machines

MDPI

Article

Multisensor Fault Diagnosis of Rolling Bearing with Noisy Unbalanced Data via Intuitionistic Fuzzy Weighted Least Squares Twin Support Higher-Order Tensor Machine

Shengli Dong [1,2], Yifang Zhang [3] and Shengzheng Wang [1,*]

[1] Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China; dong.shengli@coscoshipping.com
[2] Shanghai Ship and Shipping Research Institute Co., Ltd., Shanghai 200135, China
[3] School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China; zhangyifang@mail.dlut.edu.cn
* Correspondence: szwang@shmtu.edu.cn

Abstract: Aiming at the limitations of existing multisensor fault diagnosis methods for rolling bearings in real industrial scenarios, this paper proposes an innovative intuitionistic fuzzy weighted least squares twin support higher-order tensor machine (IFW-LSTSHTM) model, which realizes a breakthrough in the noise robustness, adaptability to the working conditions, and the class imbalance processing capability. First, the multimodal feature tensor is constructed: the fourier synchro-squeezed transform is used to convert the multisensor time-domain signals into time–frequency images, and then the tensor is reconstructed to retain the three-dimensional structural information of the sensor coupling relationship and time–frequency features. The nonlinear feature mapping strategy combined with Tucker decomposition effectively maintains the high-order correlation of the feature tensor. Second, the adaptive sample-weighting mechanism is developed: an intuitionistic fuzzy membership score assignment scheme with global–local information fusion is proposed. At the global level, the class contribution is assessed based on the relative position of the samples to the classification boundary; at the local level, the topological structural features of the sample distribution are captured by *K*-nearest neighbor analysis; this mechanism significantly improves the recognition of noisy samples and the handling of class-imbalanced data. Finally, a dual hyperplane classifier is constructed in tensor space: a structural risk regularization term is introduced to enhance the model generalization ability and a dynamic penalty factor is set to set adaptive weights for different categories. A linear equation system solving strategy is adopted: the nonparallel hyperplane optimization is converted into matrix operations to improve the computational efficiency. The extensive experimental results on the two rolling bearing datasets have verified that the proposed method outperforms existing solutions in diagnostic accuracy and stability.

Keywords: fault diagnosis; class imbalance; intuitionistic fuzzy; twin support higher-order tensor machine; Tucker decomposition

check for updates

Academic Editor: Ahmed Abu-Siada

Received: 20 April 2025
Revised: 17 May 2025
Accepted: 21 May 2025
Published: 22 May 2025

Citation: Dong, S.; Zhang, Y.; Wang, S. Multisensor Fault Diagnosis of Rolling Bearing with Noisy Unbalanced Data via Intuitionistic Fuzzy Weighted Least Squares Twin Support Higher-Order Tensor Machine. *Machines* **2025**, *13*, 445. https://doi.org/10.3390/machines13060445

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In rotating machinery systems, rolling bearings serve as critical components whose performance directly determines the operational stability and reliability of the entire system. Prolonged operation under varying loads exposes bearings to harsh working conditions including elevated temperatures, excessive humidity, and mechanical vibration, potentially leading to premature wear, fatigue fault, and catastrophic breakdowns [1–4]. Implementing

effective bearing fault diagnosis has become imperative for enhancing system reliability while mitigating downtime and associated economic losses [5–7]. The fault diagnosis framework typically comprises three sequential stages: (1) sensor signal acquisition through condition monitoring devices, (2) feature extraction from raw sensor data, and (3) fault classification. Notably, the nonstationary characteristics of signals induced by variable operational conditions (e.g., load fluctuations, rotational speed variations, and thermal transients) present significant challenges for feature extraction [8]. Time–frequency analysis has emerged as a predominant methodology in this context, enabling the simultaneous characterization of temporal and spectral information to reveal transient patterns in bearing vibration signatures. Recent advancements in intelligent diagnosis have revolutionized fault classification through automated pattern recognition. Representative approaches include neural network architectures: echo state networks [9], convolutional neural networks [10], graph convolutional networks [11,12], and deep belief networks [13]; transfer learning [14,15]; and support vector machines (SVMs) [16–20]. These data-driven methodologies demonstrate superior capability in deciphering complex nonlinear relationships between signal features and bearing health states.

The increasing integration of rotating machinery systems has intensified intercomponent coupling, rendering single sensor monitoring signals inadequate for accurately characterizing bearing operational status. Such unilateral data acquisition not only introduces diagnostic instability but also compromises result reliability due to information incompleteness. This inherent limitation underscores the necessity for multidimensional heterogeneous monitoring data integration, encompassing vibration signatures, acoustic emissions, rotational speed, and torque measurements, to establish a comprehensive health assessment framework. Tensor representation emerges as a mathematically rigorous solution for high-dimensional data encapsulation, offering two principal advantages over conventional feature vectors: (1) structural preservation maintains intrinsic data topology through multidimensional array organization; (2) relational modeling explicitly encodes intermodal interactions via tensor decomposition techniques. As evidenced by cross-domain applications, tensor-based approaches have demonstrated exceptional efficacy in human motion recognition [21], cardiac electrophysiological analysis [22], and mechanical fault diagnosis [23]. These empirical validations confirm the tensor paradigm's superiority in processing multisensor signals while retaining critical system structural information.

Conventional fault diagnosis methods predominantly adopt vector input patterns, necessitating vectorization of multisensor feature tensors to accommodate model requirements. This vectorization process introduces dual limitations: (1) the generated high-dimensional vectors escalate computational complexity and overfitting risks; (2) the structural integrity of original tensor data and intrinsic coupling relationships among monitoring signals are compromised, ultimately degrading diagnostic accuracy. To address these tensor classification challenges, recent research has focused on tensor space-based methodologies. In tensor-based model development, Hao et al. [24] pioneered the support higher-order tensor machine (SHTM) through tensor CP decomposition, enabling direct processing of tensor patterns. He et al. [25] integrated continuous wavelet transform with dynamic penalty factors to establish the support tensor machine with dynamic penalty factors (DC-STM), effectively resolving class imbalance issues in rotating machinery diagnosis. Ma et al. [26] further extended this paradigm through support multimode tensor machines for industrial big data multiclassification. Nevertheless, these tensor models demonstrate sensitivity to noise contamination and outliers, frequently yielding suboptimal classification hyperplanes. To mitigate noise interference, fuzzy set theory has been incorporated for enhanced uncertainty handling [27]. Sun et al. [28] developed fuzzy STM (FSTM) via singular spectrum analysis and Tucker decomposition, while Yang et al. [29] proposed

FSTM with pinball loss (Pin-FSTM) using hierarchical multiscale permutation entropy with pinball loss for robust bearing fault detection. Building upon fuzzy sets, the intuitionistic fuzzy (IF) theory provides superior uncertainty quantification [30,31]. Zhang et al. [32] consequently devised unified pinball loss intuitionistic fuzzy STM (UPIFSTM), achieving reliable mechanical fault diagnosis under noisy multisource conditions. Computational efficiency optimization represents another research frontier. Yang et al. [33] innovated twin SHTM (TwSHTM) through dual nonparallel hyperplanes in tensor space. By decomposing a single large quadratic programming problem (QPP) into two smaller QPPs, this architecture significantly enhances computational efficiency for planetary gearbox diagnostics, demonstrating practical implementation advantages.

Although the above methods have achieved good results, there are still three problems that need to be solved urgently. (1) The feature tensors used in the above tensor classifiers are mostly reconstructed from time-domain signals, without considering the nonstationary characteristics of signals under variable operating conditions. Although the feature tensor reconstructed based on CWT can capture the transient and local features of signals, the time–frequency resolution depends on the selection of windows and basis functions. However, windows and basis functions are fixed and unchanged in analysis, resulting in poor processing performance for multicomponent time-varying signals. (2) Although the IF set contains distance information between input sample points and two class centers, it ignores the position information of input sample points in the feature space, which can easily mistake edge support tensors for noise. Although TwSHTM solves for two smaller QPPs, it still results in considerable computational costs and remains sensitive to noise. Moreover, the above methods cannot fully utilize the potential correlation or similarity information between any pair of data points with the same label, which may be important for classification performance. In addition, there are relatively few sample data on different fault states of rolling bearings in practice, which leads to the widespread occurrence of class imbalance phenomena. Due to the dominance of data points in one class by data points in another class, the resulting decision hyperplane may not be optimal. Although DC-STM can handle class-imbalanced samples, it does not consider the impact of noise and has lower computational efficiency. (3) In practical applications, tensor data may not be linearly separable in the input space, but most of the above methods do not involve the nonlinear learning problem of tensor data. Although dual structure-preserving kernels based on CP decomposition are used in DC-STM, which can fully capture the multidirectional structure of tensor data, it cannot reflect the relevant information inside the tensor data.

The above discussion motivates this paper to propose the intuitionistic fuzzy weighted least squares twin support higher-order tensor machine (IFW-LSTSHTM) to achieve the following breakthrough advantages for the noise interference, nonstationary multisensor data fusion, class imbalance, and nonlinear classification challenges in rolling bearing fault diagnosis: (1) Adoption of the fourier synchro-squeezed transform (FST) [34] to replace the traditional continuous wavelet transform for accurate resolution of nonstationary signals, and realize high-resolution time–frequency feature modeling. The three-dimensional time–frequency tensor (time $\times$ frequency $\times$ sensor channel) is constructed to preserve the coupling relationship of multimodal signals and provide a high-fidelity representation for the analysis of complex nonstationary signals. (2) Designing an intuitionistic fuzzy membership scoring assignment scheme based on the global–local information and assigning weights to each sample by combining the a priori information of the sample's neighborhood and the penalty factor, which improves the model under the condition of category imbalance against noise and outliers. In addition, a regularization term is introduced into the model to achieve the principle of minimizing structural risk. The traditional quadratic programming is transformed into a linear system of equation solutions

to improve the training efficiency. (3) The feature tensor is mapped to a high-dimensional space by Tucker decomposition to explicitly model the higher-order correlations between time × frequency × sensor modes and to capture the nonlinear interaction features under variable operating conditions. IFW-LSTSHTM, through high-resolution time–frequency tensor representation, global–local dynamic anti-noise mechanism, kernel nonlinear mapping with Tucker decomposition, and efficient least squares solving strategy, aims to solve the bottleneck problems of existing methods in noise robustness, nonstationary multisensor data fusion, class imbalance processing, and nonlinear classification.

The contributions of this paper are summarized as follows:

1. This paper proposes a new tensor space nonlinear classifier called the IFW-LSTSHTM model. The classifier first designs two nonparallel classification hyperplanes in tensor space and solves a system of linear equations to improve computational efficiency. This classifier introduces regularization terms to achieve the principle of minimizing structural risk, and assigns penalty factors to samples of different categories to mitigate the impact of class imbalance.
2. This paper proposes a global–local-based intuitionistic fuzzy membership score assignment scheme and a sample-weighting scheme based on prior information of intraclass local neighborhood structure. By assigning appropriate weights to each sample, the robustness of the model to noise and outliers is improved.
3. Using FST time–frequency analysis to convert multisensor signals into time–frequency images can capture the spectral characteristics of multiple sensor signals that change over time. Then, the time–frequency images are reconstructed into feature tensors, which contain richer data structure information, better capture the relationships between multidimensional data, ensure the authenticity and integrity of sample patterns, improve the stability and certainty of the model, and provide an effective signal processing and representation method under variable working conditions.
4. This paper derives a tensor kernel function with Tucker decomposition, which maps the decomposed core tensor and factor matrix of the feature tensor to a high-dimensional feature space, ensuring that the interaction relationships between components in different modes can be reflected during the nonlinear mapping process, thereby maximizing the preservation of the potential structural information of the feature tensor.

The remainder of this article is organized as follows. Section 2 states the related works. Section 3 describes the proposed methods. Section 4 describes the feature tensor reconstruction process. Section 5 reports the two case studies and obtained results. Section 6 gives the conclusion of this article.

2. Related Work

2.1. Tensor Theory

The multidimensional arrays in three dimensions and above is also known as an N-order tensor, which is a higher-order generalization of a vector $\boldsymbol{a}$ (first-order tensor) and a matrix $\boldsymbol{A}$ (second-order tensor). It is denoted as $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, and its $i_1 i_2 \ldots i_N$-th entry is denoted by $\boldsymbol{\mathcal{A}}_{i_1 i_2 \ldots i_N}$ or $\boldsymbol{a}_{i_1 i_2 \ldots i_N}$, $1 \leq i_n \leq I_n, 1 \leq n \leq N$.

Definition 1. *The inner product between two tensors* $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *and* $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is defined as*

$$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} \boldsymbol{\mathcal{A}}_{i_1 i_2 \ldots i_N} \cdot \boldsymbol{\mathcal{B}}_{i_1 i_2 \ldots i_N}. \tag{1}$$

Definition 2. *The Frobenius norm of tensors* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is the square root of the sum of the squares of all its elements, i.e.,*

$$\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \sqrt{\sum_{i_1 \cdots i_N} |\mathcal{A}_{i_1 i_2 \ldots i_N}|^2}. \tag{2}$$

Definition 3. *The distance between two tensors* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *and* $\mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *is defined as*

$$d(\mathcal{A}, \mathcal{B}) = \|\mathcal{A} - \mathcal{B}\|_F. \tag{3}$$

Definition 4. *The n-mode product of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ *with a matrix* $\boldsymbol{A} \in \mathbb{R}^{J \times I_n}$ *is denoted by* $(\mathcal{A} \times_n \boldsymbol{A}) \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$*. In addition, the* $i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N$*-th element of* $\mathcal{A} \times_n \boldsymbol{A}$ *is defined as*

$$(\mathcal{A} \times_n \boldsymbol{A})_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} \mathcal{A}_{i_1 i_2 \ldots i_N} \cdot \boldsymbol{A}_{j i_n}. \tag{4}$$

Definition 5. *For an N-order tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$*, the Tucker decomposition of the tensor* $\mathcal{A}$ *can be expressed as*

$$\begin{aligned} \mathcal{A} &\approx \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} \mathcal{G}_{r_1 r_2 \cdots r_N} \left(\boldsymbol{a}_{r_1}^{(1)} \circ \cdots \circ \boldsymbol{a}_{r_N}^{(N)} \right) \\ &= \mathcal{G} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \cdots \times_N \boldsymbol{A}^{(N)}, \end{aligned} \tag{5}$$

where $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \ldots \times R_N}$ *denotes the core tensor, which retains the main information of the original tensor* $\mathcal{A}$*, the symbol "*$\circ$*" denotes the outer product of vectors. For* $n \in \{1, 2, \ldots, N\}$*,* $\boldsymbol{a}_k^{(n)} \in \mathbb{R}^{I_n}$ *is the k-th* $(k = 1, 2, \ldots, R_n)$ *column vector of the mode-n factor matrices* $\boldsymbol{A}^{(n)}$*, i.e.,* $\boldsymbol{A}^{(n)} = \left[\boldsymbol{a}_1^{(n)}, \boldsymbol{a}_2^{(n)}, \ldots, \boldsymbol{a}_{R_n}^{(n)} \right] \in \mathbb{R}^{I_n \times R_n}$*, which represents the principal component of the original tensor* $\mathcal{A}$ *in each mode [35]. Note that the CP can be considered as a special case of the Tucker decomposition, in which the core tensor has nonzero elements only on main diagonal and* $R_1 = R_2 = \ldots = R_N$*.*

