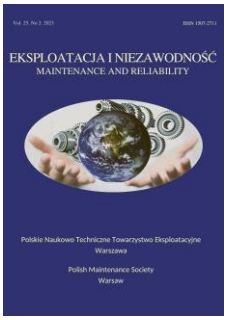


# Eksploatacja i Niezawodnosc – Maintenance and Reliability

**Volume 28 (2026), Issue 3**

journal homepage: http://www.ein.org.pl

Article citation info:
Huang C, Li T, Guo R, A frequency-adaptive feature extraction framework for bearing remaining useful life prediction, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(3) http://doi.org/10.17531/ein/218118


# A frequency-adaptive feature extraction framework for bearing remaining useful life prediction


Indexed by: Web of Science Group



**Chao Huang[a], Tianyu Li[b], Runxia Guo[b,*]**

[a] Sino-European Institute of Aviation Engineering, Civil Aviation University of China, China
[b] College of Electronic Information and Automation, Civil Aviation University of China, China


## Highlights

- Frequency-adaptive feature extraction aligned with sensor-specific degradation patterns.
- Dilated temporal extraction captures multi-scale degradation dynamics.
- Transformer-based spatial fusion enables effective cross-sensor dependency modeling.
- Improved RUL prediction across bearing datasets.


## Abstract

Accurate prediction of bearing remaining useful life (RUL) is essential for reliable rotating machinery. However, multi-sensor degradation signals exhibit diverse temporal and spectral patterns that are often insufficiently captured by feature extractors using uniform processing, limiting their ability to model signal-specific degradation behavior and affecting prediction accuracy. This study proposes a frequency-adaptive feature extraction framework for bearing RUL prediction. The framework includes a Temporal Feature Extraction Network (TFEN) that employs dilated convolutions with adaptive configurations to capture degradation dynamics across multiple temporal scales, and a Transformer-based Spatial Feature Extraction Network (SFEN) to model inter-sensor dependencies. By aligning feature extraction with the dominant frequency characteristics of each sensor channel, the proposed method improves the representation of degradation features. Experiments on two bearing datasets demonstrate its effectiveness, showing consistently enhanced prediction accuracy relative to existing models.



**Keywords**
feature extraction, bearing, multi-sensor, remaining useful life (RUL) prediction, frequency-adaptive.




## 1. Introduction

Rolling bearings are critical components in rotating machinery and are also among its most failure-prone elements, accounting for nearly 30% of mechanical faults [5,32]. Their degradation can lead to reduced efficiency and, in severe cases, catastrophic system failure. To mitigate such risks, prognostics and health management (PHM) has been widely developed, with remaining useful life (RUL) prediction serving as one of its core tasks [3,10,11]. Accurate RUL prediction enables early fault detection and supports condition-based maintenance, thereby improving equipment reliability and reducing maintenance costs.

Despite substantial progress, developing a unified and reliable framework for RUL prediction remains challenging due to the complex dynamics, strong non-stationarity, and varying operating conditions of mechanical systems. In practical bearing prognostics, the monitored signals are simultaneously influenced by load/speed fluctuations, environmental disturbances, sensor placement, and multi-source noise,


(*) Corresponding author.
E-mail addresses: C. Huang (ORCID: 0000-0002-5433-9574) chhuang@cauc.edu.cn, T. Li (ORCID: 0009-0007-1332-0395) 2024021056@cauc.edu.cn, R. Guo (ORCID: 0000-0002-9486-8606) rxguoblp@163.com

resulting in time-varying degradation patterns and distribution shifts across operating regimes. This makes it difficult to learn a single health-evolution mapping that is stable, transferable, and robust throughout the full life cycle, especially in the early degradation stage where fault signatures are weak and the signal-to-noise ratio is low. Existing approaches can generally be classified into physical model-based [1,19], statistical model-based [4,22], and data-driven methods [7,24,33]. Physical modeling of complex systems is often difficult due to structural complexity and uncertain conditions. Statistical modeling also faces challenges in extracting reliable patterns from large and stochastic datasets. In contrast, data-driven approaches utilize large-scale monitoring data to automatically extract degradation features and learn patterns of latent health evolution through machine learning or deep learning models. These methods have become increasingly dominant in predictive maintenance due to their flexibility and modeling capability. Physical model-based methods typically rely on first-principle dynamics and damage accumulation laws to extrapolate degradation trajectories. Although they provide certain interpretability and can embed prior knowledge, physical modeling of complex systems is often difficult due to structural complexity, uncertain boundary conditions, unmodeled dynamics, and parameter identifiability issues; furthermore, model mismatch may become more pronounced when the operating condition changes. Statistical model-based methods, such as stochastic processes and state-space formulations, attempt to characterize degradation evolution and uncertainty using probabilistic assumptions. However, they also face challenges in extracting reliable and consistent patterns from large-scale, highly stochastic monitoring data, and their effectiveness can degrade when assumptions are violated under variable working conditions.In contrast, data-driven approaches utilize large-scale monitoring data to automatically extract degradation features and learn patterns of latent health evolution through machine learning or deep learning models. Benefiting from representation learning, modern architectures can jointly model multi-scale temporal dependencies and complex nonlinear relationships, and can be extended to multi-sensor fusion scenarios to capture cross-channel degradation interactions. Consequently, these methods have become increasingly dominant in predictive maintenance due to their flexibility and modeling capability. Nevertheless, recent bearing RUL studies also suggest that achieving reliable data-driven prognostics requires more than simply increasing model capacity: the learned representations should be robust to operating-condition shifts, insensitive to irrelevant noise, and sensitive to weak yet progressive degradation cues, while maintaining stability across the full degradation trajectory. Therefore, building a unified framework remains an open problem, calling for feature extraction and learning strategies that balance accuracy, generalization, and physical consistency under realistic variable conditions.

Among various approaches, data-driven methods have achieved remarkable success in RUL prediction; however, their performance often depends heavily on the accurate extraction and representation of degradation features [6,21]. Since the degradation patterns are inherently tied to the physical nature of the signals, understanding their temporal and spectral characteristics is essential for designing effective feature extraction strategies. Time-domain analysis effectively captures local waveform variations and transient patterns. Yan et al. [28] utilized the root mean square (RMS) of vibration acceleration signals to construct hybrid degradation models; Yao et al. [30] adopted raw vibration signals as input to deep learning-based RUL prediction networks; and Ruan et al. [17] derived multiple time-domain features to build bearing health indicators. In addition, frequency-domain analysis reveals the spectral distribution and periodic characteristics of signals, offering deeper insight into the physical mechanisms of degradation. Yang et al. [29] proposed Fourier-transform-based modules to extract periodic features from vibration data, while Muhammad Gibran et al. [2] decomposed vibration signals into distinct frequency bands using empirical mode decomposition (EMD). These studies suggest that frequency-domain information can effectively distinguish the intrinsic characteristics of degradation signals and provide a foundation for developing adaptive feature extraction strategies.

Recently, foundation-style architectures for multivariate time series forecasting have attracted increasing attention. iTransformer adopts an inverted modeling paradigm that treats each sensor variable as a token and applies self-attention along the variable dimension while using feed-forward networks to model temporal dynamics within each token [13]. In the field of

prognostics and health management, iTransformer-based architectures have also started to be introduced into RUL prediction of rotating machinery. Lyu et al. [15] proposed an adaptive BiGRU-ASSA-iTransformer framework that combines handcrafted statistical features with bidirectional gated recurrent units (BiGRU) and an inverted Transformer backbone to improve bearing RUL estimation in aerospace manufacturing. Li et al. [12] further developed a variational mode decomposition–based adaptability feature perception fusion model (VMD-AFPFM), which employed a transposed iTransformer module to estimate the remaining useful life.

To effectively model temporal dependencies in sequential degradation signals, a flexible and efficient feature extraction method is required. Recent studies have explored adaptive feature extraction methods to enhance the representation of degradation information. Wan et al. [20] proposed multi-branch networks that combine convolutional neural networks (CNNs) and convolutional LSTM (CLSTM) to extract shallow features and deep temporal features; Xu et al. [27] developed an adaptive feature extraction approach that considers both time-domain and frequency-domain modalities, using a bidirectional long short-term memory network (Bi-LSTM) to learn temporal dependencies for accurate RUL estimation; Wang et al. [25] employ an adaptive shrinkage subnetwork to suppress noise while preserving informative features. However, most existing approaches either employ data-driven gating strategies that lack explicit correspondence with the physical characteristics of sensor signals [18,28,31] or apply uniform network architectures across all channels [14,23], which may overlook channel-specific signal behaviors and limit the effectiveness of feature extraction.

Since degradation signals from different sensors possess diverse spectral and temporal characteristics, aligning the feature extraction process with their intrinsic frequency properties is essential for capturing complete degradation information, motivating the use of a frequency-adaptive mechanism. Such a frequency-adaptive mechanism enables the temporal feature extraction network to align its receptive field and representational capacity with the intrinsic dynamics of each signal, preserving critical degradation patterns while maintaining computational efficiency. Beyond temporal modeling, degradation behaviors across sensors often exhibit interdependencies that cannot be captured by independent processing. Attention-based modules, such as transformers, can effectively model these cross-sensor interactions, facilitating global feature fusion and enhancing the overall predictive performance.

Building upon these insights, this study develops a frequency-guided, structurally adaptive feature extraction framework to achieve accurate RUL prediction for bearing degradation. The proposed framework explicitly accounts for the diverse temporal and spectral behaviors of multi-sensor signals by tailoring the feature extraction process to the frequency characteristics of each input. Specifically, a frequency-adaptive temporal feature extractor based on dilated convolutional networks is employed to capture temporal dependencies at multiple scales, while a spatial feature extractor built upon a transformer-based architecture models inter-sensor correlations. The two modules form the backbone of the proposed RUL prediction framework, achieving an effective balance between representational efficiency and physical interpretability.

The main contributions of this study are as follows:

(1) A unified noise reduction method and a frequency domain indicator are developed to categorize sensor signals into frequency bands, providing the basis for frequency-adaptive feature extraction.
(2) A feature extraction architecture is proposed, where the Temporal Feature Extraction Network (TFEN) is customized to capture temporal degradation features for each sensor data, and the Spatial Feature Extraction Network (SFEN) fuses multi-sensor information to provide a comprehensive degradation representation.
(3) A frequency-adaptive mechanism is introduced to align feature extraction with the intrinsic frequency characteristics of each sensor signal, enhancing adaptability and robustness in RUL prediction.

The remainder of this paper is organized as follows. Section 2 details the proposed RUL prediction framework, highlighting the design of the frequency-adaptive feature extraction network. Section 3 evaluates the proposed method on two bearing degradation datasets and discusses the results. Section 4 concludes the paper and outlines potential directions for future work.

## 2. The proposed method

Figure 1 illustrates the overall RUL prediction framework based on the proposed frequency-adaptive feature extraction method. The framework consists of four major stages: (1) Unified noise reduction, which suppresses sensor noise and harmonizes signal quality; (2) Average frequency computation and frequency band partitioning, which categorize sensor signals into distinct frequency bands to guide adaptive feature extraction; (3) Feature extraction, which employs a frequency-aware and structurally adaptive feature extraction network to obtain temporal-spatial representations; and (4) RUL prediction, where the fused features are mapped to the final remaining useful life prediction through multi-layer perceptron (MLP).

