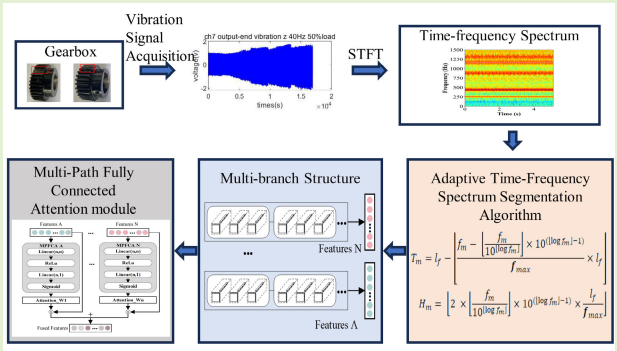
# Multibranch Time-Frequency Spectrum Segmentation Approach for Gearbox Fault Diagnosis

Hongliang Song, Yi Sun, Hongli Gao, Liang Guo, *Member, IEEE*, and Yucheng Chen

***Abstract*—In gearbox fault diagnosis research, fully utilizing the time-frequency information at meshing frequencies is crucial for enhancing the accuracy and robustness of fault diagnosis models. This study proposes a multibranch time-frequency spectrum segmentation network (TFSSN) model for gearbox fault diagnosis, employing deep learning to fully explore and integrate the time-frequency information from multiple meshing frequencies of multistage gearboxes. An adaptive time-frequency spectrum segmentation algorithm is proposed within this model to segment the time-frequency spectrum at various meshing frequencies of the gearbox. A multibranch structure is then used to extract features from each segmented time-frequency spectrum. Finally, the designed multipath fully connected attention (MPFCA) module fuses the features extracted from different time-frequency spectrum branches to accomplish fault diagnosis of multistage gearboxes. The performance of TFSSN was validated using gearbox datasets under various operating conditions. The experiments demonstrate that this method effectively extracts features strongly correlated with faults and significantly mitigates the overfitting issue in neural network models under conditions of sparse fault samples. Compared to conventional gearbox fault diagnosis methods, the proposed approach significantly improves the accuracy and robustness of fault diagnosis under conditions with limited fault samples. Specifically, the method achieves an accuracy of 0.982 and an *F*1 score of 0.99, demonstrating superior performance over existing methods.**




***Index Terms*— Attention mechanism, gearbox fault diagnosis, machine learning, neural networks.**

## I. Introduction

GEARBOXES are commonly used for power transmission in mechanical systems. Serious financial and safety consequences can occur due to the failure of vital gears, whose reliability is critical to many industrial applications. Fault detection and diagnosis of gearboxes can prevent unplanned downtime, expensive repairs, damage to equipment system performance, and even casualties [1]. Therefore, accurately identifying the health status of the gearbox is crucial to ensure the reliability and safety of rotating machinery system operations.

Received 25 August 2024; accepted 2 September 2024. Date of publication 17 September 2024; date of current version 14 November 2024. This work was supported in part by the Natural Science Foundation of Sichuan Province under Grant 2023YFG0030 and in part by the Key Research and Development Program of Liaoning Province under Grant 2022JH1/10400031. The associate editor coordinating the review of this article and approving it for publication was Dr. Te Han. *(Corresponding author: Hongli Gao.)*

The authors are with the College of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610000, China (e-mail: cdsonghl@my.swjtu.edu.cn; yisun@my.swjtu.edu.cn; hongli_gao@home.swjtu.edu.cn; guoliang@swjtu.edu.cn; cycle7@my.swjtu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2024.3457878

In the research on fault diagnosis of gearboxes, various sensor signals can characterize their health status to a certain extent. Notably, vibration-based fault diagnosis technology is the most mature and effective method for structural damage detection [2] and fault diagnosis [3] in rotating machinery. As a result, it has been widely applied.

Signal analysis techniques for fault diagnosis can be roughly divided into time-domain analysis, frequency-domain analysis, and time-frequency analysis. Time-frequency analysis can consider the frequency components of signals at different time points and combine the advantages of time-domain analysis and frequency-domain analysis. It can better capture transient changes caused by faults and provide more comprehensive and multidimensional information. Recently, it has been widely used to diagnose and monitor mechanical equipment faults. Traditional methods include linear time-frequency representations such as the short-time Fourier transform (STFT) [4], wavelet transform (WT) [5], and quadratic time-frequency distributions. Researchers have integrated higher-order statistics with time-frequency analysis to address signals' nonlinearity and non-Gaussian nature, constructing time-varying higher-order spectrum methods [6], [7].

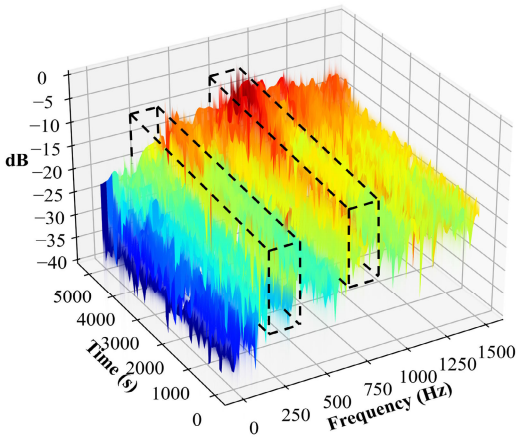



Fig. 1. Spectrum of gear wear at full life cycle, the range circled by the black box is the meshing frequency and its sidebands, the meshing frequency is 440 and 1160 Hz.

Simultaneously, numerous studies have indicated that the vibration energy of gears is primarily concentrated at the meshing frequency [8]. Gear meshing frequency refers to the frequency at which the gear teeth engage and disengage when two gears mesh and transmit power. The gear meshing frequency and sidebands are the most critical components in the gear vibration spectrum for gear fault diagnosis. As shown in Fig. 1, more energy is concentrated around the end-of-life meshing frequency and its associated sidebands. Research has revealed that amplitude modulation occurs when gear mesh eccentricity or misalignment occurs. Additionally, local gear faults can cause variations in gear angular speed during rotation, leading to frequency modulation. In numerous cases, gear fault signals exhibit the coexistence of both amplitude and frequency modulation, increasing the number and amplitude of sidebands [9].

Several scholars have investigated gearbox fault diagnosis utilizing machine-learning approaches [10]. Cui et al. [11] used a simple linear support vector machine (SVM) to classify normal and faulty data. Then, the FD-SAE was proposed for fault classification. Wang et al. [12] proposed a novel hybrid approach of a random forests (RFs) classifier for the fault diagnosis in rolling bearings. Soares et al. [13] proposed a combination of Wavelet Packet Decomposition and Gradient Boosting Decision Tree for feature extraction and classification for bearing faults and surface wear. Zhang et al. [14] developed a diagnostic method for compound faults in gearboxes based on multifeature and adaptive boost (AdaBoost), demonstrating that their algorithm surpasses traditional fault diagnosis methods in accuracy.