2.2. Support Higher-Order Tensor Machine

For a dataset containing m tensor samples $\{\mathcal{X}_i, y_i\}_{i=1}^m$, $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ represents the input sample and $y_i \in \{-1, 1\}$ represents the label corresponding to the sample. The SHTM model be simplified into the following optimization problem:

$$\begin{aligned} &\min_{(\mathcal{W}, b, \xi_i|_{i=1}^m)} \ \frac{1}{2}\|\mathcal{W}\|_F^2 + c_0 \sum_{i=1}^m \xi_i \\ &s.t. \begin{cases} y_i(\langle \mathcal{W}, \mathcal{X}_i \rangle + b) \geqslant 1 - \xi_i, \\ \xi_i \geqslant 0, \qquad i = 1, 2, \cdots, m, \end{cases} \end{aligned} \tag{6}$$

where $\mathcal{W}$ denotes the weight vector of the classification hyperplane $\mathcal{H} : \langle \mathcal{W}, \mathcal{X} \rangle + b = 0$. c_0 denotes the penalty factor, ξ_i denotes the slack variable, and b denotes the unknown bias. The Lagrangian function of the problem (5) is as follows:

$$\begin{aligned} L(\mathcal{W}, b, (\xi_i, \alpha_i, \beta_i)|_{i=1}^m) &= \frac{1}{2}\|\mathcal{W}\|_F^2 + c_0 \sum_{i=1}^m \xi_i \\ - \sum_{i=1}^m \alpha_i [y_i(\langle \mathcal{W}, \mathcal{X}_i \rangle + b) - 1 + \xi_i] &- \sum_{i=1}^m \beta_i \xi_i, \end{aligned} \tag{7}$$

where α_i and β_i represent Lagrange multipliers. Calculate the partial derivative of all variables, and set them to 0 to obtain the Karush–Kuhn–Tucker condition:

$$\begin{aligned} &\frac{\partial L}{\partial \boldsymbol{\mathcal{W}}} = \boldsymbol{\mathcal{W}} - \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{\mathcal{X}}_i = 0, \\ &\frac{\partial L}{\partial b} = \sum_{i=1}^{m} \alpha_i y_i = 0, \\ &\frac{\partial L}{\partial \xi_i} = c_0 - \alpha_i - \beta_i = 0, \quad i = 1, 2, \cdots, m. \end{aligned} \tag{8}$$

Substitute (8) into the problem (5) to obtain the dual problem as follows:

$$\begin{aligned} &\max_{(\alpha_i|_{i=1}^m)} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_j \rangle \\ &s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le c_0, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{9}$$

To obtain a more compact and meaningful representation of tensor objects, perform CP decomposition on tensors $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$, and rewrite (9) as follows:

$$\begin{aligned} &\max_{(\alpha_i|_{i=1}^m)} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{p=1}^{R} \sum_{q=1}^{R} \alpha_i \alpha_j y_i y_j \prod_{n=1}^{N} \left\langle \boldsymbol{x}_{ip}^{(n)}, \boldsymbol{x}_{jq}^{(n)} \right\rangle \\ &s.t. \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \ 0 \le \alpha_i \le c_0, i = 1, 2, \ldots, m. \end{aligned} \tag{10}$$

By solving (10), the solution vector $\boldsymbol{a}^* = \left(\alpha_1^*, \alpha_2^*, \cdots, \alpha_m^*\right)$ and bias b^* can be obtained. The final decision function is expressed as follows:

$$f(\boldsymbol{\mathcal{X}}) = sign\left(\sum_{i=1}^{m} \sum_{p=1}^{R} \sum_{q=1}^{R} \alpha_i^* y_i \prod_{n=1}^{N} \left\langle \boldsymbol{x}_{ip}^{(n)}, \boldsymbol{x}_{q}^{(n)} \right\rangle + b^* \right). \tag{11}$$

3. Proposed Methods

3.1. Global–Local-Based IF Membership Score Assignment

In this section, we propose a novel global–local-based IF membership assignment scheme that considers both the global contribution information of the sample to two classes and the local contribution information of points in the sample neighborhood.

Considering a binary classification scenario in the tensor space $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, we suppose that $\boldsymbol{\mathcal{A}} = \{\boldsymbol{\mathcal{X}}_i\}_{i=1}^{l}$ represents the set of positive class training samples (label $y_i = +1$), and $\boldsymbol{\mathcal{B}} = \{\boldsymbol{\mathcal{X}}_i\}_{i=l+1}^{m}$ represents the set of negative class training samples (label $y_i = -1$). Let $\boldsymbol{\mathcal{C}} = \{\boldsymbol{\mathcal{X}}_i\}_{i=1}^{m}$ represent the set containing all positive and negative training samples. Firstly, we introduce a nonlinear feature mapping function $\varphi(\cdot)$ mapping the tensor sample $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ in the original space to a high-dimensional Hilbert space, i.e., $\varphi : \boldsymbol{\mathcal{X}} \to \varphi(\boldsymbol{\mathcal{X}}) \in \mathbb{R}^{H_1 \times H_2 \times \ldots H_P}$. Then, we design membership and nonmembership functions for each sample in a high-dimensional feature space.

(1) Membership Function: The membership degree of a sample is calculated based on the distance between the sample and the class center to which it belongs (global information), as well as the number of samples with the same label in its neighborhood (local information). For sample $\boldsymbol{\mathcal{X}}_i$, $i \in [0, m]$, the ratio of samples with the same label as $\boldsymbol{\mathcal{X}}_i$ in its neighborhood to the total number of samples in its neighborhood is represented as

$$\theta(\boldsymbol{\mathcal{X}}_i) = \frac{\left|\left\{\boldsymbol{\mathcal{X}}_j \middle| \left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \varphi(\boldsymbol{\mathcal{X}}_j)\right\| \le \lambda, y_j = y_i \right\}\right|}{\left|\left\{\boldsymbol{\mathcal{X}}_j \middle| \left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \varphi(\boldsymbol{\mathcal{X}}_j)\right\| \le \lambda \right\}\right|}. \tag{12}$$

Then, the membership function of $\boldsymbol{\mathcal{X}}_i$ can be described as

$$\mu(\boldsymbol{\mathcal{X}}_i) = \begin{cases} \theta(\boldsymbol{\mathcal{X}}_i) \cdot \exp(-\dfrac{\left\|(\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^+)/(R^+ + h)\right\|_F^2}{2h^2}), y_i = +1, \\ \theta(\boldsymbol{\mathcal{X}}_i) \cdot \exp(-\dfrac{\left\|(\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^-)/(R^- + h)\right\|_F^2}{2h^2}), y_i = -1, \end{cases} \tag{13}$$

where $h > 0$ and $\lambda > 0$ are adjustable parameters, R^+ and $\boldsymbol{\mathcal{D}}^+$ represent the radius and class center of the positive class, R^- and $\boldsymbol{\mathcal{D}}^-$ represent the radius and class center of the negative class, i.e., $R^\pm = \max_{y_i=\pm1}\left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^\pm\right\|$, $\boldsymbol{\mathcal{D}}^\pm = \sum_{y_i=\pm1}\varphi(\boldsymbol{\mathcal{X}}_i)/m_\pm$, m_+ (m_-) is the number of positive (negative) samples.

(2) Nonmembership Function: The nonmembership degree of a sample is calculated based on the distance between the sample and the center of another class (global information), as well as the number of samples with the different label in its neighborhood (local information). For sample $\boldsymbol{\mathcal{X}}_i$, $i \in [0, m]$, the ratio of samples with the different label as $\boldsymbol{\mathcal{X}}_i$ in its neighborhood to the total number of samples in its neighborhood is represented as

$$\eta(\boldsymbol{\mathcal{X}}_i) = \frac{\left|\{\boldsymbol{\mathcal{X}}_j \middle| \left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \varphi(\boldsymbol{\mathcal{X}}_j)\right\| \le \lambda, y_j \ne y_i\}\right|}{\left|\{\boldsymbol{\mathcal{X}}_j \middle| \left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \varphi(\boldsymbol{\mathcal{X}}_j)\right\| \le \lambda\}\right|}. \tag{14}$$

Then, the nonmembership function of $\boldsymbol{\mathcal{X}}_i$ can be described as

$$v(\boldsymbol{\mathcal{X}}_i) = \begin{cases} \eta(\boldsymbol{\mathcal{X}}_i) \cdot \exp(-\dfrac{\left\|(\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^-)/(S^+ + h)\right\|_F^2}{2h^2}), y_i = +1, \\ \eta(\boldsymbol{\mathcal{X}}_i) \cdot \exp(-\dfrac{\left\|(\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^+)/(S^- + h)\right\|_F^2}{2h^2}), y_i = -1, \end{cases} \tag{15}$$

where $S^\pm = \max_{y_i=\pm1}\left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^\mp\right\|$. Next, to highlight the different contributions of each sample, we define a global–local-based intuitionistic fuzzy score function as shown in (16) and assign a score to each sample. By combining membership and nonmembership functions, the score of $\boldsymbol{\mathcal{X}}_i$ is expressed as

$$s_i = \left[p\mu(\boldsymbol{\mathcal{X}}_i)^q + (1-p)(1 - v(\boldsymbol{\mathcal{X}}_i))^q\right]^{\frac{1}{q}}, \tag{16}$$

where q is a positive parameter and $p \in [0, 1]$. Therefore, the score of the positive class samples $\boldsymbol{s}_1 = [s_1, \ldots, s_l]^T$ and the score of the negative class samples $\boldsymbol{s}_2 = [s_{l+1}, \ldots, s_m]^T$ can be obtained.

In this paper, membership and nonmembership functions are established based on tensor distances in high-dimensional space. By introducing the tensor kernel function $K(\cdot)$, we derive the following Theorem 1.

Theorem 1. *Assuming the kernel function between tensor samples $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{X}}'$ is $K(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}')$, the distance between $\varphi(\boldsymbol{\mathcal{X}})$ and $\varphi(\boldsymbol{\mathcal{X}}')$ is calculated by*

$$\left\|\varphi(\boldsymbol{\mathcal{X}}) - \varphi(\boldsymbol{\mathcal{X}}')\right\|_F = \sqrt{K(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}) + K(\boldsymbol{\mathcal{X}}', \boldsymbol{\mathcal{X}}') - 2K(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}')}. \tag{17}$$

Proof.

$$\begin{aligned} &\left\|\varphi(\boldsymbol{\mathcal{X}}) - \varphi(\boldsymbol{\mathcal{X}}')\right\|_F = \sqrt{\langle(\varphi(\boldsymbol{\mathcal{X}}) - \varphi(\boldsymbol{\mathcal{X}}')), (\varphi(\boldsymbol{\mathcal{X}}) - \varphi(\boldsymbol{\mathcal{X}}'))\rangle} \\ &= \sqrt{\langle\varphi(\boldsymbol{\mathcal{X}}), \varphi(\boldsymbol{\mathcal{X}})\rangle + \langle\varphi(\boldsymbol{\mathcal{X}}'), \varphi(\boldsymbol{\mathcal{X}}')\rangle - 2\langle\varphi(\boldsymbol{\mathcal{X}}), \varphi(\boldsymbol{\mathcal{X}}')\rangle} \\ &= \sqrt{K(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}) + K(\boldsymbol{\mathcal{X}}', \boldsymbol{\mathcal{X}}) - 2K(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}')}. \end{aligned} \tag{18}$$

Through Theorem 1, we can further derive that

$$\begin{aligned}
&\left\|\varphi(\boldsymbol{\mathcal{X}}_i)-\boldsymbol{\mathcal{D}}^{+}\right\|_F^2=\left\|\varphi(\boldsymbol{\mathcal{X}}_i)-\textstyle\sum_{y_j=1}\varphi(\boldsymbol{\mathcal{X}}_j)/m_+\right\|_F^2\\
&=\left\langle\left(\varphi(\boldsymbol{\mathcal{X}}_i)-\frac{1}{m_+}\sum_{y_j=1}\varphi(\boldsymbol{\mathcal{X}}_j)\right),\left(\varphi(\boldsymbol{\mathcal{X}}_i)-\frac{1}{m_+}\sum_{y_j=1}\varphi(\boldsymbol{\mathcal{X}}_j)\right)\right\rangle\\
&=\langle\varphi(\boldsymbol{\mathcal{X}}_i),\varphi(\boldsymbol{\mathcal{X}}_i)\rangle-\frac{2}{m_+}\cdot\left\langle\varphi(\boldsymbol{\mathcal{X}}_i),\sum_{y_j=1}\varphi(\boldsymbol{\mathcal{X}}_j)\right\rangle\\
&+\frac{1}{{m_+}^2}\cdot\left\langle\sum_{y_j=1}\varphi(\boldsymbol{\mathcal{X}}_j),\sum_{y_j=1}\varphi(\boldsymbol{\mathcal{X}}_j)\right\rangle\\
&=K(\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_i)-\frac{2}{m_+}\sum_{y_j=1}K(\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_j)+\frac{1}{{m_+}^2}\sum_{y_a=1}\sum_{y_b=1}K(\boldsymbol{\mathcal{X}}_a,\boldsymbol{\mathcal{X}}_b).
\end{aligned}\tag{19}$$

Similarly, $\left\|\varphi(\boldsymbol{\mathcal{X}}_i)-\boldsymbol{\mathcal{D}}^{-}\right\|_F^2$ can also be calculated by

$$\begin{aligned}
&\left\|\varphi(\boldsymbol{\mathcal{X}}_i)-\boldsymbol{\mathcal{D}}^{-}\right\|_F^2\\
&=K(\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_i)-\frac{2}{l_-}\sum_{y_j=-1}K(\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_j)+\frac{1}{{l_-}^2}\sum_{y_a=-1}\sum_{y_b=-1}K(\boldsymbol{\mathcal{X}}_a,\boldsymbol{\mathcal{X}}_b).
\end{aligned}\tag{20}$$

□

It is not difficult to find that the solutions (19) and (20) mainly depend on the kernel function $K(\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_j)$. The commonly used kernel functions include linear kernel functions (21) and Gaussian kernel functions (22):

$$K_L(\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_j)=\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_j\rangle,\tag{21}$$

$$K_G(\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{X}}_j)=\exp\left(-\left\|\boldsymbol{\mathcal{X}}_i-\boldsymbol{\mathcal{X}}_j\right\|_F^2/(2\sigma^2)\right),\tag{22}$$

where σ is the control parameter in the Gaussian kernel function, which is used to set the bandwidth. We considering that tensor Tucker decomposition can preserve the relevant core information in tensor data. Therefore, we will introduce tensor Tucker decomposition to enable the kernel function to retain more tensor structural information in the mapping process. According to the definition of Tucker decomposition (5), the tensors $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$ can be expressed as

$$\begin{aligned}
\boldsymbol{\mathcal{X}}_i&\approx\sum_{r_1=1}^{R_1}\cdots\sum_{r_N=1}^{R_N}\boldsymbol{\mathcal{G}}_{ir_1r_2\cdots r_N}\left(\boldsymbol{a}_{ir_1}^{(1)}\circ\cdots\circ\boldsymbol{a}_{ir_N}^{(N)}\right),\\
\boldsymbol{\mathcal{X}}_j&\approx\sum_{t_1=1}^{T_1}\cdots\sum_{t_N=1}^{T_N}\boldsymbol{\mathcal{G}}_{jt_1t_2\cdots t_N}\left(\boldsymbol{a}_{jt_1}^{(1)}\circ\cdots\circ\boldsymbol{a}_{jt_N}^{(N)}\right).
\end{aligned}\tag{23}$$