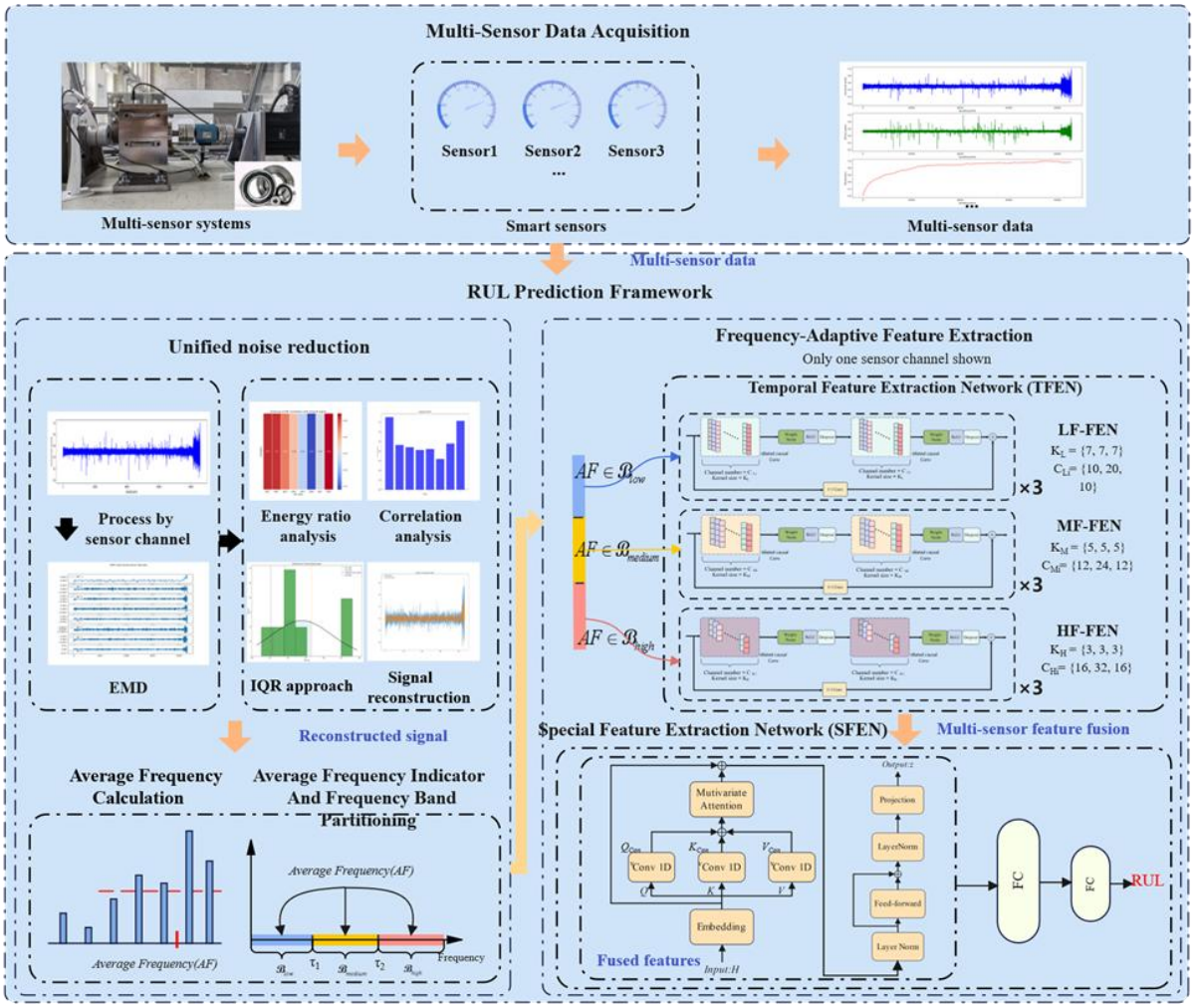


Figure 1. The overall RUL prediction framework.

The adaptive feature extraction network serves as the core of the proposed framework. It comprises two complementary components: a temporal feature extraction network (TFEN) that captures degradation dynamics within each signal channel, and a spatial feature extraction network (SFEN) that models cross-sensor dependencies and joint degradation evolution.

This hierarchical design allows the framework to integrate both intra-channel temporal dependencies and inter-channel spatial correlations, thereby enhancing the accuracy and interpretability of RUL prediction.

### 2.1. Unified noise reduction method

Given the collected data from $N$ sensor: $X = \{X^{(1)} \cdots, X^{(i)}, \cdots X^{(N)}\}^{\top}, X \in \mathbb{R}^{N \times T}$, where $T$ denotes the total number of time samples, and $X^{(i)} \in \mathbb{R}^{T}$ represents the one-dimensional signal from the $i$-th sensor. Each signal is expressed as: $X^{(i)} = x_1^{(i)} \cdots, x_t^{(i)}, \cdots x_T^{(i)}, x_t^{(i)} \in \mathbb{R}$.

To suppress random fluctuations and environmental noise while preserving degradation-related trends, a unified noise-reduction method is first applied to all sensor signals. This step ensures a consistent signal quality across sensors and provides

a cleaner basis for subsequent frequency analysis and feature extraction. Each sensor signal $X^{(i)}(t)(i = 1, \dots, N)$ is decomposed into $M$ intrinsic mode functions (IMFs) and a residual component using Empirical Mode Decomposition (EMD):

$$X^{(i)}(t) = \sum_{j=1}^{M} IMF_j^{(i)}(t) + R^{(i)}(t) \quad (1)$$

where $M$ is the total number of IMFs, and $R^{(i)}(t)$ denotes the residual component. EMD adaptively separates the signal into quasi-orthogonal modes, which provides a solid foundation for subsequent energy and correlation analyses.

To effectively suppress noise while preserving informative dynamics, the contribution of each IMF is assessed from both energy and correlation perspectives. The energy of each mode $E_j^{(i)}$ and its relative ratio $ER_j^{(i)}$ are computed as:

$$E_j^{(i)} = \sum_{t=1}^{T} (IMF_j^{(i)}(t))^2 \quad (2)$$

$$ER_j^{(i)} = \frac{E_j^{(i)}}{\sum_{m=1}^{M} E_m^{(i)}} \quad (2)$$

Among the measurement methods based on correlation, the Pearson correlation coefficient remains the primary and easily understandable choice for quantifying the linear dependency and shape similarity in time series. Referring to the existing research on fault diagnosis based on EMD, most of these studies use the Pearson correlation coefficient to select the IMFs with information content. Finally, we choose to adopt the Pearson correlation coefficient to measure the shape similarity between each intrinsic mode function component and the original signal [8,9,26].

The Pearson correlation coefficient $CA_j^{(i)}$ between the $j$-th IMF and the original signal quantifies their morphological similarity:

$$CA_j^{(i)} = \frac{cov(IMF_j^{(i)}, X^{(i)})}{\sigma_{IMF_j^{(i)}} \sigma_{X^{(i)}}} \quad (3)$$

Low-energy and weakly correlated modes are typically dominated by noise. To achieve a balanced evaluation of mode significance, an adaptive weighting strategy is introduced to integrate energy ratio and correlation analyses. The normalized weighting coefficients $w_E^{(i)}$ and $w_C^{(i)}$ are defined as:

$$w_E^{(i)} = \frac{\sum ER_j^{(i)}}{\sum \left( ER_j^{(i)} + CA_j^{(i)} \right)} \quad (4)$$

$$w_C^{(i)} = \frac{\sum CA_j^{(i)}}{\sum \left( ER_j^{(i)} + CA_j^{(i)} \right)} \quad (5)$$

A composite score is then calculated for each IMF:

$$S_j^{(i)} = w_E^{(i)} ER_j^{(i)} + w_C^{(i)} CA_j^{(i)} \quad (6)$$

Instead of using a fixed cutoff, which is sensitive to signal variation, an interquartile range (IQR)-based adaptive threshold is introduced:

$$S_{Threshold}^{(i)} = Q_1^{(i)} - \alpha \cdot IQR^{(i)} \quad (7)$$

where $Q_1^{(i)}$ is the first quartile of the score distribution, $IQR^{(i)}$ is the interquartile range, and $\alpha$ is a tunable parameter that controls the stringency of noise rejection. IMFs with $S_j^{(i)} > S_{Threshold}^{(i)}$ are retained and summed to reconstruct the denoised signal:

$$X_D^{(i)} = \sum_{j \in \text{selected IMF}} IMF_j^{(i)} \quad (8)$$

This adaptive fusion and thresholding strategy effectively suppresses stochastic noise while maintaining the signal's intrinsic dynamics, offering improved robustness over conventional energy-only or correlation-only selection schemes. In this study, several hyperparameters are involved in the spectral analysis and decomposition process. The number of dominant frequencies $K$ is chosen to represent the main spectral characteristics of the signal while avoiding excessive sensitivity to minor spectral fluctuations. In practice, $K$ is selected to ensure that the retained frequency components account for the majority of the spectral energy, while keeping the computational cost manageable.

For the mode decomposition process, the number of modes $M$ is fixed to balance decomposition stability and noise suppression. A moderate value is adopted based on commonly used settings in empirical mode decomposition – based methods and the overall spectral complexity of the datasets, which helps avoid over-decomposition and redundant high-frequency components.

### 2.2. Average frequency indicator and frequency band partitioning

This section describes the computation of frequency indicators and the definition of frequency bands, which serve as the basis

for configuring frequency-adaptive feature extraction networks in subsequent stages.

After noise reduction, the reconstructed signal $X_D^{(i)}(t)$ is transformed into the frequency domain using the Fast Fourier Transform (FFT). The resulting normalized amplitude spectrum $A^{(i)}(f)$ captures the frequency distribution characteristics of the sensor data. Due to the inherent sparsity of vibration spectra, only the top-*K* amplitude peaks are preserved to represent the dominant frequency set: $F^{(i)} = f_1^{(i)},\cdots,f_K^{(i)}$. The corresponding amplitudes $\{A_{f_1}^{(i)},\cdots,A_{f_K}^{(i)}\}$ describe the energy contributions of each dominant component. Considering spectral conjugacy, only frequencies within $1,\cdots,\left[\frac{T}{2}\right]$ are analyzed.

### 2.2.1. Frequency indicators

The local maximum and local minimum frequencies are defined as:

$$f_{max}^{(i)} = \max_{k=1,\ldots K} f_k^{(i)} \tag{9}$$

$$f_{min}^{(i)} = \min_{k=1,\ldots K} f_k^{(i)} \tag{10}$$

Across all sensors, the global frequency bounds are given by:

$$f_{max} = \max_{i=1,\ldots N} f_{max}^{(i)} \tag{11}$$

$$f_{min} = \min_{i=1,\ldots N} f_{min}^{(i)} \tag{12}$$

A new indicator in the frequency domain is formulated to better reflect the intrinsic characteristics of different inputs and guide the subsequent feature extraction mechanism. For each sensor data$X^{(i)}$, the "Average Frequency (AF)" is defined as:

$$\bar{f}^{(i)} = \left(\sum_{k=1}^{K} f_k^{(i)}\right)/K \tag{13}$$

Since Average Frequency$\bar{f}^{(i)}$is derived from the dominant FFT peaks (top-K), in this work it is used solely as a concise spectral descriptor to support frequency-based grouping of channels with consistent acquisition and preprocessing settings. It characterizes the dominant spectral concentration associated with the operational state, but is not intended to correspond to any physically interpretable defect-related fault frequency.