In recent years, data-driven deep learning methods for fault diagnosis have significantly developed [15], [16]. Convolutional neural networks (CNNs), long short-term memory (LSTM), stacked sparse autoencoder (SSAE), and deep belief networks (DBNs), among other deep learning models, have been widely used in fault diagnosis systems. CNN, in particular, has proven to be a powerful diagnostic tool because it can automatically extract features, handle large-scale and multimodal data, possess hierarchical feature learning capabilities, and adapt to hardware acceleration. Jing et al. [17] proposed a gear fault diagnosis method based on CNN. Experimental results demonstrated that CNN achieved high diagnostic accuracy by extracting features from raw data adaptively. Azamfar et al. [18] introduced a method that uses 2-D-CNN to merge data from multiple current sensors. They stacked the spectrum vectors of current signals collected by each sensor into a 2-D matrix and utilized 2-D-CNN to extract matrix features for fault classification.

Recent studies have highlighted significant challenges in applying deep CNN in fault diagnosis domains. Notably, while CNN excel in extracting intricate features, this capability renders them susceptible to overfitting, especially in scenarios characterized by a paucity of fault samples. The primary mechanism of CNN in achieving elevated classification accuracy is through gradient backpropagation. However, this process may inadvertently incorporate noncritical information into the classification schema, such as distinct signal noise or interference signals. Consequently, when these CNN models are exposed to new datasets that lack these particular signal characteristics, there is a substantial risk of a marked deterioration in the model's accuracy for fault diagnosis.

In this study, we investigate and address the prevalent issues of overfitting and compromised robustness in neural network-based models, particularly within gearbox fault diagnosis tasks characterized by a limited number of fault samples. The article proposes a multibranch time-frequency spectrum segmentation network (TFSSN) by leveraging the information from the meshing frequency and its sidebands in the time-frequency spectrum of gearbox vibrations. The method introduces the following innovations.

1) *Adaptive Time-Frequency Spectrum Segmentation Algorithm:* A novel algorithm is proposed for adaptive segmentation of time-frequency spectrum for different scales of meshing frequencies and their sidebands. This algorithm achieves domain transformation and noise reduction in the data, thereby enhancing the accuracy and robustness of the fault diagnosis model.
2) *Attention-Based Fusion of Multibranch Structure:* The proposed structure can adaptively accommodate multiple meshing frequencies and their sidebands in the time-frequency spectrum. Leveraging an attention mechanism, the model intelligently fuses the output features from numerous branches. This attention-based fusion not only enhances the model's ability to capture diverse and robust features but also ensures a comprehensive exploration of the information contained within the time-frequency spectrum.

The structure of this article is as follows: Section II presents the theoretical decomposition of the multibranch TFSSN method. In Section III, specific model parameters and the experimental dataset were introduced. Section IV discussed the experimental results. The final section provided a summary and conclusions.

## II. Methods

This section outlines the procedure for constructing the proposed TFSSN. As depicted in Fig. 2, it consists of two

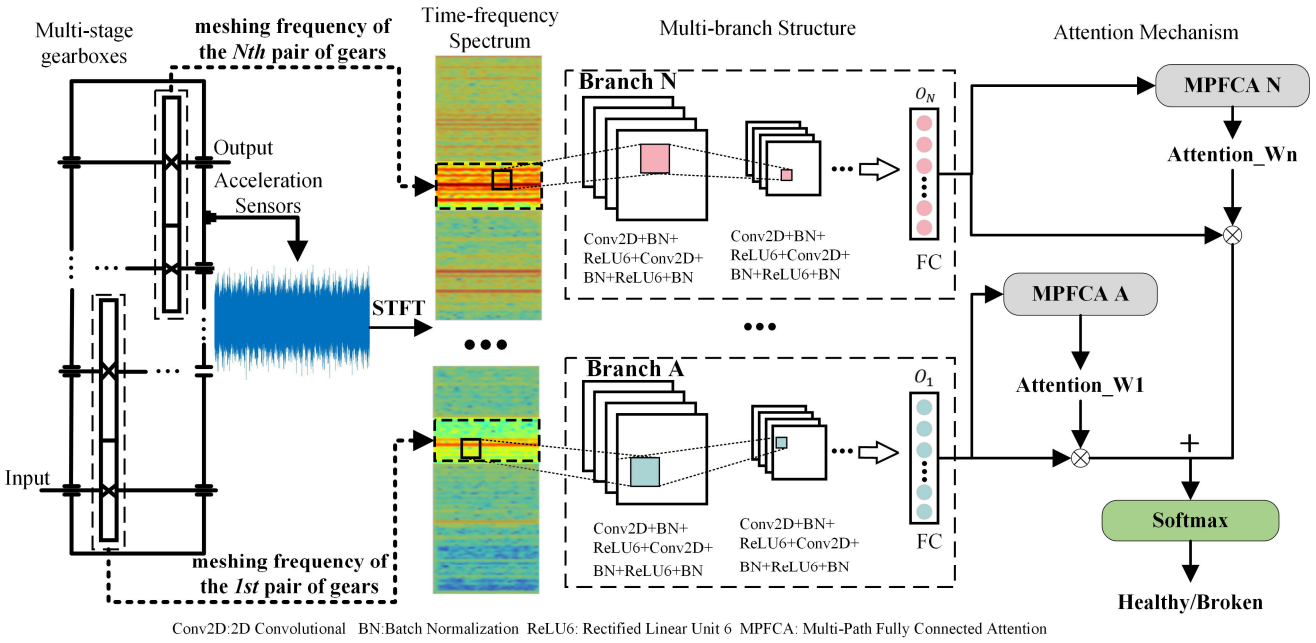



Fig. 2. Proposed multistage gearbox fault diagnosis model.

stages: an adaptive time-frequency spectrum segmentation algorithm and a multibranch shallow structure based on attention fusion. Initially, the vibration data undergo STFT to obtain the time-frequency spectrum. Subsequently, the adaptive time-frequency spectrum segmentation algorithm is applied to extract spectrum corresponding to different meshing frequencies and sidebands. Following this, the multibranch shallow structure selectively extracts features from different spectra related to various meshing frequencies. Finally, an attention mechanism is introduced to fuse the features extracted by the multibranch shallow structure, thereby achieving fault diagnosis.

