Measurement 256 (2025) 118236

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Residual feature-based semi-supervised anomaly detection for Gearbox under varying rotational speeds

Hongliang Song [a], Yi Sun [a], Hongli Gao [a], Zhichao You [b,*], Hao Xu [a]

[a] School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China
[b] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 610031, China

ARTICLE INFO

Keywords:
Anomaly detection
Semi-supervised learning
Time-varying rotational speed
Gearbox
Machine learning

ABSTRACT

Anomaly detection is of great significance for ensuring the operational safety of gearboxes and for boosting production efficiency. However, most existing anomaly detection algorithms presume constant operating speeds, which cannot adapt to actual industrial scenarios. The scarcity of anomaly samples also hampers the deployment of certain models in industrial settings. This study addresses these issues by proposing (i) a residual feature modelling approach, (ii) a new anomaly detection model, and (iii) a dedicated alarm mechanism. By modelling features strongly correlated with rotational speed, the proposed approach extracts residual features—a novel representation that effectively mitigates the effects of varying operational conditions. A new Adaptive Weighted Histogram-Based Outlier Score (AWHBOS) model is proposed. In this model, the baseline is constructed exclusively on health state data, with critical residual features being dynamically weighted. Our experiments show that this approach significantly improves overall anomaly detection performance. Furthermore, we propose a novel Tolerance-aware Queue-based Alarm Strategy to reduce false alarm rates and improve detection accuracy. Experimental results on two datasets (across multiple fault types) demonstrate that our proposed method achieves superior performance in gearbox anomaly detection under varying rotational speeds. It outperforms six mainstream semi-supervised models across key evaluation metrics.

1. Introduction

Gearboxes, as critical transmission components in modern mechanical systems, have a direct impact on the operational efficiency and safety of the entire machine [1]. Gearbox failures can result in substantial economic losses due to unplanned downtime. In real-world industrial settings, high-speed rotating machinery often operates under harsh, dynamic conditions, subjecting gear components to time-varying rotational speeds and fluctuating loads [2]. These non-stationary operating conditions make various defects and failures virtually unavoidable over time. As the frontline of equipment health management, anomaly detection is vital for identifying early signs of abnormal gearbox behaviour during operation. Early detection helps prevent catastrophic failures and lays the foundation for subsequent fault diagnosis and predictive maintenance. Therefore, advancing research in gearbox anomaly detection is critically important from both economic and safety perspectives [3].

In contrast to fault diagnosis [4], which typically relies on supervised learning methods requiring labelled fault data for model training, anomaly detection is more suitable for industrial scenarios where fault patterns are unknown, diverse, and complex. In many practical industrial environments, fault data are unavailable or extremely limited [5,6]. Simulating a comprehensive range of fault conditions within a limited time is also impractical. Given these challenges, semi-supervised or unsupervised data-driven approaches offer greater applicability for industrial anomaly monitoring [7]. In particular, developing effective anomaly detection models using only healthy data without labelled fault samples has emerged as a critical yet challenging research focus.

In the past, equipment anomaly detection models were primarily developed based on expert knowledge and physical laws [8]. With advances in sensor technology and data processing techniques, data-driven approaches that exploit historical monitoring data have begun to exhibit outstanding performance. These methods offer improved adaptability and generalizability in complex industrial environments. Data-driven anomaly detection techniques can be categorized into distance- or density-based, model-based, and statistical approaches. Distance- or density-based methods identify outliers by evaluating disparities in spatial proximity or local data density, such as Local Outlier Factor

* Corresponding author.
E-mail address: youzc@swjtu.edu.cn (Z. You).

https://doi.org/10.1016/j.measurement.2025.118236
Received 12 April 2025; Received in revised form 18 June 2025; Accepted 20 June 2025
Available online 21 June 2025

(LOF) [9,10]. Model-based methods learn expected system behaviour through predictive modelling and detect anomalies by measuring deviations between observed sensor data and the model's predictions. Examples include one-class classification algorithms, such as One-Class Support Vector Machine (OCSVM) [11–13] and tree ensemble methods like Isolation Forest (IForest) [14,15], as well as robust covariance estimators, like Minimum Covariance Determinant (MCD) [16]. Statistical methods [17,18] assume that data conform to a specific probability distribution and flag anomalies based on the degree of deviation from that distribution. Among statistical techniques, the Histogram-Based Outlier Score (HBOS) algorithm [19,20] has emerged as a widely used and computationally efficient approach. HBOS constructs histograms for each feature of the operational data independently, enabling the identification of low-density regions in each dimension. Owing to its scalability and high computational efficiency, HBOS has demonstrated strong applicability in real-time anomaly detection for mechanical systems.