Next, the tensor in the original space is transformed into a high-dimensional linearly separable feature space through nonlinear mapping. According to mapping theory, the mapping process involves transforming the Tucker decomposition (23) as follows:

$$\begin{aligned}
&\varphi:\sum_{r_1=1}^{R_1}\cdots\sum_{r_N=1}^{R_N}\boldsymbol{\mathcal{G}}_{ir_1r_2\cdots r_N}\left(\boldsymbol{a}_{ir_1}^{(1)}\circ\cdots\circ\boldsymbol{a}_{ir_N}^{(N)}\right)\rightarrow\\
&\sum_{r_1=1}^{R_1}\cdots\sum_{r_N=1}^{R_N}\varphi\left(\boldsymbol{\mathcal{G}}_{ir_1r_2\cdots r_N}\right)\left(\varphi\left(\boldsymbol{a}_{ir_1}^{(1)}\right)\circ\cdots\circ\varphi\left(\boldsymbol{a}_{ir_N}^{(N)}\right)\right),
\end{aligned}\tag{24}$$

$$\begin{aligned}\varphi : &\sum_{t_1=1}^{T_1} \cdots \sum_{t_N=1}^{T_N} \mathcal{G}_{jt_1t_2\cdots t_N}\left(\boldsymbol{a}_{jt_1}^{(1)} \circ \cdots \circ \boldsymbol{a}_{jt_N}^{(N)}\right) \rightarrow \\ &\sum_{t_1=1}^{T_1} \cdots \sum_{t_N=1}^{T_N} \varphi\left(\mathcal{G}_{jt_1t_2\cdots t_N}\right)\left(\varphi\left(\boldsymbol{a}_{jt_1}^{(1)}\right) \circ \cdots \circ \varphi\left(\boldsymbol{a}_{jt_N}^{(N)}\right)\right).\end{aligned} \tag{25}$$

Therefore, the kernel function of tensors $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$ under the tensor Tucker decomposition is derived as follows:

$$\begin{aligned}&K(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_j) = (\varphi(\boldsymbol{\mathcal{X}}_i), \varphi(\boldsymbol{\mathcal{X}}_j)) \\ &= \left(\sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} \varphi\left(\mathcal{G}_{ir_1r_2\cdots r_N}\right)\left(\varphi\left(\boldsymbol{a}_{ir_1}^{(1)}\right) \circ \cdots \circ \varphi\left(\boldsymbol{a}_{ir_N}^{(N)}\right)\right),\right. \\ &\left.\sum_{t_1=1}^{T_1} \cdots \sum_{t_N=1}^{T_N} \varphi\left(\mathcal{G}_{jt_1t_2\cdots t_N}\right)\left(\varphi\left(\boldsymbol{a}_{jt_1}^{(1)}\right) \circ \cdots \circ \varphi\left(\boldsymbol{a}_{jt_N}^{(N)}\right)\right)\right) \\ &= \sum_{r_1=1}^{R_1}\sum_{t_1=1}^{T_1} \cdots \sum_{r_N=1}^{R_N}\sum_{t_N=1}^{T_N} \left(\varphi\left(\mathcal{G}_{ir_1r_2\cdots r_N}\right)\left(\varphi\left(\boldsymbol{a}_{ir_1}^{(1)}\right) \circ \cdots \circ \varphi\left(\boldsymbol{a}_{ir_N}^{(N)}\right)\right),\right. \\ &\left.\varphi\left(\mathcal{G}_{jt_1t_2\cdots t_N}\right)\left(\varphi\left(\boldsymbol{a}_{jt_1}^{(1)}\right) \circ \cdots \circ \varphi\left(\boldsymbol{a}_{jt_N}^{(N)}\right)\right)\right) \\ &= \sum_{r_1=1}^{R_1}\sum_{t_1=1}^{T_1} \cdots \sum_{r_N=1}^{R_N}\sum_{t_N=1}^{T_N} \cdot \prod_{n=1}^{N} K\left(\mathcal{G}_{ir_1r_2\cdots r_N}\boldsymbol{a}_{ir_n}^{(n)}, \mathcal{G}_{jt_1t_2\cdots t_N}\boldsymbol{a}_{jt_n}^{(n)}\right).\end{aligned} \tag{26}$$

Further rewrite (21) and (22) as

$$\begin{aligned}K_L(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_j) \approx &\sum_{r_1=1}^{R_1}\sum_{t_1=1}^{T_1} \cdots \sum_{r_N=1}^{R_N}\sum_{t_N=1}^{T_N} \cdot \\ &\prod_{n=1}^{N} \left\langle \mathcal{G}_{ir_1r_2\cdots r_N}\boldsymbol{a}_{ir_n}^{(n)}, \mathcal{G}_{jt_1t_2\cdots t_N}\boldsymbol{a}_{jt_n}^{(n)} \right\rangle,\end{aligned} \tag{27}$$

$$\begin{aligned}K_G(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_j) \approx &\sum_{r_1=1}^{R_1}\sum_{t_1=1}^{T_1} \cdots \sum_{r_N=1}^{R_N}\sum_{t_N=1}^{T_N} \cdot \\ &\exp(-\frac{1}{2\sigma^2} \cdot \sum_{n=1}^{N} \left\| \mathcal{G}_{ir_1r_2\cdots r_N}\boldsymbol{a}_{ir_n}^{(n)}, \mathcal{G}_{jt_1t_2\cdots t_N}\boldsymbol{a}_{jt_n}^{(n)} \right\|^2).\end{aligned} \tag{28}$$

Due to the decomposed core tensor and factor matrix simulating the potential complex patterns of interactions between components under different modes, (26)–(28) map the decomposed tensor to a high-dimensional feature space, and each decomposed factor is interrelated, so the kernel function of each factor is also interrelated, which can reflect more structural information.

3.2. IFW-LSTSHTM Model

The objective of linear IFW-LSTSHTM is to find the following two nonparallel hyperplanes in $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ for the binary class problems:

$$\begin{aligned}\boldsymbol{\mathcal{H}}_1 : \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{W}}_1 \rangle + b_1 = 0, \\ \boldsymbol{\mathcal{H}}_2 : \langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{W}}_2 \rangle + b_2 = 0,\end{aligned} \tag{29}$$

where $\boldsymbol{\mathcal{W}}_1$ and b_1 are the weight vector and bias of the first hyperplane $\mathcal{H}_1$, respectively. Similarly, $\boldsymbol{\mathcal{W}}_2$ and b_2 are the weight vector and bias of the second hyperplane $\mathcal{H}_2$, respectively. By employing the kernel trick, we extend IFW-LSTSHTM to the nonlinear case. We denote the vector composed of kernel functions between sample $\boldsymbol{\mathcal{X}}_i$ and samples in set $\mathcal{C} = \{\boldsymbol{\mathcal{X}}_i\}_{i=1}^m$ as $\mathcal{K}(\boldsymbol{\mathcal{X}}_i, \mathcal{C}) = [K(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_1), \ldots, K(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_m)]$. The goal of nonlinear IFW-LSTSHTM is to search the following two nonparallel hyperplanes:

$$\begin{aligned}\boldsymbol{\mathcal{H}}_1 : \mathcal{K}(\boldsymbol{\mathcal{X}}, \mathcal{C})z_1 + b_1 = 0, \\ \boldsymbol{\mathcal{H}}_2 : \mathcal{K}(\boldsymbol{\mathcal{X}}, \mathcal{C})z_2 + b_2 = 0.\end{aligned} \tag{30}$$

If we choose the linear kernel $K(\mathcal{X}_i, \mathcal{X}_j) = \langle \mathcal{X}_i, \mathcal{X}_j \rangle$ and assume $\exists\ z_1, z_2 \in \mathbb{R}^m$, let $[\langle \mathcal{X}, \mathcal{X}_1 \rangle, \ldots, \langle \mathcal{X}, \mathcal{X}_m \rangle] z_1 = \langle \mathcal{X}, \mathcal{W}_1 \rangle$, $[\langle \mathcal{X}, \mathcal{X}_1 \rangle, \ldots, \langle \mathcal{X}, \mathcal{X}_m \rangle] z_2 = \langle \mathcal{X}, \mathcal{W}_2 \rangle$. Then, we can obtain the linear planes in (29). These planes are obtained by solving two QPPs, which are similar in formula to SHTM problems. To be consistent with the matrix formula in traditional least squares twin support vector machine optimization problems, we have the following abbreviations:

$$\begin{aligned} \mathcal{K}(\mathcal{A}, \mathcal{C}) &= \begin{bmatrix} \mathcal{K}(\mathcal{X}_1, \mathcal{C}) \\ \vdots \\ \mathcal{K}(\mathcal{X}_l, \mathcal{C}) \end{bmatrix} \\ &= \begin{bmatrix} K(\mathcal{X}_1, \mathcal{X}_1) & \cdots & K(\mathcal{X}_1, \mathcal{X}_m) \\ \vdots & \ddots & \vdots \\ K(\mathcal{X}_l, \mathcal{X}_1) & \cdots & K(\mathcal{X}_l, \mathcal{X}_m) \end{bmatrix}, \end{aligned} \tag{31}$$

$$\begin{aligned} \mathcal{K}(\mathcal{B}, \mathcal{C}) &= \begin{bmatrix} \mathcal{K}(\mathcal{X}_{l+1}, \mathcal{C}) \\ \vdots \\ \mathcal{K}(\mathcal{X}_m, \mathcal{C}) \end{bmatrix} \\ &= \begin{bmatrix} K(\mathcal{X}_{l+1}, \mathcal{X}_1) & \cdots & K(\mathcal{X}_{l+1}, \mathcal{X}_m) \\ \vdots & \ddots & \vdots \\ K(\mathcal{X}_m, \mathcal{X}_1) & \cdots & K(\mathcal{X}_m, \mathcal{X}_m) \end{bmatrix}. \end{aligned} \tag{32}$$

Note that we use the tensor kernel function with the tensor Tucker decomposition in (31) and (32). Specifically, for linear kernel function, (31) and (32) become

$$\mathcal{K}(\mathcal{A}, \mathcal{C}) = \begin{bmatrix} \langle \mathcal{X}_1, \mathcal{X}_1 \rangle & \cdots & \langle \mathcal{X}_1, \mathcal{X}_m \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathcal{X}_l, \mathcal{X}_1 \rangle & \cdots & \langle \mathcal{X}_l, \mathcal{X}_m \rangle \end{bmatrix}, \tag{33}$$

$$\mathcal{K}(\mathcal{B}, \mathcal{C}) = \begin{bmatrix} \langle \mathcal{X}_{l+1}, \mathcal{X}_1 \rangle & \cdots & \langle \mathcal{X}_{l+1}, \mathcal{X}_m \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathcal{X}_m, \mathcal{X}_1 \rangle & \cdots & \langle \mathcal{X}_m, \mathcal{X}_m \rangle \end{bmatrix}. \tag{34}$$

Firstly, by minimizing the sum of squared distances between one class of samples and their corresponding hyperplane, and ensuring that another class of samples is at least one distance away from the hyperplane, two QPP problems can be obtained as follows:

$$\begin{aligned} &\min_{z_1, b_1, \xi_2} \frac{1}{2} \| \mathcal{K}(\mathcal{A}, \mathcal{C}) z_1 + e_1 b_1 \|^2 + c_1 e_2^T \xi_2 \\ &s.t. \quad -(\mathcal{K}(\mathcal{B}, \mathcal{C}) z_1 + e_2 b_1) + \xi_2 \geq e_2, \xi_2 \geq \mathbf{0}, \end{aligned} \tag{35}$$

$$\begin{aligned} &\min_{z_2, b_2, \xi_1} \frac{1}{2} \| \mathcal{K}(\mathcal{B}, \mathcal{C}) z_2 + e_2 b_2 \|^2 + c_2 e_1^T \xi_1 \\ &\text{s.t.} \quad (\mathcal{K}(\mathcal{A}, \mathcal{C}) z_2 + e_1 b_2) + \xi_1 \geq e_1, \xi_1 \geq \mathbf{0}, \end{aligned} \tag{36}$$

where c_1, c_2 denote nonnegative parameters, e_1 and e_2 are vectors of ones of appropriate dimensions, ξ_1 and ξ_2 denote slack variables. Here, we modify the original problems (35)

and (36) in the least squares sense to (37) and (38), with the inequality constraints replaced with equality constraints, as shown below:

$$\begin{aligned} &\min_{z_1,b_1,\xi_2} \frac{1}{2}\|\mathcal{K}(\mathcal{A},\mathcal{C})z_1 + e_1 b_1\|^2 + c_1\left\|e_2^T\xi_2\right\|^2 \\ &s.t. \quad -(\mathcal{K}(\mathcal{B},\mathcal{C})z_1 + e_2 b_1) + \xi_2 = e_2, \end{aligned} \tag{37}$$

$$\begin{aligned} &\min_{z_2,b_2,\xi_1} \frac{1}{2}\|\mathcal{K}(\mathcal{B},\mathcal{C})z_2 + e_2 b_2\|^2 + c_2\left\|e_1^T\xi_1\right\|^2 \\ &\text{s.t.} \quad (\mathcal{K}(\mathcal{A},\mathcal{C})z_2 + e_1 b_2) + \xi_1 = e_1. \end{aligned} \tag{38}$$

Also note that QPPs (37) and (38) use the square of 2-norm $\|\xi_2\|^2$ and $\|\xi_1\|^2$ instead of ξ_2 and ξ_1, which makes constraint $\xi_2 \geq 0$ and $\xi_1 \geq 0$ redundant. This very simple modification allows us to transform solving QPPs (35) and (36) into solving only two linear systems of equations. To achieve the principle of minimizing structural risk, an additional regularization term is introduced in (37) and (38), respectively, resulting in the following optimization problem:

$$\begin{aligned} &\min_{z_1,b_1,\xi_2} \frac{1}{2}\|\mathcal{K}(\mathcal{A},\mathcal{C})z_1 + e_1 b_1\|^2 + \frac{c_3}{2}(\|z_1\|^2 + b_1^2) + c_1 e_2^T\xi_2 \\ &s.t. \quad -(\mathcal{K}(\mathcal{B},\mathcal{C})z_1 + e_2 b_1) + \xi_2 = e_2, \end{aligned} \tag{39}$$

$$\begin{aligned} &\min_{z_2,b_2,\xi_1} \frac{1}{2}\|\mathcal{K}(\mathcal{B},\mathcal{C})z_2 + e_2 b_2\|^2 + \frac{c_4}{2}(\|z_2\|^2 + b_2^2) + c_2 e_1^T\xi_1 \\ &\text{s.t.} \quad (\mathcal{K}(\mathcal{A},\mathcal{C})z_2 + e_1 b_2) + \xi_1 = e_1. \end{aligned} \tag{40}$$

To extract as much potential similarity information within the samples as possible, we use a sample-weighting strategy based on intraclass local neighborhood structure information. For any two points $(\mathcal{X}_i, \mathcal{X}_j)$ in the same class, the i-th row and j-th column element μ_{ij} of the similarity matrix M can be defined as

$$M_{ij} = \begin{cases} \exp\dfrac{-\|\mathcal{X}_i - \mathcal{X}_j\|_F^2}{\delta}, & \text{if } \mathcal{X}_j \in \mathrm{Neb}_K(\mathcal{X}_i), \\ 0, & \text{otherwise}, \end{cases} \tag{41}$$

where δ is the kernel parameter, $\mathrm{Neb}_K(\mathcal{X}_i)$ represents a set of K-nearest neighbors of $\mathcal{X}_i$. Redefine the k-th nonzero element in the i-th row of M as $\bar{M}_{ik} = \left\{M_{ij} \middle| M_{ij} \neq 0\right\}$. Then, the intraclass weight of $\mathcal{X}_i$ can be given as

$$w_i = \textstyle\sum_{k=1}^{K} \bar{M}_{ik}. \tag{42}$$

Finally, intraclass weight matrices of positive class W_1 and intraclass weight matrices of negative class W_2 can be given as

$$W_1 = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_l \end{bmatrix}, W_2 = \begin{bmatrix} w_{l+1} & & & \\ & w_{l+2} & & \\ & & \ddots & \\ & & & w_m \end{bmatrix}. \tag{43}$$

Next, to address the issue of class imbalance, we define positive class imbalance penalty factor D_1 and negative class imbalance penalty factor D_2 by

$$\begin{cases} D_1 = 1 - d, \\ D_2 = 1 + d, \\ r = m_- / m_+, \end{cases} \quad and \quad d = \begin{cases} m_+/, m & m_+ > m_-, \\ 0, & m_+ = m_-, \\ -m_-/m, & m_+ < m_-, \end{cases} \tag{44}$$

where d represents a dynamic parameter, and r represents the imbalance ratio.