### A. Adaptive Time-Frequency Spectrum Segmentation Algorithm

The methodology proposed in this article necessitates extracting the time-frequency spectrum at the meshing frequency of a fixed-axis multistage transmission gear. Various time-frequency spectrum extraction methods exist for time-series signals, including STFT, WT, Wigner–Ville distribution (WVD) [19], Hilbert–Huang transform (HHT) [20], and others, as previously discussed. The STFT is a prominent time-frequency analysis tool used to examine nonstationary signals' frequency characteristics. By applying the Fourier Transform at different points in time, STFT furnishes information about the signal's time and frequency components, facilitating a comprehensive analysis of its frequency evolution over time. Compared to WT, WVD, and HHT, the time-frequency spectrum generated by STFT is more straightforward to interpret, providing a direct representation of frequency variation over time. This characteristic is advantageous for segmenting the time-frequency spectrum based on the meshing frequency.

Furthermore, STFT exhibits superior real-time performance, making it adept at dynamically analyzing the time-frequency characteristics of the signal. This feature aligns well with the later-introduced multibranch structure. Finally, the computational simplicity of STFT, typically involving the fast Fourier transform (FFT) algorithm for each time window, makes it relatively easy to implement and suitable for various programming environments and hardware platforms.

The STFT is computed by sliding a window function over the signal and applying a Fourier transform to each window. A vibration signal is sampled at a frequency $f_s$, sampling time $t$, and STFT is defined as

$$X_w(w,t) = \int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-jwt}dt \tag{1}$$

where $X_w(w,t)$ is the STFT value at time $t$ and frequency $\omega$, $x(t)$ is the original continuous time signal, $\omega(t-\tau)$ is the value of $\tau$ at the offset of the window function at time $t$, $e^{-jwt}$ is the complex exponential term used to convert the signal from the time domain to the frequency domain.

The STFT serves as a powerful time-frequency analysis tool capable of unveiling the frequency characteristics of a gearbox vibration signal. Applying STFT allows one to discern the inherent frequencies within the signal and observe how these frequency characteristics evolve over a specified time range.

Numerous studies have indicated that the vibration energy of gears is predominantly concentrated at the meshing frequency. Additionally, more relevant information about gear faults can be found in the sidebands of the meshing frequency. Analyzing the experimental dataset SWJTU Gear PHM Dataset introduced in Section III verifies that the meshing frequency and its sidebands contain more information related to gear faults. The vibration signal spectrum at each time step is obtained using the FFT. The meshing frequency and sidebands are subsequently extracted. To gauge the evolution of the spectrum bands, the Euclidean distance $\boldsymbol{D}(\boldsymbol{Y})$ between the spectrum at the current time step and the initial spectrum of the gearbox

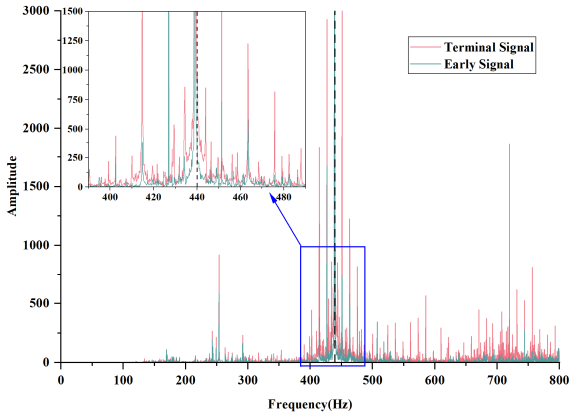



Fig. 3. Comparison of early and terminal spectrum.

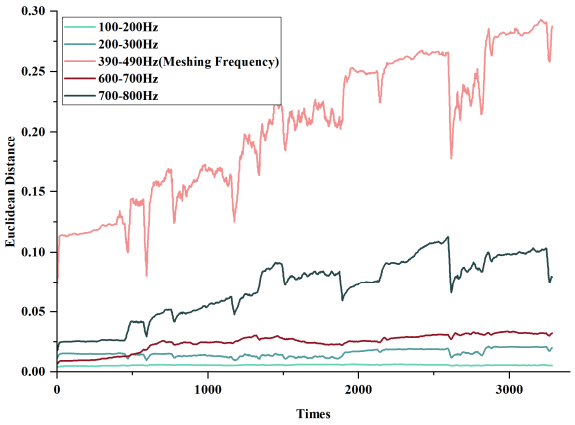



Fig. 4. Variation of Euclidean distance in different spectral intervals.

TABLE I
INDICATORS OF EUCLIDEAN DISTANCE VARIATION CURVES FOR DIFFERENT SPECTRAL INTERVALS

| | 100-200Hz | 200-300Hz | 440Hz | 600-700Hz | 700-800Hz |
|---|---|---|---|---|---|
| Spearman | 0.273 | 0.566 | **0.951** | 0.892 | 0.91 |
| Linearity (R-squared) | 0.184 | 0.421 | **0.903** | 0.75 | 0.844 |
| Peak to peak | 0.002 | 0.012 | **0.219** | 0.027 | 0.095 |

is calculated based on the following equation:

$$\boldsymbol{D}(\boldsymbol{Y}) = \sqrt{\sum_{i=2}^{n} (\boldsymbol{Y_i} - \boldsymbol{Y_1})^2} \tag{2}$$

where $\boldsymbol{Y_i}$ denotes the amplitude sequence of the frequency component at the $i$th time step, $n$ indicates the total number of time steps.

$\boldsymbol{D}(\boldsymbol{Y})$ is subjected to a moving average every 20 data points to mitigate the impact of outliers, facilitating trend analysis. The result is illustrated in Fig. 3, where 440 Hz represents the meshing frequency of the gearbox.

Spearman's coefficient [21], linearity ($R$-squared value), and peak-to-peak values were calculated to assess the monotonicity, linearity, and maximum variance of the Euclidean distances obtained for different spectra, respectively. The results are shown in Table I.

It can be observed in Fig. 4. that the Euclidean distances of the spectrum, including the meshing frequency and its sidebands, exhibit a more robust correlation and linearity with the duration of wear, indicating a higher degree of wear.

The corresponding time-frequency spectra of the multistage gearbox are extracted using the following method for the meshing spectrum of the multistage gearbox. After STFT, the time-domain width $l_t$ and frequency-domain width $l_f$ can be obtained using the following formulas:

$$l_t = \frac{f_s \cdot t}{(w - \alpha \cdot w)} + 1 \tag{3}$$

$$l_f = \frac{w}{2} \tag{4}$$

where $w$ is the STFT window size, and $\alpha$ is the window overlap ratio.

The gear meshing frequency $f$ is calculated as follows:

$$f = N_{\text{driving}} \cdot \frac{r}{60} \tag{5}$$

where $N_{\text{driving}}$ is the number of teeth of the driving gear, $r$ is the rotation speed of the driving gear (rpm).