However, these classical methods are predicated on the assumption of a static data distribution. When changes in operating conditions cause the distribution to shift significantly, detection performance typically degrades, and the false-alarm rate rises. Moreover, the independence assumption in HBOS means it does not capture inter-feature relationships and effectively treats all features as equally important. This limitation can reduce its ability to detect anomalies when only a few features (or particular combinations of features) are fault-sensitive.

With the rapid progress in deep learning, advanced neural network architectures have also been applied to anomaly detection for machine health monitoring. Recurrent neural networks [21], Deep Support Vector Data Description (DeepSVDD) [22], Variational Autoencoder (VAE) [23] and Transformers [24] have shown remarkable capability in capturing cyclical patterns and complex relationships in time-series data. Leveraging their strong temporal modeling and feature representation abilities, these models have been increasingly adopted in machine condition monitoring tasks [25,26]. Several studies [27,28] emphasize constructing and preserving health-related features by using deep networks to learn normal patterns from the data. However, deep learning approaches come with challenges: they typically entail high computational complexity and substantial hardware requirements. And they still rely on the fundamental assumption distributional consistency between training and deployment phases. In other words, these approaches generally presume that the data encountered during real-world operation follow the same or a sufficiently similar distribution as the training data.

However, in industrial environments characterized by non-stationary operating conditions, this assumption is frequently violated. The features extracted from the acquired signals exhibit pronounced non-stationarity and distributional shifts as the operating conditions change [29]. Complex working conditions lead to data distribution drift, which is a common problem in intelligent monitoring [30,31]. These shifts fundamentally challenge the core premise of data-driven models—that training and testing data follow similar distributions—and ultimately degrade model accuracy and robustness in practical anomaly detection tasks [32]. As shown in Fig. 1, under steady-state conditions, the statistical distributions of features corresponding to different health states are well-separated, indicating a certain degree of fault detection capability. However, under variable speed conditions these feature distributions overlap significantly, resulting in a significant decrease in fault distinguishability.

In recent years, many researchers have attempted to address the challenges of anomaly detection under variable operating conditions. Hu et al. [8] proposed a mutual information-based feature disentanglement network to improve anomaly detection under variable working scenarios on component- and part-level datasets. Robert George et al. [33] employed a Hidden Markov Model to extract condition-sensitive features and a Gaussian Mixture Model to capture fault-sensitive features, enabling effective fault detection across diverse operating conditions. Zhao et al. [2] proposed an adaptive inter-class convolutional neural network for intelligent gearbox fault diagnosis, which reduces distributional discrepancies across varying conditions and enhances model generalization. Xu et al. [34] developed a reconstruction-based autoencoder called the Cluster-Based Contrastive Learning Autoencoder, which incorporates cluster-guided contrastive learning to improve feature discriminability across different operating conditions and facilitate adaptive condition recognition. Li et al. [35] proposed a Knowledge Mapping-based Adversarial Domain Adaptation framework that integrates a discriminator and feature extractor to enable knowledge transfer from a labelled source domain to an unlabelled target domain, thereby supporting cross-condition diagnostics. Despite considerable advancements in data-driven gearbox anomaly detection under time-varying operating conditions, many existing approaches still impose substantial computational burdens or require extensive labelled fault data—requirements that result in high engineering costs. Consequently, the development of semi-supervised anomaly detection techniques for vibration signals collected at time-varying rotational speeds remains a research challenge of substantial practical importance.

One of the primary reasons anomaly detection methods have not yet been widely adopted in industrial environments is the high operational cost associated with downtime in mechanical systems. In many cases, the financial losses resulting from unscheduled shutdowns far exceed the cost of component replacement. Moreover, overly sensitive anomaly detection systems tend to trigger frequent false alarms, leading to system interruptions, higher maintenance costs, and disruptions to production schedules.

