aerospace

MDPI

Article

State of Health Prediction Method for the Gas Turbine Aero-Engine Fuel Metering Units Based on Inverted Stabilized LSTM-Transformer

Yingzhi Huang [1,2,*], Xiaonan Wu [2], Junwei Li [3] and Linfeng Gou [2]

[1] Shenzhen Research Institute of Northwestern Polytechnical University, Shenzhen 518057, China
[2] School of Power and Energy, Northwestern Polytechnical University, Xian 710129, China; wuxiaonan@mail.nwpu.edu.cn (X.W.); goulinfeng@nwpu.edu.cn (L.G.)
[3] Beijing Hangke Engine Control System Technology Co., Ltd., Aero Engine Corporation of China, Beijing 102200, China; ljw_bjhk@163.com
* Correspondence: huangyingzhi@nwpu.edu.cn

Abstract

As a critical actuator in aero-engine control systems, the health condition of the Fuel Metering Unit (FMU) directly influences flight safety and maintenance efficiency, making the precise prediction of its degradation process a core task in the engine's Prognostic and Health Management (PHM). This paper presents a novel inverted stabilized LSTM-Transformer (isLTransformer) approach for predicting the health state of aero-engine FMUs, addressing the limitations of existing methods in modeling long-sequence multivariate data. Firstly, a Composite Health Indicator (CHI) is constructed through semi-supervised learning (SSL), which fuses multi-sensor monitoring data to quantitatively characterize the degradation trend of the FMU throughout its operational lifecycle. Secondly, the proposed isLTransformer model is designed by replacing the feedforward network in traditional iTransformer with a stabilized LSTM module, which maintains the self-attention mechanism's capability to explicitly model dynamic correlations between multiple variables while enhancing the ability to capture nonlinear degradation within individual variables. A physical FMU test bench is designed for the real-world PHM degradation experiments, and the collected dataset was used to demonstrate the effectiveness of the proposed method. Evaluation metrics, including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), are employed to assess the prediction accuracy. The proposed method demonstrates high monotonicity and trend consistency in CHI construction. Compared to the inverted Transformer (iTransformer) and iTransformer- Bi-directional Long Short-Term Memory (BiLSTM), the proposed isLTransformer framework demonstrates significantly reduced prediction errors, validating its superiority in multivariate long-sequence prediction tasks and effectiveness for aero-engine FMU health prediction.

Keywords: aero-engine; fuel metering unit; composite health indicator; state of health prediction; isLTransformer

Check for updates

Academic Editor: Xiaodong Shao

Received: 6 February 2026
Revised: 7 March 2026
Accepted: 11 March 2026
Published: 19 March 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) license.

1. Introduction

As the power unit of the aircraft, an aero-engine constitutes a highly complex and sophisticated aerothermodynamic mechanical system operating under extreme conditions [1]. The primary objective of aero-engine control systems is to ensure stable and reliable operation across varying operational environments. The Fuel Metering Unit (FMU), comprising

critical components such as fuel pumps and fuel regulators, serves as the executive mechanism of aero-engine control systems. Its operational status directly influences aircraft safety, reliability, and economic performance. During long-term operation, the state of health of the FMU will inevitably degrade due to multiple influencing factors, potentially leading to functional failures. Prognostic and Health Management (PHM) technology has been widely applied in automotive, marine, and aerospace industries over recent decades. Unlike traditional maintenance strategies, PHM emphasizes early fault warnings before failure occurrence, thereby reducing catastrophic accident risks while lowering maintenance costs [2]. This gives the motivation for implementing health management for the safety-critical aero-engine FMU to facilitate timely replacement and maintenance before reaching failure thresholds. In order to evaluate the evolutionary process from healthy to failed conditions, a Composite Health Indicator (CHI) is usually introduced to provide a quantitative characterization of the FMU degradation [3]. As an essential component of PHM frameworks, the CHI enables the ultimate objective of developing the prediction models: forecasting the remaining time until the CHI first intersects with the predefined failure thresholds. This intersection point defines the FMU's Remaining Useful Life (RUL), a crucial parameter for predictive maintenance planning [4–6].

The aero-engine FMU exhibits complex nonlinear and strongly coupled relationships among internal subsystems and components [7,8]. Additionally, significant variations exist in initial states and degradation rates across different individual FMUs. Concurrently, with the rapid development of smart manufacturing and advanced sensor technologies, condition monitoring data exhibits trends toward multi-source, high-dimensional, and long-sequenced characteristics [9]. Within this context, degradation monitoring data acquired from single sensors is insufficient to fully characterize the stochastic evolutionary process of the health state of the aero-engine FMU.

In recent years, the PHM based on multi-sensor monitoring data has gradually become a research highlight, with numerous studies based on various data-driven methodologies for CHI construction, RUL prediction, and fault diagnosis [3]. Traditional machine learning algorithms—including Support Vector Regression (SVR) [10], Gradient Boosting Decision Trees [11], and k-Nearest Neighbors [12]—have been applied to RUL prediction, while Principal Component Analysis (PCA) has been utilized for CHI construction [13]. Nevertheless, the prediction accuracy of these approaches is highly dependent on the quality of feature extraction. Given the long sequence and multivariate characteristics, state monitoring data from mechanical equipment such as aero-engine FMUs, conventional feature engineering captures only superficial information patterns. Consequently, traditional machine learning methods are usually proven inadequate in adapting to the degradation patterns of such mechanical equipment, leading to compromised model prediction accuracy [9].

With the development of neural networks and deep learning, prediction models tailored for temporal sequence processing, such as Long Short-Term Memory (LSTM) and Transformer, have drawn widespread attention for RUL prediction and CHI construction. Song et al. [13] employed a hierarchical LSTM framework for remaining useful life prediction of mechanical equipment, addressing the error accumulation problem in long-term life prediction due to neglected short-term state evolution through joint time series forecasting and RUL regression. Guo et al. [14] introduced a novel health indicator based on the hysteresis loop, employing a backlash-like particle swarm optimization model and Gaussian process regression for degradation monitoring and life prediction of flight control systems. Isbilen et al. [15] initially proposed a similarity-based method utilizing autoencoders for temporal feature learning to predict the remaining useful life of aircraft turbofan engines. Building upon this foundation, Isbilen et al. [16] further advanced the approach by integrating multiple machine learning models with a similarity-based framework, creating

health indices from sensor data to enhance prediction accuracy and enable more robust model comparisons. Tseng and Tran [17] studied remaining useful life prediction of mechanical equipment using a novel deep learning model called Attention Long Short-Term Memory Projected, addressing the issues of low utilization efficiency of long sequence dependency information. A study on intelligent health indicators construction for fatigue life prediction of composite material structures was conducted by Moradi et al. [18] using semi-supervised LSTM and ensemble learning methods (particularly Bi-LSTM). The challenge of building reliable health indicators that satisfy monotonicity, predictability, and trendability requirements under complex loading and damage mechanisms was addressed by this approach. Shen and Zhao [19] proposed a fault diagnosis method based on Empirical Mode Decomposition and Gated Recurrent Unit, which effectively addresses the fault classification problem in aircraft hydraulic systems under strong noise environments. Zhiqiang Peng et al. [20] proposed the FC-AMSLSTM method, which employs fault classification and a decline index to decouple multiple fault modes, thereby addressing the decline in prediction accuracy for High-Pressure Compressor degradation faults under multiple operating conditions. He et al. [21] proposed a hybrid model based on a dual-attention mechanism, integrating temporal feature extraction and sensor feature extraction modules to enhance the accuracy of remaining useful life prediction for aircraft engines. Xu et al. [22] proposed a Transformer-KAN-BiLSTM hybrid model, which significantly improves the accuracy and interpretability of aero-engine RUL prediction by fusing long-term dependency modeling, adaptive nonlinear mapping, and bidirectional feature extraction. Lu et al. [23] implemented a bidirectional gated recurrent unit in their investigation of aero-engine RUL prediction. This study successfully tackles the challenge of extracting meaningful features from multi-sensor data where complex operating conditions and failure modes are present. In the meantime, significant improvements in prediction accuracy and stability have been achieved. While demonstrating satisfactory performance, this class of methods faces challenges in parameter sensitivity and model interpretability, with further verification needed regarding industrial applicability and generalization across diverse datasets. Gao et al. [24] conducted research on mechanical system RUL prediction using a customized dual-Transformer framework, addressing two major limitations of existing deep learning methods: the difficulty in detecting the degradation initiation point and the inability to effectively capture multi-cycle degradation patterns. In 2025, Gao et al. [25] investigated degradation-aware RUL prediction for industrial robots using a multi-scale temporal memory Transformer framework. The method addressed two key issues commonly encountered in existing deep learning approaches for RUL prediction: the difficulty in early perception and localization of state transition points from healthy to degraded conditions, and the challenge of effectively mining and utilizing long-, medium-, and short-term multi-scale temporal dependencies for accurate predictions. These methods face limitations in their direct application to mechanical systems such as engine FMUs, which involve coupled multi-sensor signals and complex multi-stage degradation processes.

Owing to the capability of deep learning models such as Convolutional Neural Networks (CNN) and Graph Neural Networks (GNN) to extract critical signal features and automatically capture local spatial characteristics of multivariate parameters (e.g., pressure, flow, and temperature), they are frequently applied to construct hybrid architectures integrated with LSTM and Transformer to enhance prediction accuracy. Ma et al. [26] constructed an Adaptive Graph Convolutional Transformer Encoder, which leverages adaptive graph convolution to learn inter-sensor relationships and a Transformer encoder to capture temporal dependencies, jointly optimizing the prediction of aero-engine RUL. Chen et al. [27] proposed a generative adversarial network with a balanced serial CNN-Transformer architecture to address the limited samples and extreme imbalance in fault

diagnosis. Zhao et al. [28] employed an interpretable deep learning model integrating CNN-LSTM and group-level structured pruning to optimize lithium-ion battery RUL prediction. Their approach quantified critical aging features and reduced redundant data. Huang et al. [29] proposed a PHM framework integrating a Dynamic Probability model (GMM-ADPC) and an LSTM network, achieving deep fusion of fault diagnosis and RUL estimation, thereby enhancing stability and prediction accuracy. Han et al. [30] proposed a dual-attention-guided parallel CNN-LSTM model for turbofan engine RUL prediction. By synchronously extracting spatiotemporal features via parallel architecture and enhancing critical information through feature/temporal dual-attention mechanisms, the approach addressed issues of gradient vanishing and feature loss in traditional serial models, significantly improving prediction accuracy. Zhu and Liu [31] designed a Dual-Dimension Convolutional Attention (DDCA) module integrated with TCN, which effectively captures complex spatiotemporal correlations in multivariate time series data, thereby enhancing RUL prediction performance under multiple operating conditions. Xiong et al. [32] proposed a prognostic model based on a Temporal Convolutional Network (TCN) enhanced with dual-dimensional fusion attention (DDFA) to improve RUL prediction for aero-engines. Wang et al. [33] employed a deep learning ensemble method integrating CNN and Bi-LSTM to research system RUL prediction, which addressed the inefficiency of maintenance strategies under complex operating conditions. Most existing studies treat all sensor data as a single homogeneous temporal signal. However, data points at the same timestep often represent parameters with distinct physical meanings, recorded via different measurement principles. These heterogeneous parameters are embedded into one token, which effectively collapses their multivariate correlations. In this study, we aim to conduct a more in-depth analysis of the physical coupling relationships between parameters such as pressure and temperature, thereby enhancing the representation capability of degradation features and improving prediction accuracy.

Recently, Liu et al. [34] proposed the inverted Transformer (iTransformer) model for multivariate time series forecasting. This approach inverts the conventional Transformer architecture by processing each variable's complete time series as independent tokens through the embedding layer, which captures multivariate correlations by self-attention while employing feedforward networks (FFN) to learn temporal features. While being effective in modeling inter-variable dependencies, the FFN exhibits limitations in characterizing the inherent nonlinear dynamics within individual variables—particularly the temporal degradation patterns of performance parameters. In 2025, Li et al. [35] proposed a hybrid algorithm that combines temporal convolutional networks and iTransformer for estimating aircraft aerodynamic parameters. The model's high prediction accuracy and computational efficiency in complex, noisy environments were validated using real propeller aircraft flight data. Lyu et al. [36] proposed an adaptive BiGRU-ASSA-iTransformer method for predicting the RUL of bearings in aerospace manufacturing, which employs iTransformer to model temporal dynamics of handcrafted time-, frequency-, and statistical-domain features. In 2024, Beck M et al. [37] proposed the extended Long Short-Term Memory (xLSTM) framework, wherein the stabilized LSTM (sLSTM) variant introduces enhanced exponential gating and normalized state mechanisms, and this architecture strengthens temporal modeling capabilities through dynamic memory mixing and multi-head structures. Empirical studies [38,39] have validated that sLSTM significantly improves dynamic memory updating and enhances noise robustness in time series forecasting applications. A complementary relationship exists between the advantages of sLSTM and iTransformer, though no dedicated study has explored this synergy to date.

Although existing studies have achieved significant progress in aero-engine health prediction, several limitations remain: (1) traditional LSTM faces gradient vanishing issues

when processing extremely long sequences, making it difficult to capture long-term dependencies in multi-stage degradation processes; (2) while the self-attention mechanism in iTransformer can model cross-variable correlations, its FFN employs static weights and thus cannot dynamically track the nonlinear degradation trajectory of individual performance parameters; and (3) existing LSTM-Transformer hybrid models suffer from high structural redundancy and require improved computational efficiency. To address these issues, we propose a novel inverted stabilized LSTM-Transformer (isLTransformer) prediction model for the aero-engine FMU health state prediction application. In this architecture, the FFN in the conventional iTransformer is replaced with an sLSTM module. The sLSTM, equipped with exponential gating and normalized state mechanisms, enables adaptive memory control and stable tracking of degradation trajectories, thereby capturing the nonlinear dynamics inherent in FMU performance parameters, such as cumulative wear effects, abrupt cavitation phenomena, and step changes in component friction. Meanwhile, the self-attention mechanism retains its ability to learn dynamic correlations among multiple variables. This hybrid design overcomes the limitations of iTransformer in representing long-term temporal dependencies and abrupt degradation events, making it particularly suitable for the complex multi-stage degradation processes observed in aero-engine FMUs. The main contributions of this paper compared to previous research are as follows:

(1) A novel health state prediction architecture, i.e., inverted stabilized LSTM-Transformer (isLTransformer), is proposed, in which the baseline iTransformer framework effectively captures dynamic correlations between different variables, and an integrated sLSTM module is combined to learn nonlinear temporal features of individual variables.
(2) Experiments are conducted on a physical test bench equipped with real-world aero-engine FMUs to generate the FMU performance degradation dataset, which is used to verify the effectiveness and generalizability of the proposed prediction model. The experimental results demonstrate the superior performance of the proposed isLTransformer-based approach compared to existing state-of-the-art methods.
(3) Effective preprocessing procedures are designed for the prediction task, and an attention-BiLSTM algorithm based on semi-supervised learning is employed to construct the CHI for the engine FMU.