If the set of $m$ meshing frequencies of a multistage transmission gearbox is $\boldsymbol{f} = \{f_0, f_1, \ldots, f_m\}$, then for each meshing frequency, the lower bound of the spectral region $T_m$ is intercepted according to (6), the width of the coverage frequency is $H_m$ according to (7), and the width of the time domain $l_t$ is the width of the time domain generated by the STFT

$$T_m = l_f - \left\lfloor \frac{f_m - \left\lfloor \frac{f_m}{10^{\lfloor \log f_m \rfloor}} \right\rfloor \times 10^{(\lfloor \log f_m \rfloor - 1)}}{\boldsymbol{f}_{\max}} \times l_f \right\rfloor \tag{6}$$

$$H_m = \left\lfloor 2 \times \left\lfloor \frac{f_m}{10^{\lfloor \log f_m \rfloor}} \right\rfloor \times 10^{(\lfloor \log f_m \rfloor - 1)} \times \frac{l_f}{\boldsymbol{f}_{\max}} \right\rfloor. \tag{7}$$

## B. Attention-Based Fusion of Multibranch Structure

A novel multibranch structure based on an attention mechanism is proposed to address the sidebands of multiple meshing frequencies in a multistage gearbox. This unique architecture enables concurrent processing of time-frequency spectrum in their respective regions. Leveraging the influential attention mechanism, which has proven effective in recent years, the model achieves feature fusion, culminating in a more robust multibranch structure.

Numerous deep neural network models have been introduced in fault diagnosis, showcasing powerful feature extraction capabilities. However, challenges such as overfitting [22], [23] and high computational resource requirements [24], [25] persist. Overfitting becomes pronounced with increased model depth, resulting in optimal performance on training data but suboptimal performance on test data. Furthermore, the substantial computational resources required for training and inference with deep neural networks pose constraints, especially in resource-limited industrial environments.

As shown in Fig. 5, our study introduces a novel multibranch parallel structure architecture to enhance meshing

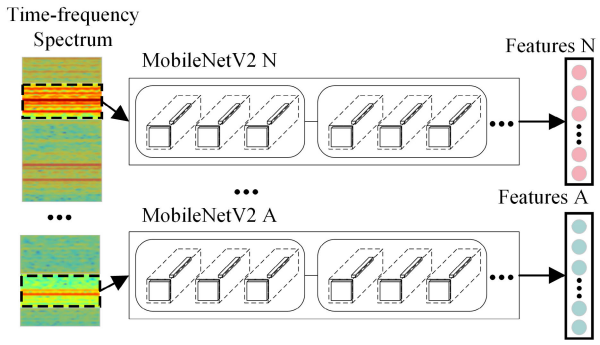



Fig. 5. Schematic of the multibranch parallel structure.

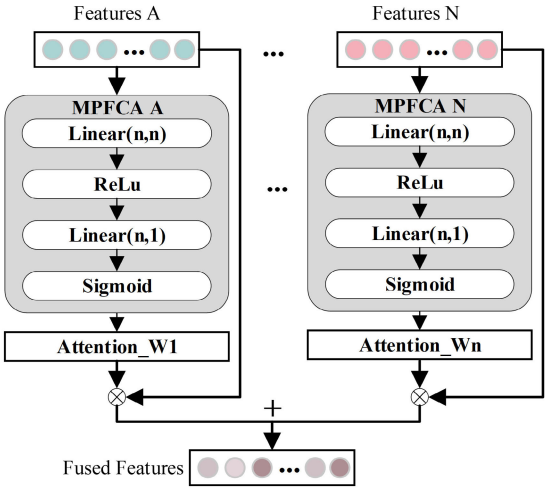



Fig. 6. Schematic of MPFCA structure.

frequency analysis. This architecture uniquely assigns dedicated layers within each branch to different meshing frequencies. This design enables precise tuning of each branch for extracting and learning meshing frequency specific features, thereby improving the extraction of critical troubleshooting data and minimizing overfitting risks. Each branch incorporates the MobileNetV2 [26], [27], [28] architecture, known for its efficiency and compactness. This choice reduces computational demands and model size without compromising accuracy. The architecture's lightweight nature makes it suitable for multibranch configurations and effective deployment on resource-constrained devices in industrial environments.

MobileNetV2, a streamlined neural network architecture, employs two primary methods to reduce model complexity and augment efficiency. The first method involves the utilization of depth-separable convolution. This technique segments the conventional convolution process into two streamlined operations: depthwise convolution and pointwise (1 × 1) convolution. Depthwise convolution is independently executed on each input channel, focusing on spatial features. Subsequently, pointwise convolution amalgamates these channel-specific features. This strategic decomposition markedly reduces the model's parameter count and computational demands. The second method is the implementation of an inverted residual structure. This structure optimizes the network's information flow efficiency by employing a unique expand-and-compress approach to feature channels. Initially, it expands the features, enhancing their representational capacity, and then compresses them, thereby improving the efficiency of the computational units within the network.

This article introduces a novel attention mechanism, the multipath fully connected attention (MPFCA) module, tailored explicitly for handling features extracted from a multibranch structure. The MPFCA module calculates and weighs each feature subset's attention scores individually. Initially, a linear transformation layer standardizes the dimensionality of the input features while introducing nonlinearity through the rectified linear unit (ReLU) activation function. This is followed by another linear layer that maps these features into distinct values indicative of attention scores. These scores are normalized within a 0–1 range using the Sigmoid activation function to derive the corresponding weights. The final step involves applying these attention scores to their feature sets via elementwise multiplication, culminating in an aggregated, integrated feature set.

As shown in Fig. 6, MPFCA incorporates two parallel attention networks, each tailored to handle separate input features, $x_1$ and $x_2$. Each network performs the following operations in a sequence.

1) *Linear and ReLU Activation:* The input feature $x$ is first transformed through a linear layer and then passed through a ReLU activation function

$$\boldsymbol{h} = \mathrm{ReLU}(\boldsymbol{W_1 x} + \boldsymbol{b_1}) \tag{8}$$

where $\boldsymbol{W_1} \in \boldsymbol{R}^{d\times d}$ and $b_1 \in \boldsymbol{R}^{d}$ are the weight matrix and bias vector, respectively.

2) *Attention Score Calculation:* The activated feature $h$ is then projected down to a single scalar, followed by a sigmoid activation to form the attention weight:

$$\alpha = \sigma(\boldsymbol{W_2 h} + b_2). \tag{9}$$

Here, $\boldsymbol{W_2} \in \boldsymbol{R}^{1\times d}$ is a weight matrix and $b_2$ is a bias scalar, with $\sigma$ denoting the sigmoid function.

Each network generates an attention weight $\alpha_1$ for $\boldsymbol{x_1}$ and $\alpha_2$ for $\boldsymbol{x_2}$, which scales the corresponding input feature. The weighted features are then summed to compute the combined feature output

$$\text{Fused\_Feature} = \alpha_1 \cdot \boldsymbol{x_1} + \alpha_2 \cdot \boldsymbol{x_2}. \tag{10}$$

This specially designed method effectively highlights features with a stronger correlation to fault occurrences and mitigates the influence of less correlated features. Such a bespoke approach ensures precise and efficient feature representation in fault detection and analysis scenarios, demonstrating the unique contribution of this research to the field.