To address the aforementioned challenges, we propose a novel

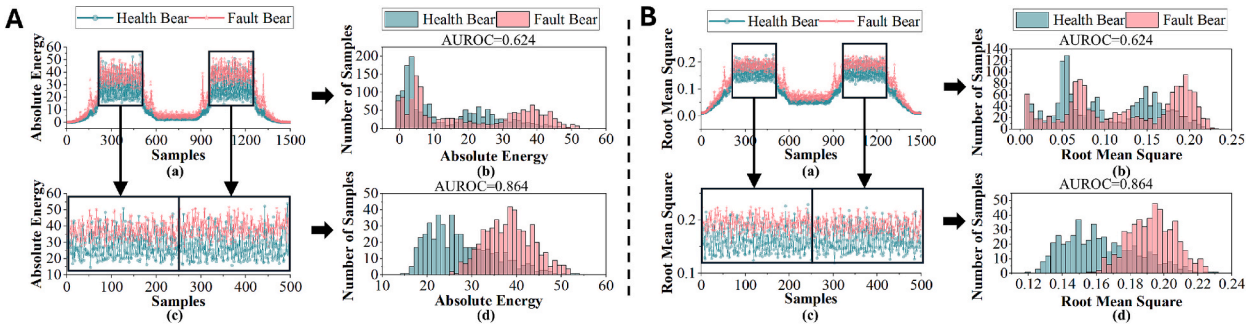


Fig. 1. Typical features are extracted segmentally over time from gearbox vibration signals. (A) Absolute Energy features. (B) Root Mean Square (RMS) features. Subplots (a) and (b) show the time-series plots and statistical histograms of the features under varying rotational speed conditions, respectively, while (c) and (d) correspond to the same types of plots under constant rotational speed conditions. The data used are obtained from the variable operating condition gearbox dataset provided by Tsinghua University. Area under receiver operating characteristic(AUROC) around 0.5 can be achieved by random guesses. A higher AUROC value indicates greater prediction power.

residual feature based semi-supervised anomaly detection method for gearboxes operating at varying rotational speeds. The main contributions of this study are summarized as follows:

a) Residual Feature Modelling: We introduce a residual feature modelling approach that transforms raw features affected by speed-variation into more stable residual representations. The method reshapes the feature distribution and improves robustness by removing speed-related trend components from the vibration features.
b) Adaptive Weighted Histogram-Based Outlier Score: We propose an improved HBOS algorithm with an adaptive weighting mechanism (AWM) to address the challenge of detecting gearbox anomalies under variable speeds using only health state data. The AWHBOS establishes an anomaly detection baseline solely from health data and dynamically increases the influence of the most fault-relevant features, thereby enabling the effective identification of multiple fault types without requiring any fault samples in training.
c) Tolerance-Aware Queue-Based Alarm Strategy (TQAS): For practical deployment where occasional anomalies can be tolerated, we design a queue-based alarm strategy to reduce false alarms. In TQAS, an alarm is not raised on a single outlier observation; instead, the system monitors a moving queue of recent anomaly scores and triggers an alarm only when the proportion of anomalies in the queue exceeds a set tolerance threshold.
d) Extensive Experimental Validation: We conduct comprehensive experiments on two publicly available variable speed gearbox datasets, covering multiple fault types. The results demonstrate that the proposed method achieves superior performance in anomaly detection under varying rotational speeds, outperforming six state-of-the-art semi-supervised models in key evaluation metrics. Furthermore, the results of the ablation studies confirm the individual effectiveness of the AWM and TQAS components.

The remainder of this paper is organized as follows. Section 2 introduces the relevant theoretical background and definitions. Section 3 provides a detailed description of the proposed AWHBOS model. Section 4 reports the experimental results on two publicly available gearbox datasets, including comparative evaluations and ablation studies. Finally, Section 5 concludes the paper.

2. Preliminary

2.1. Histogram-based outlier score

HBOS is an unsupervised anomaly detection algorithm well-suited for high-dimensional data due to its computational efficiency and scalability. HBOS operates under the assumption that feature dimensions are mutually independent. It estimates the probability density of each feature by constructing histograms independently for each dimension and then computes an outlier score for each data instance based on these individual probability estimates. The core concept of the HBOS algorithm can be summarized as follows:

a) Histogram Construction**:** For a dataset $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n\}$, where each data instance $\boldsymbol{y_i} \in \mathbb{R}^d$, each instance y_i contains d features, denoted as $\boldsymbol{y_i} = (y_{i1}, y_{i2}, \cdots, y_{id})$, HBOS constructs a histogram for each feature dimension individually. Each dimension is divided into B_j equally spaced intervals (bins), and for the j th feature, the feature space is partitioned into a sequence of bin boundaries.

$$\boldsymbol{E_j} = \left[e_j^{(0)}, e_j^{(1)}, \cdots, e_j^{(B)}\right] \tag{1}$$

where $e_j^{(b)}$ denotes the left boundary of the b th bin.

b) Density Estimation: The density estimation for each bin is given using the following formula:

$$f_j(b) = \frac{\left|\left\{\boldsymbol{y_i} \in Y | y_{ij} \in \left[e_j^{(b-1)}, e_j^{(b)}\right)\right\}\right|}{n \bullet \left(e_j^{(b)} - e_j^{(b-1)}\right)} + \mu \tag{2}$$

where n is the total number of samples in the dataset, y_{ij} denotes the value of the sample $\boldsymbol{y_i}$ in the j th dimension, and $\mu = 10^{-10}$ is a correction term.

c) Compute Univariate Anomaly Score: For the j th feature value $\boldsymbol{y_i}$ of a data instance y_{ij}, identify the corresponding bin $b_j\left(y_{ij}\right)$, and compute the univariate anomaly score.