### 2.2.2. Frequency band partitioning

To categorize signals with distinct spectral properties, the entire frequency range $[fmax_{min}$ is equally divided into three adaptive bands using two thresholds:

$$T(p, f_{min}, f_{max}) = f_{min} + p \cdot \frac{f_{max}-f_{min}}{3}, p \in \{1,2\} \tag{15}$$

$$\tau_1 = T(1, f_{min}, f_{max}) = f_{min} + \frac{f_{max}-f_{min}}{3}, when\ p = 1 \tag{16}$$

$$\tau_2 = T(2, f_{min}, f_{max}) = f_{min} + 2 \cdot \frac{f_{max}-f_{min}}{3}, when\ p = 2 \tag{17}$$

The frequency range is categorized into low-, mid-, and high-frequency bands, adhering to the established consensus within the field. Each sensor data's average frequency is then assigned to one of the following frequency bands:

$$F_{band}\left(\bar{f}^{(i)}\right) = \begin{cases} B_{low} & \text{if } \bar{f}^{(i)} < \tau_1 \\ B_{medium} & \text{if } \tau_1 \le \bar{f}^{(i)} < \tau_2 \\ B_{high} & \text{if } \bar{f}^{(i)} \ge \tau_2 \end{cases} \tag{18}$$

This categorization enables a frequency-adaptive mechanism, where inputs are grouped according to their computed dominant frequency characteristics. Building on this, the resulting partitioning serves as the basis for a frequency-adaptive mechanism, through which the downstream RUL prediction network receives temporally aligned representations tailored to the spectral properties of each input. This three-band division method is consistent with the common practice in bearing diagnosis. In this work, the average frequency (AF) and the corresponding frequency bands are used to support the adaptive selection of feature extraction structures. They are not intended to represent precise physical fault frequencies, but to reflect relative spectral characteristics for network configuration. In this work, AF-based grouping is designed for network configuration (TFEN branch assignment) rather than fault-type identification, and its absolute value is not compared across datasets with different sampling settings.

## 2.3. Temporal Feature Extraction Network (TFEN)

The Temporal Feature Extraction Network (TFEN) is designed to capture the temporal degradation dynamics embedded within multi-sensor signals. Since different sensing modalities exhibit distinct dominant frequencies and characteristic time scales, the TFEN incorporates a frequency-adaptive mechanism that aligns its internal structure with the intrinsic dynamics of each input. This design ensures that temporal receptive fields and representational capacities are properly matched to the signal's degradation behavior, enabling frequency-adaptive feature extraction.

### 2.3.1. Frequency-adaptive mechanism

Building upon the average frequency computation described in

Section 2.2, each input signal is assigned to a specific frequency band (low, medium, or high) according to its spectral characteristics. Rather than applying a uniform architecture to all signals, the proposed method introduces a frequency-adaptive assignment mechanism that dynamically selects a temporal feature extraction network aligned with the identified frequency band.

This mechanism ensures structural alignment between the temporal scale of the extractor and the intrinsic dynamics of the signal. Signals dominated by low-frequency components and long-term dependencies are processed by networks configured with broader receptive fields and larger convolution kernels, enabling global temporal modeling. Conversely, high-frequency signals with rapid variations are handled by networks employing smaller kernels and richer channel capacities, allowing finer temporal resolution and more detailed feature extraction. The medium-frequency category serves as an intermediate configuration, balancing global trend tracking and local fluctuation sensitivity.

Through this adaptive allocation, the TFEN achieves frequency-adaptive temporal feature extraction, matching its modeling capacity to the underlying spectral properties of each signal. This mechanism provides the foundation for the architecture and configurations of TFEN described in the following sections.

### 2.3.2. TFEN architecture and frequency-adaptive network configurations

The TFEN is designed to capture the temporal degradation dynamics embedded in each sensor signal while adapting to its frequency characteristics. To achieve this, the TFEN is constructed as a modular, multi-branch architecture, with each branch corresponding to a predefined frequency band (low, medium, high). Each input signal is assigned to the branch associated with its average frequency, ensuring that the temporal modeling capacity of the network aligns with the intrinsic dynamics of the signal.

The general block structure is shown in Figure 2. Within each branch, multiple residual Temporal Convolutional Network (TCN) blocks are cascaded to extract hierarchical temporal features across multiple scales progressively. Each TCN block consists of two dilated causal convolutional layers, each followed by a ReLU activation, dropout, and a cropping operation to maintain causality. The operation of a dilated causal convolution in each block is formulated as:

$$\left(\mathrm{Conv}^{(k,d,p)}z\right)[\tau] = \sum_{j=0}^{k-1} W[j]\cdot z[\tau - jd] + b, \quad p = (k-1)d \tag{14}$$

where $k$ denotes the kernel size, $d$ denotes the dilation rate, $p$ denotes the padding length, $z$ denotes the input sequence, $W$ and $b$ are the convolution weights and bias, and $\tau$ is the time index. The transformation performed by the two layers in each block can be expressed as:

$$u_1^{(l)} = D_{p_l}\left(ReLU\left(C_{p_l}\left(\mathrm{Conv}_1^{(k_l,d_l,p_l)}h^{(l-1)}\right)\right)\right), p_l = (k-1)d_l \tag{15}$$

$$u_2^{(l)} = D_{p_l}\left(ReLU\left(C_{p_l}\left(\mathrm{Conv}_2^{(k_l,d_l,p_l)}u_1^{(l)}\right)\right)\right), p_l = (k-1)d_l \tag{16}$$

$$h^{(l)} = u_2^{(l)} + h^{(l-1)} \tag{17}$$

where $l$ is the block index, $h^{(l-1)}$ is the input from the previous block, $D_{p_l}$ is the Dropout layer with a dropout rate of 0.15, and $C_{p_l}$ is the Chomp cropping operator. And $k_l, d_l, p_l$ denote the kernel size, dilation rate, and padding length, respectively.

The branch configurations follow two principles. First, the kernel size is matched to the characteristic time scale implied by the AF band: larger kernels provide a broader effective receptive field for low-frequency, slowly varying degradation patterns, whereas smaller kernels preserve temporal resolution for high-frequency fluctuations. Second, channel widths are adjusted to balance representational capacity across branches while keeping a comparable parameter budget, and the number of residual blocks is fixed to ensure structural compatibility for subsequent SFEN fusion. In practice, the specific values are selected from a small candidate set via validation performance under the same training protocol.

The architecture of each branch is identical in depth, ensuring uniform hierarchical representation, but the block parameters, including convolutional kernel sizes and channel widths, are tailored to the specific frequency band of the input. The low-frequency feature extraction branch (LF-FEN) employs larger kernels with moderate channel widths to capture long-term temporal dependencies and smooth degradation trends. The medium-frequency branch (MF-FEN) adopts intermediate kernel sizes and channels to balance local precision and contextual representation. The high-frequency

branch (HF-FEN) utilizes smaller kernels and richer channels to enhance sensitivity to rapid fluctuations and transient variations, thereby improving responsiveness to subtle high-frequency features. Table 1 summarizes the configuration of each specific network.

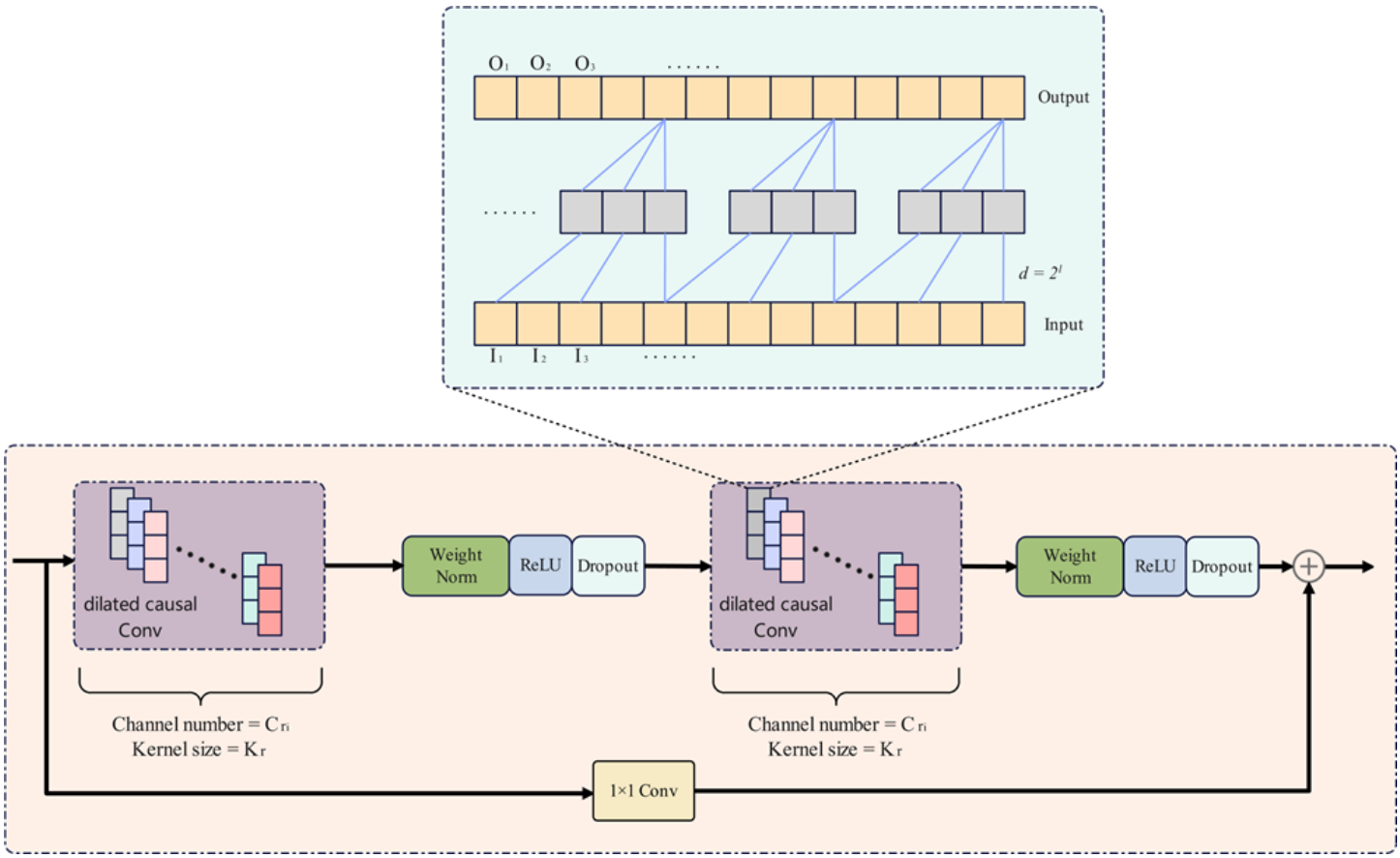


Figure 2. General block structure of a TFEN branch.

Table 1. Configurations of TFEN.

| Branch | Kernel size ($K_r$) | Block Number | Channel Number ($C_{ri}$) | | |
|---|---|---|---|---|---|
| | | | Block 1 | Block 2 | Block 3 |
| LF-FEN | $K_L = 7$ | 3 | $C_{L1} = 10$ | $C_{L2} = 20$ | $C_{L3} = 10$ |
| MF-FEN | $K_M = 5$ | 3 | $C_{M1} = 12$ | $C_{M2} = 24$ | $C_{M3} = 12$ |
| HF-FEN | $K_H = 3$ | 3 | $C_{H1} = 16$ | $C_{H2} = 32$ | $C_{H3} = 16$ |

After processing, the outputs of all branches are aligned through 1D convolution and fused to form a unified temporal feature representation $H \in \mathbb{R}^{B\times L\times F}$, which is subsequently passed to the spatial feature extraction network (SFEN) for inter-sensor modeling and RUL prediction. By coordinating kernel sizes and channel widths with frequency characteristics while maintaining uniform depth, the TFEN captures frequency-specific temporal dependencies in an efficient and structurally compatible manner.