Since the research content of this article focuses on multistage spur gearboxes, the derivation of formulas is only applicable to multistage gearboxes. Therefore, this method cannot be directly applied to other types of gearboxes, including planetary gearboxes, harmonic gear reducers, etc. These gearboxes have entirely different structural characteristics, vibration characteristics, and load distributions.

## III. Experimental Procedure

### A. Model Parameters

In this study, the window size is set to 3200, with an overlap rate of 75%, and the Kaiser window function is employed for STFT. The Kaiser window allows control over the main lobe width and sidelobe level through a parameter, $\beta$, thereby influencing the window's frequency resolution and leakage characteristics. With a $\beta$ value of 15, a larger $\beta$ yields a wider main lobe, which enhances frequency resolution at the expense of temporal resolution.

The model is constructed using the PyTorch framework, and critical hyperparameters are adjusted, including a learning rate of 0.001, a batch size of 32, and an epoch of 100. An early stopping method based on the validation set's loss function is introduced to expedite model training, achieve a more stable model selection, and prevent overfitting. The core concept involves monitoring the model's performance on the validation set and terminating training when the validation set loss shows no improvement for ten consecutive cycles. Given the potential for neural networks to converge to local optima during training, each model is trained 25× to ensure the reliability of experimental results, and method performance is characterized by its statistical outcomes.

We utilized PyTorch 2.0.0 with CUDA 11.8.0 for model construction, training, and evaluation. Pandas and Scikit-learn were employed for data processing and preprocessing. Model training was executed on an NVIDIA RTX 4060 Ti.

### B. Introduction to Comparative Modeling

This study employs comparison models, including SVM, RF, AdaBoost, Resnet-18 [29], [30], EfficientNet-B0 [31], [32], and MobileNet V2. SVM, RF, and AdaBoost are supervised machine learning classification methods that lack robust feature extraction capabilities for direct raw data classification. Therefore, various features, such as time, frequency, and wavelet features, are extracted and used to train the SVM, RF, and AdaBoost models for fault diagnosis. ResNet-18, a deep CNN, is distinguished by its residual learning framework. This framework facilitates training deeper networks by addressing issues related to vanishing gradients. In contrast, EfficientNet-B0, the foundational model of the EfficientNet family, adopts a composite scaling method. This method meticulously balances the network's depth, width, and resolution, aiming to enhance both efficiency and accuracy.

### C. Dataset

To facilitate the evaluation and validation of the proposed method, we established an experimental platform illustrated in Fig. 7 and curated a gearbox fault diagnosis dataset, referred to as the SWJTU Gear PHM Dataset. The gearbox featured in the experimental platform is a two-stage parallel-axis gearbox, where a load motor applies torque at the output end. An accelerometer is positioned at the output cover to capture the $Z$-axis vibration signal of the gearbox, with a sampling frequency of 12 800 Hz and an input frequency of 40 Hz. Essential gear parameters are detailed in Table II. According to (5), it is evident that the meshing frequencies of the gears are 440 and 1160 Hz.

TABLE II
Gearbox Parameters

| Parameters | Primary Gear Units | Secondary Gear Units |
|---|---|---|
| Number of pinion teeth | 29 | 36 |
| Number of large gear teeth | 95 | 90 |
| Pinion tooth width/mm | 15 | 15 |
| Large gear tooth width/mm | 15 | 15 |
| Modulus/mm | 1.5 | 1.5 |
| Pressure angle/° | 20 | 20 |

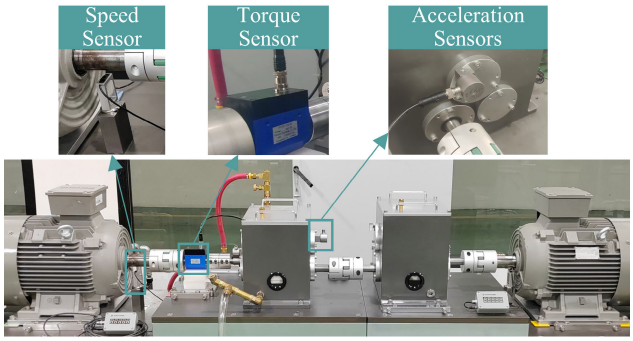



Fig. 7. Experimental platform.

TABLE III
Experimental Dataset

| No. | Gear set No. | torque load | state of health | sample size |
|---|---|---|---|---|
| 1 | set 1 | 60% | Broken pinion gear at load end | 10 |
| 2 | | | Normal | 2870 |
| 3 | set 2 | 50% | Load end gear pitting | 176 |
| 4 | | | Normal | 2930 |
| 5 | set 3 | 70% | Broken pinion gear at load end | 105 |
| 6 | | | Normal | 720 |

This dataset comprises actual gearbox operational data recorded under various load conditions, encompassing both normal and faulty instances. For further details, refer to Table III.

For the dataset collected from experiments 1–4, 80% of the data is allocated to the training set, while 20% is assigned to the validation set. The data obtained from experiments 5 and 6 are exclusively designated as the test set. It's important to note that the input and output speeds remain consistent across all experiments. However, a significant distinction arises in the test set, where a torque loading of 70% is incorporated—a condition not present in either the training or validation set. This intentional inclusion allows for comprehensively evaluating the proposed method's performance under varied operating conditions.

## IV. Result and Discussion

Two ablation experiments were designed to validate the roles of adaptive time-frequency spectrum segmentation algorithm and attention-based fusion of multibranch structure in the proposed method. The performance of the models was also compared with mainstream fault diagnosis methods.

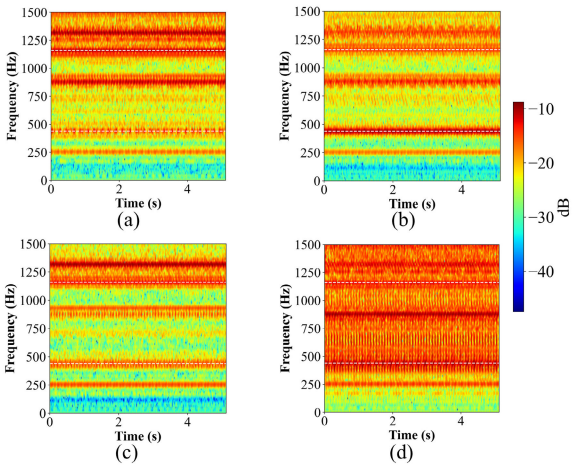



Fig. 8. Time-frequency spectra obtained by SFTF transformation, the white dotted line marks the position of the engagement frequency 440 and 1160 Hz. (a) Gearbox normal at 50% load. (b) Load-end pinion pitting at 50% load. (c) Gearbox normal at 60% load. (d) Load-end pinion tooth breakage at 60% load.