Therefore, by introducing the above intuitionistic fuzzy score (16), intraclass weight matrices (43), and class imbalance penalty factor (44), the nonlinear IFW-LSTSHTM optimization model is obtained as follows:

$$\begin{aligned} \min_{z_1, b_1, \xi_2} \ & \frac{1}{2}\|W_1(\mathcal{K}(\mathcal{A}, \mathcal{C})z_1 + e_1 b_1)\|^2 + \frac{1}{2}c_1 D_1 \left\|s_2^T \xi_2\right\|^2 + \frac{1}{2}c_3 D_1(\|z_1\|^2 + b_1^2) \\ \text{s.t.} \quad & -(\mathcal{K}(\mathcal{B}, \mathcal{C})z_1 + e_2 b_1) + \xi_2 = e_2, \end{aligned} \tag{45}$$

$$\begin{aligned} \min_{z_2, b_2, \xi_1} \ & \frac{1}{2}\|W_2(\mathcal{K}(\mathcal{B}, \mathcal{C})z_2 + e_2 b_2)\|^2 + \frac{1}{2}c_2 D_2 \left\|s_1^T \xi_1\right\|^2 + \frac{1}{2}c_4 D_2(\|z_2\|^2 + b_2^2) \\ \text{s.t.} \quad & -(\mathcal{K}(\mathcal{A}, \mathcal{C})z_2 + e_1 b_2) + \xi_1 = e_1, \end{aligned} \tag{46}$$

where c_1, c_2, c_3, c_4 denote nonnegative parameters. Substituting the equality constraints into the corresponding objective function, QPP (45) becomes

$$\begin{aligned} \min_{z_1, b_1, \xi_2} \ & \frac{1}{2}\|W_1(\mathcal{K}(\mathcal{A}, \mathcal{C})z_1 + e_1 b_1)\|^2 \\ & + \frac{1}{2}c_1 D_1 \left\|s_2^T(\mathcal{K}(\mathcal{B}, \mathcal{C})z_1 + e_2 b_1 + \xi_2)\right\|^2 + \frac{1}{2}c_3 D_1(\|z_1\|^2 + b_1^2). \end{aligned} \tag{47}$$

Setting the gradient of (47) relative to z_1 and b_1 to 0, we obtain

$$\begin{aligned} & [W_1 \mathcal{K}(\mathcal{A}, \mathcal{C})]^T W_1(\mathcal{K}(\mathcal{A}, \mathcal{C})z_1 + e_1 b_1) \\ & + c_1 D_1 \left[s_2^T \mathcal{K}(\mathcal{B}, \mathcal{C})\right]^T \left(s_2^T(\mathcal{K}(\mathcal{B}, \mathcal{C})z_1 + e_2 b_1 + e_2)\right) + c_3 D_1 z_1 = \mathbf{0}, \end{aligned} \tag{48}$$

$$\begin{aligned} & [W_1 e_1]^T W_1(\mathcal{K}(\mathcal{A}, \mathcal{C})z_1 + e_1 b_1) \\ & + c_1 D_1 \left[s_2^T e_2\right]^T \left(s_2^T(\mathcal{K}(\mathcal{B}, \mathcal{C})z_1 + e_2 b_1 + e_2)\right) + c_3 D_1 b_1 = \mathbf{0}. \end{aligned} \tag{49}$$

Rewriting (48) and (49) in matrix form, we obtain

$$\begin{aligned} & \begin{bmatrix} [W_1 \mathcal{K}(\mathcal{A}, \mathcal{C})]^T W_1 \mathcal{K}(\mathcal{A}, \mathcal{C}) & [W_1 \mathcal{K}(\mathcal{A}, \mathcal{C})]^T W_1 e_1 \\ [W_1 e_1]^T W_1 \mathcal{K}(\mathcal{A}, \mathcal{C}) & l \end{bmatrix} \begin{bmatrix} z_1 \\ b_1 \end{bmatrix} \\ & + c_1 D_1 \begin{bmatrix} \left[s_2^T \mathcal{K}(\mathcal{B}, \mathcal{C})\right]^T s_2^T \mathcal{K}(\mathcal{B}, \mathcal{C}) & \left[s_2^T \mathcal{K}(\mathcal{B}, \mathcal{C})\right]^T s_2^T e_2 \\ \left[s_2^T e_2\right]^T s_2^T \mathcal{K}(\mathcal{B}, \mathcal{C}) & m - l \end{bmatrix} \begin{bmatrix} z_1 \\ b_1 \end{bmatrix} \\ & + c_3 D_1 \mathbf{I} \begin{bmatrix} z_1 \\ b_1 \end{bmatrix} + c_1 D_1 s_2^T e_2 \begin{bmatrix} \left[s_2^T \mathcal{K}(\mathcal{B}, \mathcal{C})\right]^T \\ \left[s_2^T e_2\right]^T \end{bmatrix} = \mathbf{0}. \end{aligned} \tag{50}$$

We then make the following substitutions:

$$\begin{aligned} & R_1 = W_1 \mathcal{K}(\mathcal{A}, \mathcal{C}), R_2 = W_1 e_1, \\ & L_1 = s_2^T \mathcal{K}(\mathcal{B}, \mathcal{C}), L_2 = s_2^T e_2. \end{aligned} \tag{51}$$

Using the above substitutions, (50) can be rewritten as follows:

$$\begin{aligned}&\begin{bmatrix} \boldsymbol{R}_1^T\boldsymbol{R}_1 & \boldsymbol{R}_1^T\boldsymbol{R}_2 \\ \boldsymbol{R}_2^T\boldsymbol{R}_1 & l \end{bmatrix}\begin{bmatrix} z_1 \\ b_1 \end{bmatrix} + c_1D_1\begin{bmatrix} \boldsymbol{L}_1^T\boldsymbol{L}_1 & \boldsymbol{L}_1^T\boldsymbol{L}_2 \\ \boldsymbol{L}_2^T\boldsymbol{L}_1 & m-l \end{bmatrix}\begin{bmatrix} z_1 \\ b_1 \end{bmatrix} \\ &+c_3D_1\begin{bmatrix} z_1 \\ b_1 \end{bmatrix} + c_1D_1\boldsymbol{L}_2\begin{bmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \end{bmatrix} = 0,\end{aligned} \tag{52}$$

$$\begin{aligned}\begin{bmatrix} z_1 \\ b_1 \end{bmatrix} = -&\begin{bmatrix} \boldsymbol{R}_1^T\boldsymbol{R}_1 + c_1D_1\boldsymbol{L}_1^T\boldsymbol{L}_1 + c_2D_1 & \boldsymbol{R}_1^T\boldsymbol{R}_2 + c_1D_1\boldsymbol{L}_1^T\boldsymbol{L}_2 \\ \boldsymbol{R}_2^T\boldsymbol{R}_1 + c_1D_1\boldsymbol{L}_2^T\boldsymbol{L}_1 & l + c_1D_1(m-l) + c_3D_1 \end{bmatrix}^{-1} \\ &\cdot c_1D_1\boldsymbol{L}_2\begin{bmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \end{bmatrix},\end{aligned} \tag{53}$$

$$\begin{aligned}\begin{bmatrix} z_1 \\ b_1 \end{bmatrix} = -(&\begin{bmatrix} \boldsymbol{R}_1^T \\ \boldsymbol{R}_2^T \end{bmatrix}\begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{R}_2 \end{bmatrix} + c_1D_1\begin{bmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \end{bmatrix}\begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{L}_2 \end{bmatrix} + c_3D_1\mathrm{I})^{-1} \\ &\cdot c_1D_1\boldsymbol{L}_2\begin{bmatrix} \boldsymbol{L}_1^T \\ \boldsymbol{L}_2^T \end{bmatrix}.\end{aligned} \tag{54}$$

Defining $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{R}_2 \end{bmatrix}$ and $\boldsymbol{L} = \begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{L}_2 \end{bmatrix}$, the solution becomes

$$\begin{bmatrix} z_1 \\ b_1 \end{bmatrix} = -c_1D_1\left[\boldsymbol{R}^T\boldsymbol{R} + c_1D_1\boldsymbol{L}^T\boldsymbol{L} + c_3D_1\mathrm{I}\right]^{-1}\boldsymbol{L}^T\boldsymbol{L}_2. \tag{55}$$

Similarly, substituting the equality constraints into the corresponding objective function, QPP (46) becomes

$$\begin{aligned}&\min_{z_2,b_2,\xi_1} \frac{1}{2}\|W_2(\mathcal{K}(\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{C}})z_2 + \boldsymbol{e}_2b_2)\|^2 \\ &+ \frac{1}{2}c_2D_2\|\boldsymbol{s}_1^T(\mathcal{K}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{C}})z_2 + \boldsymbol{e}_1b_2 + \boldsymbol{e}_1)\|^2 + \frac{1}{2}c_4D_2(\|z_2\|^2 + b_2^2),\end{aligned} \tag{56}$$

then, we make the substitutions

$$\begin{aligned}&\boldsymbol{G}_1 = \boldsymbol{s}_1^T\mathcal{K}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{C}}), \boldsymbol{G}_2 = \boldsymbol{s}_1^T\boldsymbol{e}_1, \\ &\boldsymbol{H}_1 = \boldsymbol{W}_2\mathcal{K}(\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{C}}), \boldsymbol{H}_2 = \boldsymbol{W}_2\boldsymbol{e}_2.\end{aligned} \tag{57}$$

Defining $\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{G}_2 \end{bmatrix}$ and $\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_1 & \boldsymbol{H}_2 \end{bmatrix}$, we obtain the optimal solution for optimization problem (56) as follows

$$\begin{bmatrix} z_2 \\ b_2 \end{bmatrix} = c_2D_2\left[\boldsymbol{H}^T\boldsymbol{H} + c_2D_2\boldsymbol{G}^T\boldsymbol{G} + c_4D_2\mathrm{I}\right]^{-1}\boldsymbol{G}^T\boldsymbol{G}_2. \tag{58}$$

Therefore two nonparallel hyperplanes in (30) can be obtained. Finally, if the test sample $\boldsymbol{\mathcal{X}}_{test}$ is given, the following decision rule is used to give the label of the sample $\boldsymbol{\mathcal{X}}_{test}$:

$$\arg\min_{i=1,2}\left\{\frac{|\mathcal{K}(\boldsymbol{\mathcal{X}}_{test},\boldsymbol{\mathcal{C}})z_1 + b_1|}{\sqrt{z_1^T\mathcal{K}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{C}})z_1}}, \frac{|\mathcal{K}(\boldsymbol{\mathcal{X}}_{test},\boldsymbol{\mathcal{C}})z_2 + b_2|}{\sqrt{z_2^T\mathcal{K}(\boldsymbol{\mathcal{B}},\boldsymbol{\mathcal{C}})z_2}}\right\}. \tag{59}$$

Finally, the proposed IFW-LSTSHTM model is summarized in Algorithm 1.

Algorithm 1 IFW-LSTSHTM

Input: Training samples $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, $i = 1, \ldots, m$; for $i = 1, \ldots, l$, sample label $y_i = +1$, for $i = l+1, \ldots, m$, sample label $y_i = -1$; Testing sample $\boldsymbol{\mathcal{X}}_{test} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, parameters c_1, c_2, c_3, c_4, h, σ, δ.
Output: Predicted label of $\boldsymbol{\mathcal{X}}_{test}$.
Step 1. Obtain approximate tensor samples for all training samples $\boldsymbol{\mathcal{X}}_i$, $(i = 1, \ldots, m)$ by using tensor Tucker decomposition (23), and replace the original tensor samples with approximate tensor samples.
Step 2. For $i, j \in [1, \ldots, m]$, calculate $K(\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{X}}_j)$, for linear kernel function, by using (27), for nonlinear kernel function, by using (26).
Step 3. For $i \in [1, \ldots, m]$, calculate $\theta(\boldsymbol{\mathcal{X}}_i)$ and $\eta(\boldsymbol{\mathcal{X}}_i)$ by using (12) and (14). Then calculate $\left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^+\right\|_F^2$ and $\left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^-\right\|_F^2$ by using (19) and (20);
Step 4. For $i \in [1, \ldots, m]$, substitute $\theta(\boldsymbol{\mathcal{X}}_i)$, $\eta(\boldsymbol{\mathcal{X}}_i)$, $\left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^+\right\|_F^2$, and $\left\|\varphi(\boldsymbol{\mathcal{X}}_i) - \boldsymbol{\mathcal{D}}^-\right\|_F^2$ into (13) to obtain membership degree $\mu(\boldsymbol{\mathcal{X}}_i)$, and then substitute into (15) to obtain nonmembership degree $v(\boldsymbol{\mathcal{X}}_i)$.
Step 5. Obtain the score of the positive class samples $\boldsymbol{s}_1 = [s_1, \ldots, s_l]^T$ and the score of the negative class samples $\boldsymbol{s}_2 = [s_{l+1}, \ldots, s_m]^T$ by using (16).
Step 6. Calculate $\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{C}})$ and $\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}})$, for linear kernel function, by using (33) and (34), for nonlinear kernel function, by using (31) and (32).
Step 7. Obtain intraclass weight matrices of positive class and negative class $\boldsymbol{W}_1 = \mathrm{diag}(w_1, w_2, \ldots, w_l)$, $\boldsymbol{W}_2 = \mathrm{diag}(w_{l+1}, w_{l+2}, \ldots, w_m)$ by using (41)–(43).
Step 8. Obtain positive class imbalance penalty factor D_1 and negative class imbalance penalty factor D_2 by using (44).
Step 9. Let $\boldsymbol{R}_1 = \boldsymbol{W}_1\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{C}})$, $\boldsymbol{R}_2 = \boldsymbol{W}_1\boldsymbol{e}_1$, $\boldsymbol{L}_1 = \boldsymbol{s}_2^T\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}})$, $\boldsymbol{L}_2 = \boldsymbol{s}_2^T\boldsymbol{e}_2$, and obtain the optimal solutions z_1, b_1 for the first hyperplane by using (55).
Step 10. Let $\boldsymbol{G}_1 = \boldsymbol{s}_1^T\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{C}})$, $\boldsymbol{G}_2 = \boldsymbol{s}_1^T\boldsymbol{e}_1$, $\boldsymbol{H}_1 = \boldsymbol{W}_2\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{C}})$, $\boldsymbol{H}_2 = \boldsymbol{W}_2\boldsymbol{e}_2$, obtain the solutions z_2, b_2 for the second hyperplane by using (58).
Step 11. Calculate $\boldsymbol{\mathcal{K}}(\boldsymbol{\mathcal{X}}_{test}, \boldsymbol{\mathcal{C}}) = [K(\boldsymbol{\mathcal{X}}_{test}, \boldsymbol{\mathcal{X}}_1), \ldots, K(\boldsymbol{\mathcal{X}}_{test}, \boldsymbol{\mathcal{X}}_m)]$ by Step 2. Obtain the predicted label of test sample $\boldsymbol{\mathcal{X}}_{test}$ by decision rule (59).