### 2.4. Spatial Feature Extraction Network (SFEN)

The Spatial Feature Extraction Network (SFEN) is designed to model inter-sensor dependencies and integrate information across multiple sensor channels. At the core of the SFEN lies the improved iTransformer, which uses a variable-wise self-attention mechanism to model inter-sensor relationships, with an enhanced query–key–value (Q/K/V) projection that incorporates a lightweight 1D convolution to better preserve local temporal patterns. Where Q, K, and V denote the query, key, and value vectors in the self-attention mechanism.

The structure of transformer-based SFEN is shown in Figure 3. To strengthen the temporal expressiveness of the variable-wise attention, we modify the *Q-K-V* projection module of the original iTransformer. Instead of generating the query, key, and

value vectors solely through linear projections, each projection branch is preceded by a 1D convolutional layer. The convolution operates along the temporal dimension and acts as a local feature enhancer, allowing the attention module to retain short-range temporal structures that would otherwise be weakened by the variable-wise formulation. The enhanced projections are computed as:

$$Q_{Conv} = Conv_{1D}(Q), K_{Conv} = Conv_{1D}(K), V_{Conv} = Conv_{1D}(V) \quad (18)$$

where $Conv_{1D}$ denotes 1D convolution. The inserted convolution enriches the local temporal context before attention is applied, allowing the network to capture detailed temporal patterns and inter-channel relationships.

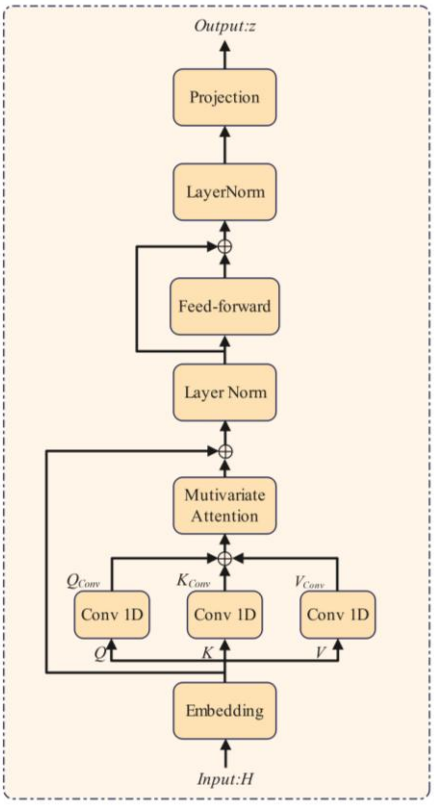


Figure 3. Structure of SFEN.

### 2.5. RUL Prediction

The final stage employs an MLP as the regression head to predict the RUL. The MLP receives the spatial representations from the SFEN and maps them to a scalar RUL estimate through fully connected layers with nonlinear activations. This stage ensures the integration of temporal-spatial information extracted by the preceding networks into a unified degradation prediction output. Specifically, the output of the SFEN$z$is first flattened into a one-dimensional feature vector, which is then passed through a two-layer MLP to produce the final scalar RUL prediction:

$$\hat{y} = \mathrm{MLP}(z) \quad (19)$$

Predictive performance is quantified using standard regression metrics, including root mean square error (RMSE), mean squared error (MSE), mean absolute error (MAE), coefficient of determination ($R^2$), and symmetric mean absolute percentage error (SMAPE). The evaluation metrics are as follows:

Mean Squared Error:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \quad (20)$$

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (21)$$

Mean Absolute Error:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \quad (22)$$

$R^2$:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (23)$$

Symmetric Mean Absolute Percentage Error:

$$SMAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{2|\hat{y}_i - y_i|}{|\hat{y}_i| + |y_i|} \quad (24)$$

The proposed framework integrates signal preprocessing, frequency-adaptive temporal feature extraction, and cross-sensor representation learning into a unified prediction process. By tailoring the TFEN to the spectral characteristics of each sensor and complementing it with the SFEN to capture inter-channel dependencies, the model constructs a degradation representation that preserves both band-dependent dynamics and multi-sensor correlations. This enriched representation is then mapped to the remaining useful life through a lightweight prediction head, enabling accurate and robust RUL estimation across heterogeneous operating conditions.

## 3. Experiments and analysis

Experiments on two datasets were conducted to evaluate the proposed prediction framework in terms of accuracy and generalization. For each dataset, two sets of experiments were conducted: the first examined the intrinsic performance of the method, while the second provided a comparative analysis against multiple state-of-the-art approaches.

### 3.1. Case study 1: experiments on FEMTO-ST dataset

#### 3.1.1. Description of the FEMTO-ST dataset

The FEMTO-ST dataset comprises online health-monitoring data for bearings collected from the PRONOSTIA platform [16]. The dataset contains two channels of vibration signals, some of which also include a temperature signal channel, each sampled at 25.6 kHz and 10 kHz. As shown in Table 2, six run-to-failure bearings' health monitoring datasets were obtained under three different operating conditions to train the model, while data from eleven test bearings' monitoring were utilized to predict the RUL.

In the following experiments, the key hyperparameters for the noise reduction process are set as follows: each channel signal is decomposed into 8 IMFs; the adjustable parameter $\alpha$ is set to 1.5; and the number of dominant frequencies $K$ is set to 5. Figure 4 presents the normalized temporal evolution of reconstructed signals obtained from frequency component analysis on bearing 1-2, clearly reflecting the variation patterns of two vibration acceleration channels and the temperature signal.

Table 2. FEMTO-ST dataset introduction.

| | Operating Condition 1 | Operating Condition 2 | Operating Condition 3 |
|---|---|---|---|
| Training dataset | Bearing 1_1 | Bearing 2_1 | Bearing 3_1 |
| | Bearing 1_2 | Bearing 2_2 | Bearing 3_2 |
| Testing dataset | Bearing 1_3 | Bearing 2_3 | Bearing 3_3 |
| | Bearing 1_4 | Bearing 2_4 | |
| | Bearing 1_5 | Bearing 2_5 | |
| | Bearing 1_6 | Bearing 2_6 | |
| | Bearing 1_7 | Bearing 2_7 | |
| Load (N) | 4000 | 4200 | 5000 |
| Speed (RPM) | 1800 | 1650 | 1500 |

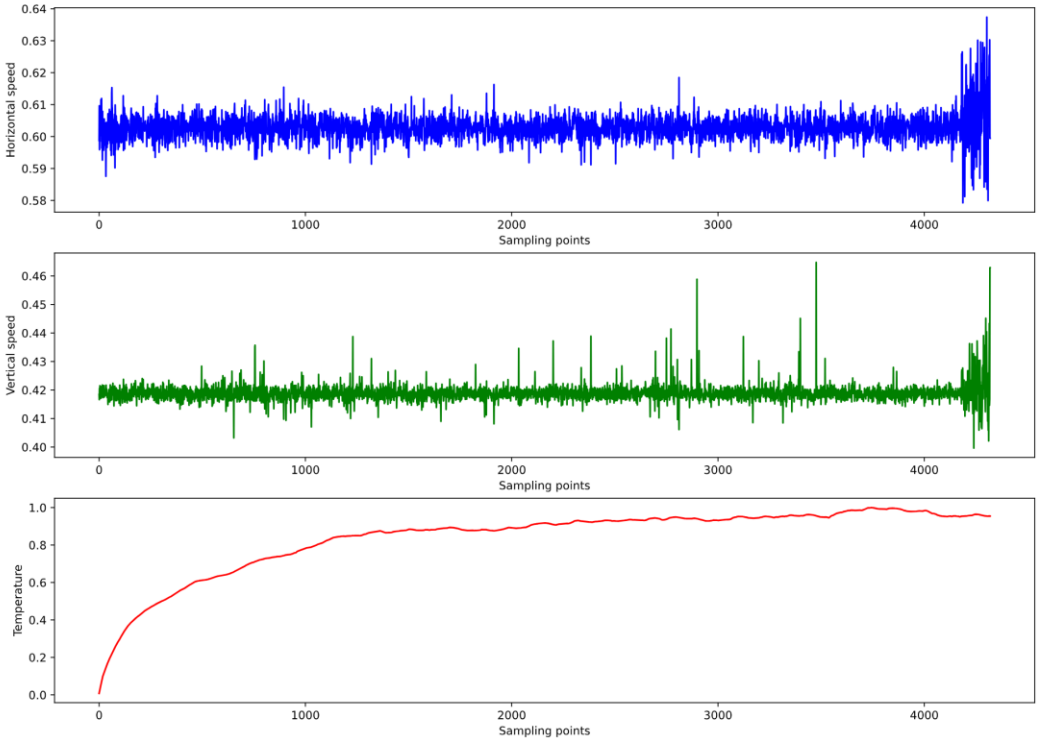


Figure 4. Trends of various sensor data of bearing 1-2.

#### 3.1.2. Evaluation of the frequency-adaptive TFEN

To evaluate the effectiveness of the proposed frequency-adaptive TFEN, we constructed several configurations for comparison. The first three use a single type of temporal feature extraction network (LF-FEN, MF-FEN, or HF-FEN), and the proposed configuration assigns TFEN based on each channel's average frequency, implementing the frequency-adaptive mechanism. In addition, two randomly assigned configurations were included to assess the impact of non-optimized TFEN allocation. To further evaluate the generalization capability under varying operating conditions, bearings 1-2, 1-4, 2-1, and 2-4 were reorganized into a four-fold cross-operating-condition validation scheme. In each fold, three bearings were used for

training, and the remaining one served as the test bearing.

All configurations were trained and tested under identical settings, and their performance was evaluated using the standard RUL metrics defined in section 2.5. The results are summarized in Table 3.

From Table 3 across all evaluation metrics, the proposed method clearly outperforms the three single-band configurations. The average RMSE is reduced by 36.6% compared with the low-frequency configuration, 37.3% compared with the medium-frequency group, and 35% compared with the high-frequency configuration. Compared with the best-performing single-band model (HF-FEN), the adaptive configuration achieves a substantial reduction in RMSE and MAE and a marked increase in $R^2$, indicating a more accurate and stable modeling of the degradation process. This improvement aligns with our design motivation: different sensor data contain distinct frequency characteristics, and forcing all channels into a single-band extractor inevitably leads to information loss, whereas the adaptive configuration captures complementary degradation patterns across bands.

The randomly assigned configurations consistently produced inferior performance, with higher error metrics. This result confirms that simply combining multiple TFEN is insufficient. Compared with the random allocation configurations, the average RMSE of the proposed method decreased by 40.5%. The proposed frequency-adaptive mechanism provides this principled allocation, enabling the network to extract the most relevant temporal patterns for each sensor data.