### A. Adaptive Time-Frequency Spectrum Segmentation Algorithm Ablation Experiments

To substantiate the efficacy of the proposed time-spectrum segmentation algorithm in this study, STFT is first used to derive the time-frequency spectrum of gears under various loads and health states. Fig. 8. illustrates that gear states exhibit notable divergences in their time-spectrum characteristics. This observation underscores the potential of CNNs to utilize these unique time-frequency features for precise fault diagnosis proficiently. Complementing this, the research by Liao et al. [33] provides empirical evidence supporting the high diagnostic accuracy of CNNs when applied to time-frequency spectrum analysis for fault detection.

In this study, the experimental design was established as follows. TFSSN model, as previously referenced, was employed as the experimental group. For the control groups, the time-frequency spectrum at specific meshing frequencies was extracted: Frequencies of 440 and 1160 Hz formed control groups 1 and 2, respectively. In alignment with (6) and (7), the time-frequency spectrum at a frequency of 200 Hz was extracted, constituting control group 3. Given that this experiment's primary frequency component of gearbox vibration lies within 1500 Hz, we also obtained the time spectrum for the 0–1500 Hz range through STFT, designated as control group 4. A singular branch network from a multibranch structure was implemented within each control group to acquire experimental data. A fully connected layer was integrated at the model's end, facilitating fault identification to ensure uniformity in network layers and parameters between the comparison and experimental groups.

As shown in Fig. 9, by comparing the accuracy changes of the validation set (left) and the loss function changes of the training set (right) during the model training process, we observe that the convergence speed of the TFSSN model is slower than that of training with a single meshing frequency, but faster than training with only nonmeshing frequencies.

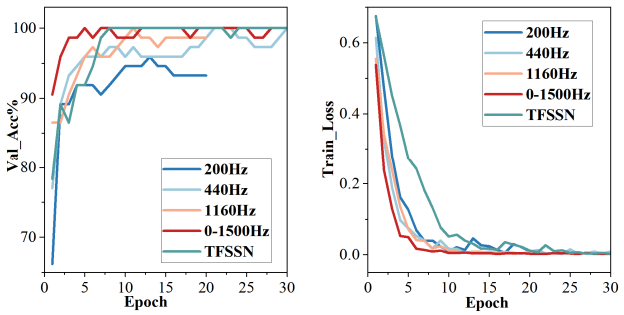
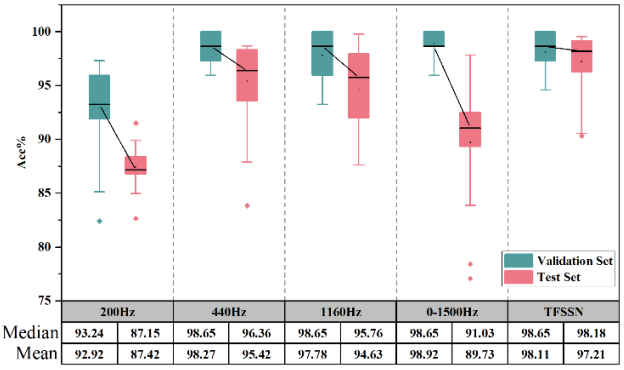



Fig. 9. Comparison of the change in accuracy of the validation set (left) and loss of the train set (right) during the model training process.




| | 200Hz | | 440Hz | | 1160Hz | | 0-1500Hz | | TFSSN | |
|---|---|---|---|---|---|---|---|---|---|---|
| Median | 93.24 | 87.15 | 98.65 | 96.36 | 98.65 | 95.76 | 98.65 | 91.03 | 98.65 | 98.18 |
| Mean | 92.92 | 87.42 | 98.27 | 95.42 | 97.78 | 94.63 | 98.92 | 89.73 | 98.11 | 97.21 |

Fig. 10. Validation set and test set results of the model for Experiments A.

It can be seen from Fig. 10. that fault diagnosis accuracy across different frequencies is consistently high in the validation set. Conversely, in the test set, diagnostics performed with the meshing frequency show marked improvement over those using nonmeshing frequencies. This disparity indicates that features derived from nonmeshing frequencies frequently lack efficacy in fault characterization, demonstrating limited robustness and cross-condition transferability. In contrast, features extracted from the time-frequency spectrum proximal to the meshing frequency encapsulate a more comprehensive understanding of gear faults.

The comparison between the number of fault samples and normal samples in Table III highlights the challenge posed by sparse fault samples in the fault diagnosis process in this experiment. The scarcity of fault data has led to a typical model overfitting issue, with the fault diagnosis accuracy on the test set consistently falling below that of the validation set. Notably, when the model operates within the 0–1500 Hz frequency range, the average accuracy gap between the validation and test sets reaches 9.19%. This indicates that the application of indiscriminate deep learning methods for fault diagnosis on time-frequency spectrograms of vibration signals exacerbates the overfitting problem. However, with the proposed TFSSN model, the average accuracy gap between the validation and test sets is minimized to 0.9%. Therefore, the TFSSN method proposed in this article can effectively mitigate the overfitting problem in neural network models under the condition of sparse fault samples.

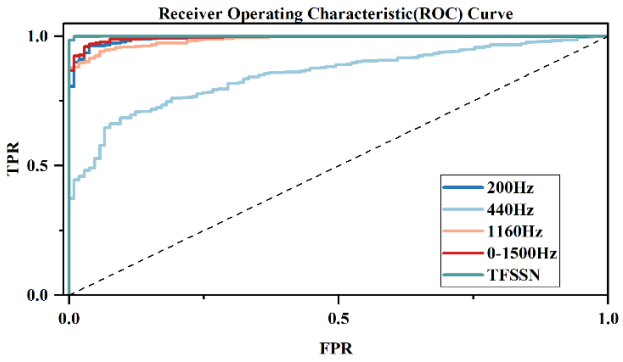



Fig. 11. Receiver operating characteristic curve for Experiments A.

Although occasionally achieving higher accuracy on the test set using the 0–1500 Hz time-frequency spectrum, the mean and median values from multiple experiments are lower than those at 440 and 1160 Hz, with a more considerable variance. This results in significant bias in model optimization, making it challenging to obtain the optimal solution stably.

The proposed time-frequency spectrum segmentation and multibranch structure in this article significantly improve mean accuracy on the test set, with the most minor variance. This indicates that the model is more accessible to train and more likely to obtain a relatively optimal solution.

Receiver operating characteristic curves are plotted in Fig. 11 to illustrate the performance for time-frequency spectrum segmentation and multibranch structure.