2. Methods

2.1. Problem Formulation

The main fuel flow rate is universally designed as the most important control input for all modern gas turbine aero-engines [40]. The fuel supply of the aircraft engines is governed by the FMU, making it a crucial part of an engine control system. Different working principles of fuel control are presented in Ref. [40]. Among these, the configuration with constant fuel differential pressure and adjustable fuel window is the most widely used and is considered in this paper. Additionally, the FMU discussed in this paper also functions to control the aero-engine vane angle. The fuel that it controls can be simply written as:

$$Q_{fm} = \mu_{fm} A_{fm} \sqrt{\frac{2(p_1 - p_2)}{\rho}} \tag{1}$$

where the subscript *fm* represents the fuel metering valve, which is the key component of FMU; μ_{fm} is the flow coefficient, A_{fm} is the opening of the metering valve, and ρ is the fuel density. p_1 and p_2 are the pressures acting on the inlet and outlet of the metering valve, respectively. The FMU consists of a metering valve, a pressure differential valve, a pressure boosting valve, a safety valve, two metering linear variable differential trans-

former sensors, and two metering electro-hydraulic servo valves, as shown in Figure 1. The degradation of components within an aero-engine fuel metering unit induces system performance deterioration through distinct mechanisms: pressure differential valve degradation (e.g., spring failure or spool sticking) destabilizes the pressure drop across the metering valve, invalidating the fuel metering equation and significantly increasing backpressure sensitivity; metering valve wear or spring degradation directly reduces the flow coefficient and opening accuracy while causing asymmetric opening/closing responses; servo valve actuation force attenuation or response delay restricts maximum metering valve opening and prolongs dynamic response time. The inherent complexity of multi-component systems limits the accuracy of physics-based life prediction models, necessitating data-driven deep learning solutions for reliable prognostications.

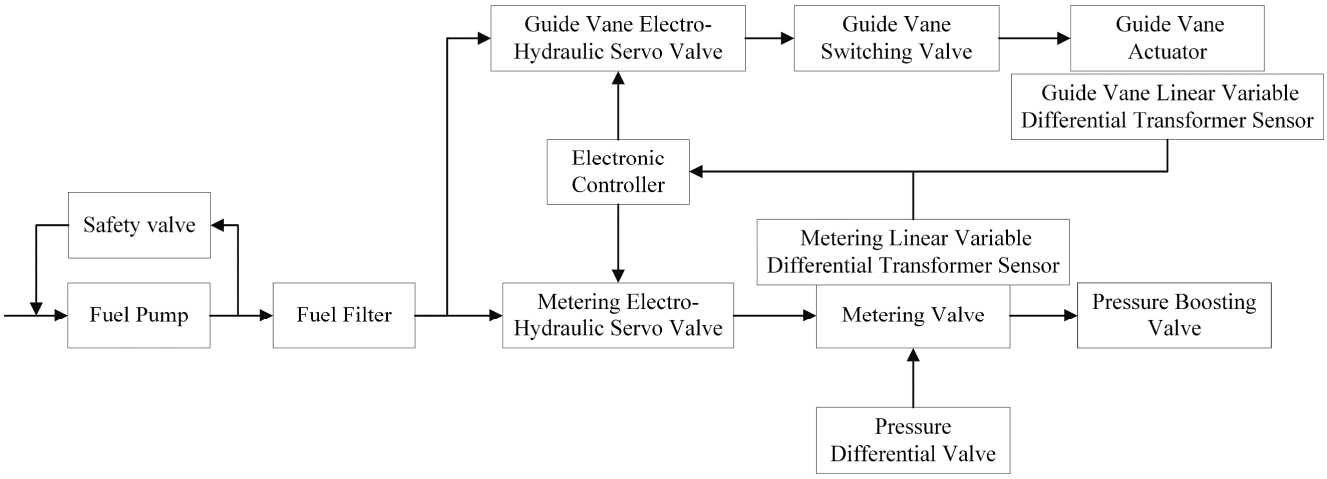


Figure 1. Components of the fuel metering unit.

An aero-engine FMU life test bench was established. Life tests were conducted on five real FMUs following a load spectrum derived from actual flight profiles. The spectrum, statistically summarizing load characteristics across various operational conditions, defined 250 test cycles. During testing, parameters such as inlet temperature/pressure, outlet temperature/flow/pressure, and metering vane Linear Variable Differential Transformer (LVDT) displacement were recorded at different rotational speeds. The prediction objective of this paper is to forecast performance parameters and CHI for future cycles by leveraging historical performance parameters, CHI, and operating parameters.

2.2. Inverted Transformer

The iTransformer [34] is an architecture specifically designed for multivariate time series forecasting with historical observation sequences as input and future sequences as output. This approach inverts the conventional Transformer [41] design by mapping the complete time series of each variable to high-dimensional vectors (termed variable tokens) in the embedding layer, thereby preserving their global temporal characteristics. This design fully leverages the attention mechanism, thereby capturing the correlations among various variables more effectively. This approach has greatly overcome the limitations of the Transformer in multivariate time series prediction that embed variables from the same timestep into mixed tokens, making it difficult to capture correlations between variables [34]. The architecture of iTransformer is depicted in Figure 2. Unlike traditional Transformers that encode multiple variables at the same timestep into a single mixed token, the iTransformer independently encodes the complete time series of each variable as a variable token (as shown in the lower-left part of the figure, where N variables generate N tokens of dimension T). These variable tokens interact through the self-attention layer, with attention weights

reflecting the correlation strength between different variables. In FMU applications, this corresponds to the physical coupling relationships between parameters such as outlet flow rate and outlet pressure. After residual connection, the self-attention output is passed to a shared FFN that applies nonlinear transformations independently to each variable token. In the figure, blue arrows indicate the flow path of variable tokens, while the red box highlights the structural differences from the standard Transformer. However, despite this design, the FFN lacks explicit temporal structure and exhibits insufficient dynamic modeling capability.

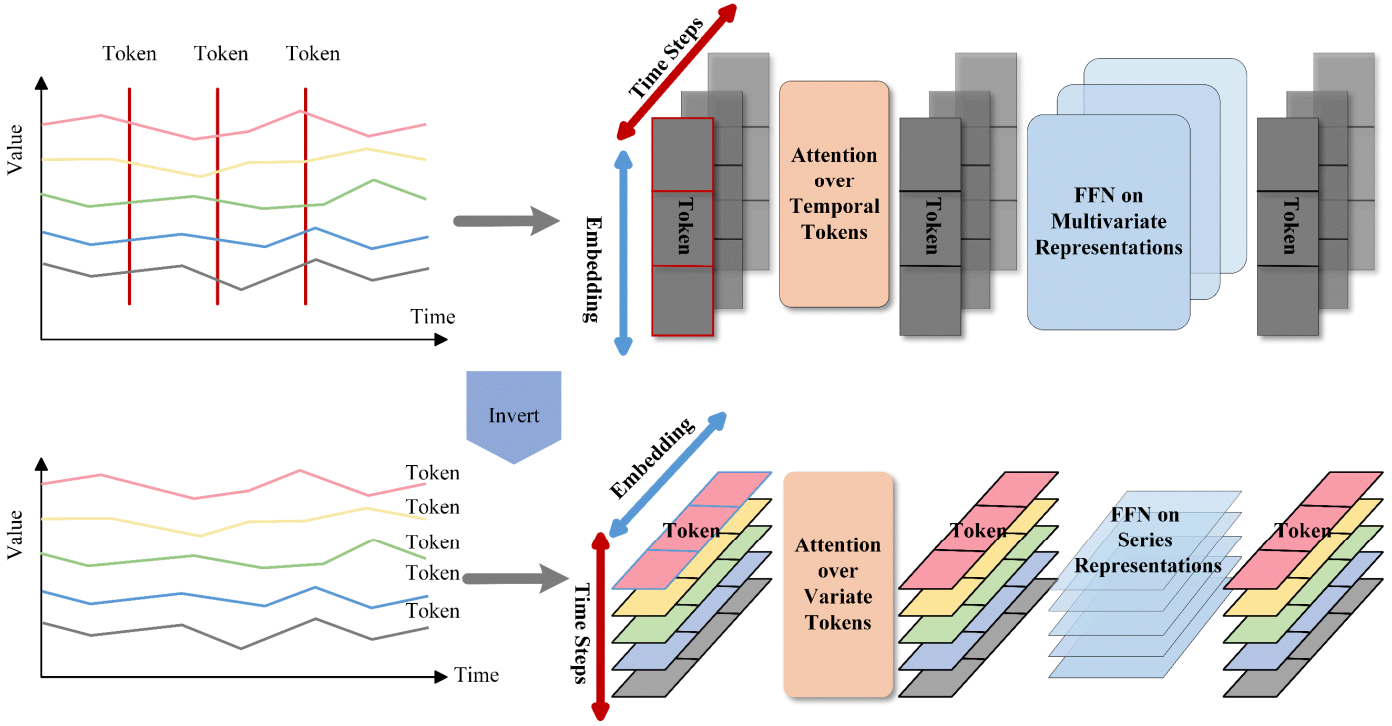


Figure 2. Architecture of iTransformer [34].

2.3. Stabilized LSTM

The sLSTM model is an advanced variant of the traditional LSTM architecture that incorporates exponential gating, memory mixing, and stabilization mechanisms. These enhancements improve the model's ability to process long contexts efficiently, make effective storage decisions, capture complex dependencies, extrapolate accurate sequences, calculate rapidly, and maintain robustness during training and inference. Therefore, sLSTM is particularly suitable for long sequence processing and tasks that require state tracking [37]. The architecture is depicted in Figure 3. The weight vectors $\boldsymbol{w_z}, \boldsymbol{w_i}, \boldsymbol{w_f}$ and $\boldsymbol{w_o}$ correspond to the input weights (connecting inputs $\boldsymbol{x_t}$ to cell input, input gate, forget gate, and output gate). The weight vectors r_z, r_i, r_f and r_o correspond to the recurrent weights (linking hidden states to cell input, input gate, forget gate, and output gate). Bias terms b_z, b_i, b_f and b_o represent corresponding biases. φ denotes the activation function for cell input (typically tanh). While σ serves as the output gate activation, $(\sigma(x) = 1/(1 + \exp(-x))$. Input and forget gates employ exponential activation functions. They can be formulated as:

$$f_t = \exp(\boldsymbol{w_f^T x_t} + r_f h_{t-1} + b_f) \tag{2}$$

$$i_t = \exp(\boldsymbol{w_i^T x_t} + r_i h_{t-1} + b_i) \tag{3}$$

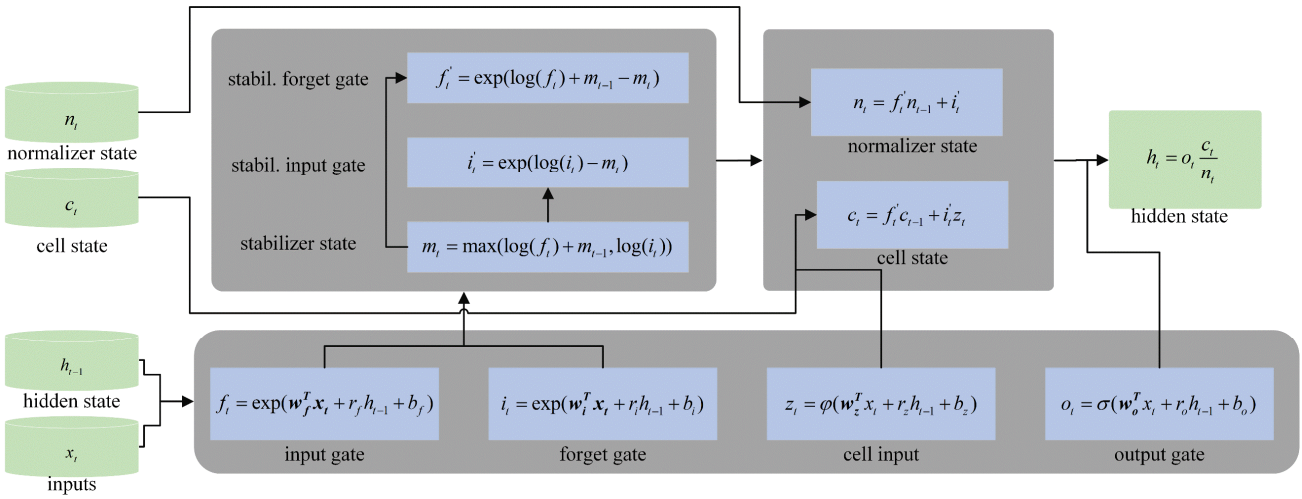


Figure 3. Architecture of sLSTM.

The exponential gate expands the dynamic range of gating mechanisms, enabling rapid adaptation of memory content to better capture abrupt information changes. State normalization n_t constrains the output range of hidden states, mitigating gradient explosion issues and enhancing training stability. It can be represented as:

$$n_t = f_t' n_{t-1} + i_t' \tag{4}$$

Formally, multiple scalar memory cells $c_t \in \mathbb{R}^d$ are organized into vector form [42], through recurrent connections $\boldsymbol{R}_z, \boldsymbol{R}_i, \boldsymbol{R}_f, \boldsymbol{R}_o$—from the hidden state vector $\boldsymbol{h}$ to cell input z, input gate i, forget gate f, and output gate o—multi-head interaction is achieved. Internal mixing within each head is enhanced via exponential gating, while no across-head interaction preserves parallelization potential, thereby increasing model expressiveness.