Table 3. Performance comparison across different configurations under Case 1.

| Method | Metric | Bearing ID | | | | |
|---|---|---|---|---|---|---|
| | | 1-2 | 1-4 | 2-1 | 2-4 | Average |
| LF-FEN | RMSE | 0.1090 | 0.1259 | 0.1206 | 0.1200 | 0.1189 |
| | MAE | 0.0753 | 0.1038 | 0.0917 | 0.0985 | 0.0923 |
| | R2 | 0.8531 | 0.8042 | 0.8189 | 0.8223 | 0.8246 |
| | MSE | 0.0119 | 0.0158 | 0.0145 | 0.0144 | 0.0142 |
| | SMAPE(%) | 24.491 | 33.479 | 27.882 | 39.767 | 31.405 |
| MF-FEN | RMSE | 0.1356 | 0.1107 | 0.1108 | 0.1234 | 0.1201 |
| | MAE | 0.1104 | 0.08129 | 0.0898 | 0.0993 | 0.0952 |
| | R2 | 0.7726 | 0.8486 | 0.8472 | 0.8120 | 0.8201 |
| | MSE | 0.0183 | 0.0122 | 0.0122 | 0.0152 | 0.0145 |
| | SMAPE(%) | 35.524 | 30.738 | 25.231 | 33.569 | 31.266 |
| HF-FEN | RMSE | 0.1072 | 0.1257 | 0.1166 | 0.1142 | 0.1159 |
| | MAE | 0.0793 | 0.1009 | 0.0920 | 0.0890 | 0.0903 |
| | R2 | 0.8580 | 0.8047 | 0.8309 | 0.8390 | 0.8332 |
| | MSE | 0.0115 | 0.0158 | 0.0135 | 0.0130 | 0.0135 |
| | SMAPE(%) | 26.503 | 33.833 | 26.945 | **25.800** | 28.270 |
| Random Allocation 1 | RMSE | 0.1246 | 0.1275 | 0.1118 | 0.1280 | 0.123 |
| | MAE | 0.0962 | 0.0991 | 0.0908 | 0.1027 | 0.0972 |
| | R2 | 0.8081 | 0.7992 | 0.8444 | 0.7978 | 0.8124 |
| | MSE | 0.0155 | 0.0163 | 0.0125 | 0.0163 | 0.0151 |
| | SMAPE(%) | 29.675 | 36.816 | 25.692 | 33.857 | 31.510 |
| Random Allocation 2 | RMSE | 0.1489 | 0.1185 | 0.1294 | 0.1248 | 0.1304 |
| | MAE | 0.1239 | 0.0945 | 0.1043 | 0.1023 | 0.1063 |
| | R2 | 0.7260 | 0.8266 | 0.7918 | 0.8079 | 0.7881 |
| | MSE | 0.0222 | 0.0140 | 0.0167 | 0.0156 | 0.0171 |
| | SMAPE(%) | 42.990 | 33.021 | 28.797 | 30.649 | 33.864 |
| Proposed Method | RMSE | **0.0727** | **0.0882** | **0.0715** | **0.0688** | **0.0753** |
| | MAE | **0.0500** | **0.0650** | **0.0539** | **0.0595** | **0.0571** |
| | R2 | **0.9346** | **0.9040** | **0.9365** | **0.9417** | **0.9292** |
| | MSE | **0.0052** | **0.0078** | **0.0051** | **0.0047** | **0.0057** |
| | SMAPE(%) | **19.702** | **25.391** | **20.386** | 32.560 | **24.510** |

The radar chart in Figure 5 provides a comprehensive overview of model performance across all metrics. The proposed method exhibits the smallest radial values for all error-related metrics, indicating reduced prediction errors and enhanced stability, while achieving the largest radial value for $R^2$, reflecting superior model accuracy.

Figures 6(a)-(d) illustrate the comparison between the predicted RUL trajectories and the corresponding actual RUL

values for each bearing. The RUL prediction trends across the six configurations further validated these findings. The proposed method closely follows the ground truth trajectory, while other configurations exhibit larger deviations, particularly at early and late stages of the bearing lifecycle.

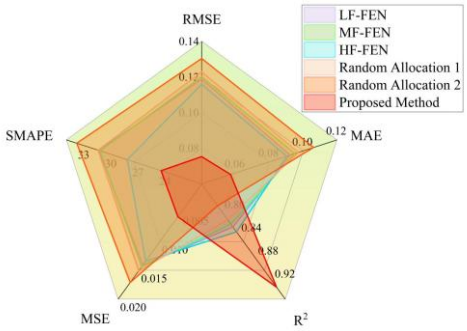


Figure 5. Performance comparison across different configurations under Case 1.

### 3.1.3. Comparison with diverse methods

To assess the advantages of the proposed model for bearing RUL prediction, it was compared with five established networks representing diverse feature extraction methods, all of which have shown strong performance in bearing RUL tasks. Controlled experiments were performed with consistent data preprocessing, training procedures, and evaluation metrics to ensure a fair comparison. In these experiments, under the proposed RUL prediction framework, the TFEN was replaced by each of the following models: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Multi-scale Dilated Convolutional Transformer (MDSCT), Temporal Attention Convolutional Transformer (TACT), and Multi-scale Wavelet Network (MWN). The comparison results are summarized in Table 4.

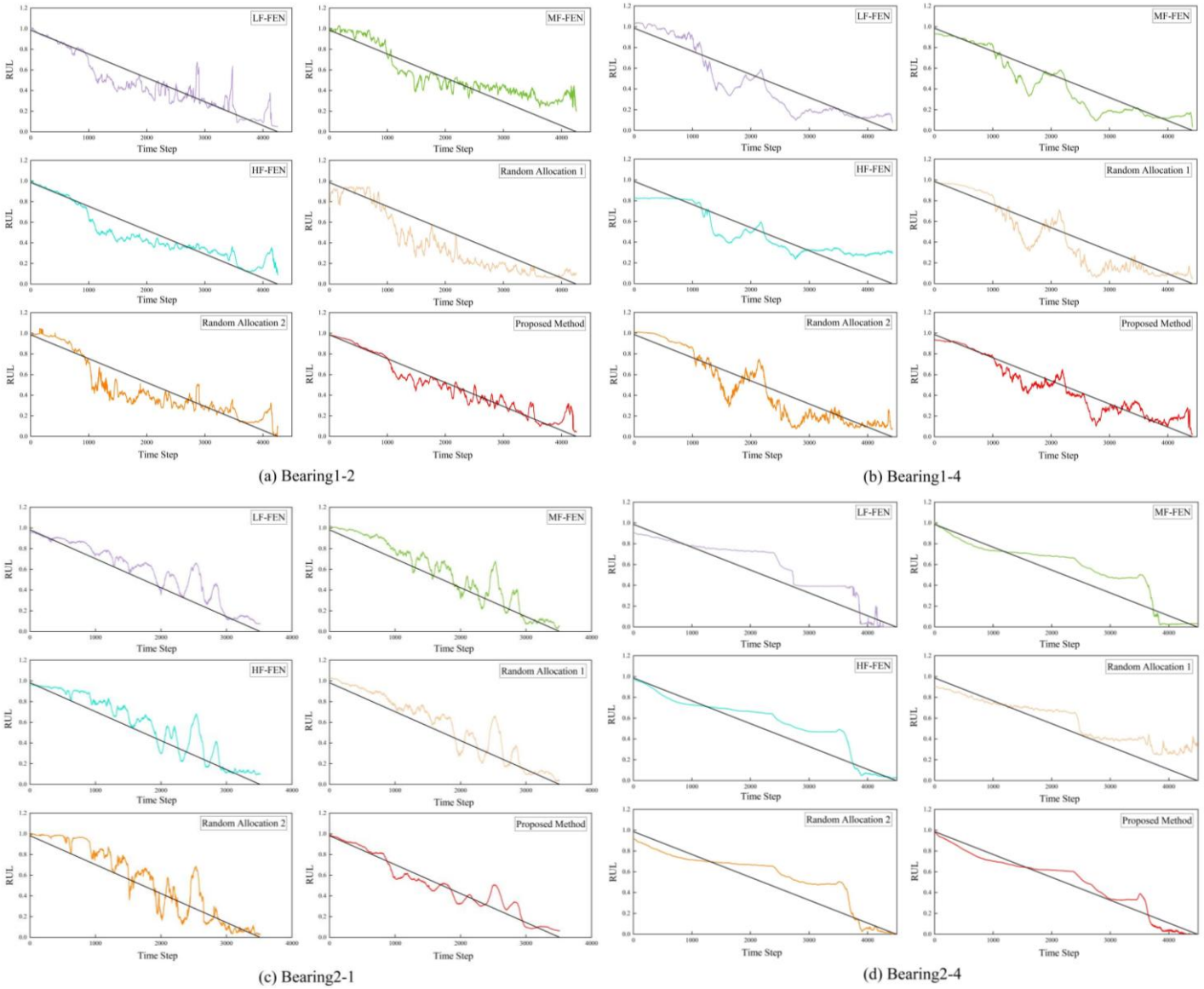


Figure 6. Predicted vs. actual RULs of four bearings using different configurations under Case 1.

As shown in Table 4, the proposed frequency-adaptive TFEN consistently outperforms all five representative methods across multiple evaluation metrics. Specifically, it achieves the lowest RMSE and MAE values, indicating higher prediction accuracy, and attains the highest $R^2$, demonstrating superior capability in capturing the underlying degradation patterns. Compared with traditional models such as CNN and LSTM, the proposed method shows marked improvements in robustness and generalization. The more advanced networks, including MDSCT, TACT, and MWN, achieve competitive results, but the proposed model still surpasses them, highlighting the effectiveness of the frequency-adaptive temporal feature extraction.

Table 4. Performance comparison between the proposed method and diverse methods under Case 1.

| Methods | Metric | Bearing ID | | | | |
|---|---|---|---|---|---|---|
| | | 1-2 | 1-4 | 2-1 | 2-4 | Average |
| CNN | RMSE | 0.1156 | 0.1277 | 0.1155 | 0.1153 | 0.1185 |
| | MAE | 0.0894 | 0.1006 | 0.0943 | 0.0911 | 0.0938 |
| | R2 | 0.8348 | 0.7986 | 0.8339 | 0.8358 | 0.8257 |
| | MSE | 0.0133 | 0.0163 | 0.0133 | 0.0133 | 0.0140 |
| | SMAPE(%) | 29.517 | 33.105 | 27.794 | **24.811** | 28.807 |
| LSTM | RMSE | 0.1353 | 0.1387 | 0.1009 | 0.1109 | 0.1214 |
| | MAE | 0.1136 | 0.1050 | 0.0742 | 0.0956 | 0.0971 |
| | R2 | 0.7738 | 0.7624 | 0.8732 | 0.8482 | 0.8144 |
| | MSE | 0.0183 | 0.0192 | 0.0101 | 0.0123 | 0.0149 |
| | SMAPE(%) | 34.994 | 37.537 | 27.470 | 49.346 | 37.336 |
| MDSCT | RMSE | 0.1124 | 0.1190 | 0.1008 | 0.1364 | 0.1171 |
| | MAE | 0.0860 | 0.0946 | 0.0745 | 0.1055 | 0.0901 |
| | R2 | 0.8438 | 0.8250 | 0.8736 | 0.7704 | 0.8282 |
| | MSE | 0.0126 | 0.0141 | 0.0101 | 0.0186 | 0.0138 |
| | SMAPE(%) | 28.335 | 34.492 | 28.856 | 46.118 | 34.450 |
| TACT | RMSE | 0.1096 | 0.0921 | 0.0871 | 0.1032 | 0.0980 |
| | MAE | 0.0823 | 0.0726 | 0.0643 | 0.0829 | 0.0755 |
| | R2 | 0.8516 | 0.8954 | 0.9058 | 0.8685 | 0.8803 |
| | MSE | 0.0120 | 0.0085 | 0.0076 | 0.0107 | 0.0097 |
| | SMAPE(%) | 28.100 | 27.814 | 22.135 | 26.099 | 26.037 |
| MWN | RMSE | 0.1125 | 0.1190 | 0.1044 | 0.1395 | 0.1188 |
| | MAE | 0.0899 | 0.0949 | 0.0810 | 0.1077 | 0.0933 |
| | R2 | 0.8435 | 0.8251 | 0.8644 | 0.7599 | 0.8232 |
| | MSE | 0.0126 | 0.0141 | 0.0109 | 0.0194 | 0.0142 |
| | SMAPE(%) | 27.572 | 32.235 | 31.638 | 27.960 | 29.851 |
| Proposed Method | RMSE | **0.0727** | **0.0882** | **0.0715** | **0.0688** | **0.0753** |
| | MAE | **0.0500** | **0.0650** | **0.0539** | **0.0595** | **0.0571** |
| | R2 | **0.9346** | **0.9040** | **0.9365** | **0.9417** | **0.9292** |
| | MSE | **0.0052** | **0.0078** | **0.0051** | **0.0047** | **0.0057** |
| | SMAPE(%) | **19.703** | **25.396** | **20.381** | 32.560 | **24.510** |

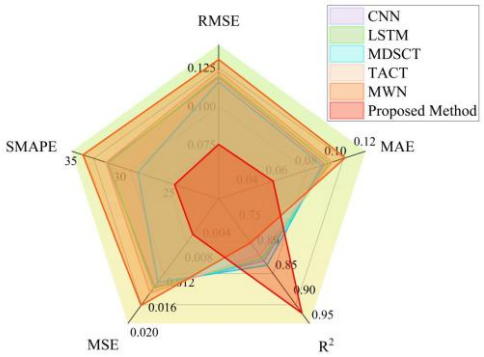


Figure 7. Radar chart of model performance across evaluation metrics under Case 1.