The time-frequency spectrum segmentation and multibranch structure demonstrated excellent performance on the test set, with an AUC value of 0.99, showcasing the model's high accuracy in classifying between normal and faulty states. The ROC curve illustrates the model's ability to balance high sensitivity and specificity.

### B. Attention-Based Fusion of Multibranch Structure Ablation Experiment

To rigorously assess the efficacy of the MPFCA in our proposed model, we conducted a series of ablation experiments. For comparative analysis, we incorporated three alternative feature fusion techniques: basic feature splicing fusion (referred to as Model A), feature averaging fusion (Model B), and feature weighting fusion (Model C). The distinct methodologies of these fusion techniques are illustrated in the Fig. 12.

Model A implements feature fusion by concatenating outputs from the branch networks. In contrast, Model B computes the fused features by summing and averaging the corresponding features across these branches. Model C, on the other hand, employs a weighted summation approach for feature fusion, utilizing two self-learning weights, denoted as $W_1$ and $W_2$, applied to the outputs of the branch networks.

The empirical outcomes demonstrate that the MPFCA component incrementally enhances the model's diagnostic accuracy. As shown in Fig. 13, MPFCA achieves the highest fault diagnosis accuracy compared to the above three feature fusion methods (Models A–C).

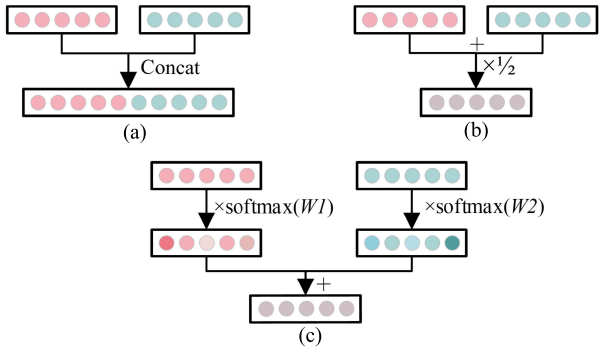



Fig. 12. Schematic representation of different feature fusion methods. (a) Model A. (b) Model B. (c) Model C.

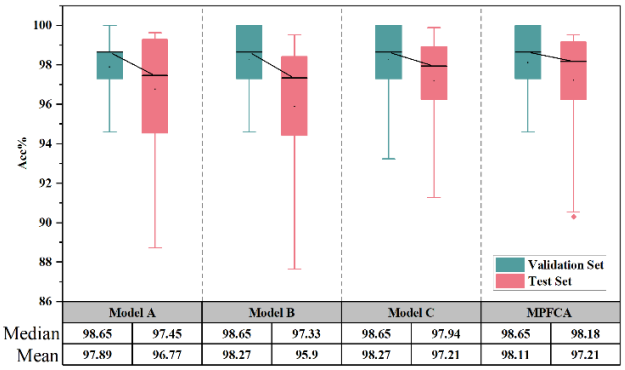



Fig. 13. Performance comparison of different feature fusion methods.

TABLE IV
PERFORMANCE COMPARISON OF TFSSN WITH MAINSTREAM MODELS FOR GEARBOX FAULT DIAGNOSIS (%)

| Method | Acc | F1 | Precision | Recall | AUC |
|---|---|---|---|---|---|
| SVM-Linear | 89.15 | 77.67 | 100 | 63.49 | 99.53 |
| RF | 90.22 | 84.47 | 80 | 89.47 | 97.73 |
| AdaBoost | 91.10 | 82.40 | 100 | 70.06 | 83.41 |
| ResNet-18 | 91.27 | 94.74 | 100 | 90 | 99.98 |
| EfficientNet-B0 | 89.57 | 94.3 | 90.23 | 98.75 | 88.13 |
| MobileNet V2 | 92.24 | 95.74 | 91.94 | 99.86 | 98.72 |
| TFSSN | **98.18** | **98.97** | 97.96 | **100** | **99.98** |

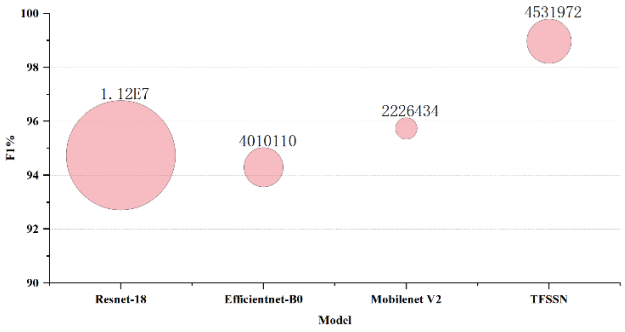



Fig. 14. Comparison of the number of parameters and $F1$ score of neural network models.

### C. Performance Comparison of Different Models

In comparison to mainstream fault diagnosis methods, the results are summarized in the Table IV.

The proposed TFSSN achieves optimal results across several metrics compared to mainstream fault diagnosis models.

Fig. 14. compares the number of parameters and $F1$ scores of the four neural network models mentioned above, demonstrating that TFSSN achieves significantly better performance using smaller model parameters.

## V. Conclusion

This article introduces a new gearbox fault diagnosis method called TFSSN. Utilizing a gearbox fault diagnosis dataset for performance testing, the proposed model achieved an accuracy of 98.18% and an $F1$ score of 98.57%. Compared to current mainstream gearbox fault diagnosis models, the TFSSN model demonstrates higher accuracy and stronger robustness primarily because of its unique design and mechanisms. These enable the TFSSN to more effectively mine and utilize the time-frequency information near multiple meshing frequencies of multistage gearboxes, facilitating fault diagnosis under varying loads. In the TFSSN model, the time-frequency spectrum segmentation algorithm partially performs signal filtering, which makes it easier for the convolutional window to capture and learn information highly correlated with gearbox faults. This feature helps mitigate overfitting issues in convolutional networks when dealing with limited data, enhancing the model's generalizability and diagnostic accuracy.

In future research, we are committed to refining and expanding the algorithm for application to other types of gearboxes, aiming to develop a more comprehensive and widely applicable gearbox fault diagnosis solution.

## References


[1] Z. Feng, M. Liang, and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mech. Syst. Signal Process.*, vol. 38, no. 1, pp. 165–205, Jul. 2013, doi: 10.1016/j.ymssp.2013.01.017.

[2] W. Xie, T. Han, Z. Pei, and M. Xie, "A unified out-of-distribution detection framework for trustworthy prognostics and health management in renewable energy systems," *Eng. Appl. Artif. Intell.*, vol. 125, Oct. 2023, Art. no. 106707, doi: 10.1016/j.engappai.2023.106707.

[3] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018, doi: 10.1016/j.ymssp.2018.02.016.

[4] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Process. Mag.*, vol. 9, no. 2, pp. 21–67, Apr. 1992, doi: 10.1109/79.127284.