4. Feature Tensor Reconstruction

In this section, we first propose a feature tensor reconstruction method based on FST, and then demonstrate the proposed multisensor fault diagnosis process. The feature tensor reconstruction process mainly includes the following:

(1) Multisensor signal preprocessing

For multiple sensor signals, $\boldsymbol{S} \in \mathbb{R}^{I_3 \times N}$ (I_3 represents the number of sensors, N represents the length of the signal sequence). Then, set the time window length to L, which means each L data point is used as a time period to form a sample. Set the sliding step size to s, which means that the first s points in the current time period are composed of the last s data points in the previous time period. Therefore, by segmenting, we can obtain m samples, and for $i = 1, 2, \ldots, m$, the i-th sample can be represented as $\boldsymbol{X}_i \in \mathbb{R}^{I_3 \times L}$.

(2) Time–frequency transform

For the i-th sample, let $\boldsymbol{x}_j \in \mathbb{R}^{1 \times L}$ represent the i-th segmentation sequence of the j-th sensor signal, and then the i-th sample can be represented as $\boldsymbol{X}_i = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \ldots, \boldsymbol{x}_{I_3}^T\right]^T \in \mathbb{R}^{I_3 \times L}$. For $j = 1, 2, \ldots, I_3$, convert $\boldsymbol{x}_j$ into a time–frequency image using FST, and crop each image into size $I_1 \times I_2$ and perform grayscale processing, ultimately normalizing the pixels to $[0, 1]$ and obtaining I_3 normalized pixel matrices.

(3) Stacking grayscale images

For the i-th sample, stacking I_3 normalized pixel matrices along the forward slice direction, we can obtain a third-order feature tensor $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, of which the three orders correspond to the feature information in the time domain and frequency domain, and the structural information in the spatial domain, respectively.

Therefore, we first use the above feature tensor reconstruction method to reconstruct the original multisensor signal into a third-order feature tensor, and then use the developed tensor-based nonlinear classifier IFW-LSTSHTM to achieve fault diagnosis. The flowchart of the proposed multisensor fault diagnosis method is shown in Figure 1.

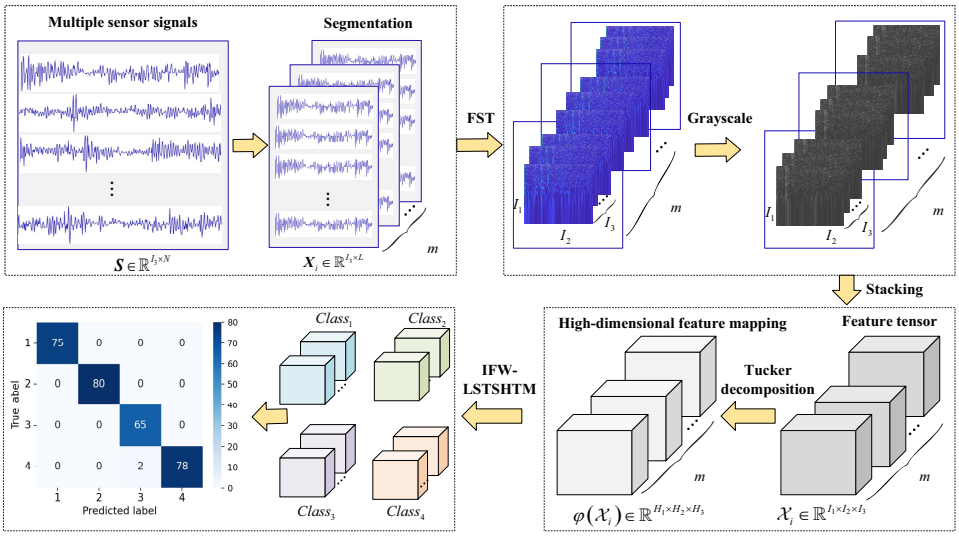


Figure 1. Flowchart of the proposed multisensor fault diagnosis method.

5. Case Studies

5.1. Case Study 1

5.1.1. Dataset Description and Settings

In this case, we use the HUST bearing dataset , being provided by Huazhong University of Science and Technology, Wuhan, China [36]. The bearing fault tests were conducted using a spectra-quest mechanical fault simulator, as depicted in Figure 2. We use four different health status bearings, including medium ball fault (label 1), medium inner race fault (label 2), medium outer race fault (label 3), and medium combination fault (label 4), as illustrated in Figure 3. It is important to note that combination fault denotes a fault in both the inner race and outer race. The type of tested bearings is ER-16K. A total of four different operating conditions were set in experiments. The operating conditions include 65 Hz, 70 Hz, 75 Hz, and 80 Hz. The sampling frequency is set to 25.6 kHz. By placing a tachometer and three-way acceleration sensor in different positions, four sensor signals are obtained, including speeed signal, and vibration signals in the x, y, and z directions.

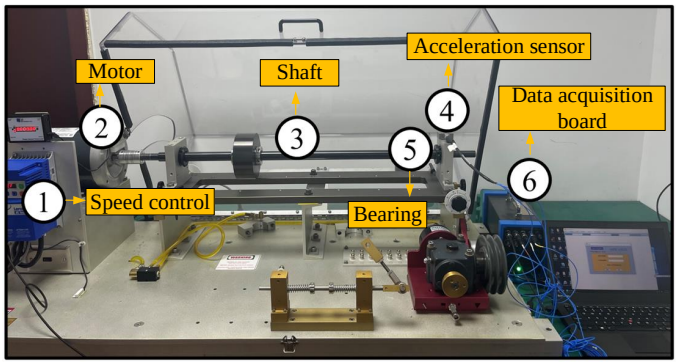
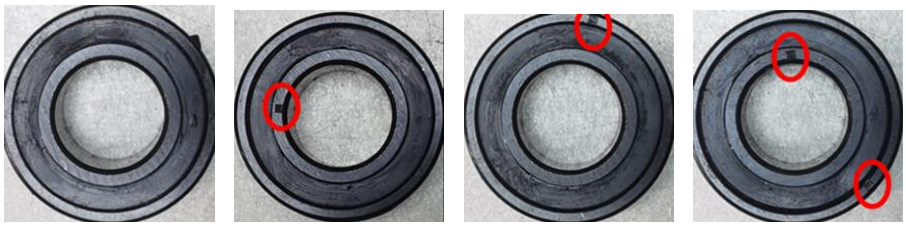


Figure 2. Test rig of HUST bearing dataset.



(**a**) (**b**) (**c**) (**d**)

Figure 3. Photographs of the HUST fault bearings. (**a**) Ball fault. (**b**) Inner race fault. (**c**) Outer race fault. (**d**) Combination fault.

Based on the tensor sample reconstruction method proposed in Section 4, 230 tensor samples are obtained for four health states, of which 150 are used for training. In the testing phase, we select 300 test samples, including 75 ball fault samples, 80 inner race fault samples, 65 outer race fault samples, and 80 combination fault samples. A detailed description of the analyzed HUST dataset is shown in Table 1. Then, we set the imbalance ratio $r = 4/5$ to verify the classification performance of the proposed method in an imbalanced data scenario. To achieve multiclassification, we adopt a one-against-one strategy and conduct six group binary classification experiments. When $r = 4/5$, there are 120 negative class training samples and 150 positive class training samples in each group experiment. After simulating nonstationary signals, we combine 82,944 signals under a 75 HZ load, 81,920 signals under a 80 HZ load, and 81,920 signals under a 70 HZ load for each sensor, resulting in 246,784 variable load signals with a trend of first increasing and then decreasing. Next, to simulate a noisy environment, we sequentially add Gaussian white noise with four different signal-to-noise ratios (SNRs) (i.e., SNR = 1 dB, SNR = 2 dB, SNR = 3 dB, SNR = 4 dB) to each merged variable load sensor signal. The window length is set to 2048, and the sliding step size is set to 1024. Then, we use the method in Section 4 to construct feature tensors with a size of $64 \times 64 \times 4$.

Table 1. Detailed description of the analyzed HUST dataset.

Label	Status Description	Train Number	Test Number
1	Ball fault	150	75
2	Inner race fault	150	80
3	Outer race fault	150	65
4	Combination fault	150	80

To solve our IFW-LSTSHTM model with lower computational complexity, we make $c_1 = c_2$, $c_3 = c_4$, h, σ, δ are both selected from the set $\{2^{-7}, 2^{-6}, \ldots, 2^6, 2^7\}$. The order of the core tensor in Tucker decomposition is set to $16 \times 32 \times 3$. We choose to use the nonlinear Gaussian kernel function in (22) and use grid search method to select all optimal parameters. We use traditional DC-STM [25], FSTM [28], Pin-FSTM [29], UPIFSTM [32], and TwSHTM [33] as comparison methods to further validate the effectiveness of the proposed model.

To further validate the comprehensive performance of the proposed method, this paper uses the following accuracy, precision, recall, and F-score as performance indicators to evaluate the classification model:

$$\begin{aligned} \text{I1}: &\quad \text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{FN}+\text{TN}} \\ \text{I2}: &\quad \text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \\ \text{I3}: &\quad \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \\ \text{I4}: &\quad \text{F-score} = 2 \times \frac{(\text{Recall} \times \text{Precision})}{(\text{Recall}+\text{Precision})} \end{aligned} \tag{60}$$

where true positive (TP) represents the number of positive class sample predicted as positive class sample, false negative (FN) represents the number of positive class sample predicted as negative class sample, false positive (FP) represents the number of negative class sample predicted as positive class sample, and true negative (TN) represents the number of negative class sample predicted as negative class sample, forming the confusion matrix shown in Table 2.

Table 2. Confusion matrix for classification problem.

Confusion Matrix		**Predicted Value**	
		Positive	**Negative**
Actual value	Positive	TP	FN
	Negative	FP	TN

5.1.2. Experiment Results and Analysis

Firstly, the FST time–frequency images of the fourth sensor signal (z-direction vibration signal) obtained from the experimental setup under four different health states are shown in Figure 4. Obviously, from Figure 4, it can be seen that different time–frequency images have different impacts, indicating significant differences in different fault characteristics.

Then, we conduct experiments on the HUST bearing dataset based on the above settings, and each experiment is independently repeated ten times. Then, we obtain the average accuracy, average precision, average recall, and average F-score of different methods, as shown in Table 3, where the value after symbol "±" represents the standard deviation obtained from 10 independent repeated experiments. According to Table 3, firstly, in the case of d = 4/5, the diagnostic performance of each method improves with the increase in SNR. For four different SNR experiments, the proposed method achieves the highest values of four indicators, with each indicator value reaching 99.48% or above. Especially for SNR = 4, the indicators accuracy, precision, and F-score of the proposed method are all as high as 99.97%, and indicator recall reaches 99.96%, significantly better than other methods. For the SNR = 1 experiment, for the accuracy results of different methods, the lowest value is DC-STM (88.77%), which may be due to its sensitivity to noise. FSTM (91.23%) considers fuzziness, slightly better than STM, but lower than Pin-FSTM (93.70%) and UPIFSTM (93.6%). Although TwSHTM (97.67%) is superior to DC-STM, it is also affected by noise and lower than the proposed method (99.50%). The values of indicator F-score, from high to low, are as follows: DC-STM (88.66%), FSTM (91.25%) Pin-FSTM (93.65%), UPIFSTM (93.67%), TwSHTM (97.71%), proposed method (99.48%). In addition, the average standard deviation of the proposed methods is below 0.32%, especially for SNR = 4: the standard deviation of the precision index value reaches the lowest value of 0.09%, and the standard deviation of the accuracy and F-score index values reaches the second lowest value of 0.10%, significantly lower than other comparative methods. Therefore, the proposed method has more advantages in resisting noise interference and stability compared to other methods.

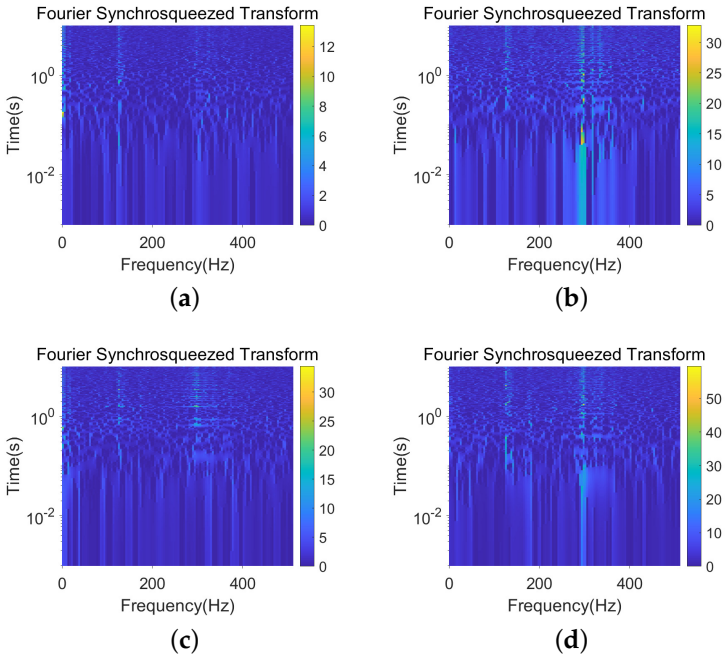


Figure 4. FST images of the z-direction vibration signal under different health states in the HUST bearing dataset. (**a**) Ball fault. (**b**) Inner race fault. (**c**) Outer race fault. (**d**) Combination fault.

Table 3. Average results (%) of different methods on the HUST bearing dataset.

