



OPEN A multi-scale cross-dimension interaction approach with adaptive dilated TCN for RUL prediction

Zhe Lu¹, Bing Li¹✉, Changyu Fu¹, Liang Xu¹, Bai Jiang², Zelong Li¹, Junbao Wu¹ & Siye Jia¹

In the domain of Prognostics and Health Management (PHM) technologies, the focus of Remaining Useful Life (RUL) prediction is on the forecasting of the time to failure by uncovering the complex correlations between equipment degradation features and RUL labels, thereby enabling effective support for predictive maintenance strategies. However, extant research primarily emphasizes single-scale and single-dimensional feature extraction, which fails to adequately capture both long- and short-term dependencies as well as the interrelationships among sensor feature dimensions. This limitation has a detrimental effect on the accuracy and robustness of RUL prediction. To address the aforementioned issues, this paper proposes an Adaptive Dilated Temporal Convolutional Network (AD-TCN) approach, incorporating a Multi-Scale Cross-Dimension Interaction Module (MSCDIM) to enhance feature extraction and interaction. First, a dynamic adaptive dilation factor is incorporated into the TCN, thereby enabling the model to adjust its receptive field dynamically, which facilitates the capture of long- and short-term dependencies across different scales, allowing a more comprehensive representation of equipment degradation patterns. Second, the MSCDIM module has been designed to perform multi-scale feature extraction, interaction, and fusion across temporal and sensor dimensions, dynamically adjusting feature weights in order to suppress redundant information and enhance the representation of critical features. Finally, contrastive and ablation experiments are conducted on the widely used C-MAPSS dataset and N-CMAPSS dataset. And the experimental results demonstrate that the proposed method achieves high performance, with the MSCDIM module exhibiting strong adaptability and significant potential for broad applications.

Keywords Relaining useful life, TCN, Deep learning, Time series forecasting

In the context of the rapid advancements witnessed within the modern manufacturing and aerospace industries, there has been a discernible increase in the demand for enhanced reliability in complex systems. The advent of large-scale data storage, pervasive adoption of sensor technologies, and substantial advancements in high-performance equipment have collectively fostered a robust technological foundation for Prognostics and Health Management (PHM) technology, thereby facilitating its extensive application and ongoing development.

The PHM technology facilitates predictive maintenance through continuous monitoring and analysis of the operational status of critical components within complex systems in real time. Utilising these evaluations, maintenance schedules are optimised to effectively prevent potential failures. A particularly critical aspect of PHM is the prediction of Remaining Useful Life (RUL), which plays a central role in PHM. By accurately assessing the present degradation state of equipment, RUL prediction enables the early identification of fault indicators and facilitates the dynamic adjustment of maintenance strategies. The implementation of these measures has been shown to have a significant impact on the mitigation of potential failures, as well as a substantial enhancement of system reliability and operational efficiency¹.

In the task of predicting the RUL of equipment, it is crucial to capture both the long-term and short-term dependencies inherent in the degradation process. Long-term dependencies involve gradual changes that follow a persistent trend, such as progressive wear or ongoing performance decline in mechanical components. In contrast, short-term dependencies manifest as rapid fluctuations, often spurred by sudden failures or abrupt environmental changes, that can substantially alter the degradation trajectory. A model that focuses on only one type of dependency may overlook critical features needed for accurately assessing degradation trends.

¹College of Intelligent Systems Science and Engineering, Harbin Engineering University, Nantong Street, 150001 Harbin, China. ²College of Mechanical and Electrical Engineering, Harbin Engineering University, Nantong Street, 150001 Harbin, China. ✉email: libing265@hrbeu.edu.cn

Consequently, achieving precise RUL predictions requires that models capture both long-term and short-term dependencies simultaneously, thereby offering a comprehensive depiction of the equipment's degradation state².

The techniques for predicting RUL can be categorized into three distinct groupings: failure mechanism-based methods, data-driven methods, and hybrid approaches. Among them, failure mechanism-based prediction methods primarily leverage domain-specific prior knowledge to develop mathematical models that reflect or simulate the degradation processes of equipment. This enables the prediction of its remaining useful life. For instance, Xue et al. combined the discrete-variable degradation model of equipment with continuous-variable boundary and initial conditions in order to dynamically describe the degradation mechanism of critical components and the validity of their approach was validated through simulations using fault data from the bearings of a nuclear circulating water pump test bench³. However, failure mechanism-based RUL prediction methods are primarily suitable for mechanical components with well-defined degradation characteristics and mechanisms that can be accurately modeled. And as equipment complexity increases, developing precise failure mechanism models becomes significantly more challenging in practical applications.

Furthermore, research on hybrid methods aims to enhance the accuracy of RUL prediction by integrating the strengths of failure mechanism models and data-driven models. For instance, Ma et al. developed a physics-based degradation model for lithium-ion batteries, leveraging domain expertise and deriving key parameters, which were then integrated with a deep neural network to predict RUL of the batteries based on a single charging curve⁴. Nevertheless, the effective integration of degradation mechanism models with deep learning remains a significant technical challenge.

In recent years, with the development of digital industry, a large number of key links deployed in various types of sensors have accumulated a large amount of industrial data, through the extraction of effective information from these multi-sensor data can be mined in-depth relationship between equipment degradation and industrial data, equipment degradation diagnosis and prediction of the state of provide a strong support. For example, Kai et al. proposed an advantageous discriminant analysis-aided collaborative alignment network (DA-CAN) that leverages multi-sensor data to enhance inter-sensor correlations, thereby enabling fault diagnosis across devices⁵. Fu et al. proposed a novel shareability-exclusivity collaborative representation (SECR) method, which partitions multi-sensor data into multiple blocks and leverages the synergistic interplay between shared information and distinctive features to extract both common and exclusive characteristics from each block, thereby significantly enhancing the diagnostic capabilities of multi-sensor systems in complex industrial processes⁶.

Data-driven methods can be further categorized into statistical learning-based methods and machine learning-based methods. Among them, statistical learning-based methods analyze equipment degradation data to extract statistical features and construct degradation models for RUL prediction. For example, Xin et al. employed a joint estimation method combining domain adaptation (DA) and unscented Kalman filtering (UKF), achieving high-precision joint estimation of the state of charge (SOC) and state of energy (SOE) for lithium-ion batteries under varying temperature conditions⁷. However, in real industrial environments, the diversity and uncertainty of critical factors such as equipment conditions and operating environments render the construction of degradation models using statistical learning methods highly complex, thereby limiting their practicality to a certain extent. Conversely, machine learning-based methods employ the utilisation of collected equipment degradation data to facilitate a profound exploration of multidimensional features and establish RUL prediction models by associating these features with the remaining useful life. These methods have been shown to excel in the handling of nonlinear and high-dimensional complex data, to be less affected by environmental factors, and to eliminate the need for extensive domain knowledge or the construction of complex degradation mechanism models. The high predictive accuracy of these methods has led to their extensive use in industrial practice.

Based on the complexity of machine learning models, they can be further divided into shallow machine learning and deep learning approaches. Specifically, shallow machine learning methods, such as support vector machines (SVM), extreme learning machines (ELM), and random forests (RF), are widely adopted traditional approaches. These methods typically involve manual feature engineering to extract key information from data and build regression models for predicting the remaining useful life of equipment. For instance, Li et al. employed Gaussian process regression (GPR) and Kalman filtering to estimate the state of health (SOH) and state of charge (SOC), combining these with least squares support vector machines (LS-SVM) to predict the RUL of lithium-ion batteries⁸. Patel et al. proposed a milling cutter RUL prediction approach based on Shannon entropy feature selection and squared exponential Gaussian process regression (SE-GPR). The method computes the Shannon entropy of features derived from vibration and force sensors to identify those that encapsulate the most state information, and subsequently employs the SE-GPR model for prediction, thereby significantly enhancing accuracy while reducing feature dimensionality⁹. However, due to the simple structure of shallow machine learning models, they heavily rely on feature engineering and domain expertise, with prediction accuracy largely dependent on the quality of manually extracted features. Conversely, deep learning methods excel in feature learning, automatically extracting deep abstract features from data without requiring manual intervention. They are particularly effective in addressing nonlinear and high-dimensional problems, leading to a growing number of studies focusing on deep learning approaches.

In the complex multivariate time series regression task of RUL prediction, traditional deep learning architectures such as RNN and CNN have been widely used to efficiently extract deep key features from time series data to capture long- and short-term dependencies. Boujamaza et al. proposed a dilated recurrent neural network (D-RNN) that embeds dilated convolutions within the RNN framework and employs random grid search for parameter optimization, ultimately achieving precise predictions of turbofan engine RUL¹⁰. However, when RNN and its variants (e.g., LSTM, GRU, etc.) networks to process time-series data, due to the model's inherent sequential dependence, the input of each time step is contingent on the output of the previous time step. Such an architecture is susceptible to issues such as gradient disappearance and gradient explosion when

performing long time-series data training to explore the long-term dependence of equipment degradation data, This complicates the effective transfer of key information to the current moment state, which hinders the model's ability to capture effective long-term dependencies in the face of extensive device degradation processes, and greatly affects the efficiency and prediction accuracy of the device RUL prediction task¹¹. To alleviate this problem, Li et al. proposed a rolling bearing RUL prediction method based on recursive exponential slow feature analyses (RESFA) and bi-directional LSTM (Bi-LSTM), which utilises an autoencoder to extract the features and effectively captures the long-term degradation trend of the system through RESFA, thus significantly improving the stability and accuracy of the prediction while coping with the noise disturbances¹². Moreover, since RNN-based models and their variants inherently depend on outputs from preceding time steps when computing the current step, they cannot efficiently process time series data in parallel, resulting in slow training, especially when capturing long-term dependencies during device degradation or handling large-scale datasets. To reduce the adverse impact of rapidly increasing time-series data on model training efficiency, Sun et al. mitigated the gradient vanishing problem in long-sequence prediction, integrates and selects multisensor monitoring data to enhance training efficiency, and more accurately captures long-term dependencies inherent in equipment degradation processes¹³.

In comparison to RNNs and their variants, CNNs have the capacity to process time series data in parallel by means of a sliding window approach. Utilising convolution operations, CNNs effectively capture the relationships between each time window and RUL labels. This approach circumvents the information loss and gradient vanishing issues that arise from conventional step-by-step processing of time series data. For instance, Guo et al. have proposed a sandglass-shaped multi-scale feature extractor (HME) based on a 1D-CNN and integrated it with a Transformer network to achieve accurate RUL prediction¹⁴. Chen et al. proposed a hybrid model that integrates a multi-scale deep convolutional neural network with a LSTM network for the extraction of features under complex and multi-fault operating conditions, thereby enabling accurate prediction of RUL for aircraft engines¹⁵. Zhang et al. proposed a parallel hybrid neural network composed of a 1D-CNN and a bidirectional gated recurrent unit (BiGRU) for predicting the RUL of aircraft engines¹⁶.

However, although CNN effectively extract short-term dependencies in time series data via convolutional operations, they are limited by a fixed receptive field when extended to capture long-term dependencies¹⁷. This limitation is particularly pronounced when processing degradation time series data under complex operational conditions across multiple time scales. In such cases, conventional CNNs cannot flexibly adjust their receptive fields to accommodate features specific to different time scales, thereby failing to fully capture the interplay between short and long term dependencies. To address this limitation, temporal convolutional networks (TCNs), a variant of convolutional neural networks, utilize causal convolutions to ensure that each time step depends solely on inputs from the current and preceding time steps. And the integration of dilated causal convolutions, which serves to expand the model's receptive field, and the employment of residual connections, which are employed to maintain stability, result in TCNs effectively alleviating the vanishing gradient issue while preserving the causality and continuity of time-series processing. Consequently, they demonstrate an enhanced capacity to capture long-term dependencies in tasks involving the analysis of equipment degradation time-series data. For instance, Zhang et al. incorporated an enhanced self-attention mechanism and a squeeze-and-excitation mechanism into TCN to capture feature correlations across different time steps and channels, thereby enabling accurate RUL prediction¹⁸. Lin et al. proposed a dual-attention framework called Channel and Temporal Attention Temporal Convolutional Network (CATA-TCN), which enhances features and extracts critical degradation characteristics to improve the accuracy of RUL prediction¹⁹. Jiang et al. integrated a multi-head self-attention (MSA) mechanism with a TCN model to capture critical features during the degradation process of rolling bearings, enabling accurate residual RUL prediction for such components²⁰. Tang et al. proposed a novel method for predicting RUL of rolling bearings, integrating health indicators and combining time-domain, frequency-domain, and time-frequency domain features through a parallel combination of TCN and Transformer, significantly improving RUL prediction accuracy by leveraging both local and global features²¹.

Although conventional TCNs have demonstrated outstanding performance in time-series forecasting, particularly in their marked ability to capture long-term dependencies, they still encounter certain challenges in practical applications. Firstly, many existing approaches to time-series analysis primarily focus on extracting and mining features along a single temporal dimension while neglecting the inter-variable correlations among data collected from multiple sensors and the coupling between features across different dimensions during the equipment degradation process, which ultimately undermines the accuracy of RUL predictions for equipment operating under complex degradation patterns. Furthermore, in the realm of temporal feature extraction, most existing approaches rely on a single time scale for feature learning, failing to fully account for the mutual influence between long-term trends and short-term fluctuations. In practice, equipment degradation typically occurs concurrently across multiple time scales, encompassing both gradual trend changes and rapid fluctuations, and neglecting the interrelationships among these scales can compromise the accuracy of RUL predictions.

In order to address the aforementioned issues, We introduce a spatiotemporal convolutional network (AD-TCN) that leverages a multi-scale cross-dimension interaction module (MSCDIM). Specifically, the AD-TCN expands the receptive field of the spatiotemporal convolutional model by introducing a dynamic, adaptive dilation factor, thereby enabling flexible adjustment according to the temporal scale of the data and effectively addressing the challenge of capturing both long-term trends and short-term fluctuations in diverse equipment degradation processes. In addition, a serial TCN architecture is employed to independently capture temporal dependencies from both the time and sensor dimensions, thereby effectively extracting multi-dimensional feature information in the equipment degradation process and further enhancing the model's representational capacity. And the proposed MSCDIM module focuses on multi-scale feature extraction along both the temporal and sensor dimensions while enabling cross-dimensional information interaction; by adjusting the weighting

parameters for each scale and dimension, the model is dynamically guided to attend to key features relevant to equipment degradation and suppress redundant information that could otherwise impede RUL prediction accuracy. Furthermore, by applying nonlinear transformations and related techniques, the extracted key degradation features are associated with RUL labels, thereby facilitating accurate predictions of remaining useful life.

The contributions of this paper are as follows:

1. A dynamic adaptive dilation factor has been designed within the TCN framework, thus enabling the convolutional receptive field to be adaptively adjusted across different time steps by learning dynamically generated dilation factors. This approach effectively captures long-, medium-, and short-term dependencies in time series data while setting a dilation factor threshold to prevent the introduction of noise.
2. The MSCDIM has been developed for the purpose of filtering information across different dimensions and scales. This module employs attention weights to adaptively focus on critical information within the features while suppressing redundant information. Furthermore, the employment of nonlinear transformations and associated operations facilitates the effective integration of diverse features, thereby enhancing the model's robustness.
3. A series of comparative and ablation experiments were conducted on the C-MAPSS dataset and the N-CMAPSS dataset in order to systematically validate the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section [Methodology](#) provides a detailed exposition of the proposed methodology. The subsequent section, Sect. [Experimental validation](#), provides a comprehensive description of the data preprocessing procedures and the experimental configurations. Section [Conclusion](#) presents a comprehensive comparison and analysis of the results. Finally, Section 5 concludes the paper and offers perspectives on future work.