Figure 7 presents a comparison of the average evaluation metrics for each model across the four bearings. Figures 8(a)-(d) illustrate the predicted versus actual RUL values for each model on the four bearings, respectively. Compared with the best-performing TACT, the proposed method still achieved an improvement of 23.1% in the RMSE metric.

### 3.2. Case study 2: experiments on accelerated degradation dataset

#### 3.2.1. Description of the accelerated degradation dataset

To further verify the advantages of the proposed method, bearing RUL prediction experiments were conducted on a mechanical component accelerated degradation test platform, which collected data throughout the full life cycle of bearings under real working conditions. The structure of the test platform

is depicted in Figure 9.

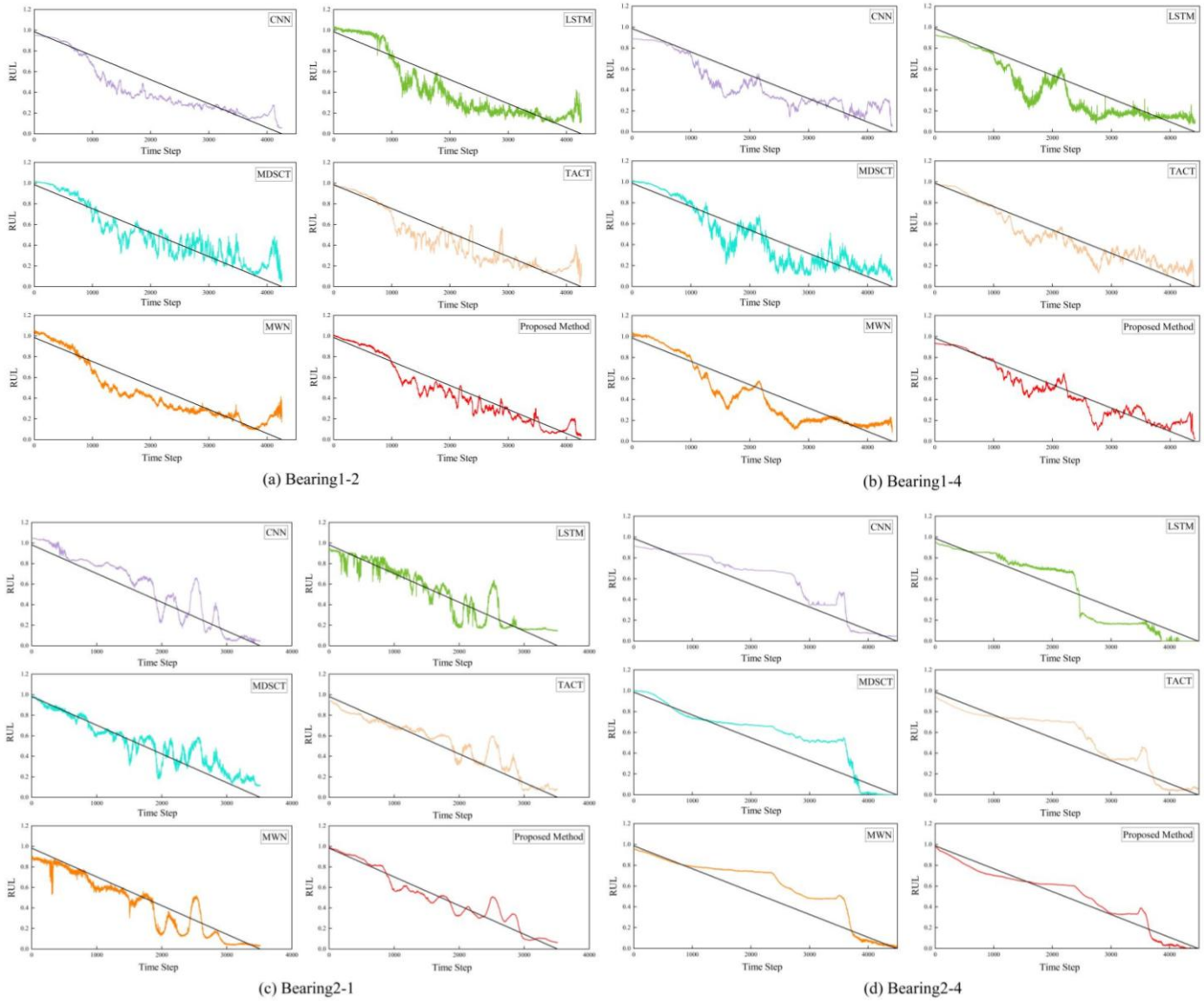


Figure 8. Predicted vs. actual RULs of four bearings using different models under Case 1.

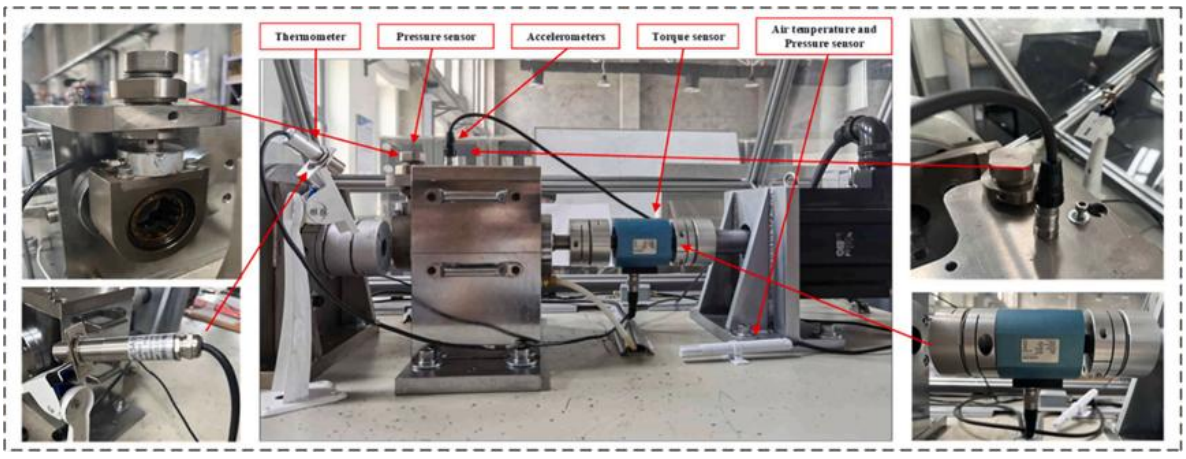


Figure 9. Accelerated degradation test platform.

The platform comprises three main components: the drive (AC) motor, drive shaft, and bearing test chamber. The vibration acceleration, radial force, operating torque, and operating temperature of the experimental bearings were measured using specific sensors: vibration acceleration sensor (PCB 603C01), tensile force sensor (Ritcl T306), torque sensor (Kistler 4520A),

and infrared thermometer (SA15CS). Additionally, ambient temperature and pressure were recorded by the thermometer and barometer (ZZ-TA3). Table 5 summarizes the working conditions of each bearing, including the corresponding service lifetimes.

### 3.2.2. Evaluation of the frequency-adaptive TFEN

Case 2 focuses on a more homogeneous data environment. Specifically, Case 2 contains multi-sensor, synchronously sampled measurements collected under a unified acquisition system. In the experiments, only four bearings from condition 2 (1400 rpm) are used for training and testing. Compared with Case 1, the data in Case 2 exhibit lower variability, providing a clearer scenario to evaluate the model's achievable performance upper bound. The experiments on the second dataset were conducted using the same configurations and procedures as described for Case 1. This approach ensured consistency in evaluating the effectiveness of the frequency-adaptive TFEN across datasets with different data characteristics. The comparison results are shown in Table 6 below.

Table 5. Introduction to the accelerated degradation dataset.

| Bearing ID | Condition | Lifetime (min) |
|---|---|---|
| Bearing 1-1 | 1 | 118.5 |
| Bearing 1-2 | 1 | 122.7 |
| Bearing 2-1 | 2 | 245.2 |
| Bearing 2-2 | 2 | 252.3 |
| Bearing 2-3 | 2 | 248.5 |
| Bearing 2-4 | 2 | 242.7 |

Table 6. Performance comparison across different configurations under Case 2.

| Methods | Metric | Bearing ID | | | | |
|---|---|---|---|---|---|---|
| | | 2-1 | 2-2 | 2-3 | 2-4 | Average |
| LF-FEN | RMSE | 0.0220 | 0.0769 | 0.0300 | 0.0249 | 0.0384 |
| | MAE | 0.0175 | 0.0667 | 0.0236 | 0.0209 | 0.0321 |
| | R2 | 0.9936 | 0.9228 | 0.9882 | 0.9919 | 0.9741 |
| | MSE | 0.0004 | 0.0059 | 0.0009 | 0.0006 | 0.0019 |
| | SMAPE(%) | 19.328 | 22.771 | 25.473 | 10.410 | 19.495 |
| MF-FEN | RMSE | 0.0194 | 0.0522 | 0.0266 | 0.0242 | 0.0306 |
| | MAE | 0.0166 | 0.0454 | 0.0250 | 0.0207 | 0.0269 |
| | R2 | 0.9951 | 0.9644 | 0.9907 | 0.9923 | 0.9856 |
| | MSE | 0.0003 | 0.0027 | 0.0007 | 0.0006 | 0.0010 |
| | SMAPE(%) | 7.317 | 16.839 | 11.732 | 12.083 | 11.992 |
| HF-FEN | RMSE | 0.0123 | 0.0516 | 0.0281 | 0.0308 | 0.0307 |
| | MAE | 0.0103 | 0.0465 | 0.0248 | 0.0252 | 0.0267 |
| | R2 | 0.9980 | 0.9653 | 0.9896 | 0.9876 | 0.9851 |
| | MSE | **0.0001** | 0.0027 | 0.0007 | 0.0009 | 0.0011 |
| | SMAPE(%) | 6.2937 | 22.951 | 13.816 | 14.570 | 14.407 |
| Random allocation 1 | RMSE | 0.0181 | 0.0647 | 0.0250 | 0.0427 | 0.0376 |
| | MAE | 0.0127 | 0.0601 | 0.0216 | 0.0366 | 0.0327 |
| | R2 | 0.9957 | 0.9452 | 0.9918 | 0.9760 | 0.9771 |
| | MSE | 0.0003 | 0.0041 | 0.0006 | 0.0018 | 0.0017 |
| | SMAPE(%) | 7.0704 | 20.784 | 13.848 | 10.923 | 13.156 |
| Random allocation 2 | RMSE | 0.0165 | 0.0565 | 0.0216 | 0.0380 | 0.0331 |
| | MAE | 0.0113 | 0.0437 | 0.0185 | 0.0339 | 0.0268 |
| | R2 | 0.9964 | 0.9582 | 0.9939 | 0.9810 | 0.9823 |
| | MSE | 0.0002 | 0.0032 | 0.0004 | 0.0014 | 0.0013 |
| | SMAPE(%) | 4.7218 | 19.583 | 10.745 | 15.735 | 12.696 |
| Proposed Method | RMSE | **0.0059** | **0.0261** | **0.0070** | **0.0126** | **0.0129** |
| | MAE | **0.0044** | **0.0225** | **0.0058** | **0.0085** | **0.0103** |
| | R2 | **0.9995** | **0.9911** | **0.9993** | **0.9979** | **0.9969** |
| | MSE | **0.0001** | **0.0006** | **0.0001** | **0.0002** | **0.0002** |
| | SMAPE(%) | **4.6230** | **11.607** | **4.1761** | **6.9012** | **6.8268** |

To provide an intuitive comparison across all evaluation metrics, a radar chart shown in Figure 10 is used to visualize the performance of the proposed method and the models based on other configurations.