Model	SNR	Accuracy	Precision	Recall	F-Score
DC-STM	1	88.77 ± 1.16	91.54 ± 0.85	89.26 ± 1.12	88.66 ± 1.27
	2	91.47 ± 1.89	92.30 ± 1.70	91.59 ± 1.90	91.54 ± 1.88
	3	92.23 ± 1.63	93.28 ± 1.30	92.38 ± 1.64	92.33 ± 1.62
	4	92.37 ± 1.04	93.28 ± 0.83	92.35 ± 1.08	92.44 ± 1.05
FSTM	1	91.23 ± 2.06	93.02 ± 1.37	91.50 ± 1.94	91.25 ± 2.12
	2	91.83 ± 1.60	92.86 ± 1.26	91.84 ± 1.58	91.93 ± 1.56
	3	92.37 ± 1.87	92.93 ± 1.68	92.36 ± 1.88	92.42 ± 1.83
	4	93.27 ± 1.12	94.00 ± 0.89	93.37 ± 1.13	93.37 ± 1.13
Pin-FSTM	1	93.70 ± 1.31	94.23 ± 1.18	93.46 ± 1.39	93.65 ± 1.36
	2	95.13 ± 1.30	95.46 ± 1.19	95.13 ± 1.34	95.22 ± 1.32
	3	95.47 ± 1.28	95.77 ± 1.14	95.40 ± 1.29	95.52 ± 1.24
	4	95.73 ± 0.79	96.16 ± 0.64	95.63 ± 0.83	95.81 ± 0.77
UPIFSTM	1	93.60 ± 1.14	94.00 ± 1.00	93.66 ± 1.11	93.67 ± 1.11
	2	95.83 ± 1.32	96.11 ± 1.18	95.82 ± 1.31	95.91 ± 1.27
	3	96.07 ± 1.10	96.28 ± 1.06	96.11 ± 1.11	96.15 ± 1.10
	4	96.77 ± 0.65	97.03 ± 0.59	96.75 ± 0.61	96.86 ± 0.61
TwSHTM	1	97.67 ± 0.80	97.78 ± 0.73	97.72 ± 0.76	97.71 ± 0.77
	2	98.03 ± 0.74	98.12 ± 0.70	98.09 ± 0.74	98.07 ± 0.73
	3	98.13 ± 0.52	98.22 ± 0.50	98.24 ± 0.50	98.19 ± 0.51
	4	98.97 ± 0.46	99.00 ± 0.43	99.02 ± 0.43	99.00 ± 0.43
Proposed	1	99.50 ± 0.27	99.48 ± 0.28	99.50 ± 0.27	99.48 ± 0.28
	2	99.63 ± 0.31	99.64 ± 0.31	99.63 ± 0.31	99.63 ± 0.31
	3	99.70 ± 0.31	99.69 ± 0.32	99.71 ± 0.31	99.70 ± 0.32
	4	99.97 ± 0.10	99.97 ± 0.09	99.96 ± 0.12	99.97 ± 0.10

In addition, based on the results of ten experiments, we plot box plots of different methods in the four different SNR experiments under $d = 4/5$ imbalanced condition, as shown in Figure 5. In the box plot, each scatter represents the accuracy result for

each experiment, and the horizontal line in the middle of the box represents the median. From Figure 5, for the four SNRs, the median of the proposed method is mostly higher than the comparison method, indicating that the average level of the proposed method is relatively high. In addition, compared to other methods, the box length of the proposed method is the shortest in most SNR experiments, indicating that the ten experimental results of the proposed method are relatively concentrated and the standard deviation is relatively small. The box plot of the proposed method has a median of 100% in the SNR = 4 experiment, and its box length is the shortest, which is a line, indicating that the proposed method has excellent diagnostic performance and high stability for imbalanced class samples.

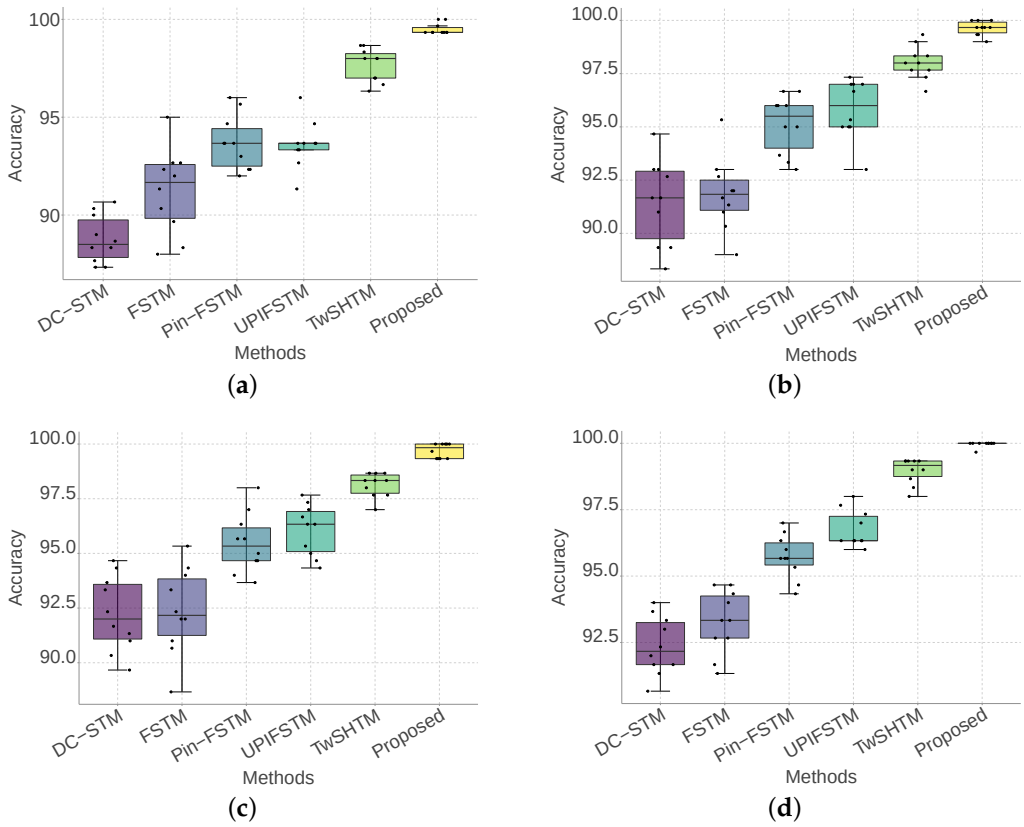


Figure 5. Box plots of different methods in HUST bearing dataset under four SNRs. (**a**) SNR = 1. (**b**) SNR = 2. (**c**) SNR = 3. (**d**) SNR = 4.

Next, to compare the predictive performance of different methods on positive and negative class samples in detail, we calculate the confusion matrice of each method in the SNR = 1 experiment. The results as shown in Figure 6. It should be noted that the results in Figure 6 are of one of the ten repeated experiments. From Figure 6, it can be seen that DC-STM, FSTM, TwSHTM, and the proposed method can correctly identify the first and second types of fault samples. However, DC-STM and FSTM cannot effectively identify the third and fourth types of fault samples, especially prone to misidentifying the fourth type of fault samples as the first and second types. For the second type of fault samples, Pin-FSTM has the worst recognition results. For the third type of fault samples, TwSHTM and the proposed method have better recognition performance than other comparison methods; specifically, the proposed method can correctly identify the third type of fault samples. In addition, for the fourth type of fault samples, the proposed method also achieved the highest recognition rate. In summary, the proposed method has the best overall recognition performance for the four different types of fault samples.

Finally, to compare the performance of various models more clearly and intuitively, in the class imbalance scenario of $d = 4/5$, we conduct Nimenyi tests on the accuracy, precision, recall, F-score indicators of the models in the four different SNR experiments. The confidence level is set to 0.05, and the results are shown in Figure 7. In Figure 7, the center of the different colored line segments represents the average ranking value of the models, and the left represents the higher ranking. The horizontal line segment centered on a circle represents the size of the critical value range, and the less overlap between different line segments, the more significant difference in model performance. From Figure 7, the proposed model is always closest to the left, indicating that the proposed model can achieve superior performance in different SNR experiments, although there is a certain degree of overlap between different line segments.

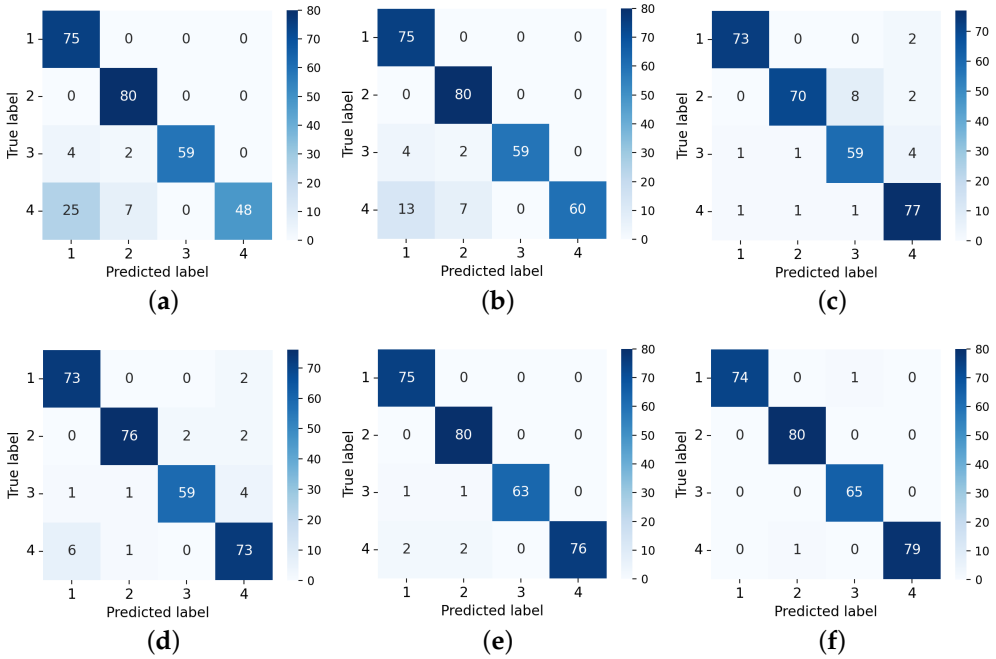


Figure 6. Confusion matrices of different methods in HUST bearing dataset. (**a**) DC-STM. (**b**) FSTM. (**c**) Pin-FSTM. (**d**) UPIFSTM. (**e**) TwSHTM. (**f**) Proposed.

5.2. Case Study 2

5.2.1. Dataset Description and Settings

In this case, we use the KAIST bearing dataset from the Korea Advanced Institute of Science and Technology [37]. The dataset includes vibration, acoustic, temperature, and drives current data of rotating machines under varying operating conditions. The dataset was acquired using four ceramic shear ICP-based accelerometers, one microphone, two thermocouples, and three current transformers (CTs) based on the international organization for standardization (ISO) standard. In this case, we use vibration data measured by four accelerometers (PCB352C34) installed in the x and y directions on two bearing seats (A, B). The rotating machinery test bench consists of a three-phase asynchronous motor, torque meter, gearbox, bearing housing A, bearing housing B, rotor, and hysteresis brake, as shown in Figure 8. In this case, we use four different health status bearings, including 0.3 mm inner race fault (label 1), 0.3 mm outer race fault (label 2), 0.1 mm shaft misalignment fault (label 3), and 583 mg rotor unbalance fault (label 4), as illustrated in Figure 9. Three different torque load conditions (0 Nm, 2 Nm, and 4 Nm) are set in the experiments. The vibration data used in this case are collected at a sampling frequency of 25.6 kHz.

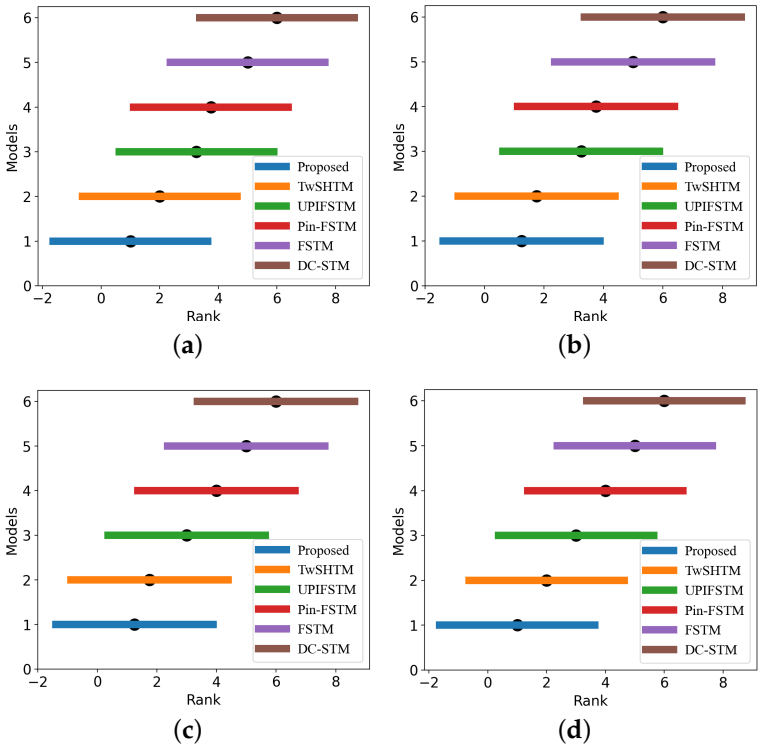


Figure 7. Nemenyi test for model performance in HUST bearing dataset under different indexes. (**a**) Accuracy. (**b**) Precision. (**c**) Recall. (**d**) F-score.

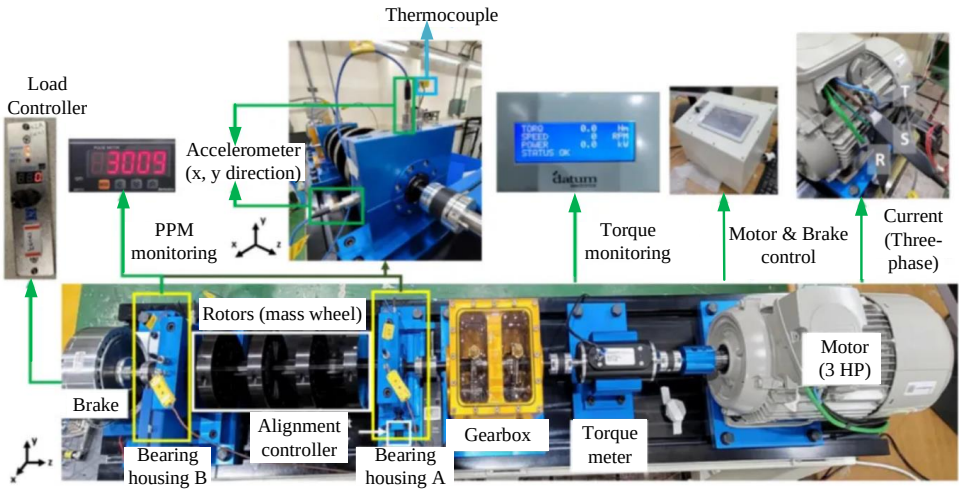


Figure 8. Layout of KAIST bearing test stand and its components.