Methodology

Problem formulation

The RUL of a system is defined as the time interval $t_{current}$ from the current time point to the point where system performance degrades to the failure threshold $t_{failure}$, expressed as follows:

$$RUL = t_{failure} - t_{current} \quad (1)$$

The core task of RUL prediction is to estimate the remaining time from the current point to the failure point based on degradation information collected during the system's operational phase. This process involves the analysis and extraction of deep degradation features from current data, combined with historical data, to provide an accurate estimate of the system's remaining useful life. Specifically, the system degradation state is represented as follows:

$$x(t) = \begin{Bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_N(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_N(t_2) \\ \dots & \dots & \dots & \dots \\ x_1(t_n) & x_2(t_n) & \dots & x_N(t_n) \end{Bmatrix} \quad (2)$$

where the x_t represents a set of data collected at time t , and $x_i(t_j)$ denotes the observation value of the i -th sensor at time t_j .

The primary objective of RUL prediction is to establish a mapping function $g(\cdot)$ that relates the input $x(t)$, which encapsulates the system's current degradation state and historical data, to the RUL label. The relationship between these variables can be expressed mathematically as follows:

$$RUL = g(x(t)) \quad (3)$$

Leveraging this mapping capability enables the accurate prediction of the system's remaining useful life, thereby providing valuable insights for maintenance and operational planning purposes²². And a typical equipment RUL degradation process is illustrated in Fig. 1.

Nevertheless, the challenge lies in the highly nonlinear and uncertain nature of the degradation process. The degradation trajectory of a system is often subject to variation due to operational conditions, environmental factors, and individual differences. Consequently, the primary research focus is on developing methodologies to comprehensively extract critical features from degradation data, integrate them with historical data, and construct a mapping function corresponding to the RUL label to enable accurate prediction of the system's future state.

Structure of the AD-TCN-MSCDIM method

This paper proposes a novel method for RUL prediction, integrating an adaptive dilated TCN with a multi-scale fusion module to enhance the accuracy of remaining useful life estimation. The proposed model's overall framework is illustrated in Fig. 2. For the complex task of RUL prediction, the process commences with data preprocessing, which incorporates data normalisation, time windowing, and RUL label construction on the C-MAPSS dataset. Subsequently, the preprocessed data is fed into the model for training, generating RUL prediction results. Finally, the model's performance is evaluated by comparing the predicted RUL with actual RUL data.

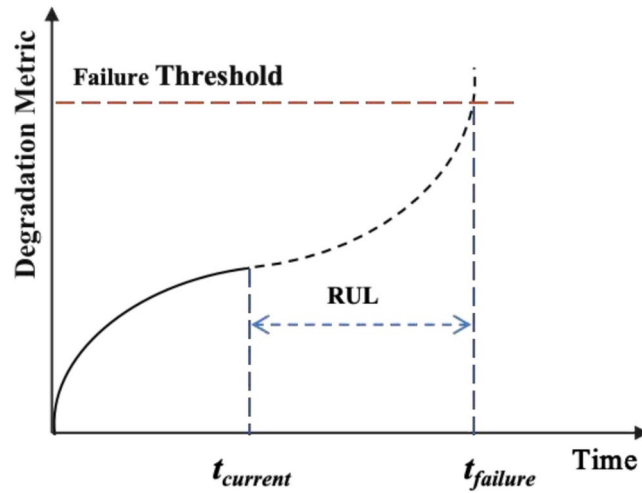


Fig. 1. Diagram of equipment RUL degradation.

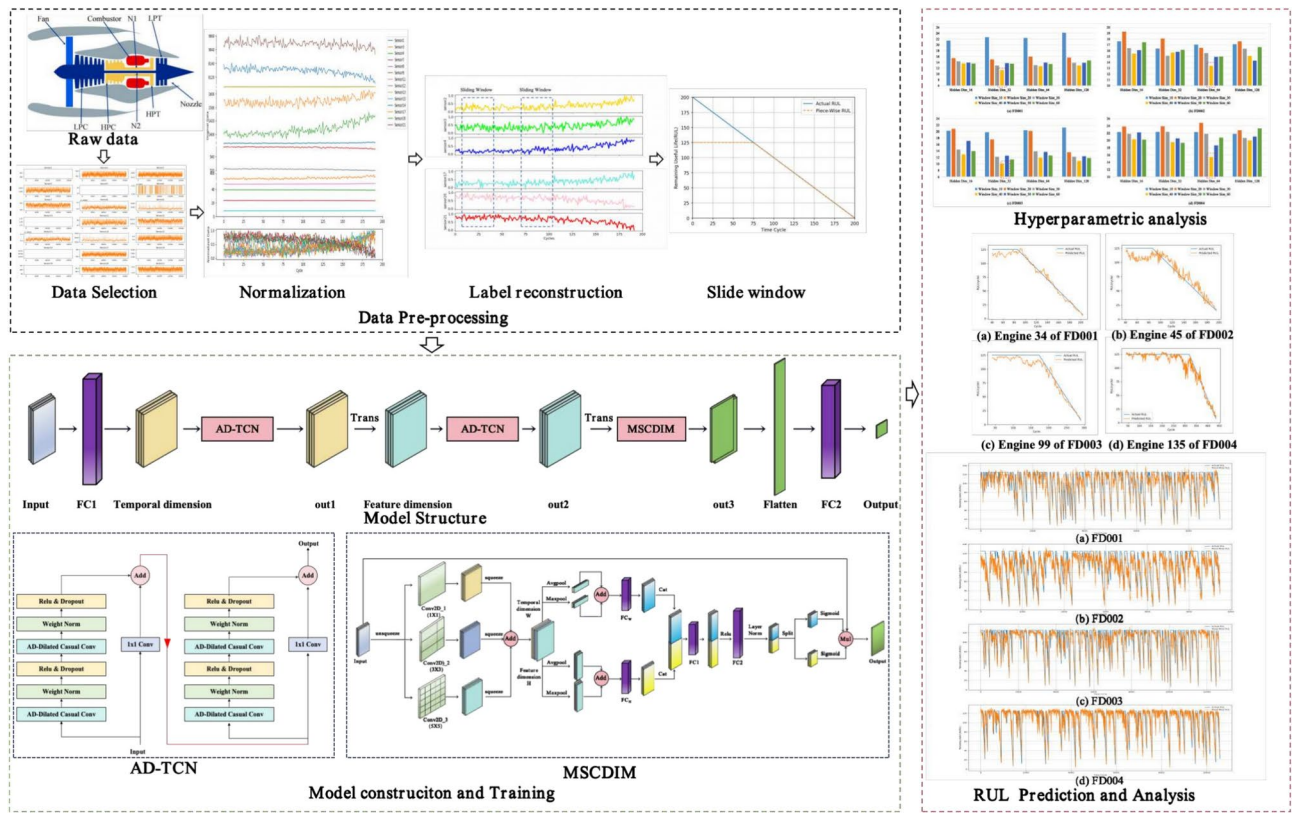


Fig. 2. The general architecture of the proposed model.

The model is composed of two primary components: the Adaptive Dilated Temporal Convolutional Network (AD-TCN) architecture and the Multi-Scale Cross-Dimension Interaction Module (MSCDIM). Among them, The AD-TCN architecture, consisting of two serially connected TCNs with adaptive dilation factors, integrates multi-layer temporal convolutions and residual connections to effectively address long-term temporal dependencies and enhance the model's ability to represent complex nonlinear degradation paths. The MSCDIM employs dynamic adjustment of feature weights through the utilisation of channel and temporal attention mechanisms, integrating multi-scale features across temporal and sensor dimensions, suppressing redundant information and enhancing the representation capabilities of features.

Adaptive dilated temporal convolutional network architecture

The TCN represents a refinement of the CNN, with a specific focus on the extraction of features from time series data. Its architecture comprises three primary components: causal convolutions, dilated causal convolutions, and residual connections.

Causal convolution

The fundamental objective of causal convolution is to construct a model that is constrained in time, thereby ensuring that the output at each time step depends solely on the current and past input data, without utilizing future information, thereby preserving true temporal causality in the model.

For a given one-dimensional input sequence $x = [x_0, x_1, \dots, x_{t-1}, x_t, x_{t+1}, \dots]$, traditional standard convolution can be expressed as a sliding operation of a convolution kernel over the sequence. Typically, zero-padding is applied at the sequence boundaries to ensure that the output length matches the input length. The convolution operation can be represented as follows:

$$y_t = \sum_{i=0}^{k-1} w_i x_{t-i} \quad (4)$$

where y_t is the output at time step t , $w = [w_0, w_1, \dots, w_{k-1}]$ is the convolution kernel of length k , and x_{t-i} represents the input value at time step $t-i$. Besides, Zero-padding is applied when $t-i < 0$ or $t-i \geq T$ (where T is the length of the input sequence) to handle cases where the index falls outside the input sequence.

However, in the standard convolution mode, the receptive field of the convolution kernel includes both past and future data, which may contravene the causality requirement in time series prediction. As demonstrated in Fig. 3a and b, causal convolution differs from standard convolution in the following way: when computing the output y_t , it only considers data from the current and past time steps, such as \dots, x_{t-1}, x_t , while excluding future information, such as x_{t+1}, x_{t+2}, \dots .

In order to ensure the strict enforcement of causality in time-series prediction, it is imperative that the convolutional kernel exclusively accesses current and past inputs. Consequently, when adopting a left-to-right time indexing ($t = 0$ as the earliest point), zero-padding is applied to the left of the input sequence (i.e., for

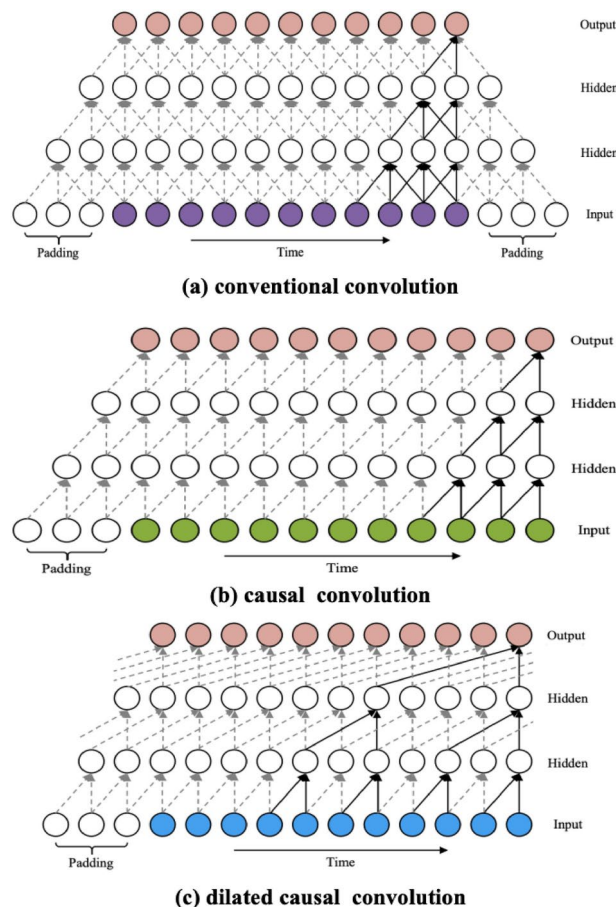


Fig. 3. Schematic of different types of convolution.

$t - i < 0$), ensuring no reference is made to future values x_{t+1}, x_{t+2} , etc. This approach prevents any information leakage from future time steps, thereby preserving strict causality during both training and inference.

Dilated causal convolution

In the context of time series prediction, causal convolution is employed with zero-padding on the left side of the input sequence to ensure that the prediction results depend solely on the current and past data. However, when attempting to capture long-term dependencies, a significant increase in the number of parameters is observed with the increase in the convolution kernel length. As shown in Fig. 3c, in order to address this issue, a dilation factor d is introduced into causal convolution, thus enabling the receptive field to be expanded without a significant increase in the number of parameters. This facilitates the capture of long-term dependencies while preserving the fundamental principles of causality.

The term 'receptive field' (RF) is employed to denote the range of input data that corresponds to a specific output position, representing the area of influence within the input data that contributes to the output at a given location. For a standard convolution with a kernel length of k , no dilation ($d = 1$), and a stride of 1, the receptive field is equal to k . And in a single-layer causal convolution, although left padding ensures that the output depends only on the current and past time steps, the receptive field remains unchanged at k . In order to capture long-term dependencies without a significant increase in the number of parameters, a dilation factor, d , is introduced into causal convolution, resulting in Dilated Causal Convolution. The output at time step t can be defined as:

$$y_t = \sum_{i=0}^{k-1} w_i x_{t-d \cdot i} \quad (5)$$

where $y(t)$ represents the output at time step t ; w_i denotes the i -th weight of the convolution kernel of length k ; $x_{t-d \cdot i}$ is the input value at time step $t - d \cdot i$; and d is the dilation factor, which determines the sampling interval of the convolution kernel along the time axis.

At this point, unlike standard convolution, each kernel element in dilated convolution no longer samples input data continuously. Instead, each kernel element samples input data at intervals of d points along the time axis. The receptive field for this operation is given by:

$$RF = 1 + (k - 1) \cdot d \quad (6)$$

where the K denotes the length of the kernel, and d denotes the dilation factor.

Although conventional TCN can extend their receptive fields to capture long-term dependencies by increasing the dilation factor d , this factor is typically fixed, inhibiting flexible adjustments based on the multi-scale requirements of time-series data. While increasing d expands the receptive field to capture extended temporal dependencies, an excessively large, fixed dilation factor is liable to introduce unnecessary parameters when the long-term dependencies in equipment degradation data are not pronounced. In such cases, short-term feature extraction may also be hampered, making it difficult to strike a balance between capturing long- and short-term dependencies. Consequently, overall model efficiency is reduced, and the accuracy of RUL predictions can be adversely affected.

Adaptive dilation factor

To overcome the limitations of conventional TCNs in capturing both long- and short-term dependencies in equipment degradation time-series data, an adaptive dilation factor d_θ is introduced. Unlike the fixed dilation factor d used in traditional TCN, d_θ can be dynamically adjusted to accommodate various input characteristics. Furthermore, by expanding or contracting the model's receptive field according to the requirements of different time steps, this adaptive factor offers greater flexibility, thereby more effectively modeling dependencies across multiple temporal scales. At this point, the convolution operation output is expressed as:

$$y_t = \sum_{i=0}^{k-1} w_i x_{t-d_\theta \cdot i} \quad (7)$$

In this formulation, d_θ serves as a dynamically trainable parameter that can be adjusted according to the characteristics of the input time-series data and is updated via backpropagation during training. Consequently, the receptive field can be automatically expanded when capturing long-term dependencies and contracted for short-term dependencies, mitigating the limitations posed by a fixed receptive field in conventional TCN. This approach also delivers enhanced flexibility in handling multi-scale time-series data under complex operating conditions.

At this point, the RF is given by: $RF = 1 + (K - 1) \cdot d_\theta$. As the value of d_θ increases, the receptive field of the convolution layer undergoes an expansion. However, d_θ is an adaptive learning parameter, excessively large values can result in an overly expansive receptive field, leading to a significant increase in the number of parameters and potentially causing over-smoothing. Conversely, overly small values of d_θ may fail to effectively expand the receptive field. Therefore, it is crucial to set a threshold for d_θ based on practical requirements to ensure that the receptive field of a single convolution layer is both effective and appropriately scaled.

In the context of d_{min} , it is imperative to ensure that the minimum receptive field of a single convolution layer is of a specified length and effective. This is to avoid the inclusion of distant historical data that may introduce noise. Consequently, d_{min} is set to 1 to ensure sufficient capability in capturing short-term dependency features from the data. For d_{max} , maximizing its value is essential for capturing long-term dependency features, as it allows the receptive field of a single convolution layer to cover a broader time span. Nevertheless, it is imperative to note that unduly large values have the potential to substantially augment computational expenses and engender instability during the training process. A review of studies related to RUL prediction on the C-MAPSS dataset reveals that the time window W is typically set within the range of 20–60²³.