2.4. Semi-Supervised Learning Based CHI Construction

In actual aero-engine FMU health monitoring, acquiring ground-truth health labels is exceptionally challenging. The CHI is constructed using an attention-based BiLSTM model with semi-supervised learning (SSL) to provide a quantitative characterization of the FMU degradation, as illustrated in Figure 4. While traditional supervised learning relies on labeled data, semi-supervised learning effectively addresses data scarcity by leveraging limited simulated labels alongside abundant unlabeled data [18]. This study establishes target values through physics-informed label simulation, with simulated labels as follows:

$$SL = \sum_{j=1}^{J} w_{t,j} \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

where SL is the simulated label at timestep t (target health indicator during training), $w_{t,j}$ denotes the weight of the j-th performance parameter at this timestep, x_t represents the current value of the performance parameter, x_{min} indicates the failure threshold value of the performance parameter, x_{max} specifies the maximum value of the performance parameter, $w_{t,j}, x_{\max}, x_{\min}$ are typically obtained from technical documentation and expert experience. During training, the semi-mean square error loss function is employed, as shown in Equation (6).

$$L_{regress} = \frac{1}{2N_j} \sum_{i=1}^{N_j} (SL - CHI)^2 \tag{6}$$

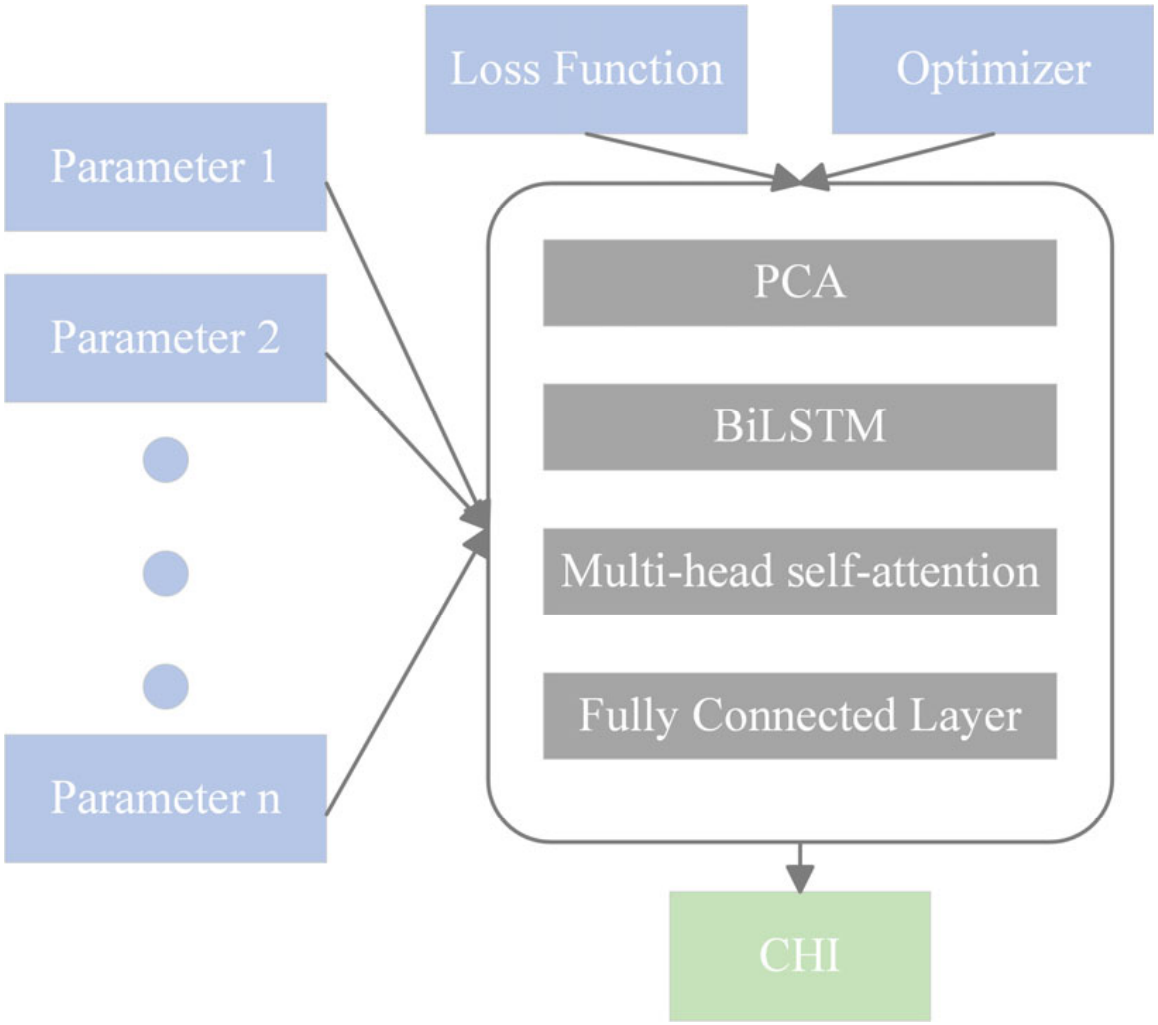


Figure 4. SSL-based CHI construction.

Principal Component Analysis (PCA) reduces feature dimensionality while preserving critical information. The attention-based Bi-LSTM captures dependencies within time series and correlations between variables. The fully connected layer outputs prediction results, constraining CHI within the range [0, 1].

2.5. Inverted Stabilized LSTM-Transformer

To address the challenges of long-range dependency modeling and degradation anomaly detection in high-dimensional sensor time series data for aero-engine FMU state of health prediction, this study proposes an inverted stabilized LSTM-Transformer (isLTransformer) architecture that replaces the traditional FFN and normalization layers in iTransformer with an sLSTM, achieving a refined model of the dynamic degradation characteristics of FMU multiple performance parameters. While the FFN module in conventional iTransformer effectively extracts temporal features via fully connected layers, its static weight mechanism presents challenges in tracking the dynamic temporal evolution of sensor parameters throughout degradation processes—particularly showing reduced responsiveness to late-stage transient performance shifts. The introduction of sLSTM addresses this deficiency. Through exponential gating and normalized state mechanisms, sLSTM constructs dynamic memory units and stabilized states based on the temporal dimension to explicitly record the cumulative effects of FMU historical health states and mutation events (such as cavitation phenomena on pump housing surfaces and

step changes in bearing wear), adaptively balancing short-term fluctuations and long-term degradation trends in sensor parameters, thereby overcoming the long-term dependency loss problems inherent in FFN-based temporal modeling. In the health state prediction of an aero-engine FMU, given historical performance parameters, CHI, and operating parameters $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\} \in \mathbb{R}^{T \times N}$ with T operating cycles and N variates, the objective is to predict M performance parameters and CHI $\boldsymbol{Y} = \{\boldsymbol{x}_{T+1}, \ldots, \boldsymbol{x}_{T+S}\} \in \mathbb{R}^{S \times M}$ $(N > M)$ for future S operating cycles (where $N > M$). Let $\boldsymbol{X}_{t,:}$ denote all performance parameters, CHI, and operating parameters at cycle t, and $\boldsymbol{X}_{:,n}$ represent the complete time series of variable n. The structure of the proposed isLTransformer prediction model is illustrated in Figure 5. The isLTransformer architecture consists of embedding layers, multivariate self-attention layers, normalization layers, sLSTM layers, and projection layers.

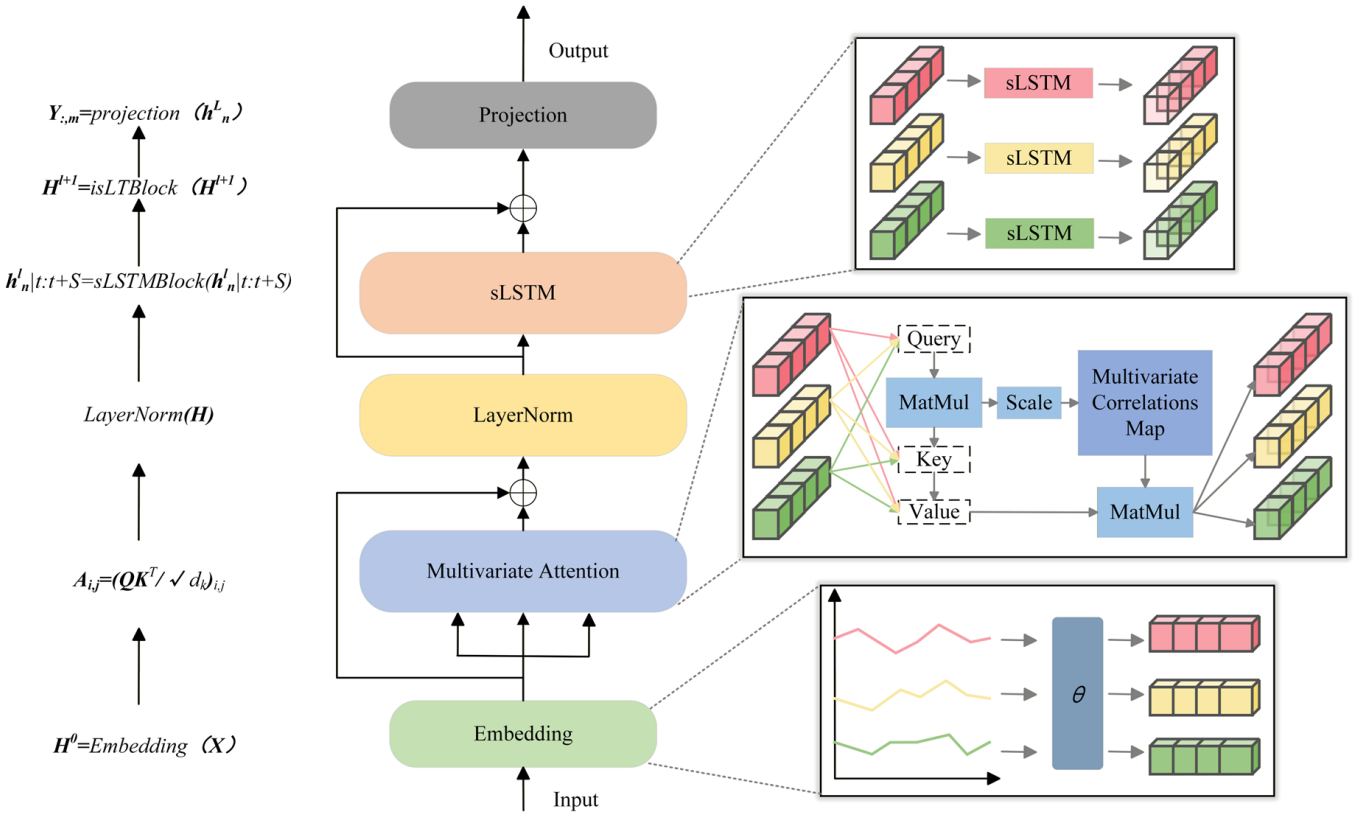


Figure 5. Architecture of isLTransformer.

The operational process of the embedding layer can be formulated as:

$$\begin{aligned} \boldsymbol{h}_{\boldsymbol{n}}^0 &= Embedding(\boldsymbol{X}_{:,\boldsymbol{n}}) \\ \boldsymbol{H}^0 &= Embedding(\boldsymbol{X}) \end{aligned} \tag{7}$$

where $\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots \boldsymbol{h}_N\} \in \mathbb{R}^{N \times D}$ represents N embedding tokens with dimension D, and the superscript indicates the layer index. The embedding process $\mathbb{R}^T \to \mathbb{R}^D$ is implemented by a multilayer perceptron. Variable tokens interact through the self-attention mechanism. After comprehensive representations $\boldsymbol{H}$ for each sequence extracted, the self-attention module obtains queries, keys, and values $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbb{R}^{N \times d_k}$ (where d_k is the projection dimension) through linear projections. The query and key for each variable token are $q_i, k_i \in \mathbb{R}^{d_k}$, and the dot product term of pre-softmax scores can be represented as:

$$\boldsymbol{A}_{i,j} = (\boldsymbol{Q}\boldsymbol{K}^T / \sqrt{d_k})_{i,j} \propto \boldsymbol{q}_i^T \boldsymbol{k}_j \tag{8}$$

Since each token is already normalized in the feature dimension, these dot product terms can reflect correlations between variables. The entire dot product matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ represents the multivariate correlations among variable tokens. Consequently, highly correlated variables receive greater weights in the interaction of values V. The isLTransformer contains only one normalization layer, positioned after the self-attention residual connection, which effectively stabilizes the feature distribution following cross-variable interaction [43,44]. The normalization applied to the single variable sequence is

$$LayerNorm(\boldsymbol{H}) = \left\{ \frac{\boldsymbol{h}_n - Mean(\boldsymbol{h}_n)}{\sqrt{Var(\boldsymbol{h}_n)}} | n = 1, 2, \ldots, N \right\} \tag{9}$$

where $Mean(\boldsymbol{h_n})$ represents the mean of $\boldsymbol{h_n}$, and $Var(\boldsymbol{h_n})$ represents the variance of $\boldsymbol{h_n}$. The sLSTM layer independently operates on the sequence representation of each variable token, extracting features from each input sequence and predicting the representation of future sequences. The working principle of sLSTM has been explained in detail in Section 2.3. Due to its unique exponential gating, normalized states, and memory mixing techniques, this module possesses dynamic storage capabilities and strong training stability, demonstrating distinct advantages in tasks requiring complex state tracking, such as state of health prediction. The representation of its operation on a single variable is

$$\boldsymbol{h_n^l} \Big| t : t + S = sLSTMBlock(\boldsymbol{h_n^l} \Big| t - T : t) \tag{10}$$

The main body of the isLTransformer can be represented as

$$\boldsymbol{H}^{l+1} = isLTBlok(\boldsymbol{H}^l), l = 0, 1, \ldots, L - 1 \tag{11}$$

The prediction output of the isLTransformer is implemented by the projection layer, and its working process can be represented as

$$\hat{\boldsymbol{Y}}_{:,\boldsymbol{m}} = projection(\boldsymbol{h_n^L})(m \in M, n \in N, M < N) \tag{12}$$

and the output layer only outputs the performance parameters and CHI of the aero-engine FMU.

2.6. State of Health Prediction Framework

This study considers a CHI that comprehensively reflects the health state of the FMU through multi-dimensional data fusion and dynamic analysis. The core objective of the aero-engine FMU state of health prediction model is to achieve precise prediction of degradation trends of multiple key performance parameters and CHI variations. The state of health prediction framework is illustrated in Figure 6, comprising three parts: data preprocessing, CHI construction, and health state prediction. Initially, a moving average algorithm eliminates abnormal interferences in sensor signals, while a standardization method eliminates the effects of different dimensions and magnitudes from multi-source sensor data, laying the foundation for cross-dimensional data fusion and feature extraction. As the core representation of the health state, CHI integrates multi-sensor feature parameters through attention-Bi-LSTM based on semi-supervised learning, comprehensively reflecting performance degradation trends and coupling relationships between key parameters. The prediction framework constructs a dual-channel input–output structure of CHI and original sensor data: the CHI channel reflects the macroscopic degradation trend of the overall system health, while the original performance parameter channel provides microscopic dynamic fluctuation information. The two channels are fused at the input layer of the isLTransformer encoder through feature concatenation. Specifically, the CHI sequence and the performance parameter matrix are directly connected along the variable

dimension to form a comprehensive input matrix. This operation integrates the global degradation information embedded in the CHI with the local dynamic fluctuations carried by the raw performance parameters, providing a holistic representation for subsequent temporal modeling. The concatenated features are directly fed into the isLTransformer encoder, allowing the model to autonomously learn the dependencies between the CHI and individual performance parameters through its self-attention mechanism and sLSTM modules. This simple yet effective fusion strategy ensures that both the overall degradation trend and fine-grained parametric variations are jointly considered during prediction, while avoiding unnecessary complexity, thereby enhancing the accuracy and robustness of the prediction results.

