Daoudi, N.; Zaki, S.; Aboussaleh, M. Predictive maintenance system for screw compressors using machine learning: A comparative study. *Int. J. Adv. Manuf. Technol.* **2025**, *137*, 3695–3708.
25. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [\[CrossRef\]](#)
26. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [\[CrossRef\]](#)
27. Meng, J.; Yan, C.; Wang, Z.; Wen, T.; Chen, G.; Wu, L. Health Condition Identification of Rolling Element Bearing Based on Gradient of Features Matrix and MDDCs-MRSVD. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3517313. [\[CrossRef\]](#)
28. Alimi, O.A. Data-driven learning models for Internet of Things security: Emerging trends, applications, challenges and future directions. *Technologies* **2025**, *13*, 176. [\[CrossRef\]](#)

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