$$HBOS_j\left(y_{ij}\right) = -\log\left(f_j\left(b_j\left(y_{ij}\right)\right)\right) \tag{3}$$

d) Overall Anomaly Score: **B**ased on the assumption of independence between dimensions, the overall anomaly score for each data instance is obtained by summing the univariate anomaly scores across all dimensions.

$$HBOS(\boldsymbol{y_i}) = -\sum_{j=1}^{d} \log\left(f_j\left(b_j\left(y_{ij}\right)\right)\right) \tag{4}$$

A higher anomaly score indicates that the instance is more likely to be an outlier. Given a predefined anomaly ratio parameter γ, the anomaly score threshold β can be determined as follows. Instances with scores exceeding β are classified as anomalies.

$$\beta = Percentile\left(\{HBOS(\boldsymbol{y_i})\}_{i=1}^{n}, 100 \bullet (1-\gamma)\right) \tag{5}$$

The HBOS algorithm is conceptually simple, computationally efficient, and highly scalable. It is well-suited for high-dimensional datasets and has demonstrated strong performance in real-time anomaly detection tasks.

3. Proposed method

3.1. Motivation

For equipment operating at a constant rotational speed, the collected vibration signal $x_{const}(t)$ during monitoring can be regarded as a superposition of multiple signal components. At time t, the vibration signal mainly consists of the following three types of components.

a) Intrinsic signal of the equipment $x_o(t)$**:** This reflects the inherent vibration patterns determined by the equipment's physical properties. It is typically related to the equipment's structural parameters and operational characteristics.
b) Health-related signal $x_h(t)$: This component contains characteristic features induced by faults such as wear, cracks, or looseness. It provides critical information for condition monitoring and fault diagnosis.
c) Multi-source noise signal $n(t)$: This includes environmental noise, sensor noise, and other external interferences.

$$x_{const}(t) = x_o(t) + x_h(t) + n(t) \tag{6}$$

However, for equipment operating under variable rotational speed conditions, an additional component is introduced:

d) Operating condition signal $x_c(t)$**:** This reflects the vibration response of the equipment under varying speed profiles.

Accordingly, the vibration signal collected under variable speed conditions, denoted as $x_{var}(t)$, can be expressed as:

$$x_{var}(t) = x_o(t) + x_h(t) + x_c(t) + n(t) \tag{7}$$

It is imperative to apply signal processing and feature extraction techniques to isolate health-related information from the composite signal (which comprises all four components) and to extract meaningful patterns from the data. Given that the intrinsic signal remains relatively

stable and multi-source noise is generally of low amplitude and limited interference, suppressing the influence of the operating condition signal $x_c(t)$ becomes crucial for accurate anomaly detection.

Under variable speed conditions, applying a feature extraction method $\mathscr{F}$ yields the extracted feature f_{var}, such that:

$$\mathscr{F}(\boldsymbol{x}(t)) = \mathscr{F}(x_o(t) + x_h(t) + x_c(t) + n(t)) = f_{var} \tag{8}$$

Should the feature f_{var} demonstrate an explicit dependency on the rotational speed, it implies that the feature may be partially governed by or functionally related to the speed variable.

$$f_{var} = g(\mathrm{s}) + \varepsilon \tag{9}$$

where $g(\mathrm{s})$ denotes a mathematical function with rotational speed s as the independent variable, capturing the portion of f_{var} that varies with speed. The term ε represents the residual, which accounts for the variation not explained by speed. This residual primarily reflects information about the equipment's health condition, other unmodeled factors and noise.

For long operation periods, feature vectors under variable operating conditions can be obtained by segment-wise feature extraction, resulting in a feature vector $\boldsymbol{f}_{var} \in \mathbb{R}^n$ and a corresponding speed vector $\mathbf{s} \in \mathbb{R}^n$. Although the exact form of $g(\mathbf{s})$ is challenging to determine explicitly, it can be approximated through a fitting process, yielding an estimated function $\hat{g}(\mathbf{s})$. The residual vector can then be expressed as follows:

$$\boldsymbol{\varepsilon} = \boldsymbol{f}_{var} - \hat{g}(\mathbf{s}) \tag{10}$$

The significance of this formulation lies in the following aspects:

a) Removal of Speed-Related Trends

The estimated function $\hat{g}(\mathbf{s})$ captures the systematic trend in the feature $\boldsymbol{f}_{var}$ caused by variations in rotational speed. By subtracting this trend, the subsequent analysis of the residual $\boldsymbol{\varepsilon}$ becomes less affected by changes in operating conditions. The residual $\boldsymbol{\varepsilon}$ primarily retains information related to the intrinsic characteristics of the equipment, potential fault signatures, or other noise components.

b) Construction of a Unified Dataset

Following this transformation, feature data collected under different rotation speeds can be regarded as originating from equivalent operating states, effectively mitigating the systematic bias introduced by varying rotational speeds. The transformation enables the construction of a unified health dataset, thereby simplifying data analysis and model development.