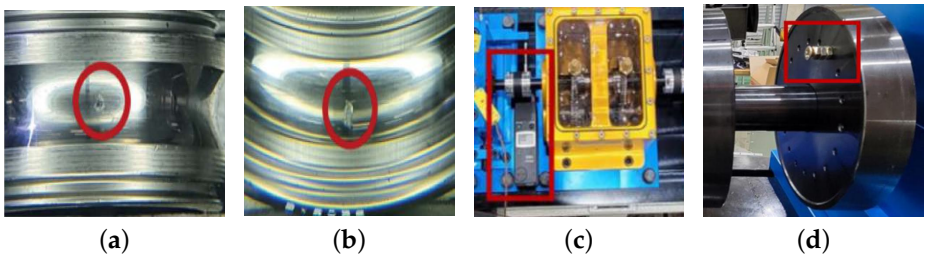


Figure 9. Photographs of the KAIST fault bearings. (**a**) Inner race fault. (**b**) Outer race fault. (**c**) Shaft misalignment fault. (**d**) Rotor unbalance fault.

Based on the tensor sample reconstruction method proposed in Section 4, 230 tensor samples are obtained for four health states, of which 150 are used for training. In the testing

phase, we select 300 test samples, including 75 inner race fault samples, 80 outer race fault samples, 65 shaft misalignment fault samples, and 80 rotor unbalance fault samples. A detailed description of the analyzed KAIST dataset is shown in Table 4. Then, we set the imbalance ratio r = 4/5 to verify the classification performance of the proposed method in an imbalanced data scenario. To achieve multiclassification, we adopt a one-against-one strategy and conduct six group binary classification experiments. When r = 4/5, there are 120 negative class training samples and 150 positive class training samples in each group experiment. Next, to simulate nonstationary signals, for each sensor, we connect the signals generated under different loads during the same time period, that is, combining 82,944 signals under a 0 Nm load, 81920 signals under a 2 Nm load, and 81,920 signals under a 4 Nm load to obtain 246,784 variable load signals. Similarly, to simulate a noisy environment, we sequentially add Gaussian white noise with four different signal-to-noise ratios (SNRs) (i.e., SNR = 1, SNR = 2, SNR = 3, SNR = 4) to each merged variable load sensor signal. The window length is set to 2048, and the sliding step size is set to 1024. Then, we use the method in Section 4 to construct feature tensors with a size of $64 \times 64 \times 4$. The selection and setting of parameters, comparison methods, and performance evaluation indicators are consistent with those in the HUST bearing experiment.

Table 4. Detailed description of the analyzed KAIST dataset.

Label	Status Description	Train Number	Test Number
1	Inner race fault	150	75
2	Outer race fault	150	80
3	Shaft misalignment fault	150	65
4	Rotor unbalance fault	150	80

5.2.2. Experiment Result and Analysis

Similarly, based on the KAIST dataset, the FST time–frequency images of vibration signal on B bearing seat x-direction under four different health states are shown in Figure 10. It can be observed from Figure 10 that there are significant differences in the time–frequency images of different faults. Therefore, we will also use FST to obtain time–frequency features and reconstruct them into feature tensors to achieve effective fault identification.

Then, we simulate four different SNR experiments on the KAIST bearing dataset, repeating each experiment 10 times, and obtain the average accuracy, precision, recall, and F-score results of different methods, as shown in Table 5. From Table 5, we can see that as the SNR gradually increases, the noise gradually decreases and the diagnostic performance of all models improves. For example, as the SNR increases, the accuracy indicators of the proposed method are 98.63%, 99.03%, 99.10%, and 99.50%, respectively, which are significantly higher than other comparative methods. The precision indicators of DC-STM, FSTM, Pin-FSTM, and UPIFSTM are higher than those of accuracy, recall, and F-score indicators, while the precision indicator of TwSHTM and the proposed method are slightly lower than other indicators. For SNR = 1, the F-score index values of different methods from high to low are as follows: the proposed method (98.58%), TwSHTM (97.58%), UPIFSTM (94.62%), Pin-FSTM (93.39%), FSTM (93.14%), DC-STM (92.60%), and other indicators also have the same trend, indicating that the proposed method still has excellent diagnostic performance, even with strong noise interference and class sample imbalance. In addition, under the same experimental setup, the proposed method has the lowest standard deviation, especially in the SNR = 3 experiment, where the standard deviations of all indicators reach their lowest values: accuracy (0.26%), precision (0.25%), recall (0.30%), F-score (0.28%), which also indicates that the proposed method has good stability.

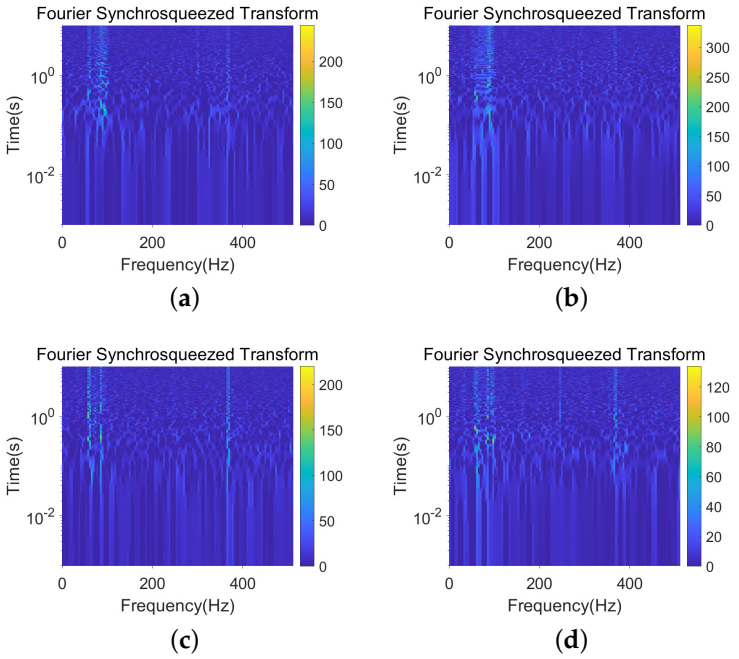


Figure 10. FST images of vibration signal on B bearing seat x-direction under four different health states in the KAIST bearing dataset. (**a**) Inner race fault. (**b**) Outer race fault. (**c**) Shaft misalignment fault. (**d**) Rotor unbalance fault.

Table 5. Average results (%) of different methods on the KAIST dataset.

Model	SNR	Accuracy	Precision	Recall	F-Score
DC-STM	1	92.87 ± 1.46	93.86 ± 0.96	93.31 ± 1.37	92.60 ± 1.54
	2	93.20 ± 1.38	94.07 ± 0.90	93.63 ± 1.29	92.95 ± 1.45
	3	93.37 ± 1.00	94.16 ± 0.68	93.78 ± 0.94	93.13 ± 1.05
	4	93.83 ± 0.82	94.48 ± 0.56	94.22 ± 0.77	93.61 ± 0.86
FSTM	1	93.40 ± 2.11	94.26 ± 1.39	93.80 ± 1.97	93.14 ± 2.23
	2	93.63 ± 2.40	94.45 ± 1.59	94.02 ± 2.24	93.38 ± 2.54
	3	94.33 ± 4.48	95.32 ± 3.12	94.67 ± 4.18	94.05 ± 4.74
	4	94.47 ± 2.84	95.04 ± 2.03	94.78 ± 2.68	94.24 ± 3.00
Pin-FSTM	1	94.20 ± 4.36	95.89 ± 2.71	93.31 ± 5.03	93.39 ± 5.26
	2	94.37 ± 2.39	95.74 ± 1.42	93.51 ± 2.76	93.80 ± 2.90
	3	94.67 ± 2.71	95.98 ± 1.73	93.86 ± 3.12	94.14 ± 3.14
	4	95.13 ± 3.10	96.06 ± 2.29	94.74 ± 3.23	94.79 ± 3.34
UPIFSTM	1	94.80 ± 0.54	95.17 ± 0.41	95.13 ± 0.51	94.62 ± 0.56
	2	94.97 ± 1.18	95.33 ± 0.92	95.28 ± 1.10	94.79 ± 1.22
	3	95.13 ± 0.91	95.28 ± 0.75	95.32 ± 0.83	94.97 ± 0.94
	4	95.63 ± 1.09	95.71 ± 0.99	95.79 ± 1.06	95.51 ± 1.12
TwSHTM	1	97.67 ± 0.94	97.59 ± 0.92	97.79 ± 0.90	97.58 ± 0.97
	2	98.20 ± 0.90	98.12 ± 0.86	98.31 ± 0.84	98.13 ± 0.93
	3	98.47 ± 0.85	98.38 ± 0.87	98.56 ± 0.80	98.41 ± 0.88
	4	99.13 ± 0.81	99.07 ± 0.84	99.19 ± 0.76	99.10 ± 0.84
Proposed	1	98.63 ± 0.48	98.53 ± 0.50	98.66 ± 0.49	98.58 ± 0.50
	2	99.03 ± 0.38	98.59 ± 0.41	99.07 ± 0.37	98.99 ± 0.39
	3	99.10 ± 0.26	99.05 ± 0.25	99.10 ± 0.30	99.06 ± 0.28
	4	99.50 ± 0.31	99.47 ± 0.34	99.50 ± 0.30	99.48 ± 0.32

In addition, based on the accuracy indicator obtained from 10 independent repeated experiments, we plot box plots of different methods under the four SNRs experiment,

as shown in Figure 11. For the SNR = 1 experiment, the box plot lengths of Pin-FSTM, FSTM, and DC-STM are relatively long, while the box plot lengths of TwSHTM, UPIFSTM, and the proposed model are relatively short, indicating that the results of the Pin-FSTM, FSTM, and DC-STM models are more dispersed, while the results of TwSHTM, UPIFSTM, and the proposed model are more concentrated, which is consistent with the standard deviation results of 4.36%, 2.11%, 1.46%, 0.94%, 0.54%, and 0.48% of the corresponding models in Table 5. In four different SNR experiments, the median of the proposed model is higher than other comparison methods, and the box length is the shortest, indicating that the average level of the 10 independent repeated experiments of the proposed model is relatively high and the results are relatively concentrated.

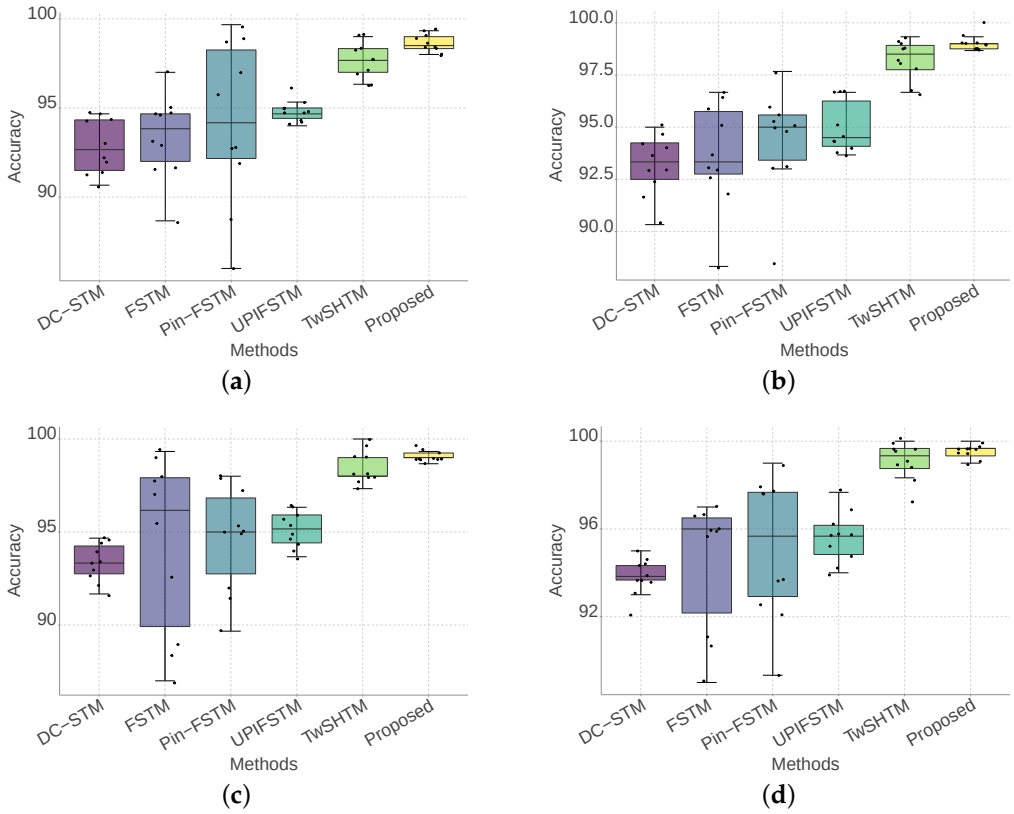


Figure 11. Box plots of different methods in KAIST bearing dataset under four SNRs. (**a**) SNR = 1. (**b**) SNR = 2. (**c**) SNR = 3. (**d**) SNR = 4.

Next, for the KAIST dataset, we also calculate the confusion matrix for each method in the four SNR experiments, as shown in Figure 12. From Figure 12, all methods can correctly identify the first and second types of samples. For the third type of sample, Pin-FSTM is prone to misidentifying it as the fourth type of sample, resulting in the worst recognition performance. Other methods can achieve correct recognition of the third type of sample. For the fourth type of sample, DC-STM, FSTM, UPIFSTM, and TwSHTM have poor diagnostic performance and are prone to misidentifying it as the third type of sample, while Pin-FSTM and the proposed method have good diagnostic performance. Overall, the proposed method can achieve global optimal diagnosis for four different types of samples. Finally, we conduct Nemenyi tests on the performance of different models under the accuracy, precision, recall, F-score indexes, and the results are shown in Figure 13. The confidence level is set to 0.05. From Figure 13, under the four indexes, the proposed model is always the leftmost, indicating the highest ranking and outstanding performance of the proposed model.