Thus, when the convolution kernel length is set to $k = 3$, to ensure coverage of a maximum time step of 60, the following condition must be satisfied:

$$RF = 1 + (k - 1) \cdot d_{max} \geq 60 \quad (8)$$

And through calculation, when $d_{max} = 30$, the receptive field becomes:

$$RF = 1 + (3 - 1) \cdot 30 > 60 \quad (9)$$

which can cover a time step of 60. Therefore, d_{max} is set to 30. Moreover, during the processes of model initialization and training, the quantity d_θ is initialized to $(d_{min} + d_{max})/2$. Thereafter, a learnable parameter ϕ and a sigmoid activation function are utilised to map d_θ within the range of d_{min} and d_{max} . This methodological approach permits the model to determine the optimal receptive field in an adaptive manner, based on the prevailing dependency requirements during the training process.

$$d_\theta = d_{min} + (d_{max} - d_{min}) \cdot \delta(\phi) \quad (10)$$

where the parameter ϕ is learnable and dynamically optimized during network training, and the $\delta(\cdot)$ represents the sigmoid activation function, which maps ϕ to the range $[0, 1]$. The value of $\delta(\cdot)$ is constrained within the range $[d_{min}, d_{max}]$ and adaptively adjusts the receptive field size during training. This ensures a balance between capturing short-term local dependency features and representing long-term global dependency features during the modeling process.

Although setting thresholds for d_θ helps prevent the receptive field from becoming too large or too small, to ensure that the input length T_{in} matches the output length T_{out} , appropriate padding and trimming operations need to be applied to the output data in conjunction with the adaptive dilation factor d_θ . The output length is calculated as follows:

$$T_{out} = \frac{T_{in} + 2p - d_\theta(k - 1) - 1}{s} + 1 \quad (11)$$

where T_{in} represents the input sequence length, T_{out} denotes the output sequence length, k is the convolution kernel size, and p refers to the amount of padding applied, s is the stride. And when the stride s is set to 1, the formula T_{out} can be further simplified as:

$$T_{out} = \frac{T_{in} + 2p - d_\theta(k - 1) - 1}{1} + 1 = T_{in} + 2p - d_\theta(k - 1) \quad (12)$$

And when $T_{out} = T_{in}$, the amount of padding p can be calculated as:

$$p = \frac{d_\theta(k - 1)}{2} \quad (13)$$

And in this case, the input and output sequence lengths are consistent. However, to ensure causality, a right-padding strategy is required. In this case, the padding p is set to:

$$p = d_\theta(k - 1) \quad (14)$$

To ensure that the output at each time step depends only on past and current input values, without incorporating future information. When substituting the formula(14) into the formula, we have:

$$T_{out} = T_{in} + 2(d_\theta(k - 1)) - d_\theta(k - 1) = T_{in} + d_\theta(k - 1) \quad (15)$$

This results in each output sequence being $d_\theta(k - 1)$ time steps longer than the input sequence, which causes a mismatch in lengths. To address this, the Chomp1d operation in TCN is used to trim the extra $d_\theta(k - 1)$ time steps from each output sequence, ensuring the output length matches the input length. The adjusted output length is expressed as:

$$T'_{out} = T_{out} - p = (T_{in} + p) - p = T_{in} \quad (16)$$

Thus, by applying right-side padding and trimming, causality is ensured for each time step, while maintaining consistency between the lengths of the output T'_{out} and input T_{in} sequences at every time step.

Residual block

The utilization of residual connections facilitates the integration of the features extracted by TCN with the input information, thereby effectively mitigating the gradient vanishing problem that arises as network depth increases during back propagation.

As shown in Fig. 4, each residual block in the AD-TCN module is composed of Adaptive Dilation Factor convolution, activation function, Weight normalization, dropout layer, and a residual connection. The mathematical formulation is as follows:

$$y_t = \text{Activation}(F(x_t) + x_t) \quad (17)$$

where, the x_t represents the input vector at time t , y_t represents the output feature vector corresponding to time t , and $F(x_t)$ denotes the output generated by processing x_t through the Adaptive Dilation Factor convolution, activation function, and other layers.

As illustrated in Fig. 4, to fully explore the dynamic relationships among various features within time-series data during equipment degradation, this study proposes a feature extraction framework composed of two serial AD-TCN modules. Firstly, the initial AD-TCN module processes the raw time-series data, leveraging multi-layer adaptive dilated causal convolutions and residual connections along the time dimension (W) to extract deep temporal dependency features that effectively capture both short- and long-term dependencies. In addition, a clipping strategy and fully connected layers are utilized for channel mapping to ensure consistency between input and output shapes, thereby enhancing the model's overall efficiency. Besides, a second AD-TCN module takes the output from the initial stage as its input, further extracting temporal correlations from the sensor feature dimension (H). This approach delves into the intricate sequential relationships between sensor data features and temporal characteristics, thereby capturing both short-term fluctuations and long-term trends.

By designing a serial architecture with two AD-TCN modules, the model independently performs temporal modeling on distinct features along both the time and sensor dimensions in equipment degradation time-series datasets, thereby effectively capturing the long- and short-term dependencies within each dimension. Besides, this architecture enhances the model ability to flexibly handle complex temporal correlations spanning both the time and sensor dimensions, while simultaneously providing high-quality feature representations for subsequent multi-scale feature extraction. As a result, it improves the accuracy and robustness of RUL predictions in equipment degradation scenarios.

Multi-scale cross-dimension interaction module

Although the dual-module AD-TCN framework initially extracts key temporal features from both time and sensor dimensions by modeling time dependencies and feature correlations in sequence data, it still faces certain limitations in deeper multi-scale feature extraction and cross-dimensional feature interaction. Specifically, the proposed AD-TCN architecture primarily focuses on extracting dependencies at a single temporal scale by introducing an adaptive dilation factor and feature-dimension transformations to dynamically expand the model's receptive field along the temporal dimension, thereby effectively capturing both short- and long-term dependencies in the data, yet it does not comprehensively model the complex multi-scale patterns arising during the degradation process, restricting the model's capacity for cross-multi-scale feature extraction. Besides, although the serial AD-TCN architecture independently captures critical features along the time and sensor dimensions, it does not sufficiently resolve cross-dimensional interactions or address the redundancy introduced by feature coupling, thereby hindering the effective integration of complex dependencies across multiple time

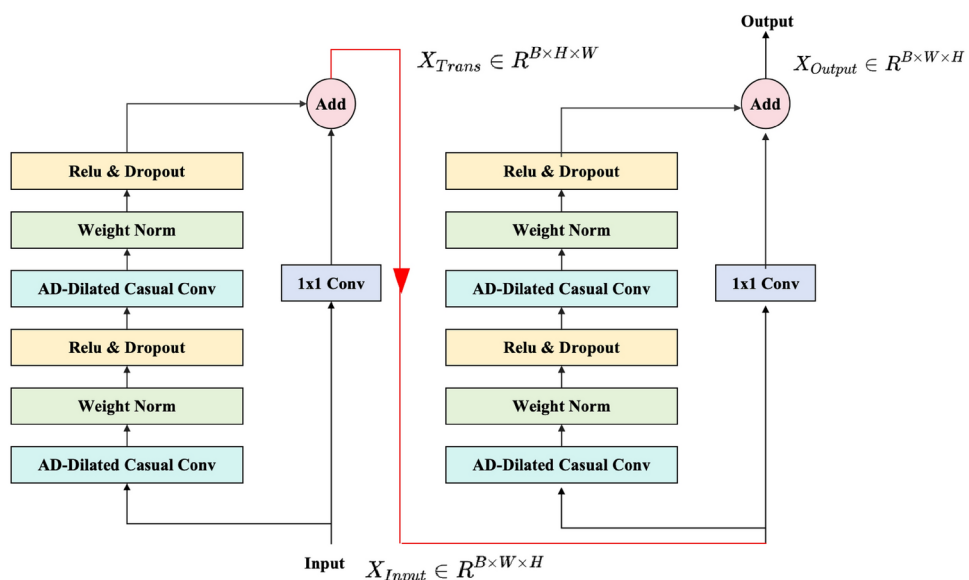


Fig. 4. Structure of AD-TCN.

scales and sensor dimensions and making it challenging to achieve feature fusion and interaction in degradation time-series data.

To address these limitations in the serial AD-TCN architecture, we propose the MSCDIM module. By employing parallel multi-scale convolutional branches, the module achieves more comprehensive and in-depth multi-scale feature extraction across both the temporal and sensor dimensions, thereby avoiding the feature stacking issues caused by information redundancy and limited dimensional interaction in serial architectures, while effectively capturing the complex interactions among features at different time scales and across various dimensions, ultimately enabling the efficient fusion of critical features. Beside, to address cross-dimensional information fusion between the temporal dimension and the sensor feature dimension, we design operations such as max pooling, average pooling, and nonlinear transformations to compute distinct attention weights for each dimension-enabling the model to adaptively emphasize pivotal feature information while suppressing redundant data-and further employ concatenation and segmentation strategies to ensure the effective integration of key features across diverse dimensions and time scales, thereby enhancing model robustness and its capacity to capture crucial patterns in complex degradation scenarios.

Therefore, by employing parallel multi-scale feature extraction and cross-dimensional information fusion, the MSCDIM module heightens the model's capacity to interpret pivotal information from both the temporal and sensor feature dimensions, providing richer and more precise feature representations for equipment degradation RUL prediction.

As shown in Fig. 5, the input data $X (X \in R^{B \times W \times H})$ undergoes an unsqueeze operation to add a channel dimension, adjusting the shape of the input data to meet the requirements for subsequent convolution operations, resulting in the new data tensor $X' (X' \in R^{B \times 1 \times W \times H})$.

$$X' = \text{Unsqueeze}(X, \text{dim} = 1) \in R^{B \times 1 \times W \times H} \tag{18}$$

By applying Conv2d operations at three distinct scales, the model provides multiple receptive fields that capture localized spatial information at varying scales within the $W \times H$ space. And by performing convolution operations at multiple scales to extract spatial features of varying levels, the model conducts local modeling on the input data across diverse ranges, capturing everything from fine details to global structures, thereby enhancing its capacity to represent multi-level spatial patterns and providing a comprehensive feature representation that facilitates subsequent feature fusion.

$$X_i = \text{Conv2D}_i(X'), i = 1, 2, 3 \tag{19}$$

Subsequently, a squeeze operation is applied to revert the data shape to its original dimensions, ensuring consistency between output and input shapes and preventing shape mismatches introduced by multi-scale convolution operations. Moreover, feature information from different scales is compressed and fused into a single output Y , thereby achieving a more comprehensive and enriched feature representation, which serves as a sufficient input for subsequent feature processing and integration.

$$Y = \sum_{i=1}^3 (\text{Squeeze}(X_i, \text{dim} = 1)) \tag{20}$$

Next, the corresponding features of the fused feature representation Y are aggregated and mined along the W and H dimensions. The features in each dimension are aggregated and linearly transformed, while pooling operations are applied to compress the features from different dimensions, reducing dimensional redundancy while preserving key global statistical features.

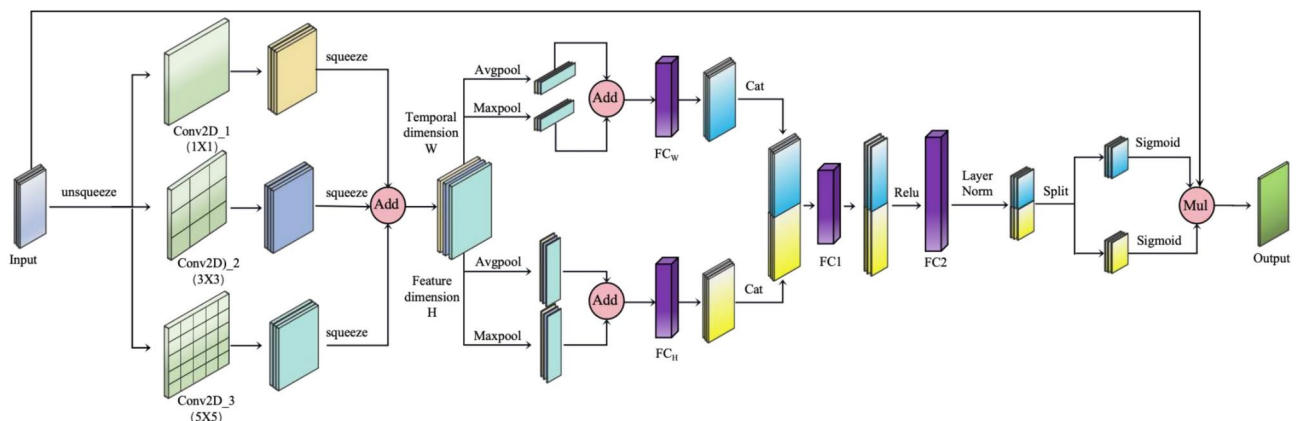


Fig. 5. Structure of MSCDIM.

In this process, pooling operations are performed along the W dimension, thereby aggregating the H -values in each column. The application of both average pooling and max pooling serves to reduce the spatial dimensions of the feature map, while ensuring the retention of the most significant features and global information at each position. Furthermore, a fully connected layer is employed to facilitate the mapping of features, thereby enhancing the model's capacity to represent spatial information.

$$Y_W = FC_W(Avgpool_H(Y) + Maxpool_H(Y)) \quad (21)$$

Specifically, $Avgpool_H$ applies average pooling along the H -dimension to extract overall trends and global features, thereby capturing common information at every position. And the $Maxpool_H$ performs max pooling along the H -dimension, highlighting extreme values within the features and enabling the model to focus on critical local information at each position. Subsequently, the FC_W (a fully connected layer) processes and transforms the pooled features along the W -dimension, learning intricate relationships among the input features and strengthening the representation capacity of W -dimensional information. Finally, the final output Y_W encapsulates the W -dimensional feature representation after pooling and processing with the fully connected layer.

In a similar manner, performing pooling along the H -dimension yields the feature representation Y_H . Next, a concatenation operation combines the key features from both the W -dimension and the H -dimension into a unified, high-dimensional representation Y_{cat} that thoroughly captures the complex correlations present in degradation data.

By means of this concatenation operation, features from the time and sensor dimensions are effectively fused, enabling cross-dimensional information integration while preserving each dimension's structural characteristics, ultimately offering a more comprehensive feature representation for subsequent integration and model training. Subsequently, the fused features undergo additional refinement through fully connected layers, nonlinear activation functions, and normalization, ensuring deeper feature representation and numerical stability; consequently, the model is better equipped to handle complex multi-dimensional features and multi-scale dependencies in equipment degradation RUL prediction tasks, while also improving its robustness and generalization capabilities.

$$Y_H = FC_H(Avgpool_W(Y) + Maxpool_W(Y)) \quad (22)$$

$$Y_{cat} = Cat(Y_W, Y_H) \quad (23)$$

$$Z = LayerNorm(FC_2(Relu(FC_1(Y_{cat})))) \quad (24)$$

Where the FC_1 , FC_2 and FC_H are fully connected layers, the $Avgpool_W$ and $Maxpool_W$ operations are applied along the W -dimension to aggregate features across the spatial dimension, the $Relu$ is the activation function, the Cat is the concatenation operation, and the $LayerNorm$ is the layer normalization operation.

Finally, to further emphasize the salient features of the time and sensor dimensions, the MSCDIM module employs a $Split$ operation to isolate these dimensional representations, then applies a $\sigma(\cdot)$ function for nonlinear mapping to derive attention weights corresponding to each dimension, thereby adaptively focusing on the key features across different dimensions. And the generated attention weights Q_W and Q_H are multiplied element-wise with the original input X to achieve cross-dimensional information fusion. By adaptively accentuating key features in both the temporal and sensor dimensions while suppressing redundant or irrelevant features, the model is able to more effectively capture deep and complex features as well as long- and short-term dependencies, ultimately producing the final output Z_{out} .

$$Z_W, Z_H = Split(Z), Z_W \in^{B \times 1 \times W}, Z_H \in^{B \times 1 \times H} \quad (25)$$

$$\sigma_W = \delta(Z_W^T), \sigma_H = \delta(Z_H) \quad (26)$$

$$Z_{out} = X \cdot \sigma_W \cdot \sigma_H \quad (27)$$

Where, the $Split$ denotes the feature splitting operation, and $\sigma(\cdot)$ represents the application of the Sigmoid activation function, which performs a nonlinear transformation on the features, compressing the output values to the range (0, 1).