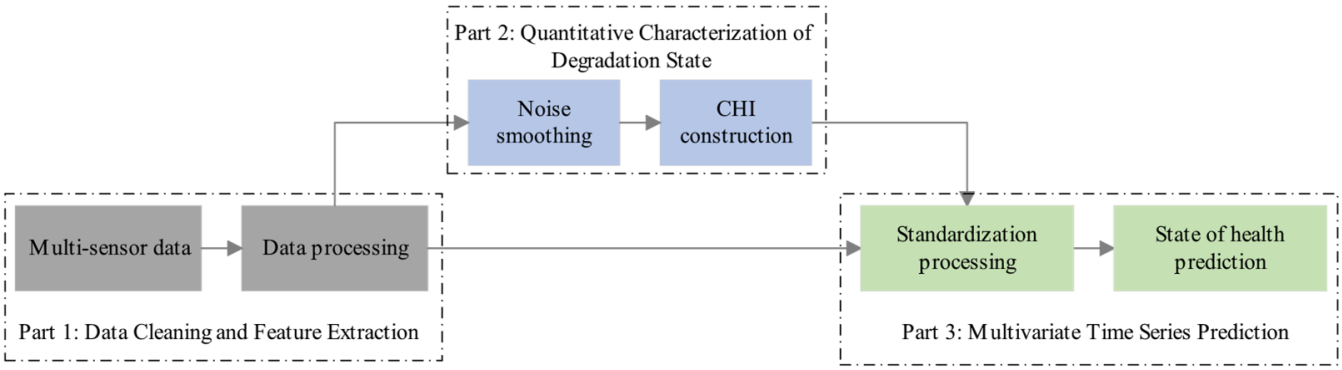


Figure 6. Framework of health state prediction.

3. Experimental Study

The experimental study is conducted to verify the proposed method for state of health prediction. The proposed isLTransformer framework is conducted with Python v3.8.20, Torch 2.0.0, and CUDA 11.8.

3.1. Dataset

A physical experiment test bench is built, and lift tests are implemented on five real aero-engine FMUs. The structure and physical appearance of the FMU life test bench are shown in Figures 7 and 8. The test bench consists of a fuel supply system, device under test (DUT) system, temperature control circulation system, lubrication oil system, and air system. The fuel supply system provides fuel and specified inlet pressure to the FMU under test, the DUT system simulates loads, the temperature control circulation system mainly functions to circulate and filter oil while controlling temperature, the lubrication oil system primarily provides lubrication, and the air system mainly serves to provide a stable source for the FMU. The load spectrum of the life test describes the variations in loads experienced by the FMU during actual operation as flight conditions change. The load spectrum is generated by statistically analyzing and summarizing the amplitude, frequency, and duration of loads, creating a load distribution pattern that characterizes the actual working conditions of the FMU. The simulated flight states of an aircraft include startup, ground idle, flight idle, takeoff hover, climb, level flight, hover, descent, landing hover, flight idle, ground idle, and shutdown. During the performance degradation tests of the aero-engine FMU, two categories of condition monitoring parameters were synchronously recorded: operational parameters and performance parameters. Operational parameters, including rotational speed, inlet pressure, inlet temperature, and outlet temperature, reflect the working conditions of the FMU. Performance parameters, such as outlet flow rate, outlet pressure, metering LVDT difference-to-sum ratio, guide vane LVDT difference-to-sum ratio, and filter circulation pump outlet pressure, directly characterize the functional status and degradation level of the FMU.

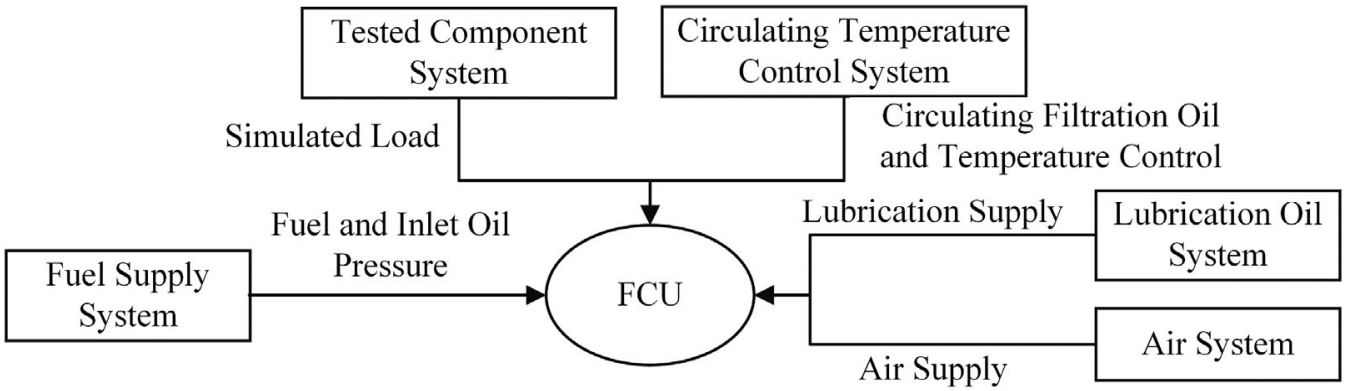


Figure 7. Components of the FMU life test bench.

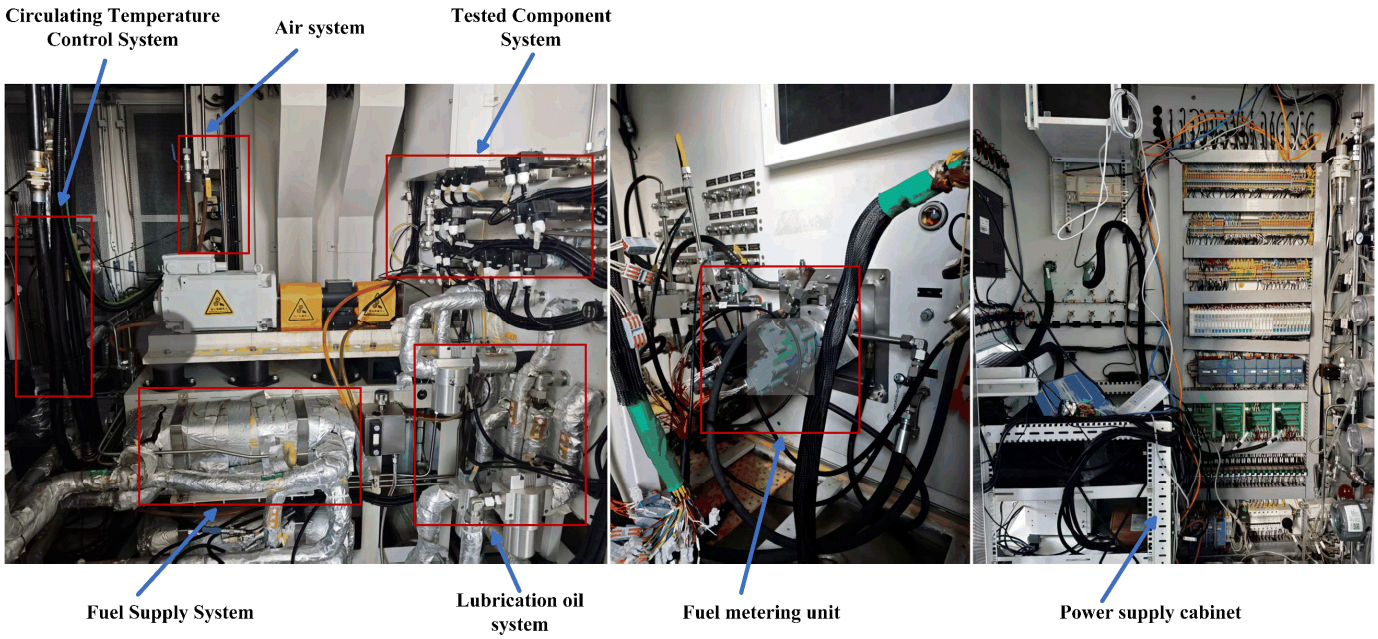


Figure 8. Physical test bench.

The procedure for the performance degradation test is as follows: pre-test inspection → FMU installation → test operation → post-test inspection → disassembly inspection → micrometer measurement → non-destructive testing. Following the above procedure and the prescribed load spectrum, performance degradation tests were conducted on five aero-engine FMUs over 250 cycles each. Although the FMUs did not reach complete failure by the end of the tests, significant performance degradation data were accumulated. All parameters were recorded continuously as time series, resulting in a multivariate, long-sequence performance degradation dataset for the FMUs. This dataset provides physically meaningful input–output samples for constructing composite health indicators and training health state prediction models, ensuring a strong correlation between the data-driven models and the actual physical degradation process.

The degradation test data of real engine FMUs serve as the validation dataset for this research. The dataset records monitoring data from each operational cycle of the aero-engine FMU over 250 cycles from its initial operational cycle. It covers key performance degradation indicators and operational parameter records. However, by the end of the test, the FMU had not yet reached a failure state. The dataset consists of multiple multivariate time series from different FMUs, which means that the data can be viewed as a fleet of the same type of FMUs. Each FMU has different initial wear and manufac-

turing variations, which are considered normal conditions rather than fault conditions. Both operational setting parameters that significantly influence the FMU performance and the performance parameters are included in the dataset. Table 1 presents the condition monitoring parameter information recorded in the dataset. Rotational speed, inlet pressure, inlet temperature, and outlet temperature are the operational setting parameters. The outlet flow rate, outlet pressure, metering LVDT displacement difference-to-sum ratio, guide vane LVDT displacement difference-to-sum ratio, and filter circulation pump outlet pressure represent the FMU performance.

Table 1. Condition monitoring parameters.

Number	Symbol	Parameter	Unit
1	N	rotational speed	r/min
2	P_{in}	inlet pressure	MPa
3	T_{in}	inlet temperature	°C
4	T_{out}	outlet temperature	°C
5	Q_{out}	outlet flow rate	L/h
6	P_{out}	outlet pressure	MPa
7	R_{met}	metering LVDT displacement difference-to-sum ratio	-
8	R_{gv}	guide vane LVDT displacement difference-to-sum ratio	-
9	P_{pump}	filter circulation pump outlet pressure	MPa

3.2. Data Preprocessing

Performance data at the maximum rotational speed in each cycle of the aero-engine FMU were selected to characterize its degradation process. Obvious outliers were removed based on data trends and the actual working process of the FMU, and missing values were filled using interpolation methods. The mean, peak, and peak-to-peak values of T_{out}, Q_{out}, P_{out}, R_{met}, and R_{gv} at the maximum rotational speed are calculated. The final input parameters for the model are shown in Table 2. Gaussian smoothing, as shown in Equation (13), is applied to smooth the original data and highlight degradation trends:

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \tag{13}$$

where σ represents the standard deviation, controlling the degree of Gaussian smoothing (larger σ results in stronger smoothing effects), and x represents the offset from the center point. To determine the optimal σ value for each monitoring parameter, a grid search combined with cross-validation was employed. Candidate σ values were selected from the set {5.0, 7.5, 10.0, 12.5, 15.0, 20.0}. Parameters derived from root-mean-square and peak values generally adopt moderate σ values (7.5–10), while peak-to-peak metrics, being more susceptible to high-frequency fluctuations, require stronger smoothing and thus adopt σ = 20 to effectively attenuate noise without distorting long-term degradation trends. Raw input signals (such as inlet pressure and temperature) adopt σ values ranging from 10 to 20 based on their inherent noise characteristics. This parameter-specific tuning ensures that subsequent feature extraction and CHI construction are based on clean, trend-consistent data.

Table 2. Input parameters.

Number	Symbol	Parameter	Unit
1	$P_{in,m}$	mean inlet pressure	MPa
2	$T_{in,m}$	mean inlet temperature	°C
3	$T_{out,m}$	mean outlet temperature	°C
4	$T_{out,p}$	outlet temperature peak value	°C
5	$T_{out,pp}$	outlet temperature peak-to-peak value	°C
6	$Q_{out,m}$	mean outlet flow rate	L/h
7	$Q_{out,p}$	outlet flow rate peak value	L/h
8	$Q_{out,pp}$	outlet flow rate peak-to-peak value	L/h
9	$P_{out,m}$	mean outlet pressure	MPa
10	$P_{out,p}$	outlet pressure peak value	MPa
11	$P_{out,pp}$	outlet pressure peak-to-peak value	MPa
12	$R_{met,m}$	mean metering LVDT displacement difference-to-sum ratio	-
13	$R_{met,p}$	metering LVDT displacement difference-to-sum ratio peak value	-
14	$R_{met,pp}$	metering LVDT displacement difference-to-sum ratio peak-to-peak value	-
15	$R_{gv,m}$	mean guide vane LVDT displacement difference-to-sum ratio	-
16	$R_{gv,p}$	guide vane LVDT displacement difference-to-sum ratio peak value	-
17	$R_{gv,pp}$	guide vane LVDT displacement difference-to-sum ratio peak-to-peak value	-
18	$P_{pump,m}$	Mean filter circulation pump outlet pressure	MPa
19	C	current cycle period	-

There are significant dimensional differences in measurements from different sensors. Through standardization, as shown in Equation (14), each feature is scaled to a standard normal distribution with mean 0 and standard deviation 1, placing all features on the same scale. Figure 9 shows the standardized performance degradation data and smoothed data from one aero-engine FMU.

$$z = \frac{x - \mu}{\sigma} \tag{14}$$

3.3. Model Training

This study employs a semi-supervised attention-based BiLSTM to construct a CHI for FMU. The model architecture consists of a bidirectional LSTM layer to capture temporal features, an attention mechanism layer to weight critical timesteps, and a fully connected layer for final prediction. The core of model training lies in the construction of sequence samples: the input is a time series segment of historical data processed via sliding time windows and PCA dimensionality reduction (retaining six principal components), with each sample having dimensions of 30 (timesteps) × 6 (feature dimensionality). The corresponding output label (i.e., training target) is not raw sensor data, but a weighted normalized value that integrates multiple key physical parameters (including $Q_{out,m}$, $P_{out,p}$, $R_{met,pp}$, and $R_{gv,pp}$). This value ranges from 0 (complete failure) to 1 (pristine condition), providing the model with a supervisory signal that reflects clear degradation trends. In terms of training configuration, the Half-Mean Squared Error (MSE) is adopted as the loss function, and the Adam optimizer (learning rate = 10^{-3}) is used to train the model for 25 epochs with a batch size of 64, enabling effective learning of degradation patterns from historical sequences. The final output of this model is the CHI, a value ranging continuously from 0 to 1. A CHI value closer to 1 indicates a healthier FMU state.