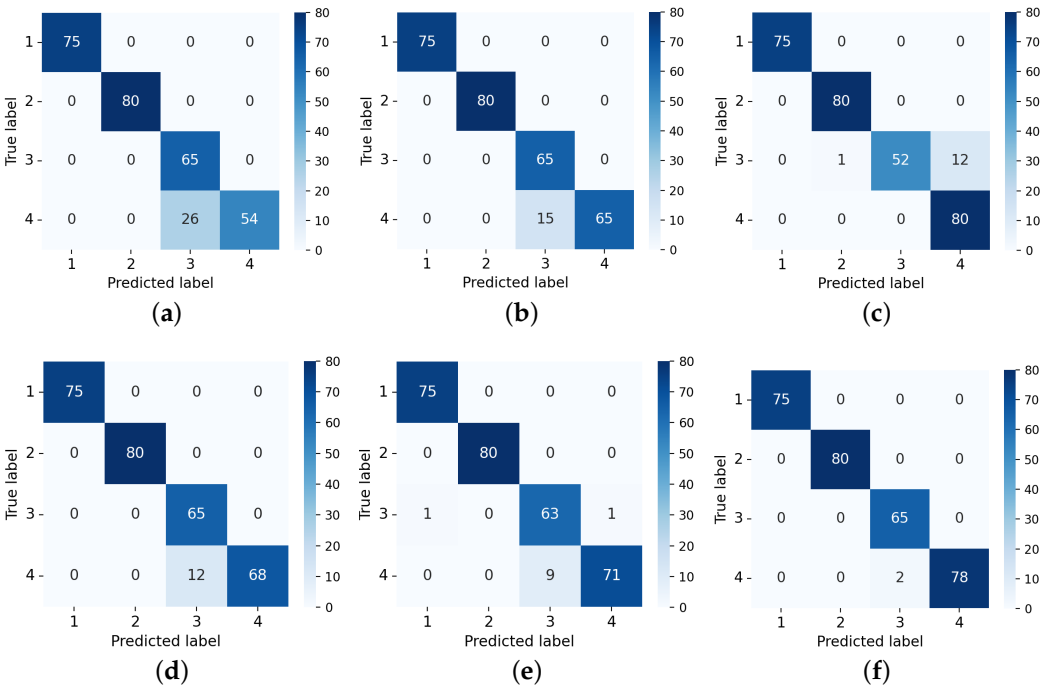


Figure 12. Confusion matrices of different methods in KAIST bearing dataset. (**a**) DC-STM. (**b**) FSTM. (**c**) Pin-FSTM. (**d**) UPIFSTM. (**e**) TwSHTM. (**f**) Proposed.

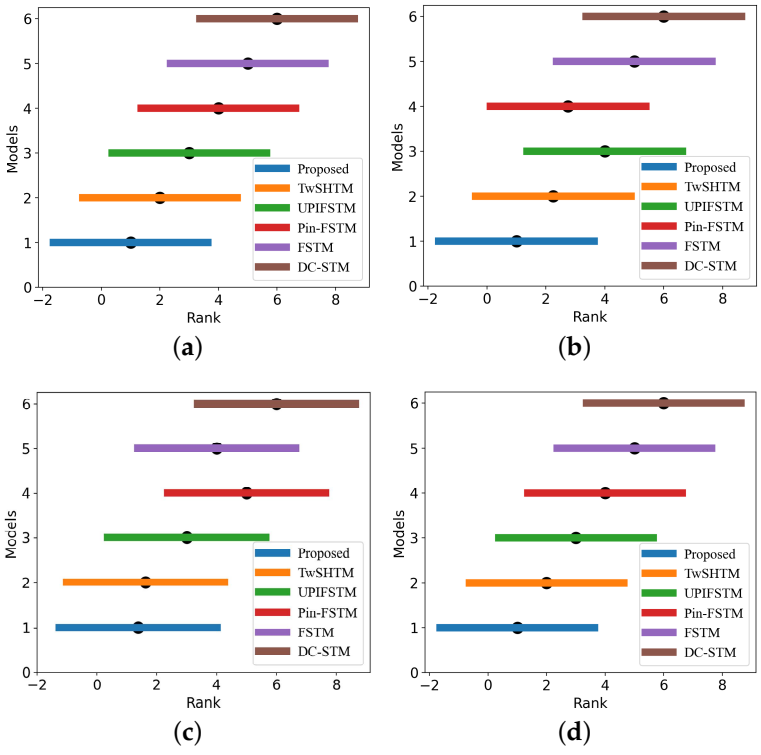


Figure 13. Nemenyi test for models performance in KAIST bearing dataset under different indexes. (**a**) Accuracy. (**b**) Precision. (**c**) Recall. (**d**) F-score.

The main differences between the proposed model and the traditional models are shown in Table 6. It is not difficult to find that the proposed model has several outstanding advantages. Firstly, assigning penalty factors to different types of samples can effectively handle imbalanced data. Secondly, in terms of computational efficiency, designing two nonparallel hyperplanes and solving a system of linear equations significantly

improves computational efficiency. Thirdly, in dealing with noise and outlier interference, the designed intuitionistic fuzzy membership score assignment scheme and neighborhood prior information weighting scheme can better distinguish the contributions of different samples. Fourthly, in terms of preserving tensor structure, Tucker decomposition can preserve more tensor structure information.

Table 6. Main differences between the proposed model and traditional models.

Models	DC-STM [25]	FSTM [28]	Pin-FSTM [29]	UPIFSTM [32]	TwSHTM [33]	IFW-LSTSHTM Proposed
Input order	*N*-order	*N*-order	2-order	3-order	*N*-order	*N*-order
Imbalance problem	√	×	×	×	×	√
Number of planes	One	One	One	One	Two	Two
Computing efficiency	Low ↓	Low ↓	Low ↓	Low ↓	High ↑	High ↑
Noise insensitivity	×	√	√	√	×	√
Intuitionistic fuzzy	×	×	×	√	×	√
Global–local information	×	×	×	×	×	√
Structural information	×	√	×	×	×	√
Decomposition method	CP	Tucker	×	×	CP	Tucker
Neighborhood information	×	×	×	×	×	√

6. Conclusions

Rolling bearings are prone to complex scenarios such as noise interference, variable operating conditions, and class imbalance. To improve the fault diagnosis performance in such scenarios, this paper reconstructs multisensor monitoring signals as feature tensors through FST and implements nonlinear learning through tensor kernel functions with Tucker decomposition. We design an intuitionistic fuzzy membership score assignment scheme and a sample-weighting scheme based on local neighborhood prior information and penalty factors, and we developing an intuitionistic fuzzy weighted least squares twin support higher-order tensor machine model for ship rolling bearing fault diagnosis. Finally, we conduct fault diagnosis experiments on the HUST bearing dataset and KAIST bearing dataset under variable operating conditions. For each dataset, we simulate fault classification experiments under four different SNR groups with class imbalance. The experimental results show that the method achieves the best average level in the four evaluation indicators of average accuracy, precision, recall, and F-score, with relatively small standard deviations.

Future work will explore lightweight fuzzy tensor models to further extend their application in real-time fault diagnosis, and through hierarchical processing architecture, model lightweighting techniques, and edge sensor co-design, we will adapt modern surveillance systems with integrated MEMS sensors and test the performance on devices to verify their industrialization feasibility.

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

Funding: This work was funded by Shanghai Science and Program of Shanghai Academic/Technology Research Leader, China, grant number 23XD1431000. Fund recipient Wang Shengzheng. The APC was funded by Wang Shengzheng.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available in [36,37].

Conflicts of Interest: Shengli Dong is employed by Shanghai Ship and Shipping Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Chi, H.; Chen, H. Research on Rolling Bearing Fault Diagnosis Method Based on MPE and Multi-Strategy Improved Sparrow Search Algorithm Under Local Mean Decomposition. *Machines* **2025**, *13*, 336 . [CrossRef]
2. Yuan, B.; Lei, L.; Chen, S. Optimized Variational Mode Decomposition and Convolutional Block Attention Module-Enhanced Hybrid Network for Bearing Fault Diagnosis. *Machines* **2025**, *13*, 320. [CrossRef]
3. Wang, S.; Qiu, S.; Sun, Z.; Hsieh, T.H.; Qian, F.; Xiao, Y. Single-View 3D Object Perception Based on Vessel Generative Adversarial Network for Autonomous Ships. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 9238–9252. [CrossRef]
4. Yu, Y.; Karimi, H.R.; Gelman, L.; Cetin, A.E. MSIFT: A novel end-to-end mechanical fault diagnosis framework under limited & imbalanced data using multi-source information fusion. *Expert Syst. Appl.* **2025**, *274*, 126947.
5. Zhang, Y.F.; Han, B.; Han, M. A Novel Distributed Data-Driven Strategy for Fault Detection of Multi-Source Dynamic Systems. *IEEE Trans. Circuits Syst. II Express Briefs* **2022**, *69*, 4379–4383. [CrossRef]
6. Zhu, J.; Zhu, L. Fault Diagnosis for Imbalanced Datasets Based on Deep Convolution Fuzzy System. *Machines* **2025**, *13*, 326. [CrossRef]
7. Huang, Y.; Huang, W.; Hu, X.; Liu, Z.; Huo, J. UDDGN: Domain-Independent Compact Boundary Learning Method for Universal Diagnosis Domain Generation. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3531220. [CrossRef]
8. Shen, C.; Wang, X.; Wang, D.; Li, Y.; Zhu, J.; Gong, M. Dynamic Joint Distribution Alignment Network for Bearing Fault Diagnosis Under Variable Working Conditions. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3510813. [CrossRef]
9. Liu, M.R.; Sun, T.; Sun, X.M. Brain-Inspired Spike Echo State Network Dynamics for Aero-Engine Intelligent Fault Prediction. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3516313. [CrossRef]
10. Tang, Y.; Zhang, C.; Wu, J.; Xie, Y.; Shen, W.; Wu, J. Deep Learning-Based Bearing Fault Diagnosis Using a Trusted Multiscale Quadratic Attention-Embedded Convolutional Neural Network. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3513215. [CrossRef]
11. Mao, Z.; Wang, H.; Jiang, B.; Xu, J.; Guo, H. Graph Convolutional Neural Network for Intelligent Fault Diagnosis of Machines via Knowledge Graph. *IEEE Trans. Ind. Inform.* **2024**, *20*, 7862–7870. [CrossRef]
12. Yu, Y.; He, Y.; Karimi, H.R.; Gelman, L.; Cetin, A.E. A two-stage importance-aware subgraph convolutional network based on multi-source sensors for cross-domain fault diagnosis. *Neural Netw.* **2024**, *179*, 106518. [CrossRef]
13. Liao, X.; Wang, D.; Qiu, S.; Ming, X. ReDBN: An Interpretable Deep Belief Network for Fan Fault Diagnosis in Iron and Steel Production Lines. *IEEE/ASME Trans. Mechatronics* **2024**, 1–12. [CrossRef]
14. Qian, Q.; Zhang, B.; Li, C.; Mao, Y.; Qin, Y. Federated transfer learning for machinery fault diagnosis: A comprehensive review of technique and application. *Mech. Syst. Signal Process.* **2025**, *223*, 111837. [CrossRef]
15. Li, Y.; Xu, X.; Hu, L.; Sun, K.; Han, M. A centroid contrastive multi-source domain adaptation method for fault diagnosis with category shift. *Measurement* **2025**, *248*, 116801. [CrossRef]
16. Rezvani, S.; Wang, X.; Pourpanah, F. Intuitionistic Fuzzy Twin Support Vector Machines. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 2140–2151. [CrossRef]
17. Sun, B.; Liu, X. Significance Support Vector Machine for High-Speed Train Bearing Fault Diagnosis. *IEEE Sens. J.* **2023**, *23*, 4638–4646. [CrossRef]
18. Han, T.; Zhang, L.; Yin, Z.; Tan, A.C. Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. *Measurement* **2021**, *177*, 109022. [CrossRef]
19. Rezvani, S.; Wang, X. Intuitionistic fuzzy twin support vector machines for imbalanced data. *Neurocomputing* **2022**, *507*, 16–25. [CrossRef]
20. Tanveer, M.; Tiwari, A.; Choudhary, R.; Jalan, S. Sparse pinball twin support vector machines. *Appl. Soft Comput.* **2019**, *78*, 164–175. [CrossRef]
21. He, H.; Tan, Y.; Zhang, W. A wavelet tensor fuzzy clustering scheme for multi-sensor human activity recognition. *Eng. Appl. Artif. Intell.* **2018**, *70*, 109–122. [CrossRef]
22. He, H.; Tan, Y.; Xing, J. Unsupervised classification of 12-lead ECG signals using wavelet tensor decomposition and two-dimensional Gaussian spectral clustering. *Knowl.-Based Syst.* **2019**, *163*, 392–403. [CrossRef]
23. Ge, M.; Lv, Y.; Ma, Y. Research on Multichannel Signals Fault Diagnosis for Bearing via Generalized Non-Convex Tensor Robust Principal Component Analysis and Tensor Singular Value Kurtosis. *IEEE Access* **2020**, *8*, 178425–178449. [CrossRef]

24. Hao, Z.; He, L.; Chen, B.; Yang, X. A Linear Support Higher-Order Tensor Machine for Classification. *IEEE Trans. Image Process.* **2013**, *22*, 2911–2920.
25. He, Z.; Shao, H.; Cheng, J.; Zhao, X.; Yang, Y. Support tensor machine with dynamic penalty factors and its application to the fault diagnosis of rotating machinery with unbalanced data. *Mech. Syst. Signal Process.* **2020**, *141*, 106441. [CrossRef]
26. Ma, Z.; Yang, L.T.; Zhang, Q. Support multimode tensor machine for multiple classification on industrial big data. *IEEE Trans. Ind. Inform.* **2020**, *17*, 3382–3390. [CrossRef]
27. Ding, W.; Zhou, T.; Huang, J.; Jiang, S.; Hou, T.; Lin, C.T. FMDNN: A Fuzzy-Guided Multigranular Deep Neural Network for Histopathological Image Classification. *IEEE Trans. Fuzzy Syst.* **2024**, *32*, 4709–4723. [CrossRef]
28. Sun, T.; Sun, X.M. New Results on Classification Modeling of Noisy Tensor Datasets: A Fuzzy Support Tensor Machine Dual Model. *IEEE Trans. Syst. Man, Cybern. Syst.* **2022**, *52*, 5188–5200. [CrossRef]
29. Yang, C.; Jia, M. Hierarchical multiscale permutation entropy-based feature extraction and fuzzy support tensor machine with pinball loss for bearing fault identification. *Mech. Syst. Signal Process.* **2021**, *149*, 107182. [CrossRef]
30. Rezvani, S.; Wang, X. Class imbalance learning using fuzzy ART and intuitionistic fuzzy twin support vector machines. *Inf. Sci.* **2021**, *578*, 659–682. [CrossRef]
31. Liang, Z.; Zhang, L. Intuitionistic fuzzy twin support vector machines with the insensitive pinball loss. *Appl. Soft Comput.* **2022**, *115*, 108231. [CrossRef]
32. Zhang, Y.; Han, B.; Han, M. Mechanical Fault Diagnosis with Noisy Multisource Signals via Unified Pinball Loss Intuitionistic Fuzzy Support Tensor Machine. *IEEE Trans. Ind. Inform.* **2024**, *20*, 62–72. [CrossRef]
33. Yang, C.; He, Q.; Li, Z.; Jia, M.; Gabbouj, M.; Peng, Z. Multichannel Fault Diagnosis of Planetary Gearbox via Difference-Average Symbol Transition Entropy and Twin Support Higher-Order Tensor Machine. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3507210. [CrossRef]
34. Gundewar, S.K.; Kane, P.V. Bearing fault diagnosis using time segmented Fourier synchrosqueezed transform images and convolution neural network. *Measurement* **2022**, *203*, 111855. [CrossRef]
35. Kolda, T.G.; Bader, B.W. Tensor Decompositions and Applications. *SIAM Rev.* **2009**, *51*, 455–500. [CrossRef]
36. Zhao, C.; Zio, E.; Shen, W. Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study. *Reliab. Eng. Syst. Saf.* **2024**, *245*, 109964. [CrossRef]
37. Jung, W.; Kim, S.H.; Yun, S.H.; Bae, J.; Park, Y.H. Vibration, acoustic, temperature, and motor current dataset of rotating machine under varying operating conditions for fault diagnosis. *Data Brief* **2023**, *48*, 109049. [CrossRef]

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