Experimental validation

In this section, the effectiveness of the proposed method is validated through experiments conducted using the widely adopted C-MAPSS and N-CMAPSS datasets, with the results then being compared with those of state-of-the-art methods in the field. The structure of this chapter is organized as follows: Sect. [Dataset description](#) introduces the fundamental information and salient features of the C-MAPSS and N-CMAPSS datasets. Section [The data preprocessing](#) delineates the fundamental steps involved in the process of data preprocessing. Section [The hyperparameter experiments](#) presents hyperparameter simulation experiments and their analysis. Section [The comparison experiments](#) involves the execution of comparative experiments with other state-of-the-art methods, with the aim of validating the effectiveness of the proposed model. Section [The ablation experiment](#) performs ablation experiments to further verify the effectiveness of individual modules.

Furthermore, all experiments in this study were conducted on an experimental platform configured with PyTorch 2.2.2, Python 3.9.19, and an NVIDIA GeForce RTX 4080. To reduce the impact of randomness, all

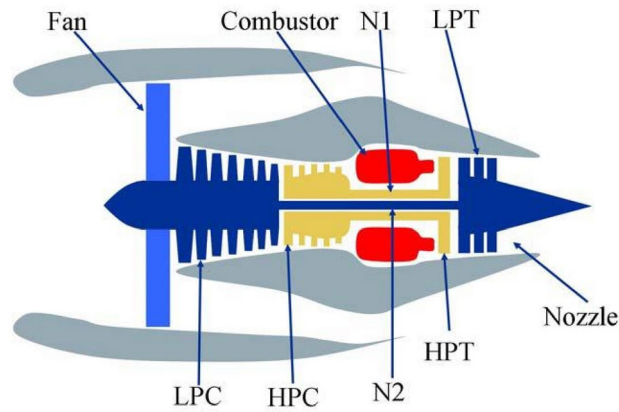


Fig. 6. Schematic diagram of the engine in C-MAPSS.

| Component | Description |
|-----------|--------------------------|
| Fan | Fan blades |
| LPC | Low-pressure compressor |
| HPC | High-pressure compressor |
| Combustor | Combustion chamber |
| N1 | Low-pressure rotor |
| N2 | High-pressure rotor |
| HPT | High-pressure turbine |
| LPT | Low-pressure turbine |
| Nozzle | Exhaust nozzle |

Table 1. Structure variables in the C-MAPSS.

experiments were repeated multiple times, and hyperparameter settings were optimized using the grid search method to identify the best configuration.

Dataset description

The C-MAPSS (Commercial Modular Aero-Propulsion System Simulation) dataset, provided by the NASA Ames Research Center, is a baseline dataset generated using commercial modular aero-propulsion system simulation software to simulate the full lifecycle of turbofan engines. And It is employed extensively for the validation of RUL prediction models for aircraft engines.

As demonstrated in Fig. 6, the aircraft engine is composed of components such as the fan and low-pressure compressor, with the detailed structure presented in Table 1. Furthermore, 21 distinct sensor types are installed at various locations on the engine to collect parameters throughout its lifecycle, with the specific sensor configurations listed in Table 2.

As shown in Table 3, the C-MAPSS dataset is comprised of four subsets: FD001, FD002, FD003, and FD004. Each subset is designed to simulate distinct operating conditions and failure modes, incorporating both training and testing sets. The training sets contain multi-sensor data collected throughout the entire lifecycle of the engines, while the testing sets include operational state data collected before failure points, along with the corresponding RUL labels. FD004 is regarded as the most complex subset due to its incorporation of the most variability in operating conditions and failure modes.

The N-CMAPSS is an updated iteration of the C-MAPSS dataset, capturing the full operational trajectories of multiple engines from normal function to failure. As shown in Table 4, it is divided into eight subsets (DS01 through DS08), featuring additional engine units, longer run cycles, and more complex fault modes, while also providing richer sensor data and more diverse operating conditions.

Furthermore, in order to accurately evaluate the performance of the proposed model, the Root Mean Square Error (RMSE) and Score metrics are adopted, which are widely used in such evaluations. The RMSE is utilised to quantify the discrepancy between the predicted and actual RUL values, and its definition is as follows:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (28)$$

| Number | Symbol | Description | Units |
|--------|-----------|---------------------------------|---------|
| S1 | T2 | Total temperature at fan inlet | °R |
| S2 | T24 | Total temperature at LPC outlet | °R |
| S3 | T30 | Total temperature at HPC outlet | °R |
| S4 | T50 | Total temperature at LPT outlet | °R |
| S5 | P2 | Pressure at fan inlet | psia |
| S6 | P15 | Total pressure in bypass duct | psia |
| S7 | P30 | Total pressure at HPC outlet | psia |
| S8 | Nf | Physical fan speed | rpm |
| S9 | Nc | Physical core speed | rpm |
| S10 | epr | Engine pressure ratio (P50/P2) | – |
| S11 | Ps30 | Static pressure at HPC outlet | psia |
| S12 | phi | Ratio of fuel flow to Ps30 | pps/psi |
| S13 | NRf | Corrected fan speed | rpm |
| S14 | NRc | Corrected core speed | rpm |
| S15 | BPR | Bypass Ratio | – |
| S16 | farB | Burner fuel–air ratio | – |
| S17 | htBleed | Bleed Enthalpy | – |
| S18 | Nf_dmd | Demanded fan speed | rpm |
| S19 | PCNfR_dmd | Demanded corrected fan speed | rpm |
| S20 | W31 | HPT coolant bleed | lbm/s |
| S21 | W32 | LPT coolant bleed | lbm/s |

Table 2. schematic diagram of the engine in the C-MAPSS.

| Dataset | FD001 | FD002 | FD003 | FD004 |
|------------------------------------|-------|-------|-------|-------|
| The number of training engines | 100 | 260 | 100 | 249 |
| The number of testing engines | 100 | 259 | 100 | 248 |
| The number of operating conditions | 1 | 6 | 1 | 6 |
| The number of fault modes | 1 | 1 | 2 | 2 |

Table 3. Description of C-MAPSS Dataset.

| Name | Units | Flight classes | Failure modes | Size |
|------|-------|----------------|---------------|-------|
| DS01 | 10 | 1,2,3 | 1 | 7.6M |
| DS02 | 9 | 1,2,3 | 2 | 6.5M |
| DS03 | 15 | 1,2,3 | 1 | 9.8M |
| DS04 | 10 | 2,3 | 1 | 10.0M |
| DS05 | 10 | 1,2,3 | 1 | 6.9M |
| DS06 | 10 | 1,2,3 | 1 | 6.8M |
| DS07 | 10 | 1,2,3 | 1 | 7.2M |
| DS08 | 54 | 1,2,3 | 1 | 35.6M |

Table 4. Description of N-CMAPSS Dataset.

Where \hat{y}_i represents the actual RUL of the i -th current sample, y_i represents the i -th predicted RUL, and N is the total number of samples. Additionally,, in the Score metric, a penalty mechanism is introduced in order to account for discrepancies between predictions and actual values. In instances where the predicted value deviates from the actual value, a penalty is applied, with greater penalties imposed when the predicted RUL exceeds the actual RUL. The metric is defined as follows:

$$Score = \begin{cases} \sum_{i=1}^N e^{-\frac{y_i - \hat{y}_i}{13} - 1}, & y_i < \hat{y}_i \\ \sum_{i=1}^N e^{\frac{y_i - \hat{y}_i}{10} - 1}, & y_i \geq \hat{y}_i \end{cases} \tag{29}$$

Where \hat{y}_i represents the actual RUL of the i -th current sample, y_i represents the i -th predicted RUL, and N is the total number of samples. Additionally, as demonstrated in Fig. 7, when the predicted RUL corresponds to the actual RUL, both RMSE and Score are equivalent to zero. However, as the prediction error increases, the RMSE grows linearly, while the Score exhibits exponential growth. Consequently, although both RMSE and Score are related to prediction errors, Score is more sensitive, tolerating smaller errors well but amplifying larger errors significantly²⁴.

The data preprocessing

Data selection and normalization

The C-MAPSS dataset comprises measurement data collated by 21 sensors. However, the correlation between the degradation data from different sensors and the RUL labels varies significantly. To address this, a correlation analysis is conducted using visualisation techniques to identify and exclude sensors with weak correlations to RUL. The specific correlations are illustrated in Fig. 8. And it can be observed that sensors S1, S5, S6, S10, S16, S18, and S19 maintain constant values throughout the engine lifecycle. In order to provide further validation of this hypothesis, the Pearson correlation coefficients between the sensor data and the RUL labels are computed. The outcomes, illustrated in Fig. 9, substantiate that the aforementioned sensors bear no correlation to the engine RUL labels. Consequently, the sensors identified as S2, S3, S4, S7, S8, S9, S11, S12, S13, S14, S15, S17, S20, and S21 are selected as inputs to the model²⁵.

To provide further validation of the model performance, we supplement our analysis with the N-CMAPSS dataset, which offers comprehensive coverage of fault modes, extended run cycles, and richer sensor data²⁶. Among its eight subsets, DS02 has been widely employed in numerous studies; accordingly, this research selects DS02 for additional verification. Specifically, DS02 comprises six training sets (Units 2, 5, 10, 16, 18, and 20) and three test sets (Units 11, 14, and 15). These three test sets are merged into a single combined test set, referred to as the WT test set, to facilitate a more robust evaluation. The specific parameters of this dataset are presented in Table 5, where Unit Num denotes the subset unit index, Sample Size refers to the number of samples (in millions), and Failure Mode indicates the type of degradation, categorized as either HPT (High Pressure Turbine Degradation) or LPT (Low Pressure Turbine Degradation). In addition, the 20 sensor parameters listed in Table 6 are selected as inputs to the model.

Besides, the discussion in this section is based on the C-MAPSS dataset, which includes fewer parameters than the N-CMAPSS dataset and is thus more suitable for detailed analysis and illustration. Despite the difference in scale, the two datasets exhibit a high degree of similarity in terms of data structure, degradation patterns, and sensor configuration, allowing insights derived from C-MAPSS to be transferable and representative for the more complex scenarios encountered in N-CMAPSS²⁷.

Furthermore, as shown in Fig. 10, in order to mitigate the impact of differences in magnitude and unit inconsistencies among sensors, a min-max normalization strategy is applied. This method involves mapping the data collected by each sensor to a unified range of [0, 1], thereby eliminating proportional differences between sensors²⁸. The efficacy of this normalization is twofold: it accelerates the model convergence process and improves training efficiency. The formula employed is as follows:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (30)$$

where X_{norm} represents the normalized data, X is the original data, X_{max} and X_{min} represent the maximum and minimum values in the data.

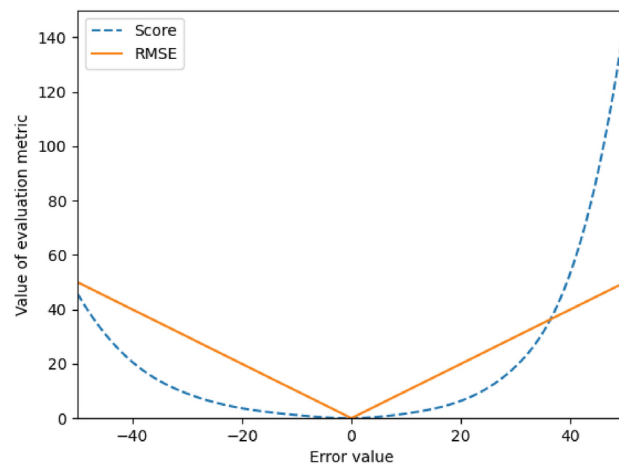


Fig. 7. The relationship between RMSE and Score.

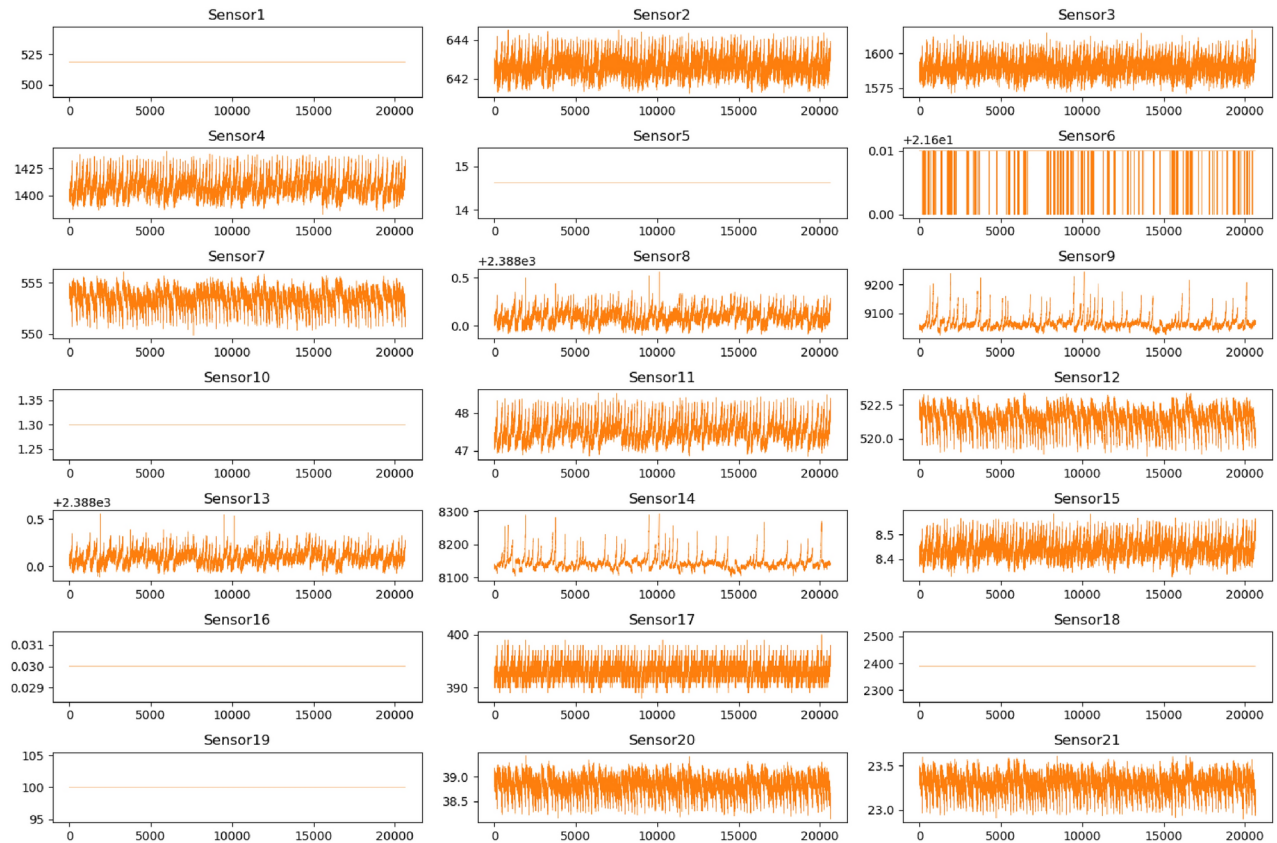


Fig. 8. The sensor value trends across the aircraft engine lifecycle.

The slide window

In the context of RUL prediction tasks, multi-time-step data of equivalent duration has been shown to contain a greater quantity of information than single-time-step data. Consequently, as shown in Fig. 11, when processing time-series data, the utilisation of a sliding time window method along the time axis is a prevailing practice for the segmentation of time-series data into multiple samples of equal length. These samples are then matched with their corresponding RUL labels²⁹. This methodological approach facilitates the capture of more detailed degradation trends in the data, thereby enhancing the model's predictive performance.