In summary, by applying the feature extraction method $\mathscr{F}$ and fitting the speed-related function $\hat{g}(\mathbf{s})$, the trend induced by rotational speed can be effectively removed.

The overall framework is illustrated in Fig. 2.

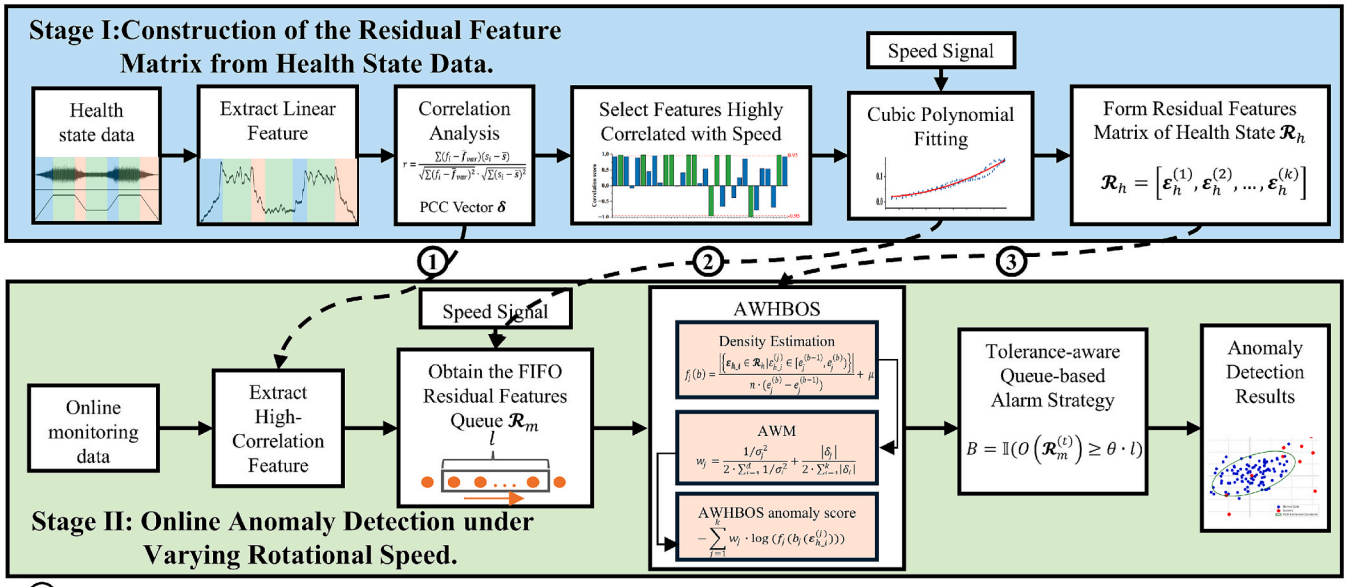
3.2. Construction of the residual feature matrix from health state data

Manual feature extraction is a fundamental technique in equipment anomaly detection and fault diagnosis. In our research on gear vibration analysis, many manually extracted features have demonstrated strong correlations with overall equipment health and specific fault types. In comparison with features learned through neural networks, manually engineered features have the key advantage of being independent of data-driven model training. Consequently, they are typically more computationally efficient and interpretable, qualities that render them particularly well suited to real-time, online anomaly detection. The time-domain features extracted in this study are shown in Table. 1.

In addition to time-domain features, nine frequency-domain features were also extracted, including Spectral Decrease, Spectral Skewness, Spectral Distance, Spectral Spread, Spectral Kurtosis, Spectral Entropy, Median Frequency, Spectral Centroid, and Power Bandwidth [36]. Due to the complexity of their mathematical formulations, detailed definitions are not presented in this paper.

After residual feature extraction, the Pearson Correlation Coefficient (PCC) [37] measures the linear relationship between each feature and the rotational speed. The calculation formula is as follows:

$$\delta = \sum (f_i - \frac{\overline{\boldsymbol{f}_{var}})(s_i - \bar{\mathbf{s}})}{\sqrt{\sum (f_i - \overline{\boldsymbol{f}_{var}})^2} \bullet \sqrt{\sum (s_i - \bar{\mathbf{s}})^2}} \tag{11}$$



(1) Features exhibiting a PCC greater than 0.95 with rotational speed are selected from the healthy state features.

(2) The residual features of the online monitoring data are computed based on the health-status cubic polynomial fitting function.

(3) The AWHBOS model is trained with the residual feature matrix derived from health state data.