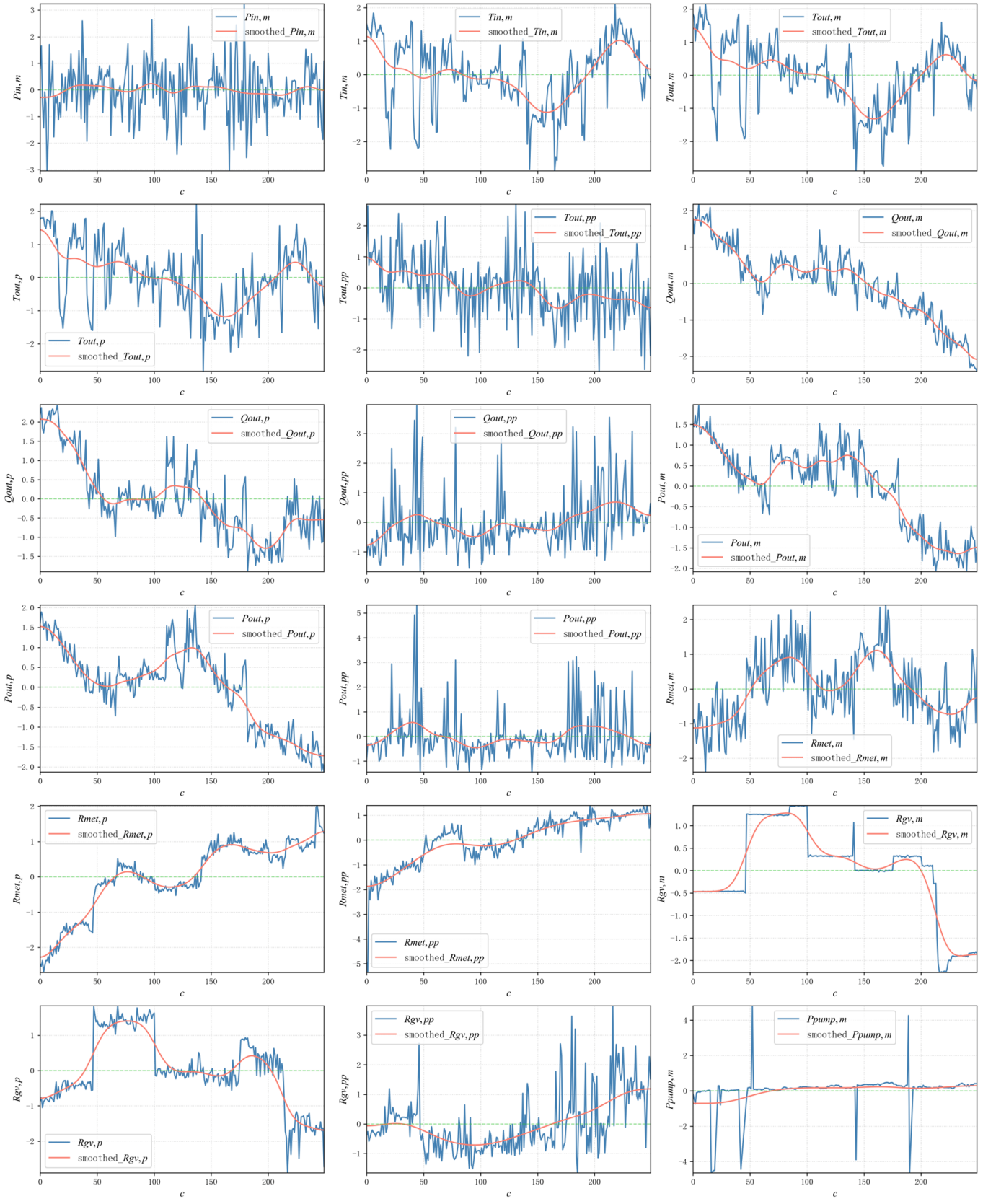


Figure 9. Standardized performance degradation data of a certain FMU.

https://doi.org/10.3390/aerospace13030290

This research strictly divided the dataset into training, validation, and test sets in a 0.7:0.1:0.2 ratio in chronological order to avoid data leakage. Sliding time window lengths were set to {5, 5, 10, 10, 15}, with the window sliding one timestep each time to ensure data continuity, and prediction lengths were set to {1, 5, 5, 10, 10}. The prediction process and detailed settings are shown in Table 3. The parameter settings for isLTransformer include model dimension d_model of 512, fully connected layer dimension d_ff of 512, encoder number e_layers of 2, attention head number n_heads of 8, activation function as gelu function, loss function as mean squared error loss function, Adam optimizer with dynamic learning rate adjustment, initial learning rate lr of 0.0001, and regularization parameter dropout of 0.1. The experiment was conducted on a PC equipped with an Intel Core i7-12700H CPU, 16GB of RAM, an NVIDIA GeForce RTX 3060 GPU, and a 2TB SSD. The dynamic learning rate adjustment strategy follows a step decay schedule, where the learning rate is multiplied by 0.5 at the beginning of each epoch. This exponential decay effectively balances convergence speed and stability during training. To prevent overfitting, early stopping is employed with a patience of 7 epochs. After each epoch, the validation loss is monitored; if it fails to decrease for 7 consecutive epochs, training is terminated, and the best model weights are restored. The training process runs for a maximum of 25 epochs.

Table 3. Prediction process and detailed settings.

Algorithm 1 SSL-based attention-BiLSTM Algorithm 2 isLTransformer	
Require: Input multivariate time series of FMU performance $\boldsymbol{X} \in \mathbb{R}^{T \times N-1}$; input length T; predicted length S; variate number N; token dimension D; and isLTransformer block number L.	
1. Data fusion, construct CHI.	$X \in \mathbb{R}^{T \times N}$
2. $\boldsymbol{X} = \boldsymbol{X}.transpose$	$X \in \mathbb{R}^{N \times T}$
3. MLP embeds series into variate tokens. $\boldsymbol{H}^0 = MLP(\boldsymbol{X})$	$H^0 \in \mathbb{R}^{N \times D}$
4. Run through the isLTransformer blocks.	
5. Self-attention layer is applied to variate tokens.	
6. $\boldsymbol{H}^l = LayerNorm(\boldsymbol{H}^{l-1} + Atten(\boldsymbol{H}^l))$	$\boldsymbol{H}^l \in \mathbb{R}^{N \times D}$
7. SLSTM updates the sequence representation of each token.	
8. End for	
9. $\hat{\boldsymbol{Y}} = MLP(\boldsymbol{H}^L)$	$\hat{\boldsymbol{Y}} \in \mathbb{R}^{M \times S}$
10. $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{Y}}.transpose$	$\hat{\boldsymbol{Y}} \in \mathbb{R}^{S \times M}$
11. Return the prediction result.	

3.4. Metrics

The CHI is used to quantitatively characterize the degradation state of aero-engine FMUs, reflecting their evolution process from healthy to failure states. Its key characteristics include monotonicity (Mo), trendability (Tr) [18], and correlation (Corr) [3]. Mo requires that CHI present an overall increasing or decreasing trend during the FMU degradation phase. Tr refers to the correlation between CHI curves of different engine FMUs, ensuring consistent damage evolution trends across all FMUs, typically validated using the minimum Pearson Correlation Coefficient (PCC) between CHIs of different FMUs, as shown in Equation (15). The PCC range is [−1,1], with a stronger correlation as the absolute value of PCC approaches 1. Corr indicates that CHI is highly related to the actual degradation state

of the equipment, also validated using PCC, with the PCC for a certain sensor monitoring data of one FMU shown in Equation (16).

$$Tr = \min \left| \frac{\sum_{c=1}^{C} (x_{m,c} - \overline{x}_m)(x_{n,c} - \overline{x}_n)}{\sqrt{\sum_{c=1}^{C} (x_{m,c} - \overline{x}_m)^2} \sqrt{\sum_{c=1}^{C} (x_{n,c} - \overline{x}_n)^2}} \right| \tag{15}$$

where $x_{m,c}, x_{n,c}$ represent the CHI values of two different FMUs at cycle c, and C is the total number of operational cycles of the FMU.

$$r_{i,j} = \frac{\sum_{k=0}^{K_i} (c_{i,k} - \overline{c}_i)(x_{i,j}(c_{i,k}) - \overline{x}_{i,j})}{\sqrt{\sum_{k=0}^{K_i} (c_{i,k} - \overline{c}_i)^2} \sqrt{\sum_{k=0}^{K_i} (x_{i,j}(c_{i,k}) - \overline{x}_{i,j})^2}} \tag{16}$$

where $c_i = \{c_{i,0}, c_{i,1}, c_{i,2}, \ldots, c_{i,K_i}\}$ represents the monitoring cycles corresponding to the i-th $(1 \leq i \leq P)$ FMU, $x_{i,j}(c)$ is the performance degradation monitoring data collected by the j-th $(1 \leq j \leq N)$ sensor of the i-th FMU at cycle c, and $K_i + 1$ is the number of monitoring data points for the i-th FMU.

The state of health prediction of mechanical equipment is usually regarded as a regression problem. One of the evaluation metrics used for aero-engine FMU health state prediction is Root Mean Squared Error (RMSE), as shown in Equation (17). RMSE is the square root of the mean squared error between observed and predicted values, amplifying larger errors and exhibiting high sensitivity to outliers.

$$RMSE = \sqrt{\frac{\sum_{q=1}^{Q} (\hat{\boldsymbol{Y}}_q - \boldsymbol{Y}_q)^2}{Q}} \tag{17}$$

where Q is the total number of samples, $\boldsymbol{Y}_q$ is the actual health state of the q-th sample, and $\hat{\boldsymbol{Y}}_q$ is the predicted health state of the q-th sample. The second evaluation metric for FMU health state prediction is Mean Absolute Error (MAE), as shown in Equation (18). MAE computes the mean absolute error with linear weighting, demonstrating greater robustness against outliers.

$$MAE = \frac{1}{Q} \sum_{q=1}^{Q} \left| \boldsymbol{Y}_q - \overline{\boldsymbol{Y}}_q \right| \tag{18}$$

In the practical application of FMU health prediction, the numerical differences between RMSE and MAE carry distinct implications for maintenance decision-making. RMSE assigns higher weights to larger errors and is therefore more sensitive to significant individual deviations in predictions, whereas MAE averages all errors linearly, providing a more robust reflection of overall prediction accuracy. When RMSE is significantly higher than MAE, it indicates that the model achieves high prediction accuracy at most time points but exhibits substantial deviations at a few critical moments. These moments often correspond to sudden degradation phases in FMU performance, such as early warning signs like accelerated wear of components or anomalies in control response. In engineering practice, accurately capturing these abrupt transition points holds significant early warning value for preventive maintenance. If performance turning points can be identified in advance, maintenance personnel can schedule targeted inspections or component replacements in a timely manner, thereby avoiding mission interruptions or safety risks caused by unexpected failures. Conversely, when RMSE and MAE are close, it suggests a relatively smooth

degradation process where the model can stably track the overall trend, making it suitable for routine condition monitoring. Therefore, by jointly analyzing the differences between RMSE and MAE, it is possible not only to evaluate the overall accuracy of prediction models but also to assess their capability to capture degradation transition events, providing more targeted insights for optimizing maintenance strategies.

4. Results and Discussion

4.1. Composite Health Indicator (CHI)

The degradation stage characteristics of the considered CHI are shown in Figure 10. After integrating multi-sensor data, the one-dimensional CHI exhibits an overall monotonically decreasing trend, effectively characterizing the degradation process of the engine FMU. The PCC values between CHIs of different FMUs in the same model are displayed in Table 4, in which all the absolute PCC values exceed 0.98. This indicates that the Tr of CHIs is 0.98, and the CHI curves of all FMUs show similarity in both trends and pattern changes, which qualify them for group health monitoring. Table 5 shows the PCC values between multiple parameters (including single sensors and CHI) and cycle periods of each FMU. The absolute PCC values of different sensors range from 0.21 to 0.89, indicating their limited correlation with the degradation states. The integrated CHI, through Bi-LSTM with the attention mechanism based on semi-supervised learning, achieves efficient fusion of multi-sensor data, with absolute PCC values all exceeding 0.93.

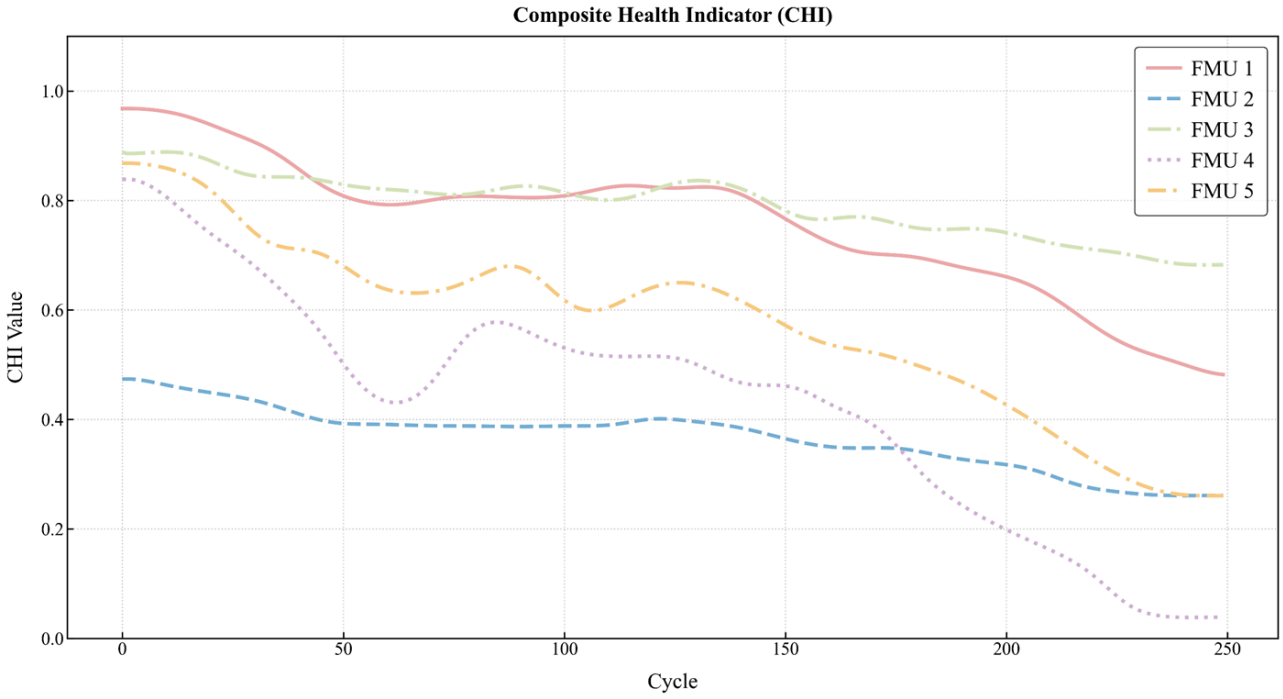


Figure 10. Composite health indicators of the five FMU.