Thus, to achieve this objective, a range of time window sizes W have been configured in this study. The purpose of this configuration is to determine the optimal time window size for the proposed model, with the objective of achieving the best possible performance. Further details can be found in the hyperparameter experiment section.

The RUL label reconstruction

During the operation of the equipment, the RUL label is set to 0 at the time of failure and calculated backwards as the remaining useful life for each time point relative to the failure time. However, within the context of the equipment's lifecycle, it is generally assumed that the initial operation phase represents a healthy state. And as the equipment operates, the ageing process accelerates, leading to a reduction in the equipment's remaining lifespan. As shown in Fig. 12 In order to address this issue, a piecewise linear model is employed, with the RUL label threshold set at 125 during the healthy state phase²³. This methodological approach facilitates the model's capacity to prioritise degradation characteristics in subsequent phases of operation, enhancing its resilience and prediction precision.

The hyperparameter experiments

In this section, the investigation focuses on two key hyperparameters in the model: the window_size (W) and the hidden_dim (H). By this way to provide preliminary validation of the model's effectiveness, and explore the impact of hyperparameter variations under a range of complex operating conditions.

In order to ascertain the impact of differing window_size (W) on the RUL prediction task within the confines of the experimental setup, reference was made to analyses conducted in related studies that experimented with varying window sizes (e.g. 1, 10, 20, 30, 40, 50, 60) on the dataset³⁰. Consequently, we set the range of w to 10, 20, 30, 40, 50, 60. Moreover, analogous research investigated the impact of the hyperparameter hidden_dim (H), evaluating values of $H = 16, 32, 64, 128$ ³¹. Consequently, the range of H was set to 16, 32, 64, 128. Thus, we

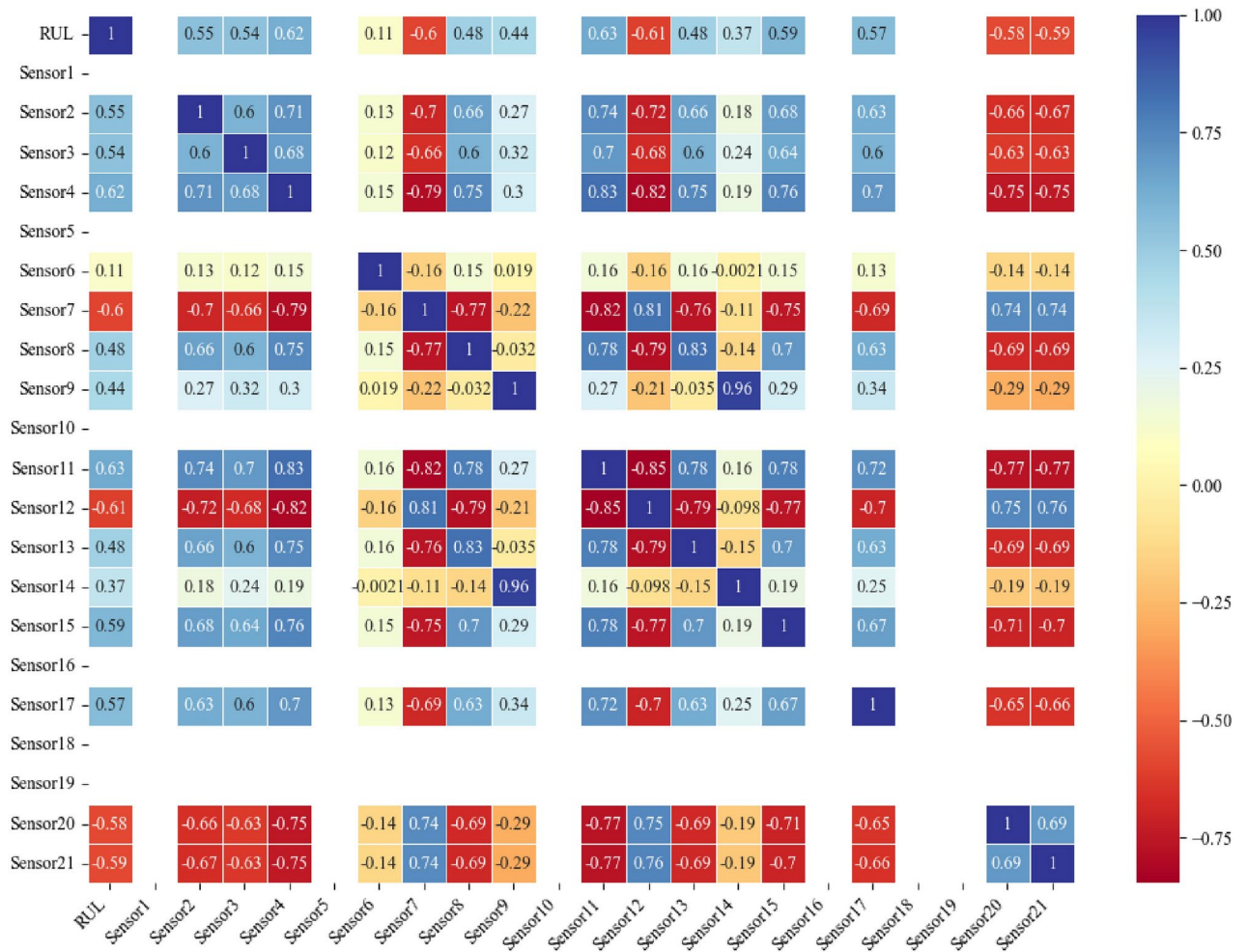


Fig. 9. Pearson coefficients between sensor readings and RUL.

| Dataset | Unit num | Sample size | Failure mode |
|--------------|----------|-------------|--------------|
| Training set | 2 | 0.85 | HPT |
| | 5 | 1.03 | HPT |
| | 10 | 0.95 | HPT |
| | 16 | 0.77 | HPT LPT |
| | 18 | 0.89 | HPT LPT |
| | 20 | 0.77 | HPT LPT |
| Test set | 11 | 0.66 | HPT LPT |
| | 14 | 0.16 | HPT LPT |
| | 15 | 0.43 | HPT LPT |

Table 5. The statistics of DS02 Dataset in N-CMAPSS.

conducted experiments within the parameter ranges of $w = [10, 20, 30, 40, 50, 60]$ and $H = [16, 32, 64, 128]$ to determine the optimal hyperparameter configuration for our model.

The size of the window is the primary factor in determining the length of historical data that the model is able to observe along the temporal dimension, enabling it to capture the dynamic evolution characteristics of the time series. And the hidden_dim primarily maps the 14 original sensor features from their raw space, which may contain redundancy and noise, to a more expressive latent representation space. Because the direct input of raw sensor feature data into the model not only makes it challenging to accurately capture the complex nonlinear relationships between features but also limits the model's ability to represent complex equipment degradation patterns due to fixed low-dimensional feature representations. This has the potential to compromise learning efficiency and adversely affect prediction performance.

| Num | Symbol | Description | Units |
|-----|--------|------------------------------------|-------|
| 1 | alt | Altitude | ft |
| 2 | XM | Flight Mach number | - |
| 3 | TRA | Throttle-resolver angle | % |
| 4 | Wf | Fuel flow | pps |
| 5 | Nf | Physical fan speed | rpm |
| 6 | Nc | Physical core speed | rpm |
| 7 | T2 | Total temperature at fan inlet | °R |
| 8 | T24 | Total temperature at LPC outlet | °R |
| 9 | T30 | Total temperature at HPC outlet | °R |
| 10 | T40 | Total temperature at burner outlet | °R |
| 11 | T48 | Total temperature at HPT outlet | °R |
| 12 | T50 | Total temperature at LPT outlet | °R |
| 13 | P15 | Total pressure in bypass duct | psia |
| 14 | P2 | Total pressure at fan inlet | psia |
| 15 | P21 | Total pressure at fan outlet | psia |
| 16 | P24 | Total pressure at LPC outlet | psia |
| 17 | Ps30 | Static pressure at HPC outlet | psia |
| 18 | P30 | Total pressure at HPC outlet | psia |
| 19 | P40 | Total pressure at burner outlet | psia |
| 20 | P50 | Total pressure at LPT outlet | psia |

Table 6. Sensor descriptions in the N-CMAPSS Dataset.

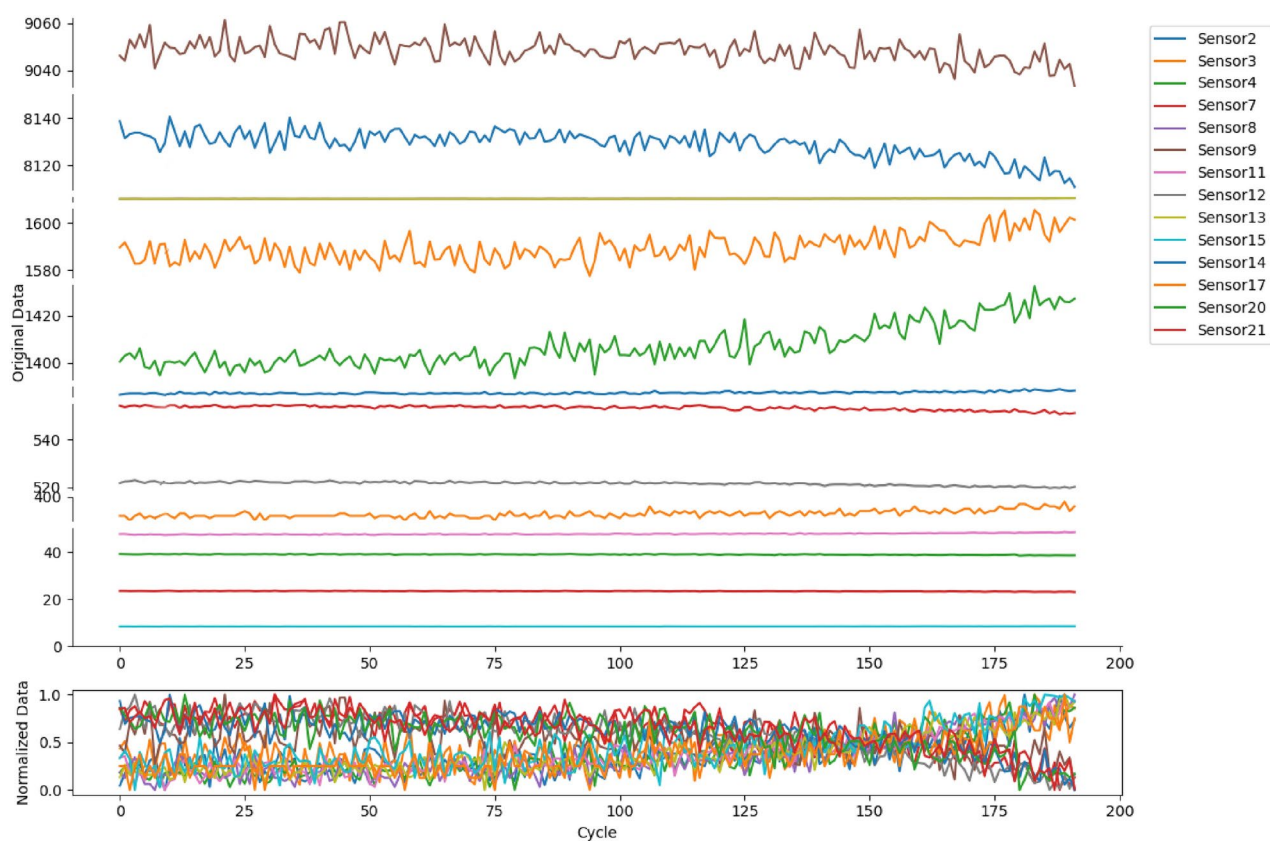


Fig. 10. The normalization of C-MAPSS.

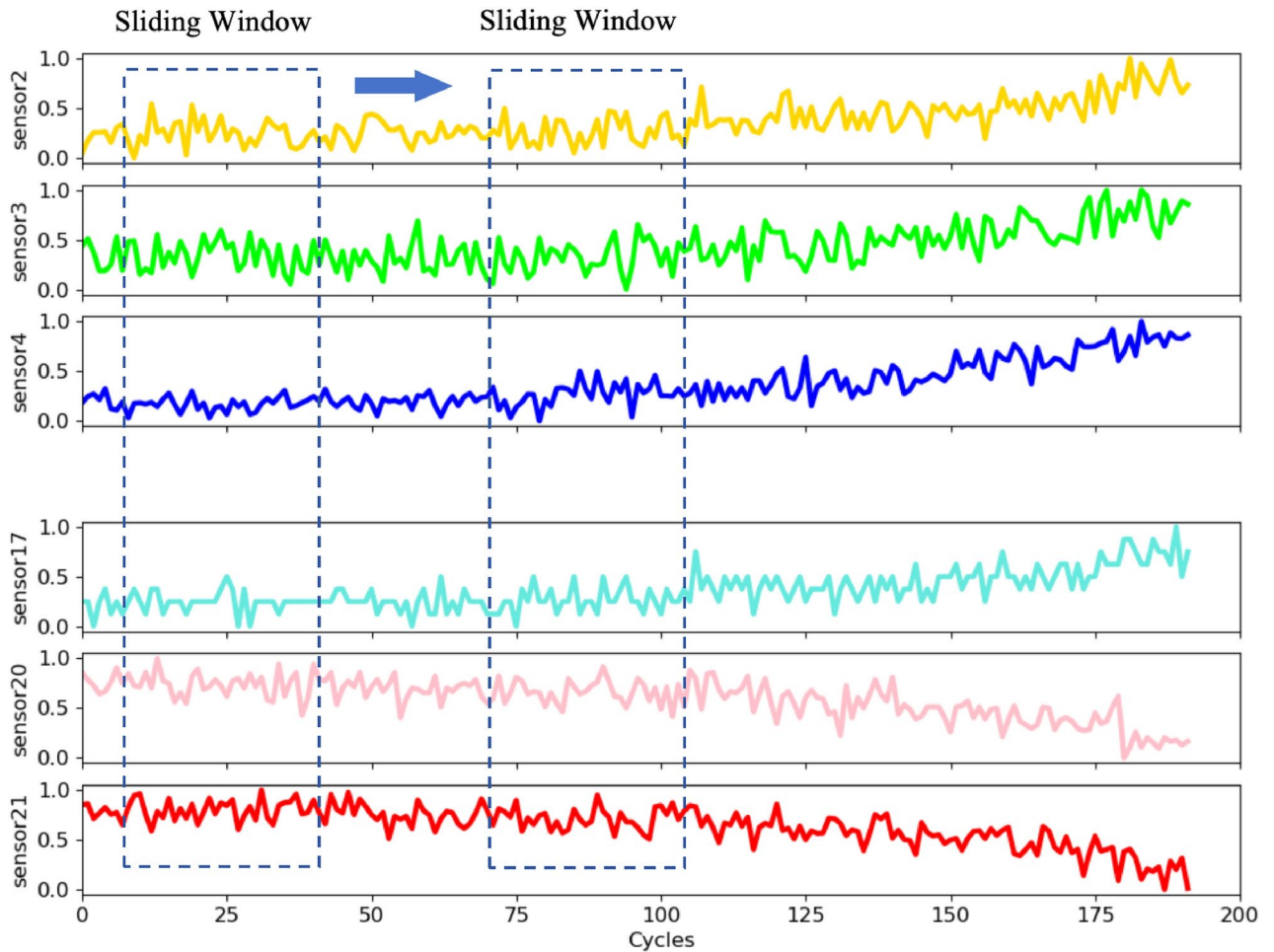


Fig. 11. The process of the slide window.