Across all four bearings, the adaptive TFEN outperformed single-frequency band configurations. The average RMSE decreased to 0.0129, while $R^2$ increased to 0.9969. Compared to the optimal single-frequency band configuration MF-FEN, the adaptive TFEN achieved an average RMSE reduction of approximately 57.8%, and an $R^2$ improvement from 0.9856 to 0.9969. Compared with the random allocation configurations, the average RMSE decreased by 63.5%.

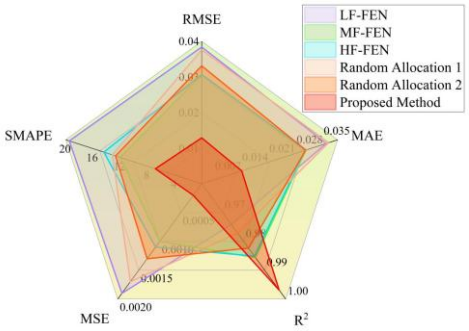


Figure 10. Performance comparison across different configurations under Case 2.

The performance improvements highlight the effectiveness of the frequency-adaptive assignment in enhancing RUL prediction accuracy. These two bearings (2-2 and 2-4) are further examined in the RUL trend plots presented in Figure 11, where the adaptive TFEN also shows the most stable and accurate tracking of the degradation trajectory.

It is evident that Case 2 achieves noticeably higher prediction accuracy than Case 1, this improvement is primarily attributed to the more consistent degradation dynamics and uniform operating conditions in Case 2. Under the more stable setting of Case 2, the proposed method achieves the closest alignment between predicted and true RUL trajectories among all configurations, demonstrating its strong capacity to exploit a consistent degradation pattern.

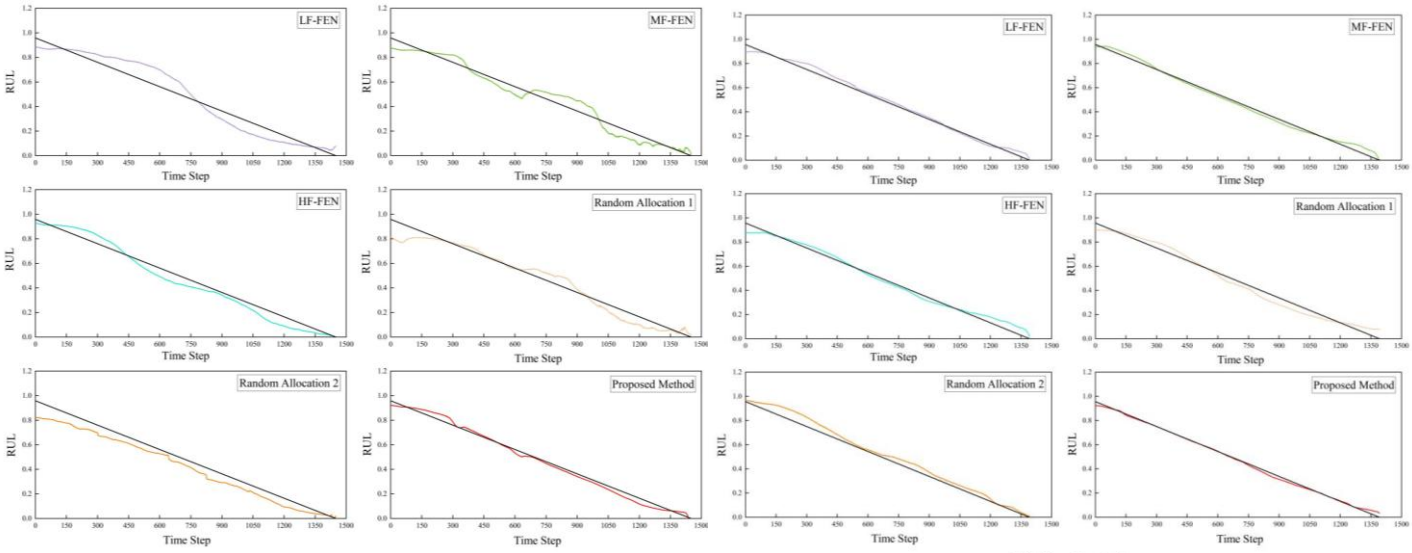


Figure 11. Predicted vs. actual RULs of two bearings using different configurations under Case 2.

### 3.2.3. Comparison with diverse methods

To evaluate model performance under homogeneous operating conditions, Case 2 follows the same comparison protocol as Case 1, involving the proposed method and five representative deep models. On the accelerated degradation test platform under condition 2, four bearings were tested at 1400 rpm to evaluate the performance of the frequency-adaptive TFEN. The comparison results are shown in Table 7 below.

The experimental results in Table 7 clearly show that the proposed model exhibits the best performance on all four test bearings. Under the relatively homogeneous operating condition of Case 2, the model can learn a more stable mapping between sensor measurements and RUL. Compared to the best-performing comparison model MDSCT, the RMSE is reduced by approximately 58.1%, and other metrics also show an optimization trend.

A radar chart is then presented in Figure 12 to summarize the overall error-based metrics across the four bearings, followed by RUL trend plots for bearings 2-2 and 2-4 to illustrate the prediction performance of each method.

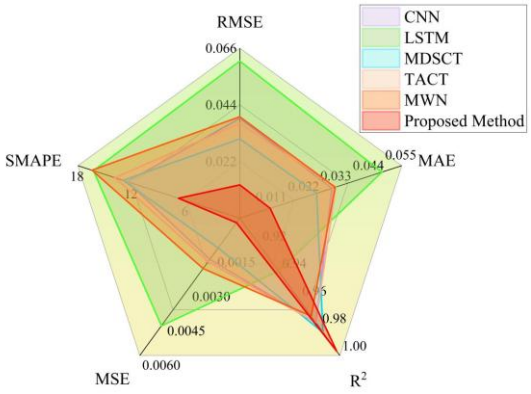


Figure. 12. Radar chart of model performance across evaluation metrics under Case 2.

From Figure 13, the proposed frequency-adaptive TFEN consistently outperforms all comparison methods, demonstrating that aligning feature extractors to channel-specific spectral characteristics improves representation quality in homogeneous settings, highlighting the effectiveness of the proposed method in enhancing RUL prediction accuracy.

Table 7. Performance comparison between the proposed method and diverse methods under Case 2.

| Methods | Metric | Bearing ID | | | | |
|---|---|---|---|---|---|---|
| | | 2-1 | 2-2 | 2-3 | 2-4 | Average |
| CNN | RMSE | 0.0212 | 0.0670 | 0.0242 | 0.0426 | 0.0388 |
| | MAE | 0.0185 | 0.0530 | 0.0205 | 0.0330 | 0.0313 |
| | R2 | 0.9941 | 0.9413 | 0.9923 | 0.9762 | 0.9760 |
| | MSE | 0.0004 | 0.0045 | 0.0006 | 0.0018 | 0.0018 |
| | SMAPE(%) | 9.7431 | 13.861 | 9.5682 | 16.893 | 12.516 |
| LSTM | RMSE | 0.0406 | 0.1099 | 0.0283 | 0.0651 | 0.0610 |
| | MAE | 0.0322 | 0.0861 | 0.0228 | 0.0521 | 0.0483 |
| | R2 | 0.9784 | 0.8421 | 0.9895 | 0.9444 | 0.9386 |
| | MSE | 0.0017 | 0.0121 | 0.0008 | 0.0042 | 0.0047 |
| | SMAPE(%) | 13.140 | 19.657 | 9.5756 | 22.738 | 16.277 |
| MDSCT | RMSE | 0.0161 | 0.0623 | 0.0191 | 0.0257 | 0.0308 |
| | MAE | 0.0139 | 0.0526 | 0.0167 | 0.0211 | 0.0261 |
| | R2 | 0.9966 | 0.9493 | 0.9952 | 0.9913 | 0.9831 |
| | MSE | 0.0003 | 0.0039 | 0.0004 | 0.0007 | 0.0013 |
| | SMAPE(%) | 10.277 | 21.938 | 9.5381 | 10.198 | 12.987 |
| TACT | RMSE | 0.0186 | 0.0727 | 0.0189 | 0.0368 | 0.0368 |
| | MAE | 0.0153 | 0.0660 | 0.0155 | 0.0335 | 0.0326 |
| | R2 | 0.9955 | 0.9310 | 0.9953 | 0.9822 | 0.9760 |
| | MSE | 0.0003 | 0.0053 | 0.0004 | 0.0014 | 0.0019 |
| | SMAPE(%) | 9.5246 | 22.417 | 9.9156 | 13.720 | 13.894 |
| MWN | RMSE | 0.0173 | 0.0754 | 0.0138 | 0.0511 | 0.0394 |
| | MAE | 0.0155 | 0.0602 | 0.0122 | 0.0415 | 0.0324 |
| | R2 | 0.9961 | 0.9258 | 0.9975 | 0.9657 | 0.9713 |
| | MSE | 0.0003 | 0.0057 | 0.0002 | 0.0026 | 0.0022 |
| | SMAPE(%) | 9.0901 | 29.894 | 7.1677 | 19.619 | 16.442 |
| Proposed Method | RMSE | **0.0059** | **0.0261** | **0.0070** | **0.0126** | **0.0129** |
| | MAE | **0.0044** | **0.0225** | **0.0058** | **0.0085** | **0.0103** |
| | R2 | **0.9995** | **0.9911** | **0.9993** | **0.9979** | **0.9969** |
| | MSE | **0.0001** | **0.0006** | **0.0001** | **0.0002** | **0.0002** |
| | SMAPE(%) | **4.6223** | **11.600** | **4.1770** | **6.9062** | **6.8263** |