Table 4. PCC values of CHIs from different FMUs.

PCC	FMU1	FMU2	FMU3	FMU4	FMU5
FMU1	1	0.99	0.98	0.99	0.99
FMU2	-	1	0.99	0.99	0.99
FMU3	-	-	1	0.98	0.99
FMU4	-	-	-	1	0.99
FMU5	-	-	-	-	1

Table 5. PCC values between multiple parameters and cycles of five FMUs.

PCC	CHI	$Q_{out,m}$	$P_{out,p}$	$R_{met,pp}$	$R_{gv,pp}$
C_{FMU1}	−0.94	−0.88	−0.76	0.84	0.40
C_{FMU2}	−0.93	−0.81	−0.43	0.47	0.21
C_{FMU3}	−0.94	−0.62	−0.50	0.69	0.44
C_{FMU4}	−0.93	−0.89	−0.57	0.43	0.26
$C_{\mathbf{FMU5}}$	−0.95	−0.87	−0.78	0.83	0.54

4.2. Prediction Performance

This section validates the prediction performance of the proposed isLTransformer on the aero-engine FMU performance degradation dataset. The experiments cover both long-term and short-term prediction tasks, verifying the effectiveness of isLTransformer. Figure 11 shows the comparison between the predicted and actual CHI values in the test sets of five FMUs with different initial health states. The predicted values track the actual values precisely, which indicates the effectiveness of the prediction architecture under different initial health conditions. Figures 12–14 show the comparison between predicted and actual values of key parameters in three test sets of these FMUs when the time window length is 5 and the prediction length is 1. The proposed prediction model demonstrates good prediction throughout the entire operational process of the FMUs in the test set. Figures 15–17 show the comparison between predicted and actual values of key parameters in three test sets of these FMUs when the time window length is 10 and the prediction length is 10. The proposed model maintained good predictions throughout the operation of the FMU test set. However, the prediction accuracy was lower than that when the time window length was 5 and the prediction length was 1. Table 6 gives the total prediction errors of the architecture under different time window lengths (WL), prediction lengths (PL), and batch sizes. The RMSE ranges from 0.446 to 0.760, and the MAE ranges from 0.176 to 0.415, with overall small errors. When the sliding time window length increases, prediction errors decrease; when the prediction length decreases, prediction errors increase.

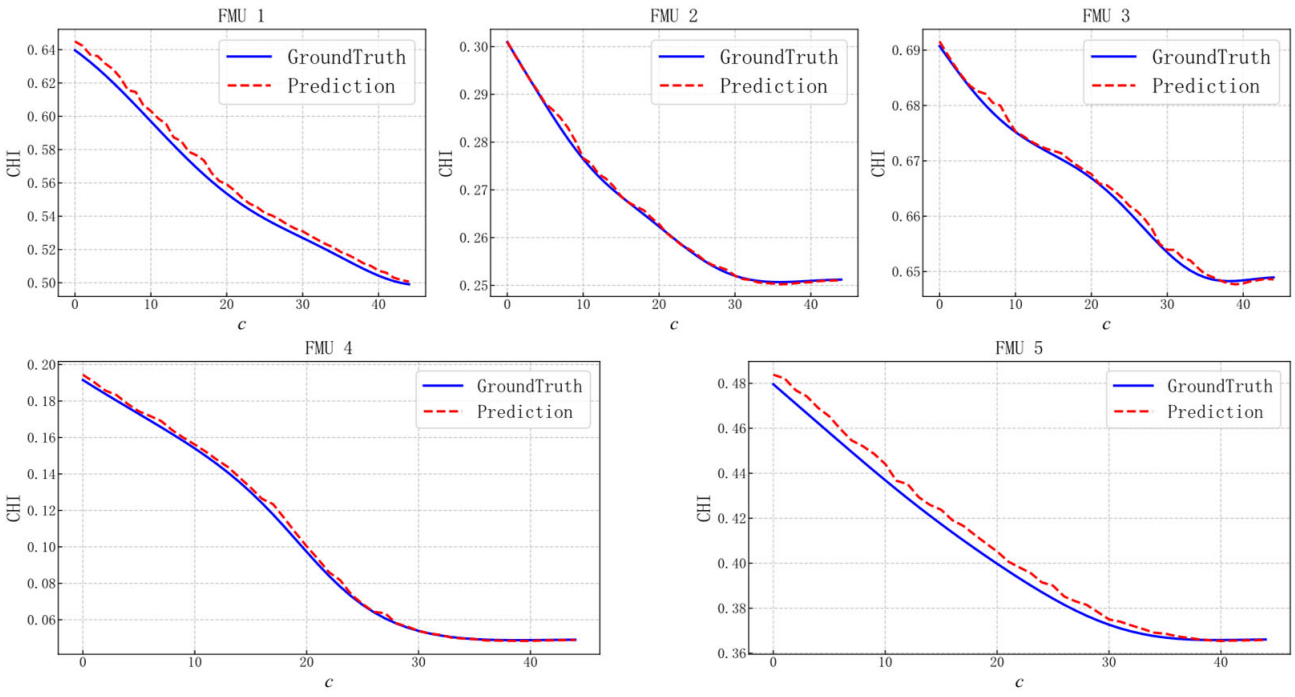


Figure 11. Comparison between actual and predicted CHI values of five FMU.

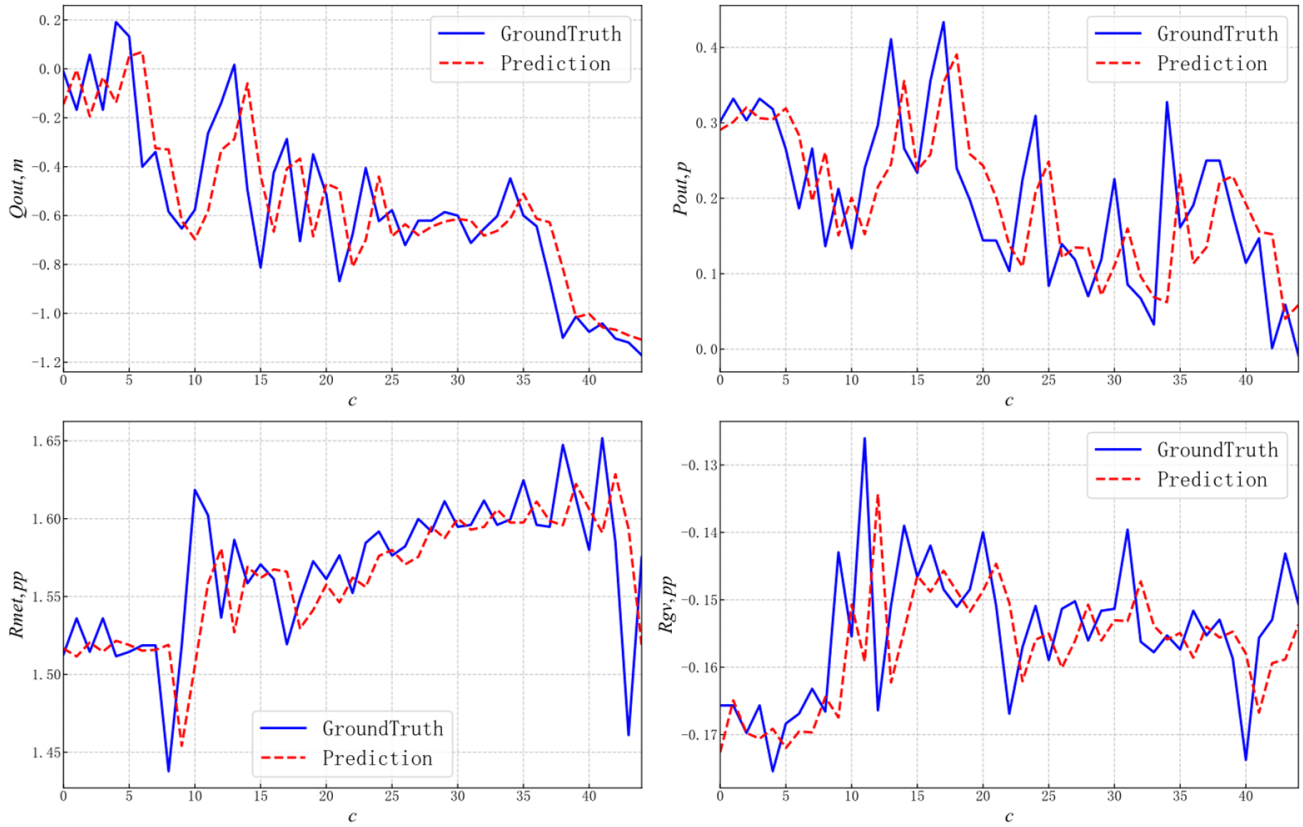


Figure 12. Prediction results for FMU No. 1 (window length = 5, prediction length = 1).

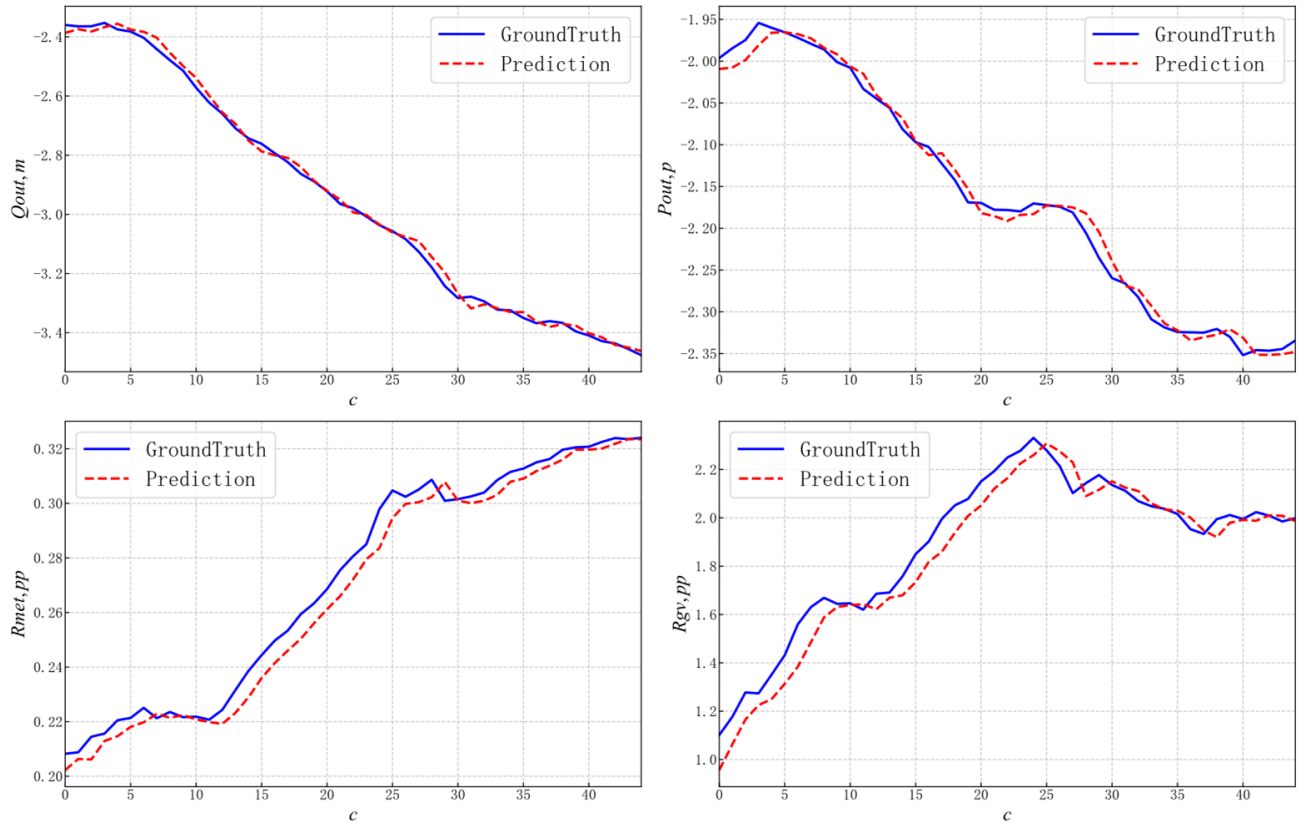


Figure 13. Prediction results for FMU No. 4 (window length = 5, prediction length = 1).