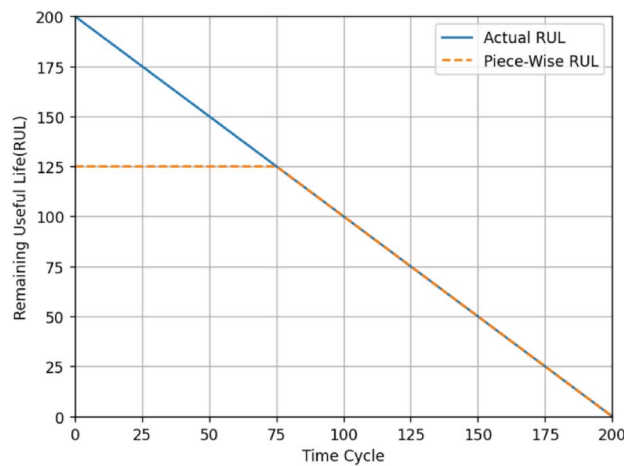


Fig. 12. Segmented linear function of RUL.

The introduction of a fully connected layer facilitates the mapping of the 14-dimensional raw sensor features into a new H -dimensional space, thereby enabling the effective extraction of more expressive high-dimensional features. Besides, the model is able to adaptively adjust the weight coefficients through backpropagation, thereby ensuring the comprehensive capture of relevant information for the prediction task whilst enhancing the model's representational capacity without a significant increase in computational overhead. Therefore, by tuning the

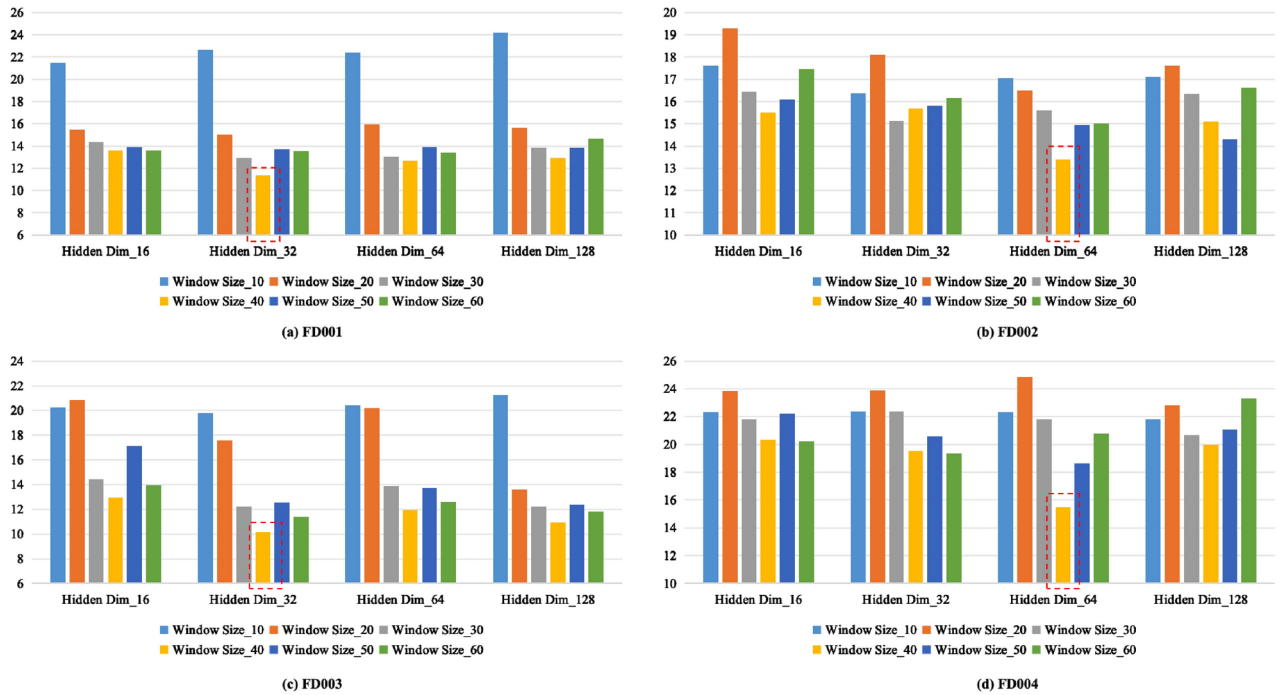


Fig. 13. The RMSE with different hyperparameters on the four datasets.

| Hyperparameter | Value | Params |
|----------------------------|-------------|--------|
| Sliding time window size | 40 | W |
| Sliding time window stride | 1 | sl |
| Hidden dim | 32 64 32 64 | H |
| Batch size | 256 | bs |
| Optimizer | Adam | opt |
| Learning rate | 0.001 | lr |
| The number of epochs | 300 | epoch |

Table 7. Experimental hyperparameter configuration.

window size and hidden dimension hyperparameters, the model can achieve improved key information capture and enhanced representational capacity across both the temporal and sensor feature dimensions.

As demonstrated in Fig. 13, across the four subsets of the C-MAPSS dataset, the model attains optimal overall performance when $W=40$, with the lowest RMSE values. Furthermore, the selection of the hyperparameter H is contingent upon the complexity of the dataset. In FD001 and FD003, where the operation conditions are set to 1 and the data is relatively simple, a smaller feature space ($H = 32$) suffices to capture the relevant features. Conversely, in FD002 and FD004, with six operation conditions and more complex sensor data, a higher-dimensional feature space ($H = 64$) is required to capture richer feature distributions and engine degradation patterns.

Therefore, all the hyperparameter settings adopted in the experiments of this paper are shown in Table 7.

The comparison experiments

In this chapter, we compare our method with related state-of-the-art approaches in the field using the C-MAPSS dataset and N-CMAPSS dataset to validate the effectiveness of the proposed model.

- MSIDSN³²: A multi-scale integrated deep self-attention network (MSIDSN) is proposed for RUL prediction, leveraging multi-scale feature extraction, enhanced recurrent modules, and efficient loss functions to process aero-engine degradation data and deliver accurate predictions.
- AGATT³³: An RUL prediction model integrating attention mechanisms, graph attention networks (GATs), and transformers, employs spatial and temporal channels to extract and fuse dependencies across sensors and time steps for accurate predictions.
- Res-HSA³⁴: A residual hybrid network with a self-attention mechanism for RUL prediction, enhancing feature extraction and mitigating abnormal health fluctuations for improved accuracy in rotating machinery maintenance.

- **IMDSSN²³**: An integrated multi-head dual sparse self-attention network that combines ProbSparse and Log-Sparse mechanisms for efficient and accurate remaining useful life prediction, addressing computational complexity and enhancing feature extraction from temporal data.
- **ABGRU³⁵**: A model integrating an attention mechanism with a Gate Recurrent Unit (GRU) network to predict RUL, improving feature weighting and enhancing prediction accuracy on the NASA C-MAPSS dataset.
- **MSDCNN¹⁵**: A hybrid approach combining a multi-scale deep convolutional neural network with Long Short-Term Memory (LSTM) to predict the RUL of aircraft engines.
- **PHNN¹⁶**: A novel method combining a 1-D Convolutional Neural Network (1-DCNN) and Bidirectional Gated Recurrent Unit (BiGRU) for predicting the RUL, effectively handling spatial and temporal features.
- **SA-MSTCN-GFA³⁶**: An Attention-based Multi-Scale Temporal Convolutional Network (MSTCN) integrating Self-Attention and Global Fusion Attention mechanisms.
- **ATCN¹⁸**: An attention-based temporal convolutional network (ATCN) leveraging an improved self-attention mechanism and squeeze-and-excitation mechanism for aero-engine RUL prediction.
- **MHT¹⁴**: A multiscale Hourglass-Transformer (MHT) incorporating hourglass-shaped multiscale feature extraction and pyramid self-attention mechanisms for aero-engine RUL prediction.
- **CATA-TCN¹⁹**: A dual attention framework, CATA-TCN, integrating channel attention and temporal attention within a temporal convolutional network to enhance feature extraction.
- **Transformer-RW³⁷**: A transformer-based RUL estimation method combined with an Adaptive RUL-wise Re-weighting (ARR) technique to address data imbalance.
- **DL-PRP³⁸**: A deep learning-based probabilistic RUL prediction model that optimizes prior distribution parameters to enhance prediction accuracy and robustness.
- **DVGTformer³⁹**: A dual-view graph Transformer (DVGTformer) model for RUL prediction that integrates temporal and spatial information from multi-sensor signals to enhance prediction accuracy and robustness.
- **MAGCN¹⁵**: A multi-scale adaptive graph neural network that captures scale-specific temporal and inter-variable dependencies for improved multivariate time series forecasting.
- **LOGO⁴⁰**: A local-global correlation fusion framework that integrates dynamic and stable sensor relationships using graph neural networks to jointly model spatial and temporal dependencies for accurate RUL prediction.
- **PI-STHNN²⁷**: A physics-informed spatio-temporal hybrid neural network that integrates prior physical knowledge and parallel spatio-temporal feature extraction to enhance RUL prediction of aircraft engines.

As demonstrated in Table 8, the proposed method attains RMSE values of 11.39, 13.41, 10.17, and 15.28 across the four subsets of the C-MAPSS dataset, with an average RMSE of 12.56. The corresponding Score values are 170.52, 563.03, 134.87, and 877.44, with an average Score of 436.47. And in order to further evaluate the performance differences between our model and state-of-the-art methods, the average values of the metrics were calculated and a comparative analysis conducted. The findings indicate that, in comparison to the second-best method, our model attains RMSE enhancements of 0.74%, 3.78%, and 1.16% in FD002, FD003, and FD004, correspondingly. Furthermore, in FD002 and FD004, Score enhances by 24.53% and 20.77%, respectively, while maintaining a consistently high level of performance in other cases. Overall, compared to the second-best method, the proposed model has been shown to achieve an average RMSE improvement of 3.83% and an average Score improvement of 25.36%. These results demonstrate the strong competitiveness of the proposed method in RUL prediction tasks, particularly in the more complex FD002 and FD004 datasets. The superior performance of the proposed model can be attributed to its ability to map raw sensor features from their original space into a more expressive H -dimensional space. This facilitates the model's capture of complex linear relationships among sensor features, thereby enhancing its predictive performance.

| Model | RMSE | | | | | Score | | | | |
|--------------|-------|-------|-------|-------|-------|--------|---------|--------|---------|---------|
| | FD001 | FD002 | FD003 | FD004 | Avg | FD001 | FD002 | FD003 | FD004 | Avg |
| MSIDSN | 11.74 | 18.26 | 12.04 | 22.48 | 16.13 | 205.00 | 2046.00 | 196.00 | 2910.00 | 1339.25 |
| AGATT | 11.45 | 13.87 | 10.88 | 16.43 | 13.16 | 204.00 | 836.00 | 205.00 | 1567.00 | 703.00 |
| Res-HSA | 11.91 | 17.27 | 11.88 | 17.43 | 14.62 | 227.00 | 1199.00 | 272.00 | 2508.00 | 1051.50 |
| IMDSSN | 12.14 | 17.40 | 12.35 | 19.78 | 15.42 | 206.11 | 1775.15 | 229.54 | 2852.81 | 1265.90 |
| ABGRU | 12.83 | 17.97 | 13.23 | 21.55 | 16.40 | 221.54 | 2072.21 | 279.18 | 3625.77 | 1549.68 |
| MSDCNN | 12.96 | 18.70 | 11.78 | 21.57 | 16.25 | 256.59 | 1873.86 | 211.99 | 2699.34 | 1260.45 |
| PHNN | 12.99 | 16.10 | 12.48 | 18.29 | 14.97 | 223.30 | 1201.35 | 268.83 | 2304.08 | 999.39 |
| SA-MSTCN-GFA | 10.62 | 15.60 | 10.57 | 15.46 | 13.06 | 165.00 | 1096.00 | 178.00 | 1115.00 | 638.50 |
| ATCN | 11.18 | 15.82 | 11.34 | 17.80 | 14.04 | 194.25 | 1210.57 | 249.19 | 1934.86 | 897.22 |
| MHT | 11.92 | 13.70 | 10.63 | 17.73 | 13.50 | 215.00 | 746.00 | 150.00 | 1572.00 | 670.75 |
| CATA-TCN | 12.80 | 17.61 | 13.16 | 21.04 | 16.15 | 234.31 | 1361.23 | 290.63 | 2303.42 | 1047.40 |
| Transorm-RW | 11.36 | 18.97 | 11.28 | 20.69 | 15.58 | 192.22 | 2433.15 | 133.41 | 2842.44 | 1400.31 |
| DL-PRP | 11.96 | 13.51 | 11.40 | 17.58 | 13.61 | 233.40 | 902.13 | 255.60 | 1704.59 | 773.93 |
| DVGTformer | 11.33 | 14.28 | 11.89 | 15.50 | 13.25 | 179.75 | 797.26 | 254.55 | 1107.50 | 584.77 |
| Ours | 11.39 | 13.41 | 10.17 | 15.28 | 12.56 | 170.52 | 563.03 | 134.87 | 877.44 | 436.47 |

Table 8. Comparative experiment results on C-MAPSS dataset.

Table 9 shows that our method achieves RMSE values of 4.98, 8.33, 4.24, and 5.76 for Units 11, 14, 15, and the WT test set on the N-CMAPSS dataset, with corresponding Score values of 3545.67, 1281.17, 1381.75, and 6687.09. Compared with the second-best method, our model improves the RMSE metric by 2.68%, 1.39%, and 1.69% on Units 11, 15, and WT, respectively, and achieves Score improvements of 17.25%, 24.06%, 24.39%, and 22.64% on Units 11, 14, 15, and WT, respectively. These results suggest that our method offers a consistent performance advantage across multiple test scenarios. The improvements in both RMSE and Score demonstrate the model's effectiveness in learning degradation trends and capturing key temporal-spatial features relevant to RUL prediction.

To visually demonstrate the predictive capability of the proposed method on the C-MAPSS dataset, we compared the actual RUL and predicted RUL of selected engines through visualization (as shown in Fig. 14), and further a typical engine prediction result is selected in each dataset for further demonstration (as shown in Fig. 15). It is evident that, with the exception of minor deviations at specific time points, the predicted curves demonstrate a high degree of alignment with the actual RUL curves in the majority of cases. This finding suggests that the proposed method demonstrates a high degree of accuracy and stability in engine RUL prediction tasks.

In addition, due to the large number of samples in the N-CMAPSS test datasets, we randomly selected 150 samples from unit11, 14, 15 and WT for visualization to assess the effectiveness of the proposed model in the RUL prediction task. As illustrated in Fig. 16), the predicted RUL values closely follow the ground truth across most time steps, indicating that the model is capable of accurately capturing the degradation trend.

The ablation experiment

In this section, the effectiveness of the proposed method is validated and the generalizability of the MSCDIM module within other typical approaches is further evaluated through the design of four sets of ablation experiments as follows:

- Baseline TCN Model: A standard TCN model is used as the baseline for comparison.
- AD-TCN: A TCN model incorporating the dynamic adaptive dilation factor, designed to evaluate the effectiveness of the adaptive factor in temporal modeling tasks.
- TCN-M: A model that integrates the MSCDIM module into the baseline TCN model to validate the effectiveness of the module in feature extraction and fusion.
- AD-TCN-M: A model that integrates the MSCDIM module into the AD-TCN model, representing the complete architecture of the proposed method.