Fig. 2. Model Construction Workflow.

Table 1
The time-domain features used in this study. x_i denotes the time-domain signal. N is the number of samples. f_s is the sampling frequency. p_i is the number of occurrences of each value, and k is the number of distinct values.

Time-Domain Features List			
Mean	$\frac{1}{N}\sum_{i=1}^{N} x_i$	Variance	$\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_x)^2$
Standard Deviation	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_x)^2}$	Skewness	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_x)^3}{\sigma_x^3}$
Kurtosis	$\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_x)^4}{\sigma_x^4}$	Root Mean Square	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}$
Mean Diff	$\frac{1}{N-1}\sum_{i=1}^{N-1}(x_{i+1}-x_i)$	Mean Absolute Deviation	$\frac{1}{N}\sum_{i=1}^{N}\lvert x_i-\mu_x\rvert$
Mean Absolute Diff	$\frac{1}{N-1}\sum_{i=1}^{N-1}\lvert x_{i+1}-x_i\rvert$	Average Power	$\frac{1}{N}\sum_{i=1}^{N}{x_i}^2$
Area Under the Curve	$\sum_{i=1}^{N-1}\frac{(x_{i+1}+x_i)}{2}\bullet\frac{1}{f_s}$	Absolute energy	$\sum_{i=1}^{N}{x_i}^2$
Max	$max(x)$	Min	$min(x)$
Peak to Peak Distance	$max(x)-min(x)$	Interquartile Range	Q_3-Q_1 Q_3 is the third quartile, Q_1 is the first quartile.
Median	$median(x)$	Median Diff	$median(x_{i+1}-x_i)$
Median Absolute Diff	$median(\lvert x_{i+1}-x_i\rvert)$	Median Absolute Deviation	$median(\lvert x_i-median(x)\rvert)$
Centroid	$\frac{\sum_{i=0}^{N-1}\frac{i}{f_s}\lvert x_i\rvert^2}{\sum_{n=0}^{N-1}\lvert x_i\rvert^2}$	Entropy	$-\frac{\sum_{i=1}^{k}p_i\log_2(p_i)}{\log_2(N)}$

where f_i denotes the i th feature sample in $\boldsymbol{f_{var}}$, and s_i represents the corresponding rotational speed in $\mathbf{s}$. The terms $\overline{\boldsymbol{f_{var}}}$ and $\bar{\boldsymbol{s}}$ are the mean values of $\boldsymbol{f_{var}}$ and $\mathbf{s}$, respectively.

Among these features, those with a PCC greater than 0.95 are selected. The correlation coefficients of these k selected features are then used to construct the PCC Vector, denoted as $\boldsymbol{\delta}$.

After comparing multiple functional forms, we modelled $\hat{g}(s)$ using the least squares method. The results demonstrate that the cubic polynomial balances fitting accuracy, stability, and computational efficiency. Accordingly, the cubic polynomial model is adopted to characterize the relationship between rotational speed and feature behaviour to enhance the accuracy and robustness of subsequent analysis.

As shown in Eq. (9), we can obtain the residual feature vector under health state, denoted as $\boldsymbol{\varepsilon_h}$, and further construct a residual feature matrix for the health state, $\mathscr{R}_h \in \mathbb{R}^{q\times k}$, where q is the number of samples and k is the number of selected features.

$$\mathscr{R}_h=\left[\boldsymbol{\varepsilon}_h^{(1)},\boldsymbol{\varepsilon}_h^{(2)},\cdots,\boldsymbol{\varepsilon}_h^{(k)}\right]=\begin{bmatrix}\varepsilon_{h\_1}^{(1)},\varepsilon_{h\_1}^{(2)},\cdots,\varepsilon_{h\_1}^{(k)}\\ \varepsilon_{h\_2}^{(1)},\varepsilon_{h\_2}^{(2)},\cdots,\varepsilon_{h\_2}^{(k)}\\ \cdots\\ \varepsilon_{h\_q}^{(1)},\varepsilon_{h\_q}^{(2)},\cdots,\varepsilon_{h\_q}^{(k)}\end{bmatrix} \tag{12}$$

where $\boldsymbol{\varepsilon}_h^{(k)}$ represents the k th feature vector extracted under the health state.