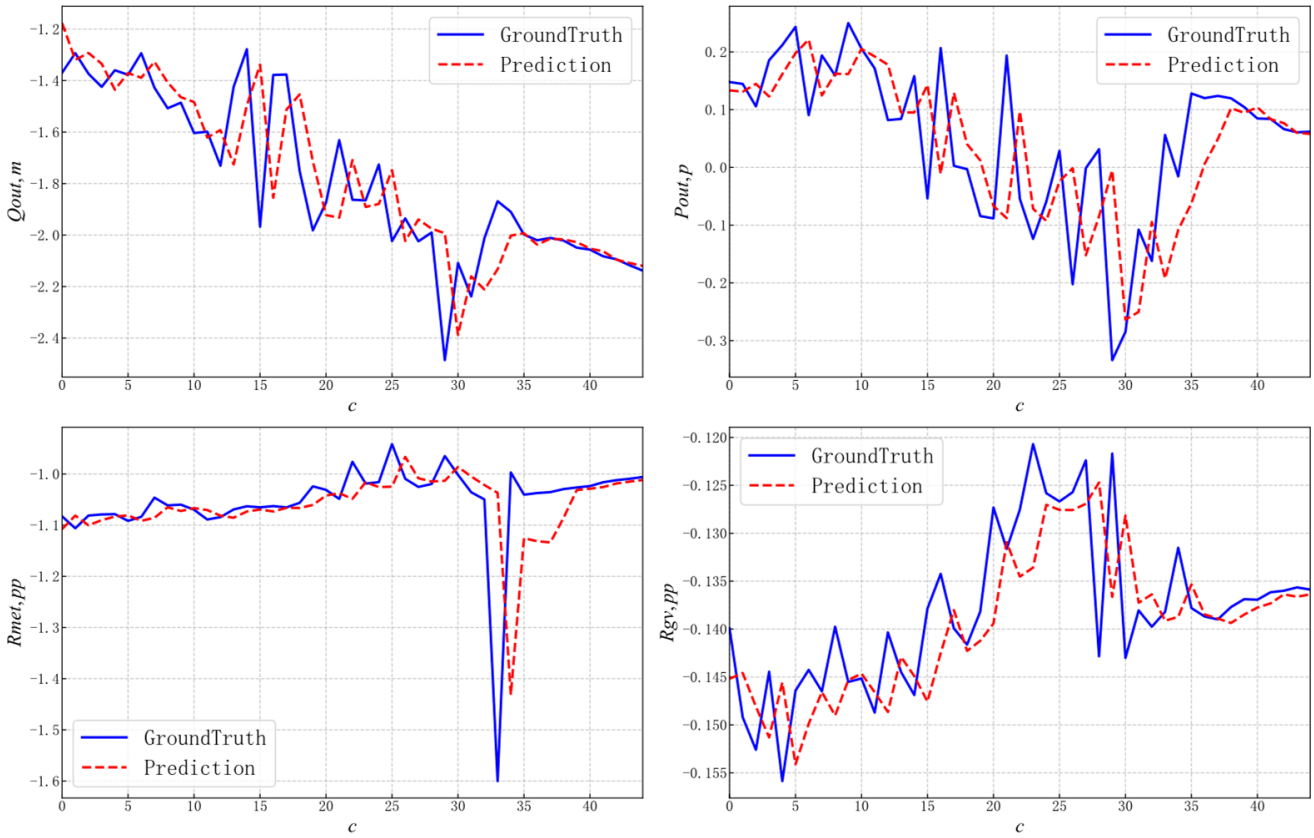


Figure 14. Prediction results for FMU No. 5 (window length = 5, prediction length = 1).

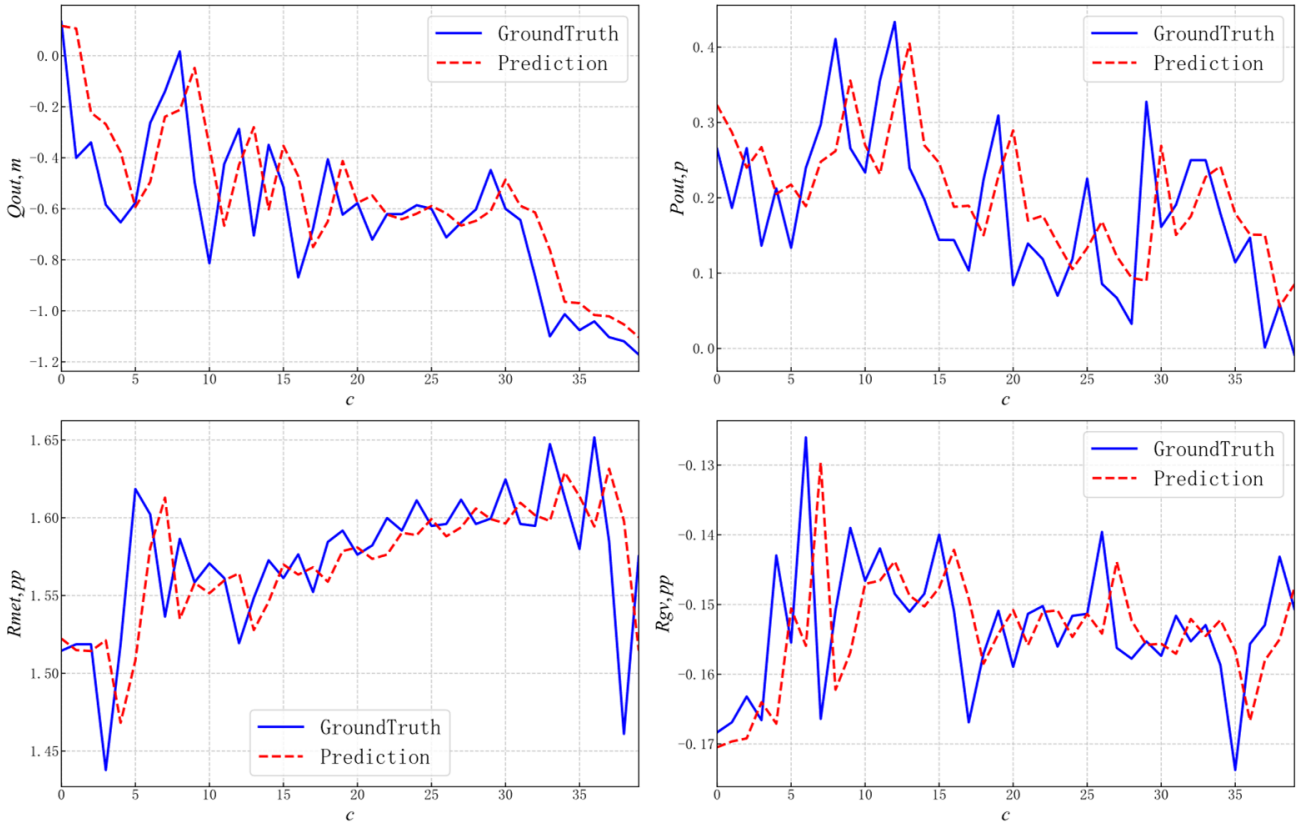


Figure 15. Prediction results for FMU No. 1 (window length = 10, prediction length = 10).

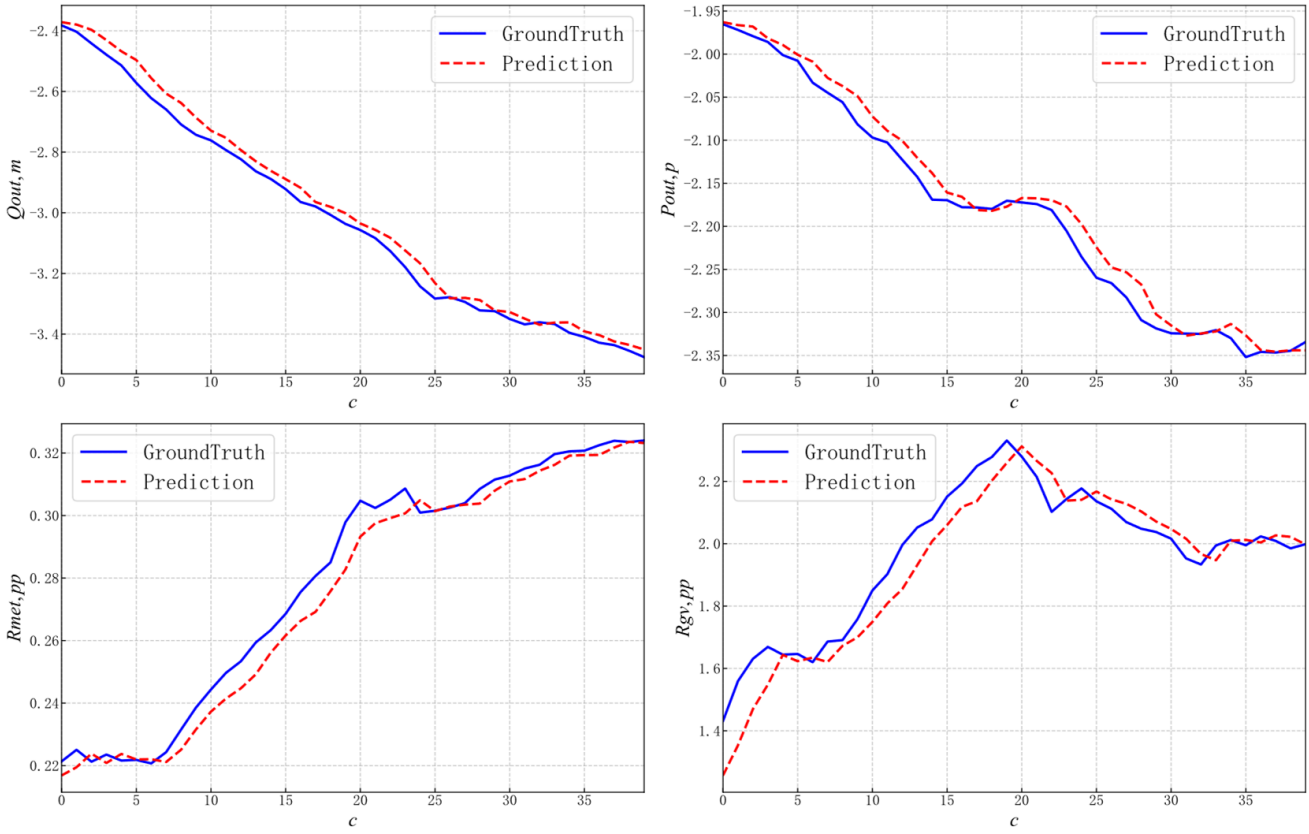


Figure 16. Prediction results for FMU No. 4 (window length = 10, prediction length = 10).

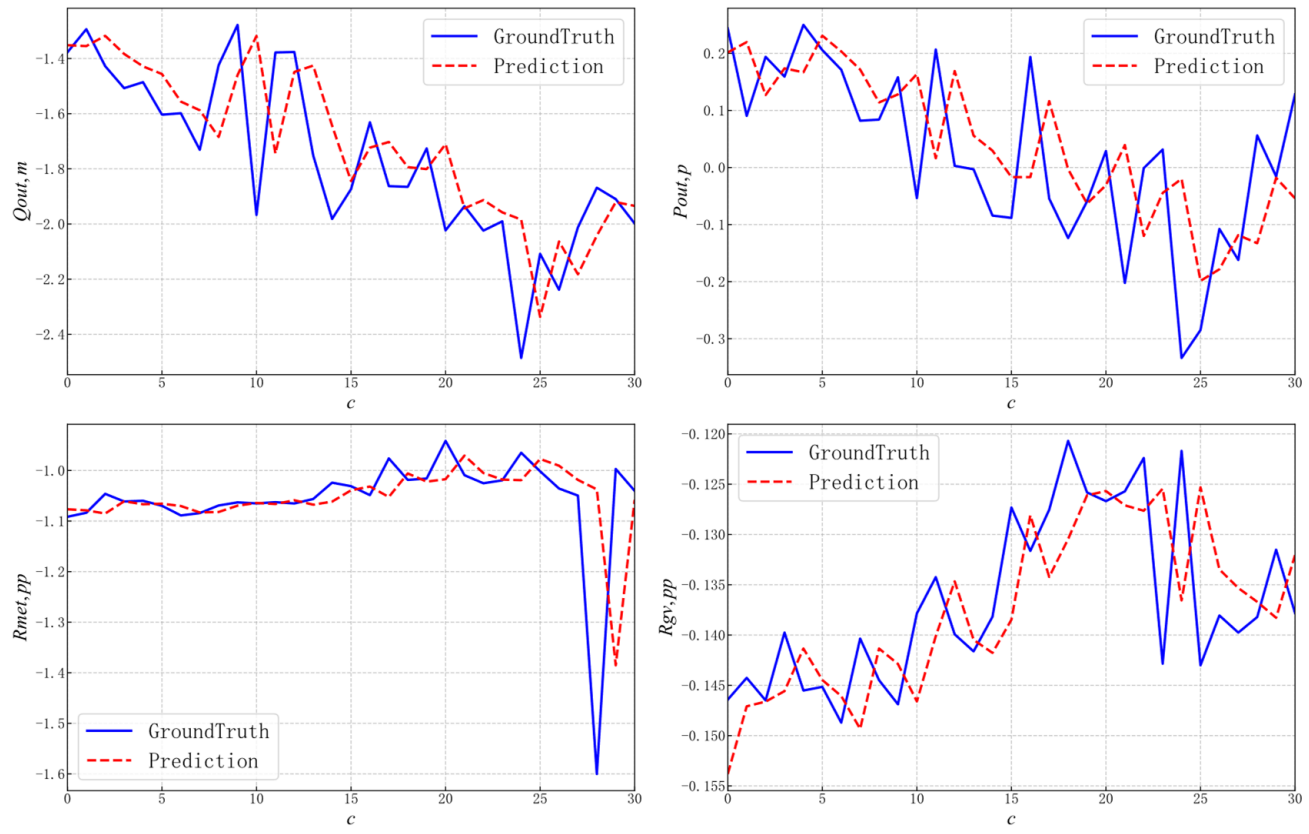


Figure 17. Prediction results for FMU No. 5 (window length = 10, prediction length = 10).

Table 6. Errors under different time window lengths (WL), prediction lengths (PL), and batch sizes.

WL-PL	Batch Size	RMSE	MAE
5-1	1	0.446	0.176
	3	0.451	0.176
	5	0.461	0.190
5-5	1	0.660	0.302
	3	0.653	0.296
	5	0.656	0.298
10-5	1	0.628	0.298
	3	0.632	0.302
	5	0.630	0.306
10-10	1	0.750	0.393
	3	0.750	0.400
	5	0.760	0.415
15-15	1	0.742	0.402
	3	0.743	0.407
	5	0.743	0.407

4.3. Comparative Experiments

The health state prediction performance of the proposed isLTransformer is compared to that of the state-of-the-art iTransformer and iTransformer-BiLSTM models, and the prediction errors on the aero-engine FMU dataset of these three models are summarized in Table 7. Under most configurations of time windows and prediction lengths, isLTransformer achieves the best performance with the prediction errors of RMSE and MAE. Compared with the iTransformer, the prediction RMSE of the proposed isLTransformer decreases by an average of 1%, and MAE decreases by an average of 4%. Similarly, an average decrease of 1.2% on the RMSE and 3.5% on the MAE can be observed when comparing the proposed method to the iTransformer-BiLSTM.

Table 7. Comparison of prediction errors across different architectures.