As demonstrated in Table 10, the incorporation of the dynamic adaptive dilation factor into the baseline TCN model resulted in a reduction of RMSE by 4.54%, 12.75%, 11.85%, and 12.53% across the four subsets, with an average improvement of 10.42%. A similar enhancement was observed in the Score metric, which demonstrated improvements of 10.64%, 38.22%, 23.05%, and 48.05%, with an average improvement of 29.99%. These results suggest that incorporating the dynamic adaptive dilation factor enables the model to adjust its focus on different time steps adaptively, effectively capturing both long-term and short-term dependencies in the data. Furthermore, the incorporation of the MSCDIM module into the baseline TCN model resulted in RMSE improvements of 2.37%, 10.37%, 8.88%, and 15.46% across the four subsets, with an average improvement of 9.27%. For the Score metric, the improvements were 13.83%, 45.78%, 17.89%, and 46.64%, with an average improvement of 29.99%. And the incorporation of the MSCDIM module into the AD-TCN model resulted in an enhancement of RMSE by 4.86%, 7.02%, 7.26%, and 23.68% across the subsets, with an average improvement of 10.70%. A similar trend was observed in the Score metric, which exhibited enhancements of 16.23%, 18.29%, 23.02%, and 65.44%, yielding an average improvement of 30.74%. These results demonstrate that the MSCDIM module effectively suppresses redundant information and integrates features in multi-scale cross-dimensional interactions. In addition, the module demonstrated superior performance on the more complex FD002 and FD004 datasets, indicating its strong applicability to complex time-series data modelling tasks.

Besides, the Table 11 reports the ablation study results on the N-CMAPSS dataset. When the baseline TCN is enhanced with an adaptive dilation factor, the RMSE improvements of 7.36%, 2.61%, 10.86%, and 5.62% are observed for Units 11, 14, 15, and the WT test set, respectively, along with Score improvements of 10.65%,

| Model | RMSE | | | | Score | | | |
|------------|--------|--------|--------|------|---------|---------|---------|----------|
| | unit11 | unit14 | unit15 | WT | unit11 | unit14 | unit15 | WT |
| CNNGRU | 6.56 | 13.57 | 5.74 | 7.54 | 5578.30 | 3507.92 | 2356.28 | 11442.49 |
| IMDSSN | 5.68 | 9.82 | 4.57 | 6.02 | 4245.76 | 1581.97 | 1569.84 | 7397.57 |
| DAMCNN | 6.09 | 11.82 | 5.27 | 6.82 | 4698.23 | 2154.13 | 1827.57 | 8679.93 |
| MLP | 6.92 | 10.47 | 5.76 | 7.10 | 6092.26 | 2238.15 | 2616.44 | 10946.84 |
| Dual-Mixer | 5.64 | 12.46 | 5.72 | 6.88 | 4284.98 | 2462.82 | 1895.84 | 8643.63 |
| MAGCN | 8.33 | 10.52 | 5.11 | 7.37 | 13835 | 4566 | 3623 | 20821 |
| LOGO | 5.73 | 6.72 | 4.54 | 6.07 | 7509 | 1940 | 3017 | 15127 |
| PI-STHNN | 5.12 | 6.08 | 4.30 | 5.86 | 6413 | 1687 | 2795 | 13672 |
| Ours | 4.98 | 8.33 | 4.24 | 5.76 | 3545.67 | 1281.17 | 1381.75 | 6687.09 |

Table 9. Comparative experiment results on N-CMAPSS dataset.

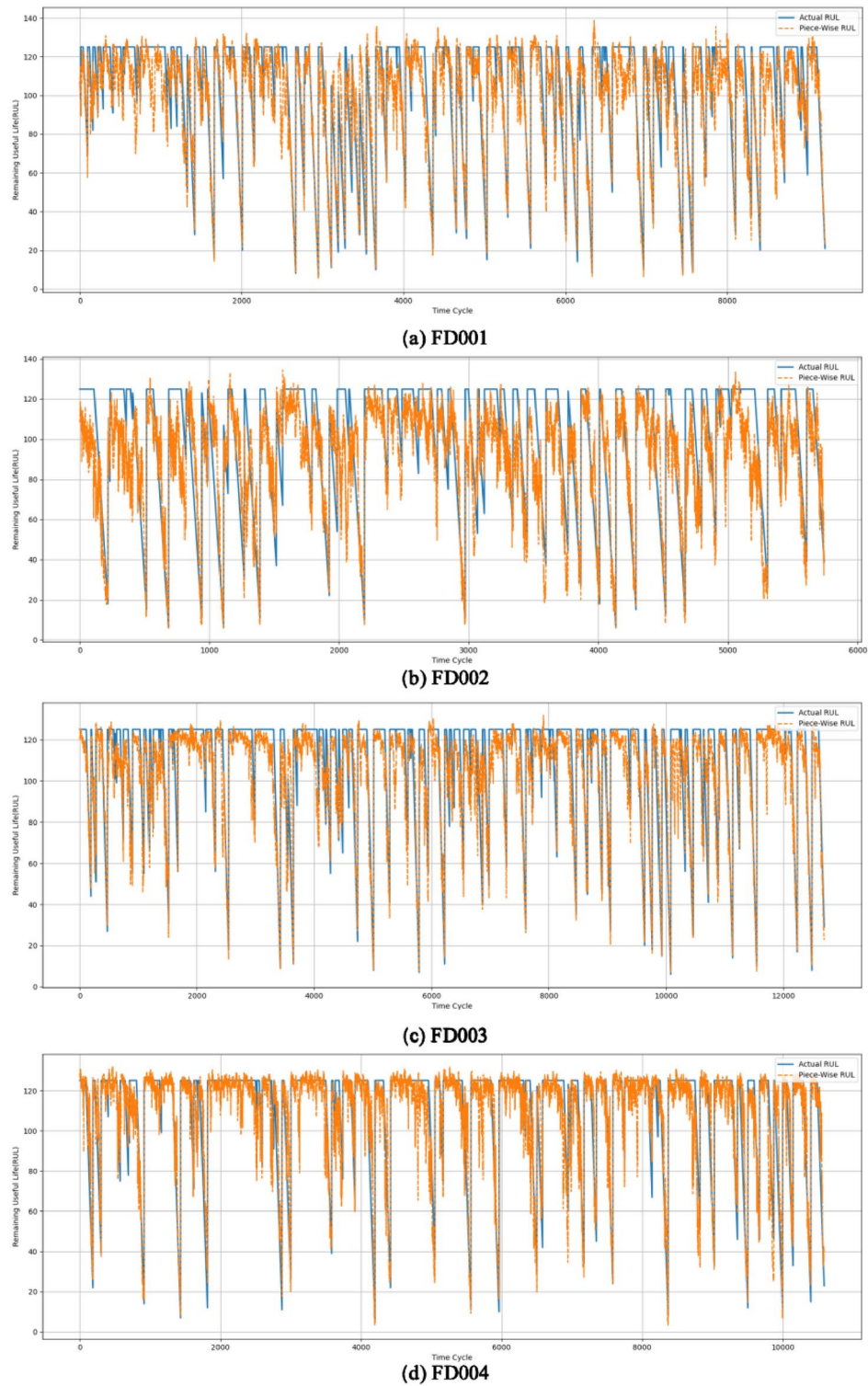


Fig. 14. Remaining useful life prediction on the C-MAPSS dataset.

5.06%, 17.83%, and 10.03% . Further integrating the MSCDIM module into the TCN yields additional RMSE improvements of 10.13%, 30.15%, 14.10%, and 20.44% and Score enhancements of 13.29%, 49.66%, 16.15%, and 27.38%. Overall, our complete method surpasses the baseline TCN by achieving RMSE improvements of 12.16%, 45.69%, 26.79%, and 24.12% and Score enhancements of 18.72%, 66.38%, 33.87%, and 34.83% on Units 11, 14, 15, and WT, respectively. These results suggest that both the adaptive dilation factor and the MSCDIM module contribute to improved performance, demonstrating the generality and effectiveness of our approach for RUL prediction.

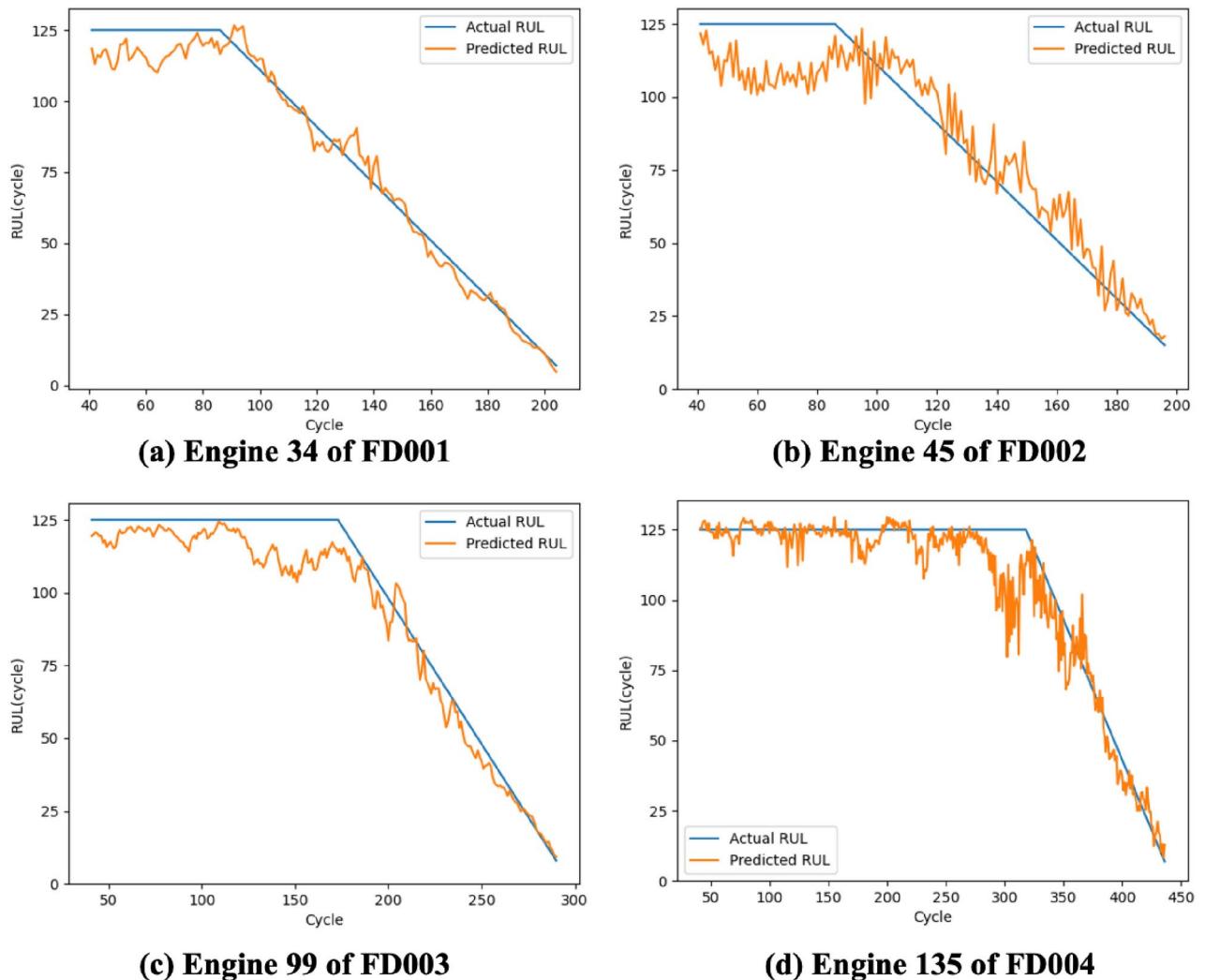


Fig. 15. Partial engine prediction results in four subsets of C-mapss.

To further explore the generalizability of the proposed MSCDIM module, we integrated it into widely used models in the field and conducted experimental validation. The results are as follows:

- CNN-GRU¹⁶: A multi-dimensional feature fusion network that combines Convolutional Neural Networks and Gated Recurrent Units.
- DAMCNN⁴¹: A Deep Attention Mechanism Convolutional Neural Network designed to capture long-term dependencies in time series data.
- IMDSSN²³: An Integrated Multi-head Dual Sparse Self-Attention Network (IMDSSN) based on an improved Transformer.
- MLP⁴²: A deep network architecture for multidimensional feature extraction that can be used for time series data.
- Dual-Mixer³¹: A dual-path feature extraction network that leverages the MLP-Mixer framework for parallel temporal and feature dimension interaction and extraction.

Where, $-M$ is the model after adding the MSCDIM module. And *Improvement* is the model enhancement after adding the MSCDIM module.

As shown in Table 12, the implementation of the MSCDIM module in five widely utilised models within the field resulted in an average enhancement of 9.97% in RMSE and 23.85% in Score. And a subsequent analysis of the performance on the four subsets reveals that the RMSE improved by an average of 7.39%, 20.47%, 5.63%, and 6.39%, while the Score improved by 10.27%, 33.70%, 22.13%, and 29.29%, respectively. These improvement results indicate that the MSCDIM module enhances the ability to capture multi-dimensional feature information and perform dynamic feature fusion. Particularly in scenarios with higher data complexity, the significant improvement in the Score metric further demonstrates the module's advantages in suppressing redundant information and enhancing the expression of key features through its multi-scale and cross-dimensional

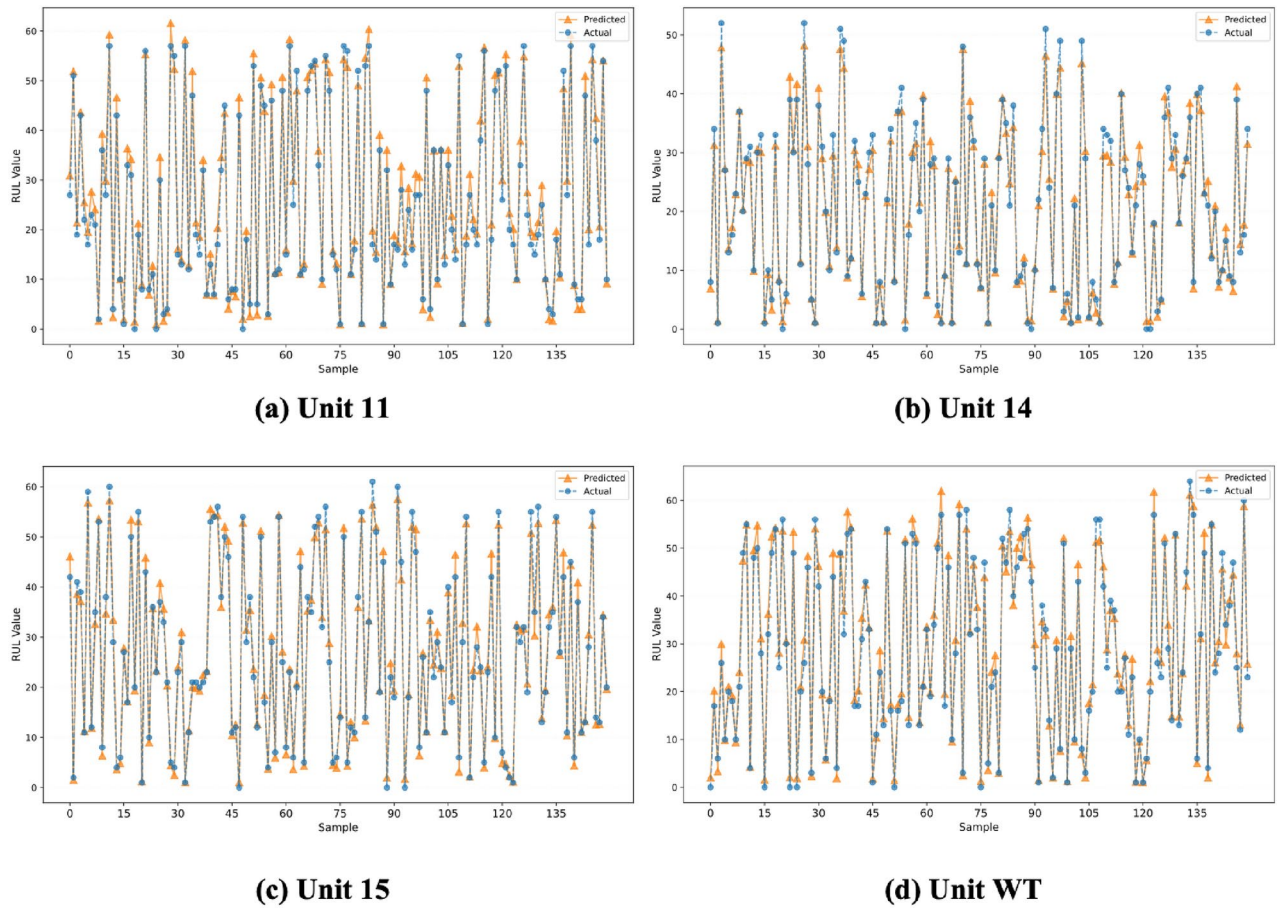


Fig. 16. Partial engine prediction results in four subsets of N-Cmaps.