3.3. AWHBOS algorithm

The features identified in Section 3.2 as having high linear correlation with rotational speed are extracted from the online monitoring signals. Using the fitted cubic speed–feature functions, the residual features $\boldsymbol{\varepsilon_m}$ of the monitoring signal are obtained. During this process, a first-in-first-out residual feature queue $\mathscr{R}_m$ is constructed, with its length l determined by the sampling frequency and the desired sensitivity of monitoring. Specifically, at time t, $\mathscr{R}_m^{(t)} \in \mathbb{R}^{l\times k}$ consists of the residual features from the current and the previous l-1 samples.

$$\mathscr{R}_m^{(t)}=\begin{bmatrix}\varepsilon_{t-l}^{(1)},\varepsilon_{t-l}^{(2)},\cdots,\varepsilon_{t-l}^{(k)}\\ \varepsilon_{t-l+1}^{(1)},\varepsilon_{t-l+1}^{(2)},\cdots,\varepsilon_{t-l+1}^{(k)}\\ \cdots\\ \varepsilon_{t}^{(1)},\varepsilon_{t}^{(2)},\cdots,\varepsilon_{t}^{(k)}\end{bmatrix} \tag{13}$$

According to Eq. (2), the density estimation for each residual feature in each bin b can be computed using the following formula.

$$f_j(b)=\frac{\left|\left\{\boldsymbol{\varepsilon_{h\_i}}\in\mathscr{R}_h|\varepsilon_{h\_i}^{(j)}\in\left[e_j^{(b-1)},e_j^{(b)}\right)\right\}\right|}{n\bullet\left(e_j^{(b)}-e_j^{(b-1)}\right)}+\mu \tag{14}$$

where $\boldsymbol{\varepsilon_{h\_i}}$ denotes the residual feature vector of the i th sample in $\mathscr{R}_h$.

To improve the performance of the HBOS model, we propose a variant integrated with an adaptive weighting mechanism. Specifically, each residual feature is assigned a weight $\boldsymbol{w}$ based on two factors: the inverse of its variance within the health state matrix $\mathscr{R}_h$, and its corresponding value in the PCC Vector $\boldsymbol{\delta}$. The adaptive weight for the j th feature is computed as follows:

$$w_j=\frac{1/\sigma_j^2}{2\bullet\sum_{i=1}^{d}1/\sigma_i^2}+\frac{|\delta_j|}{2\bullet\sum_{i=1}^{k}|\delta_i|} \tag{15}$$

where σ_j^2 denotes the variance of the j th feature, δ_j represents the correlation score of the j th feature from the Pearson correlation vector, k is the total number of feature dimensions.

The weighting calculation comprises two components: variance-based weights and correlation-based weights. The variance-based weights emphasize residual features with minor variance, as such features are typically more stable and, therefore, more reliable for anomaly detection. The correlation-based weights accentuate features that exhibit stronger associations with rotational speed because those most affected by operating conditions produce residuals that more effectively suppress speed-induced effects. The absolute value of each correlation coefficient is employed in the weighting process to ensure that both positive and negative correlations contribute symmetrically to the model.

The anomaly score of $\boldsymbol{\varepsilon_{h\_i}}$ in AWHBOS is calculated as follows:

$$AWHBOS(\boldsymbol{\varepsilon_{h\_i}})=-\sum_{j=1}^{k}w_j\bullet\log\left(f_j\left(b_j\left(\boldsymbol{\varepsilon}_{h\_i}^{(j)}\right)\right)\right) \tag{16}$$

And anomaly score of $\boldsymbol{\varepsilon_{h\_i}}$ is then used to define the detection threshold.

$$\beta = Percentile\left(\{AWHBOS(\boldsymbol{\varepsilon_{h\_i}})\}_{i=1}^{n}, 100 \bullet (1-\gamma)\right) \tag{17}$$

For the online constructed queue $\mathscr{R}_m^{(t)}$, anomaly scores are calculated using Eq. (16). By comparing these scores against the anomaly detection threshold β, we can determine how many of the samples in $\mathscr{R}_m^{(t)}$ are identified as anomalies.

Although a number of researchers have recognized that the standard HBOS assigns equal importance to every feature and have therefore proposed various weighting strategies, these studies differ fundamentally from our work. For example, Nerijus et al. [38] employ a supervised feedback weighting scheme in which the statistical contribution of key attributes to the anomaly score is iteratively learned from labelled data. Binzat et al. [39] adjust the density of individual histogram bins within the same feature to realize intra-feature weighting. Both techniques either rely on labelled anomalies or modify only the distribution of a single attribute and thus cannot satisfy the fully unsupervised feature-weighting requirement of the present study.

3.4. Tolerance-aware queue-based alarm strategy

In practical applications of gearbox anomaly detection, some transmission systems can tolerate transient gearbox anomalies without immediately leading to severe failures. Overly sensitive alarm mechanisms may result in unnecessary interruptions or maintenance activities. Accordingly, confirming an anomaly through several consecutive detections before issuing an alarm is often a more reliable and effective strategy. Furthermore, since vibration signals are sampled at high rates, a suitable alarm strategy will not cause an unduly long delay in triggering alerts.