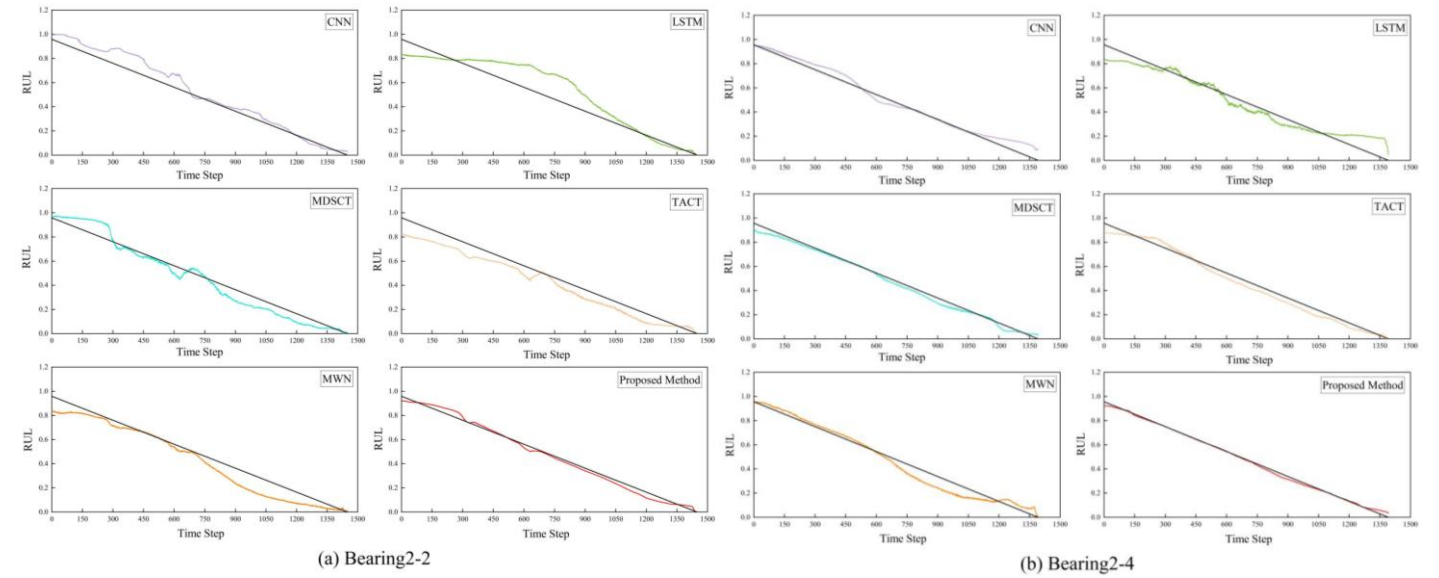


Figure 13. Predicted vs. actual RULs of two bearings using different models under Case 2.

## 4. Conclusion

This paper presented a frequency-adaptive framework for bearing remaining useful life prediction. A tailored TFEN was developed with three specialized branch configurations and a frequency-adaptive mechanism that assigns each sensor channel to the most appropriate feature extractor based on its frequency characteristics. Combined with the SFEN, the

framework effectively captures both band-specific degradation patterns and inter-sensor dependencies.

Extensive experiments on two bearing degradation datasets demonstrated that the proposed method consistently outperformed single-band designs, random allocation strategies, and five representative deep models. The results verified its robustness under distribution shifts and its strong predictive capability in different environments.

In the absence of explicit physical priors for heterogeneous sensors, we employ an equal-division scheme to ensure a deterministic and dataset-agnostic partition rule. This choice is consistent with the three-branch TFEN architecture and reduces additional boundary-related hyperparameters, thereby mitigating overfitting risk under limited run-to-failure observations. However, the currently designed prediction network models still have limitations. The current frequency allocation strategy and AF index are essentially statistical designs, lacking the detailed physical knowledge that explicitly incorporates bearing dynamics. In future work, we will investigate physics-informed partitioning by integrating rotating speed, bearing characteristic frequencies, and resonance-band information, so as to adaptively adjust band boundaries across varying operating conditions. Future research will explore semi-supervised and constrained remaining useful life learning methods to reduce reliance on the labels of running until failure, and develop more frequency partition schemes based on physical principles. These schemes can adaptively adjust the band boundaries according to operating conditions and fault characteristics to further enhance the reliability of the remaining useful life prediction model.

**References**


1. Aizpurua J I, Catterson V M, Abdulhadi I F, Garcia M S. A Model-Based Hybrid Approach for Circuit Breaker Prognostics Encompassing Dynamic Reliability and Uncertainty. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2018; 48(9): 1637–1648, https://doi.org/10.1109/TSMC.2017.2685346.
2. Alfarizi M G, Tajiani B, Vatn J, Yin S. Optimized Random Forest Model for Remaining Useful Life Prediction of Experimental Bearings. IEEE Transactions on Industrial Informatics 2023; 19(6): 7771–7779, https://doi.org/10.1109/TII.2022.3206339.
3. Chen Z S, Yang Y M, Hu Z. A Technical Framework and Roadmap of Embedded Diagnostics and Prognostics for Complex Mechanical Systems in Prognostics and Health Management Systems. IEEE Transactions on Reliability 2012; 61(2): 314–322, https://doi.org/10.1109/TR.2012.2196171.
4. Chiachío J, Jalón M L, Chiachío M, Kolios A. A Markov chains prognostics framework for complex degradation processes. Reliability Engineering & System Safety 2020; 195: 106621, https://doi.org/10.1016/j.ress.2019.106621.
5. El-Thalji I, Jantunen E. A summary of fault modelling and predictive health monitoring of rolling element bearings. Mechanical Systems and Signal Processing 2015; 60–61: 252–272, https://doi.org/10.1016/j.ymssp.2015.02.008.
6. Feng J, Cai F, Li H et al. A data-driven prediction model for the remaining useful life prediction of lithium-ion batteries. Process Safety and Environmental Protection 2023; 180: 601–615, https://doi.org/10.1016/j.psep.2023.10.042.
7. Guo R, Li H, Huang C. Operation stage division and RUL prediction of bearings based on 1DCNN-ON-LSTM. Measurement Science and Technology 2024; 35(2): 025035, https://doi.org/10.1088/1361-6501/ad0e3a.
8. Gupta S, Thakar U, Tokekar S. A comprehensive survey on techniques for numerical similarity measurement. Expert Systems with Applications 2025; 277: 127235, https://doi.org/10.1016/j.eswa.2025.127235.
9. d'Hondt J E, Li H, Yang F et al. A Structured Study of Multivariate Time-Series Distance Measures. Proceedings of the ACM on Management of Data 2025; 3(3): 1–29, https://doi.org/10.1145/3725258.
10. Hu A, Zhu Y, Liu S et al. A novel vision transformer network for rolling bearing remaining useful life prediction. Measurement Science and Technology 2024; 35(2): 025106, https://doi.org/10.1088/1361-6501/ad0705.
11. Hu Y, Liu S, Lu H, Zhang H. Remaining Useful Life Model and Assessment of Mechanical Products: A Brief Review and a Note on the State Space Model Method. Chinese Journal of Mechanical Engineering 2019; 32(1): 15, https://doi.org/10.1186/s10033-019-0317-y.
12. Li J, Wang Z, Tian D et al. A novel RUL prediction framework based on the adaptability feature perception fusion model method. Journal of Energy Storage 2025; 119: 116322, https://doi.org/10.1016/j.est.2025.116322.
13. Liu Y, Hu T, Zhang H et al. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. 2024.

doi:10.48550/arXiv.2310.06625, https://doi.org/10.48550/arXiv.2310.06625.

14. Luo X, Bu W, Liang H, Zheng M. Convolutional Neural Network - Gated Recurrent Unit combined with Error Correction for Lithium Battery State of Health Estimation. Eksploatacja i Niezawodność – Maintenance and Reliability 2025. doi:10.17531/ein/202184, https://doi.org/10.17531/ein/202184.
15. Lyu Y, Qiu Q, Chu Y, Zhang J. An Adaptive BiGRU-ASSA-iTransformer Method for Remaining Useful Life Prediction of Bearing in Aerospace Manufacturing. Actuators 2025; 14(5): 238, https://doi.org/10.3390/act14050238.
16. Nectoux P, Gouriveau R, Medjaher K et al. PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests. 2012.
17. Ruan D, Ma L, Yang Y et al. Improvement by Monte Carlo for Trajectory Similarity-Based RUL Prediction. IEEE Transactions on Instrumentation and Measurement 2024; 73: 1–11, https://doi.org/10.1109/TIM.2024.3353866.
18. Soydaner D. Attention mechanism in neural networks: where it comes and where it goes. Neural Computing and Applications 2022; 34(16): 13371–13385, https://doi.org/10.1007/s00521-022-07366-3.
19. Thelen A, Li M, Hu C et al. Augmented model-based framework for battery remaining useful life prediction. Applied Energy 2022; 324: 119624, https://doi.org/10.1016/j.apenergy.2022.119624.
20. Wan S, Li X, Zhang Y et al. Bearing remaining useful life prediction with convolutional long short-term memory fusion networks. Reliability Engineering & System Safety 2022; 224: 108528, https://doi.org/10.1016/j.ress.2022.108528.
21. Wang C, Zhang L, Chen L et al. Remaining useful life prediction of nuclear reactor control rod drive mechanism based on dynamic temporal convolutional network. Reliability Engineering & System Safety 2025; 253: 110580, https://doi.org/10.1016/j.ress.2024.110580.
22. Wang D, Tsui K-L. Statistical Modeling of Bearing Degradation Signals. IEEE Transactions on Reliability 2017; 66(4): 1331–1344, https://doi.org/10.1109/TR.2017.2739126.
23. Wang T, Li X, Wang W et al. A spatiotemporal feature learning-based RUL estimation method for predictive maintenance. Measurement 2023; 214: 112824, https://doi.org/10.1016/j.measurement.2023.112824.
24. Wang Y, Deng L, Zheng L, Gao R X. Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics. Journal of Manufacturing Systems 2021; 60: 512–526, https://doi.org/10.1016/j.jmsy.2021.07.008.
25. Wang Z, Guan Y. Multiscale convolutional neural-based transformer network for time series prediction. Signal, Image and Video Processing 2024; 18(2): 1015–1025, https://doi.org/10.1007/s11760-023-02823-5.
26. Wu Q, Zong T, Cheng W et al. Network Construction for Bearing Fault Diagnosis Based on Double Attention Mechanism. Computational Intelligence and Neuroscience 2022; 2022: 1–17, https://doi.org/10.1155/2022/3987480.
27. Xu Z, Chen X, Yang L et al. Multi-modal adaptive feature extraction for early-stage weak fault diagnosis in bearings. Electronic Research Archive 2024; 32(6): 4074–4095, https://doi.org/10.3934/era.2024183.
28. Yan M, Wang X, Wang B et al. Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. ISA Transactions 2020; 98: 471–482, https://doi.org/10.1016/j.isatra.2019.08.058.
29. Yang S, Tang B, Wang W et al. Physics-informed multi-state temporal frequency network for RUL prediction of rolling bearings. Reliability Engineering & System Safety 2024; 242: 109716, https://doi.org/10.1016/j.ress.2023.109716.
30. Yao X, Zhu J, Jiang Q et al. RUL prediction method for rolling bearing using convolutional denoising autoencoder and bidirectional LSTM. Measurement Science and Technology 2024; 35(3): 035111, https://doi.org/10.1088/1361-6501/ad123c.
31. Zhang Q, Liu Q, Ye Q. An attention-based temporal convolutional network method for predicting remaining useful life of aero-engine. Engineering Applications of Artificial Intelligence 2024; 127: 107241, https://doi.org/10.1016/j.engappai.2023.107241.
32. Zhao B, Zhang X, Li H, Yang Z. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. Knowledge-Based Systems 2020; 199: 105971, https://doi.org/10.1016/j.knosys.2020.105971.
33. Zhao H, Liu H, Jin Y et al. Feature Extraction for Data-Driven Remaining Useful Life Prediction of Rolling Bearings. IEEE Transactions on Instrumentation and Measurement 2021; 70: 1–10, https://doi.org/10.1109/TIM.2021.3059500.