| Models | FD001 | | FD002 | | FD003 | | FD004 | |
|----------|-------|--------|-------|---------|-------|--------|-------|---------|
| | RMSE | Score | RMSE | Score | RMSE | Score | RMSE | Score |
| TCN | 12.54 | 227.79 | 16.53 | 1115.36 | 12.44 | 227.66 | 22.89 | 4887.50 |
| AD-TCN | 11.97 | 203.56 | 14.42 | 689.03 | 10.96 | 175.20 | 20.02 | 2539.00 |
| TCN-M | 12.24 | 196.30 | 14.81 | 604.79 | 11.33 | 186.94 | 19.35 | 2608.02 |
| AD-TCN-M | 11.39 | 170.52 | 13.41 | 563.03 | 10.17 | 134.87 | 15.28 | 877.44 |

Table 10. Comparative ablation experiment results in C-MAPSS dataset.

| Models | Unit11 | | Unit14 | | Unit15 | | UnitWT | |
|----------|--------|---------|--------|---------|--------|---------|--------|----------|
| | RMSE | Score | RMSE | Score | RMSE | Score | RMSE | Score |
| TCN | 5.67 | 4362.14 | 15.34 | 3810.19 | 5.79 | 2089.47 | 7.59 | 10261.77 |
| AD-TCN | 5.26 | 3897.66 | 14.94 | 3617.57 | 5.16 | 1717.00 | 7.17 | 9232.21 |
| TCN-M | 5.10 | 3782.28 | 10.71 | 1918.13 | 4.98 | 1751.98 | 6.04 | 7452.38 |
| AD-TCN-M | 4.98 | 3545.67 | 8.33 | 1281.17 | 4.24 | 1381.75 | 5.76 | 6687.09 |

Table 11. Comparative ablation experiment results in N-CMAPSS dataset.

interaction mechanism. Consequently, the MSCDIM module demonstrates considerable applicability and potential for RUL prediction tasks based on the C-MAPSS dataset.

Conclusion

In this paper, a Multi-Scale Cross-Dimension Interaction Approach with Adaptive Dilated TCN (AD-TCN-MSCDIM) for RUL prediction is proposed. The integration of a dynamic adaptive dilation factor into the

| Models | FD001 | | FD002 | | FD003 | | FD004 | |
|--------------|--------|--------|--------|---------|-------|--------|-------|---------|
| | RMSE | Score | RMSE | Score | RMSE | Score | RMSE | Score |
| CNNGRU | 12.99 | 223.3 | 16.10 | 1201.35 | 12.48 | 268.83 | 18.29 | 2304.08 |
| CNNGRU-M | 12.97 | 224.30 | 15.00 | 674.35 | 11.63 | 185.05 | 17.66 | 1659.20 |
| Improvement | 0.13% | -0.45% | 6.86% | 43.87% | 6.77% | 31.17% | 3.44% | 27.99% |
| DAMCNN | 15.57 | 241.90 | 18.97 | 1316.15 | 15.62 | 867.70 | 19.05 | 1981.66 |
| DAMCNN-M | 13.36 | 186.62 | 17.25 | 1064.67 | 14.58 | 615.16 | 18.52 | 1583.06 |
| Improvement | 14.19% | 22.85% | 9.09% | 19.11% | 6.62% | 29.10% | 2.77% | 20.11% |
| IMDSSN | 12.14 | 206.11 | 17.40 | 1775.15 | 12.35 | 229.54 | 19.78 | 2852.81 |
| IMDSSN-M | 12.29 | 217.08 | 15.03 | 643.07 | 11.81 | 196.84 | 18.13 | 1501.14 |
| Improvement | -1.19% | -5.32% | 13.59% | 63.77% | 4.37% | 14.25% | 8.35% | 47.38% |
| MLP | 17.39 | 355.24 | 16.32 | 1052.97 | 13.06 | 224.28 | 19.45 | 2075.41 |
| MLP-M | 14.07 | 267.87 | 15.47 | 798.36 | 12.70 | 180.50 | 17.68 | 1224.17 |
| Improvement | 19.08% | 24.60% | 5.22% | 24.18% | 2.69% | 19.52% | 9.12% | 41.02% |
| Dual-Mixer | 12.51 | 220.50 | 15.06 | 938.31 | 12.79 | 248.04 | 16.10 | 969.67 |
| Dual-Mixer-M | 11.91 | 199.13 | 4.88 | 773.47 | 11.80 | 206.80 | 14.77 | 873.15 |
| Improvement | 4.73% | 9.69% | 67.57% | 17.57% | 7.71% | 16.62% | 8.29% | 9.95% |

Table 12. Comparative ablation experiment results in C-MAPSS dataset.

traditional TCN architecture, in conjunction with the construction of a serial AD-TCN framework, facilitates the incorporation of the MSCDIM module. This integration enables the model to dynamically adjust its receptive field based on the complex temporal data characteristics, thereby focusing on key features and suppressing the impact of redundant information on the model's predictive performance. And an extensive series of tests were conducted on the widely utilised C-MAPSS dataset and N-CMPASS dataset in the domain of RUL prediction. The findings demonstrate the efficacy of the proposed method in surpassing advanced approaches in the RUL prediction task. This conclusion was further validated through a series of ablation experiments, which demonstrated the efficacy of the dynamic adaptive dilation factor on TCN. The experiments also confirmed the versatility and effectiveness of the MSCDIM module.

Nevertheless, the model exhibits certain limitations. The current configuration, though effective, incurs relatively high computational costs. Consequently, subsequent endeavours will focus on the optimisation of the model while preserving its predictive accuracy, thereby enhancing its adaptability to varied operational settings. Furthermore, the exploration of the transferability of the algorithm to other industrial domains, such as the prediction of key components in rice milling equipment, is planned. This will enhance the model's generalizability and expand its potential applications.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 7 February 2025; Accepted: 29 April 2025

Published online: 02 June 2025

References

- Wang, L., Zhu, Z. & Zhao, X. Dynamic predictive maintenance strategy for system remaining useful life prediction via deep learning ensemble method. *Reliab. Eng. Syst. Safety* **245**, 110012 (2024).
- Chen, L. et al. Remaining useful life prediction of milling cutters based on long-term data sequence and parallel fully convolutional feature learning. *Journal of Intelligent Manufacturing* 1–23 (2024).
- Liu, X. et al. Optimized online remaining useful life prediction for nuclear circulating water pump considering time-varying degradation mechanism. *IEEE Transactions on Industrial Informatics* (2024).
- Ma, L., Tian, J., Zhang, T., Guo, Q. & Hu, C. Accurate and efficient remaining useful life prediction of batteries enabled by physics-informed machine learning. *J. Energy Chem.* **91**, 512–521 (2024).
- Zhong, K. et al. Discriminant analysis-guided alignment network for multi-machine fault collaborative learning and diagnosis. *IEEE Transactions on Instrumentation and Measurement* (2024).
- Fu, Y., Ding, J. & Xu, X. Shareability-exclusivity collaborative representation and its application for distributed process monitoring. *IEEE Transactions on Instrumentation and Measurement* (2024).
- Bao, X. et al. Joint estimation of state-of-charge and state-of-energy of lithium-ion batteries at different ambient temperatures based on domain adaptation and unscented kalman filter. *Electric Power Syst. Res.* **231**, 110284 (2024).
- Li, D., Liu, X. & Cheng, Z. The co-estimation of states for lithium-ion batteries based on segment data. *J. Energy Storage* **62**, 106787 (2023).
- Patel, N., Rai, R. N. & Sahu, P. K. Effective rul prediction of the milling tool based on shannon entropy feature selection method and squared exponential-gaussian process regression model. *Int. J. Adv. Manufact. Technol.* **136**, 693–715 (2025).
- Boujamza, A. & Elhaq, S. L. Optimizing remaining useful life predictions for aircraft engines: A dilated recurrent neural network approach. *IFAC-PapersOnLine* **58**, 811–816 (2024).
- Zhang, Z. et al. A novel local enhanced channel self-attention based on transformer for industrial remaining useful life prediction. *Eng. Appl. Artif. Intell.* **141**, 109815 (2025).

12. Li, X., Zhang, W., Ding, Y., Cai, J. & Yan, X. Rolling bearing degradation stage division and rul prediction based on recursive exponential slow feature analysis and bi-lstm model. *Reliab. Eng. Syst. Safety* **259**, 110923 (2025).
13. Sun, S., Wang, J., Xiao, Y., Peng, J. & Zhou, X. Few-shot rul prediction for engines based on cnn-gru model. *Sci. Rep.* **14**, 16041 (2024).
14. Guo, J., Lei, S. & Du, B. Mht: A multiscale hourglass-transformer for remaining useful life prediction of aircraft engine. *Eng. Appl. Artif. Intell.* **128**, 107519 (2024).
15. Chen, W., Liu, C., Chen, Q. & Wu, P. Multi-scale memory-enhanced method for predicting the remaining useful life of aircraft engines. *Neural Comput. Appl.* **35**, 2225–2241 (2023).
16. Zhang, J. et al. A parallel hybrid neural network with integration of spatial and temporal features for remaining useful life prediction in prognostics. *IEEE Trans. Instrum. Meas.* **72**, 1–12 (2022).
17. Liang, P., Yu, Z., Wang, B., Xu, X. & Tian, J. Fault transfer diagnosis of rolling bearings across multiple working conditions via subdomain adaptation and improved vision transformer network. *Adv. Eng. Inform.* **57**, 102075 (2023).
18. Zhang, Q., Liu, Q. & Ye, Q. An attention-based temporal convolutional network method for predicting remaining useful life of aero-engine. *Eng. Appl. Artif. Intell.* **127**, 107241 (2024).
19. Lin, L. et al. Channel attention & temporal attention based temporal convolutional network: A dual attention framework for remaining useful life prediction of the aircraft engines. *Adv. Eng. Inform.* **60**, 102372 (2024).
20. Jiang, G., Duan, Z., Zhao, Q., Li, D. & Luan, Y. Remaining useful life prediction of rolling bearings based on tcn-msa. *Meas. Sci. Technol.* **35**, 025125 (2023).
21. Tang, Y., Liu, R., Li, C. & Lei, N. Remaining useful life prediction of rolling bearings based on time convolutional network and transformer in parallel. *Meas. Sci. Technol.* **35**, 126102 (2024).
22. Li, H., Zhang, Z., Li, T. & Si, X. A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities. *Mech. Syst. Signal Process.* **209**, 111120 (2024).
23. Zhang, J., Li, X., Tian, J., Luo, H. & Yin, S. An integrated multi-head dual sparse self-attention network for remaining useful life prediction. *Reliab. Eng. Syst. Safety* **233**, 109096 (2023).
24. Keshun, Y., Guangqi, Q. & Yingkui, G. A 3d attention-enhanced hybrid neural network for turbofan engine remaining life prediction using cnn and bilstm models. *IEEE Sensors Journal* (2023).
25. Hu, G. et al. Remaining useful life prediction of mechanical equipment based on time-series auto-correlation decomposition and cnn. *Meas. Sci. Technol.* **35**, 105104 (2024).
26. Chao, M. A., Kulkarni, C. S., Goebel, K. & Fink, O. Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics. *Data* (2021).
27. Zhou, M., Li, Y., Cao, Y., Ma, X. & Xu, Z. Physics-informed spatio-temporal hybrid neural networks for predicting remaining useful life in aircraft engine. *Reliab. Eng. Syst. Safety* **256**, 110685 (2025).
28. Zhao, C., Shi, H., Huang, X. & Zhang, Y. A multiple conditions dual inputs attention network remaining useful life prediction method. *Eng. Appl. Artif. Intell.* **133**, 108160 (2024).
29. Fu, S. et al. Mca-dtcn: A novel dual-task temporal convolutional network with multi-channel attention for first prediction time detection and remaining useful life prediction. *Reliab. Eng. Syst. Safety* **241**, 109696 (2024).
30. Zhang, Z., Song, W., Li, Q. & Gao, H. Multiscale global and local self-attention-based network for remaining useful life prediction. *Meas. Sci. Technol.* **34**, 125154 (2023).
31. Fu, E., Hu, Y., Peng, K. & Chu, Y. Supervised contrastive learning based dual-mixer model for remaining useful life prediction. *Reliab. Eng. Syst. Safety* **251**, 110398 (2024).
32. Zhao, K., Jia, Z., Jia, F. & Shao, H. Multi-scale integrated deep self-attention network for predicting remaining useful life of aero-engine. *Eng. Appl. Artif. Intell.* **120**, 105860 (2023).
33. Zeng, M., Wu, F. & Cheng, Y. Remaining useful life prediction via spatio-temporal channels and transformer. *IEEE Sensors Journal* (2023).
34. Zhu, J., Jiang, Q., Shen, Y., Xu, F. & Zhu, Q. Res-hsa: Residual hybrid network with self-attention mechanism for rul prediction of rotating machinery. *Eng. Appl. Artif. Intell.* **124**, 106491 (2023).
35. Lin, R., Wang, H., Xiong, M., Hou, Z. & Che, C. Attention-based gate recurrent unit for remaining useful life prediction in prognostics. *Appl. Soft Comput.* **143**, 110419 (2023).
36. Xu, Z., Zhang, Y. & Miao, Q. An attention-based multi-scale temporal convolutional network for remaining useful life prediction. *Reliability Engineering & System Safety* 110288 (2024).
37. Kim, G., Choi, J. G. & Lim, S. Using transformer and a reweighting technique to develop a remaining useful life estimation method for turbofan engines. *Eng. Appl. Artif. Intell.* **133**, 108475 (2024).
38. Keshun, Y., Guangqi, Q. & Yingkui, G. Optimizing prior distribution parameters for probabilistic prediction of remaining useful life using deep learning. *Reliab. Eng. Syst. Safety* **242**, 109793 (2024).
39. Wang, L., Cao, H., Ye, Z., Xu, H. & Yan, J. Dvgtformer: A dual-view graph transformer to fuse multi-sensor signals for remaining useful life prediction. *Mech. Syst. Signal Process.* **207**, 110935 (2024).
40. Wang, Y. et al. Local-global correlation fusion-based graph neural network for remaining useful life prediction. *IEEE Transactions on Neural Networks and Learning Systems* (2023).
41. Jiang, F. et al. Dual-attention-based multiscale convolutional neural network with stage division for remaining useful life prediction of rolling bearings. *IEEE Trans. Instrum. Meas.* **71**, 1–10 (2022).
42. Tolstikhin, I. O. et al. Mlp-mixer: An all-mlp architecture for vision. *Adv. Neural. Inf. Process. Syst.* **34**, 24261–24272 (2021).

Acknowledgements

The authors would like to thank the National Natural Science Foundation of China (No. 32472000) for providing support for this paper.

Author contributions

Conceptualization, Zhe Lu.; methodology, Zhe Lu.; validation, Changyu Fu, Jiang Bai.; formal analysis, Changye Fu, Junbao Wu.; data curation, Liang Xu.; writing—original draft preparation, Zhe Lu.; writing—review and editing, Siye Jia.; visualization, Li Zelong.; project administration and funding acquisition Bing Li.; All authors have read and agreed to the published version of the manuscript.

Declarations

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical approval

Not applicable.

institutional review

Not applicable.

Informed consent

Not applicable.

Additional information

Correspondence and requests for materials should be addressed to B.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025