WL-PL	Models	RMSE	MAE
5-1	isLTransformer	0.451	0.176
	iTransformer	0.458	0.189
	iTransformer-BiLSTM	0.451	0.181
5-5	isLTransformer	0.660	0.302
	iTransformer	0.658	0.302
	iTransformer-BiLSTM	0.656	0.297
10-5	isLTransformer	0.628	0.298
	iTransformer	0.641	0.321
	iTransformer-BiLSTM	0.633	0.312
10-10	isLTransformer	0.750	0.393
	iTransformer	0.757	0.413
	iTransformer-BiLSTM	0.770	0.427
15-15	isLTransformer	0.742	0.402
	iTransformer	0.752	0.414
	iTransformer-BiLSTM	0.761	0.412

4.4. Ablation Experiments

iTransformer contains two normalized layers. The first one is positioned after the self-attention residual connection, responsible for stabilizing feature distributions after cross-

variable interactions. The second layer is located at the end of the FFN residual connection, controlling feature scaling after temporal dynamic modeling. Since the sLSTM component inherently possesses temporal dimension normalization characteristics, isLTransformer retains only a single normalized layer. Table 8 demonstrates the prediction accuracy and training speed of isLTransformer with single-layer and dual-layer normalization (LayerNorm). The prediction errors (RMSE and MAE) of the two architectures exhibit minimal differences, but the simplified architecture with single-layer normalization significantly reduces the computational complexity by shortening the average training duration by 18.66% and improving inference speed.

Table 8. Prediction errors and training speed with different numbers of LayerNorm.

WL-PL	LayerNorm	RMSE	MAE	Training Time(s)
5-1	1	0.451	0.176	1362
	2	0.448	0.178	1506
5-5	1	0.660	0.302	1331
	2	0.660	0.295	2400
10-5	1	0.628	0.298	2606
	2	0.629	0.297	2677
10-10	1	0.750	0.393	2574
	2	0.751	0.395	2677
15-15	1	0.742	0.402	1709
	2	0.741	0.401	2539

5. Conclusions

This paper proposed a novel inverted stabilized LSTM-Transformer (isLTransformer) architecture for the health state prediction of the aero-engine FMU. The real-world tests of FMUs on a physical test bench are conducted to generate a performance degradation dataset, and the CHI of the FMU based on an attention Bi-LSTM architecture is constructed and validated from the dataset. The health state of FMUs is comprehensively characterized from both overall and local perspectives through CHI and multiple FMU performance parameters. The isLTransformer architecture is proposed and applied to the prediction of the state of health of the aero-engine FMUs. This model explicitly captures dynamic correlations between multiple variables (such as the coupling between outlet flow rate and pressure) through self-attention, while capturing long and short-term dependencies in time series through sLSTM. Prediction test performance based on experiment data demonstrates that isLTransformer achieves performance improvements in both prediction accuracy and computational efficiency compared with other state-of-the-art iTransformer and iTransformer-BiLSTM. The proposed method provides theoretical support and technical solutions for the health state prediction of PHM in the aviation field. In the future, we aim to extend this hybrid architecture to a wider range of mechanical equipment PHM tasks.

Author Contributions: Conceptualization, Y.H. and X.W.; methodology, Y.H.; software, X.W.; validation, X.W.; formal analysis, Y.H.; investigation, X.W.; resources, J.L.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.W., Y.H., L.G. and J.L.; visualization, X.W.; supervision, L.G.; project administration, L.G.; funding acquisition, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study is co-supported by the National Natural Science Foundation of China (No. 52305036) and Guangdong Basic and Applied Basic Research Foundation (No. 2022A1515110005).

Data Availability Statement: The data that support the findings of this study are available upon reasonable request from the corresponding author.

Conflicts of Interest: Author Junwei Li was employed by the Beijing Hangke Engine Control System Technology Co., Ltd., AEEC. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

FMU	Fuel Metering Unit
PHM	Prognostic and Health Management
CHI	Composite Health Indicator
SSL	Semi-supervised Learning
iTransformer	inverted Transformer
BiLSTM	Bi-directional Long Short-Term Memory
RUL	Remaining Useful Life
SVR	Support Vector Regression
PCA	Principal Component Analysis
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
GNN	Graph Neural Network
FFN	Feedforward Network
sLSTM	stabilized LSTM
isLTransformer	inverted stabilized LSTM-Transformer
LVDT	Linear Variable Differential Transformer
DUT	Device Under Test
MSE	Mean Squared Error
Mo	Monotonicity
Tr	Trendability
Corr	Correlation
PCC	Pearson Correlation Coefficient
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
WL	Window Lengths
PL	Prediction Lengths

References

1. Liu, M.R.; Sun, T.; Sun, X.M. Brain-inspired spike echo state network dynamics for aero-engine intelligent fault prediction. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3516313. [CrossRef]
2. Javed, K.; Gouriveau, R.; Zerhouni, N. State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *Mech. Syst. Signal Process.* **2017**, *94*, 214–236. [CrossRef]
3. Yang, X.Q.; Gao, X.Q.; Zheng, H.Y.; Yang, M.S.; Liu, Y. A hybrid prognosis method based on health indicator and wiener process: The case of multi-sensor monitored aero-engine. *Eng. Appl. Artif. Intell.* **2025**, *144*, 110099. [CrossRef]
4. Sikorska, J.Z.; Hodkiewicz, M.; Ma, L. Prognostic modelling options for remaining useful life estimation by industry. *Mech. Syst. Signal Process.* **2011**, *25*, 1803–1836. [CrossRef]
5. Li, T.Z.; Chen, J.; Yuan, S.F.; Cadini, F.; Sbarufatti, C. Particle filter-based damage prognosis using online feature fusion and selection. *Mech. Syst. Signal Process.* **2023**, *203*, 110713. [CrossRef]
6. Guo, L.; Li, N.P.; Jia, F.; Lei, Y.G.; Lin, J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **2017**, *240*, 98–109. [CrossRef]
7. Saleh, A.; Montazeri-Gh, M. Black box modeling of a turboshaft gas turbine engine fuel control unit based on neural NARX. *Proc. Inst. Mech. Eng. Part M-J. Eng. Marit. Environ.* **2019**, *233*, 949–956. [CrossRef]
8. Yuan, Y.; Zhang, T.H.; Lin, Z.L.; Zhang, J.M. An investigation into factors determining the metering performance of a fuel control unit in an aero engine. *Flow Meas. Instrum.* **2020**, *71*, 101672. [CrossRef]

9. Zhao, K.; Jia, Z.; Jia, F.; Shao, H.D. Multi-scale integrated deep self-attention network for predicting remaining useful life of aero-engine. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105860. [CrossRef]
10. Khelif, R.; Chebel-Morello, B.; Malinowski, S.; Laajili, E.; Fnaiech, F.; Zerhouni, N. Direct remaining useful life estimation based on support vector regression. *IEEE Trans. Ind. Electron.* **2017**, *64*, 2276–2285. [CrossRef]
11. Deng, K.Y.; Zhang, X.Y.; Cheng, Y.J.; Zheng, Z.Y.; Jiang, F. A remaining useful life prediction method with long-short term feature processing for aircraft engines. *Appl. Soft Comput.* **2020**, *93*, 106344. [CrossRef]
12. Zhao, L.J.; Xie, T.; Wei, Y.H.; Liu, Y.J.; Qin, Y. Overview of digital twin-driven rotating machinery fault diagnosis: Status and trends. *Meas. Sci. Technol.* **2025**, *36*, 052001. [CrossRef]
13. Song, T.; Liu, C.; Wu, R.; Jin, Y.F.; Jiang, D.X. A hierarchical scheme for remaining useful life prediction with long short-term memory networks. *Neurocomputing* **2022**, *487*, 22–33. [CrossRef]
14. Guo, Y.H.; Ma, C.B.; Dong, X.; Liang, Y.; Hu, B.Y. A novel health indicator based on hysteresis loop for health prediction of flight control systems. *Measurement* **2021**, *186*, 110076. [CrossRef]
15. Isbilen, F.; Bektas, O.; Konar, M. Similarity-Based Prediction of Remaining Useful Life for Aircraft Turbofan Engines Using Auto Encoders. In Proceedings of the 12th Ankara International Aerospace Conference (AIAC 2023), Online, Ankara, Turkey, 13–15 September 2023; pp. 1–11.
16. Isbilen, F.; Bektas, O.; Avsar, R.; Konar, M. Improved machine learning models with a similarity-based approach for remaining useful life prediction. *Aeronaut. J.* **2025**, *129*, 485–505. [CrossRef]
17. Tseng, S.H.; Tran, K.D. Predicting maintenance through an attention long short-term memory projected model. *J. Intell. Manuf.* **2024**, *35*, 807–824. [CrossRef]
18. Moradi, M.; Galanopoulos, G.; Kuiters, T.; Zarouchas, D. A novel intelligent health indicator using acoustic waves: CEEMDAN-driven semi-supervised ensemble deep learning. *Mech. Syst. Signal Process.* **2025**, *224*, 112156. [CrossRef]
19. Shen, K.N.; Zhao, D.B. An EMD-LSTM Deep Learning Method for Aircraft Hydraulic System Fault Diagnosis Under Different Environmental Noises. *Aerospace* **2023**, *10*, 55. [CrossRef]
20. Peng, Z.Q.; Wang, Q.B.; Liu, Z.R.; He, R.J. Remaining Useful Life Prediction for Aircraft Engines under High-Pressure Compressor Degradation Faults Based on FC-AMSLSTM. *Aerospace* **2024**, *11*, 293. [CrossRef]
21. He, C.; Li, Z.; Zheng, C.; Zhang, Z.; Zhang, L. A Hybrid Model Based on a Dual-Attention Mechanism for the Prediction of Remaining Useful Life of Aircraft Engines. *Sensors* **2025**, *25*, 5682. [CrossRef]
22. Xu, K.J.; Guo, Y.Q.; Zhou, Q.F. Research on the Remaining Useful Life Prediction Algorithm for Aero-Engines Based on Transformer-KAN-BiLSTM. *Aerospace* **2025**, *12*, 998. [CrossRef]
23. Lu, X.H.; Pan, H.B.; Zhang, L.X.; Ma, L.; Wan, H. A dual path hybrid neural network framework for remaining useful life prediction of aero-engine. *Qual. Reliab. Eng. Int.* **2024**, *40*, 1795–1810. [CrossRef]
24. Gao, Z.; Jiang, W.X.; Wu, J.; Wang, Y.H.; Zhu, H.P. A customized dual-transformer framework for remaining useful life prediction of mechanical systems with degraded state. *Mech. Syst. Signal Process.* **2025**, *230*, 112611. [CrossRef]
25. Gao, Z.; Wang, C.J.; Wu, J.; Wang, Y.H.; Jiang, W.X.; Dai, T.J. Degradation-aware remaining useful life prediction of industrial robot via multiscale temporal memory Transformer framework. *Reliab. Eng. Syst. Saf.* **2025**, *262*, 111176. [CrossRef]
26. Ma, M.; Wang, Z.Z.; Zhong, Z.R. Transformer Encoder Enhanced by an Adaptive Graph Convolutional Neural Network for Prediction of Aero-Engines' Remaining Useful Life. *Aerospace* **2024**, *11*, 289. [CrossRef]
27. Chen, H.L.; Wei, J.A.; Huang, H.S.; Wen, L.; Yuan, Y.G.; Wu, J.X. Novel imbalanced fault diagnosis method based on generative adversarial networks with balancing serial CNN and Transformer (BCTGAN). *Expert Syst. Appl.* **2024**, *258*, 125171. [CrossRef]
28. Zhao, B.; Zhang, W.G.; Zhang, Y.R.; Zhang, C.P.; Zhang, C.; Zhang, J.W. Lithium-ion battery remaining useful life prediction based on interpretable deep learning and network parameter optimization. *Appl. Energy* **2025**, *379*, 124713. [CrossRef]
29. Huang, Y.F.; Tao, J.; Sun, G.; Zhang, H.; Hu, Y. A Prognostic and Health Management Framework for Aero-Engines Based on a Dynamic Probability Model and LSTM Network. *Aerospace* **2022**, *9*, 316. [CrossRef]
30. Han, B.K.; Yin, P.W.; Zhang, Z.Z.; Wang, J.R.; Bao, H.Q.; Song, L.J. Remaining useful life prediction of turbofan engines based on dual attention mechanism guided parallel CNN-LSTM. *Meas. Sci. Technol.* **2025**, *36*, 016160. [CrossRef]
31. Zhu, Y.X.; Liu, Z.D. A Dual-Dimension Convolutional-Attention Module for Remaining Useful Life Prediction of Aeroengines. *Aerospace* **2024**, *11*, 809. [CrossRef]
32. Xiong, Y.X.; Guo, B.; Dian, S. Aero-engine remaining useful life prediction using a TCN prognostic model enhanced with dual-dimensional fusion attention. *Meas. Sci. Technol.* **2025**, *36*, 036116. [CrossRef]
33. Wang, L.B.; Zhu, Z.B.; Zhao, X.F. Dynamic predictive maintenance strategy for system remaining useful life prediction via deep learning ensemble method. *Reliab. Eng. Syst. Saf.* **2024**, *245*, 110012. [CrossRef]
34. Liu, Y.; Hu, T.G.; Zhang, H.R.; Wu, H.X.; Wang, S.Y.; Ma, L.T. iTransformer: Inverted Transformers are effective for time series forecasting. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
35. Li, T.Y.; Si, H.Q.; Qiu, J.X.; Li, J.Y.; Gong, Y.Q. A hybrid algorithm of TCN-iTransformer for aircraft aerodynamic parameter estimation based on dual attention mechanism. *Aerosp. Sci. Technol.* **2025**, *164*, 110350. [CrossRef]

36. Lyu, Y.; Qiu, Q.P.; Chu, Y.; Zhang, J. An adaptive BiGRU-ASSA-iTransformer method for remaining useful life prediction of bearing in aerospace manufacturing. *Actuators* **2025**, *14*, 238. [CrossRef]
37. Beck, M.; Pöppel, K.; Spanring, M.; Auer, A.; Prudnikova, O.; Kopp, M. xlstm: Extended long short-term memory. In Proceedings of the Neural Information Processing Systems, Vancouver, BC, Canada, 9–15 December 2024.
38. Alharthi, M.; Mahmood, A. xLSTMTime: Long-term time series forecasting with xLSTM. *AI* **2024**, *5*, 1482–1495. [CrossRef]
39. Kraus, M.; Divo, F.; Dhami, D.S.; Kersting, K. xLSTM-Mixer: Multivariate time series forecasting by mixing via scalar memories. *arXiv* **2024**, arXiv:2410.16928.
40. Tudosie, A.N. Aircraft gas-turbine engine's control based on the fuel injection control. In *Aeronautics and Astronautics*; Mulder, M., Ed.; IntechOpen: Rijeka, Croatia, 2011.
41. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L. Attention is all you need. In Proceedings of the Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
42. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [CrossRef]
43. Kim, T.; Kim, J.; Tae, Y.; Park, C.; Choi, J.; Choo, J. Reversible instance normalization for accurate time-series forecasting against distribution shift. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021.
44. Liu, Y.; Wu, H.X.; Wang, J.M.; Long, M.S. Non-stationary Transformers: Exploring the stationarity in time series forecasting. In Proceedings of the Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; pp. 9881–9893.

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.