[5] A. A. Jaber and R. Bicker, "Fault diagnosis of industrial robot gears based on discrete wavelet transform and artificial neural network," *Insight-Non-Destructive Test. Condition Monitor.*, vol. 58, no. 4, pp. 179–186, Apr. 2016, doi: 10.1784/insi.2016.58.4.179.

[6] L. Stanković, "S-class of time–frequency distributions," *IEE Proc.-Vis., Image, Signal Process.*, vol. 144, no. 2, p. 57, Apr. 1997, doi: 10.1049/ip-vis:19970917.

[7] J. R. Fonoliosa and C. T. Nikias, "Wigner higher order moment spectra: Definition, properties, computation and application to transient signal analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 1, p. 245, Jan. 1993, doi: 10.1109/TSP.1993.193143.

[8] I. Yesilyurt, "The application of the conditional moments analysis to gearbox fault detection—A comparative study using the spectrogram and scalogram," *NDT & E Int.*, vol. 37, no. 4, pp. 309–320, Jun. 2004, doi: 10.1016/j.ndteint.2003.10.005.

[9] X. Fan and M. J. Zuo, "Gearbox fault detection using Hilbert and wavelet packet transform," *Mech. Syst. Signal Process.*, vol. 20, no. 4, pp. 966–982, May 2006, doi: 10.1016/j.ymssp.2005.08.032.

[10] A. Dubaish and A. Jaber, "State-of-the-art review into signal processing and artificial intelligence-based approaches applied in gearbox defect diagnosis," *Eng. Technol. J.*, vol. 10, no. 9, pp. 1–16, Nov. 2023, doi: 10.30684/etj.2023.142462.1535.

[11] M. Cui, Y. Wang, X. Lin, and M. Zhong, "Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4927–4937, Feb. 2021, doi: 10.1109/JSEN.2020.3030910.

[12] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5581–5588, Sep. 2017, doi: 10.1109/JSEN.2017.2726011.

[13] J. L. L. Soares et al., "Fault diagnosis of belt conveyor idlers based on gradient boosting decision tree," *Int. J. Adv. Manuf. Technol.*, vol. 132, nos. 7–8, pp. 3479–3488, Apr. 2024, doi: 10.1007/s00170-024-13549-0.

[14] Y. Zhang, Y. Jia, W. Wu, Z. Cheng, X. Su, and A. Lin, "A diagnosis method for the compound fault of gearboxes based on multi-feature and BP-AdaBoost," *Symmetry*, vol. 12, no. 3, p. 461, Mar. 2020, doi: 10.3390/sym12030461.

[15] T. Han, J. Tian, C. Y. Chung, and Y.-M. Wei, "Challenges and opportunities for battery health estimation: Bridging laboratory research and real-world applications," *J. Energy Chem.*, vol. 89, pp. 434–436, Feb. 2024, doi: 10.1016/j.jechem.2023.10.032.

[16] A. A. Dubaish and A. A. Jaber, "Comparative analysis of SVM and ANN for machine condition monitoring and fault diagnosis in gearboxes," *Math. Model. Eng. Problems*, vol. 11, no. 4, pp. 976–986, Apr. 2024, doi: 10.18280/mmep.110414.

[17] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017, doi: 10.1016/j.measurement.2017.07.017.

[18] M. Azamfar, J. Singh, I. Bravo-Imaz, and J. Lee, "Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106861, doi: 10.1016/j.ymssp.2020.106861.

[19] *Phys. Rev. 40, 749 (1932)—On the Quantum Correction for Thermodynamic Equilibrium*. Accessed: Oct. 24, 2023. [Online]. Available: https://journals.aps.org/pr/abstract/10.1103/PhysRev.40.749

[20] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. Lond. Ser. Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998, doi: 10.1098/rspa.1998.0193.

[21] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 100, pp. 441–471, Apr. 1987, doi: 10.2307/1422689.

[22] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, Feb. 2019, Art. no. 022022, doi: 10.1088/1742-6596/1168/2/022022.

[23] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, 2021, doi: 10.1145/3446776.

[24] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," 2017, *arXiv:1710.09282*.

[25] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, Dec. 2017, doi: 10.1109/JPROC.2017.2761740.

[26] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[28] T. Xue, H. Wang, and D. Wu, "MobileNetV2 combined with fast spectral kurtosis analysis for bearing fault diagnosis," *Electronics*, vol. 11, no. 19, p. 3176, Oct. 2022, doi: 10.3390/electronics11193176.

[29] K. Zhang, B. Tang, L. Deng, and X. Liu, "A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox," *Measurement*, vol. 179, Jul. 2021, Art. no. 109491, doi: 10.1016/j.measurement.2021.109491.

[30] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Comput. Appl.*, vol. 32, pp. 6111–6124, May 2020, doi: 10.1007/s00521-019-04097-w.

[31] H. Qu, J. Yang, M. Shen, H. Chen, and D. Zhou, "Fault diagnosis of rolling bearing under time-varying speed conditions based on EfficientNetv2," *Meas. Sci. Technol.*, vol. 33, no. 6, Jun. 2022, Art. no. 065023, doi: 10.1088/1361-6501/ac57ec.

[32] B. Hu, J. Tang, J. Wu, and J. Qing, "An attention EfficientNet-based strategy for bearing fault diagnosis under strong noise," *Sensors*, vol. 22, no. 17, p. 6570, Aug. 2022, doi: 10.3390/s22176570.

[33] Y. Liao, X. Zeng, and W. Li, "Wavelet transform based convolutional neural network for gearbox fault classification," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Harbin)*, Jul. 2017, pp. 1–6, doi: 10.1109/PHM.2017.8079274.

**Hongliang Song** received the M.S. degree in mechanical and electronic engineering from Southwest Jiaotong University, Chengdu, Sichuan, China, in 2020, where he is currently pursuing the Ph.D. degree.

Throughout his academic journey, he has been deeply involved in research related to fault diagnosis and intelligent maintenance technology of electromechanical equipment.

**Yi Sun** received the M.S. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, Sichuan, China, in 2020, where he is currently pursuing the Ph.D. degree in mechatronic engineering.

His research interests include machinery condition monitoring and structural load identification.

**Hongli Gao** received the Ph.D. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, Sichuan, China, in 2005.

He is currently a Professor at Southwest Jiaotong University. His research interests include designing and the reliability analysis of complex electromechanical equipment.

**Liang Guo** (Member, IEEE) received the Ph.D. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, Sichuan, China, in 2016.

He is currently an Associate Professor at Southwest Jiaotong University. His research interests include machinery intelligent fault diagnostics, health indicator construction, and remaining useful life prediction.

**Yucheng Chen** received the Bachelor of Engineering degree from Southwest Jiaotong University, Chengdu, Sichuan, China, in 2021, where he is currently pursuing the master's degree.

His research interests include condition monitoring and fault diagnosis of intelligent machinery.