To address the practical requirements of online monitoring under such conditions, we propose a Tolerance-aware Queue-based Alarm Strategy. At each time step t, a First-In-First-Out queue $\mathscr{R}_m^{(t)}$ is constructed. The proportion of samples in the queue identified as anomalies is then calculated and compared against a predefined fault threshold to determine whether an alarm should be triggered. A fault is deemed to exist when the proportion of anomalous samples within the queue satisfies the following condition:

$$\mathscr{B} = \mathbb{I}\left(O\left(\mathscr{R}_m^{(t)}\right) \geq \theta \bullet l\right) \tag{18}$$

where $\mathscr{B}$ denotes the binary alarm output, $\mathscr{B} = 1$ indicates an alarm is triggered and $\mathscr{B} = 0$ indicates no alarm. $O\left(\mathscr{R}_m^{(t)}\right)$ represents the number of samples identified as anomalies within the queue $\mathscr{R}_m^{(t)}$. The parameter θ is a tolerance-aware sensitivity coefficient, typically recommended to be set between 0.4 and 0.7, which controls the system's sensitivity to anomaly density.

The core advantage of TQAS is its capacity to tolerate algorithmic misclassifications—and even brief or sporadic real-world equipment anomalies—by queuing detection outcomes and evaluating their proportion. This strategy significantly reduces false alarm rates while maintaining sensitivity to persistent or trending faults. Serving as a secondary validation layer for algorithm-generated anomaly detections, TQAS strikes an effective balance between alert latency and reliability, exhibiting strong practicality and robustness in continuous, high-sampling-rate industrial monitoring tasks.

4. Experiments

This section presents a series of experiments designed to evaluate the effectiveness of the proposed method. The experimental setup is organized as follows: Section 4.1 introduces an overview of experimental setup, including dataset descriptions, configuration of model parameters, comparison methods, and evaluation metrics. Section 4.2 presents the comparison results. Section 4.3 illustrates the implementation details and intermediate steps of the proposed modelling framework. Section 4.4 conducts ablation studies to validate the individual contributions of the AWM and the TQAS. Section 4.5 investigates the impact of the Tolerance-Aware Sensitivity Coefficient on anomaly detection performance.

4.1. Experimental setup

4.1.1. Dataset description

In this section, the performance of the proposed model under variable operating conditions is validated using two benchmark datasets: the Tsinghua Gearbox Dataset and the HUST Gearbox Dataset.

a) Tsinghua Gearbox Dataset

The Multi-Mode Fault Diagnosis Dataset of Gearbox under Variable Working Conditions, released by Tsinghua University in 2024, contains vibration, rotational speed, and torque signals [40]. The dataset covers various fault types, including multiple single gear faults and compound gear–bearing faults, each with different severity levels. The experimental setup is shown in Fig. 3. Fig. 4 provides a data visualization of the rotational speed, the vibration in the Z direction, and the maximum vibration amplitude. The key information for the Tsinghua Gearbox Dataset is shown in Table 2.

b) HUST Gearbox Dataset

The second dataset used in this study was collected by Huazhong University of Science and Technology for gearbox fault diagnosis [41]. Fault experiments were conducted using a Spectra-Quest mechanical fault simulator under two load conditions, denoted as L1 and L2. The experimental setup is shown in Fig. 5. During each test, the rotational speed increases linearly from 0 to 40 Hz within 2 s and then decreases back to 0 Hz at the same rate. For anomaly detection, the vertical vibration acceleration of the gearbox is again utilized. The dataset includes fault types such as tooth breakage and tooth chipping, as illustrated in Fig. 6. The key information for the HUST Gearbox Dataset is shown in Table 3.

4.1.2. Configuration of model parameters

For both datasets, vibration and status signals are preprocessed using a consistent procedure to ensure the generalizability of the proposed model. Specifically, the data are segmented using a sliding window of length 9126 samples with 50 % overlap. Feature extraction is then performed according to the procedures detailed in Section 3. Key model parameters of AWHBOS are set as follows: the number of histogram bins $B = 50$, the anomaly ratio parameter$\gamma = 0.08$. The detailed configuration of model parameters is shown in Table 4.

4.1.3. Comparison methods

To demonstrate the effectiveness of the proposed AWHBOS method, we compare it against several other anomaly detection models, all of which operate on the manually extracted features introduced in this study. It is important to emphasize that supervised methods are excluded from the comparison because we focus on unsupervised anomaly detection with health state data.

MCD [16]: a robust statistical method based on the Minimum Covariance Determinant, which identifies outliers by estimating the covariance matrix of normal data.

IForest [15]: a widely used machine learning approach that detects anomalies by constructing Isolation Forest on normal samples and isolating rare instances.

LOF [9]: a density-based method that detects anomalies by evaluating the local density deviation of each data point to its neighbours.

OCSVM [13]: A one-class classification model that separates normal

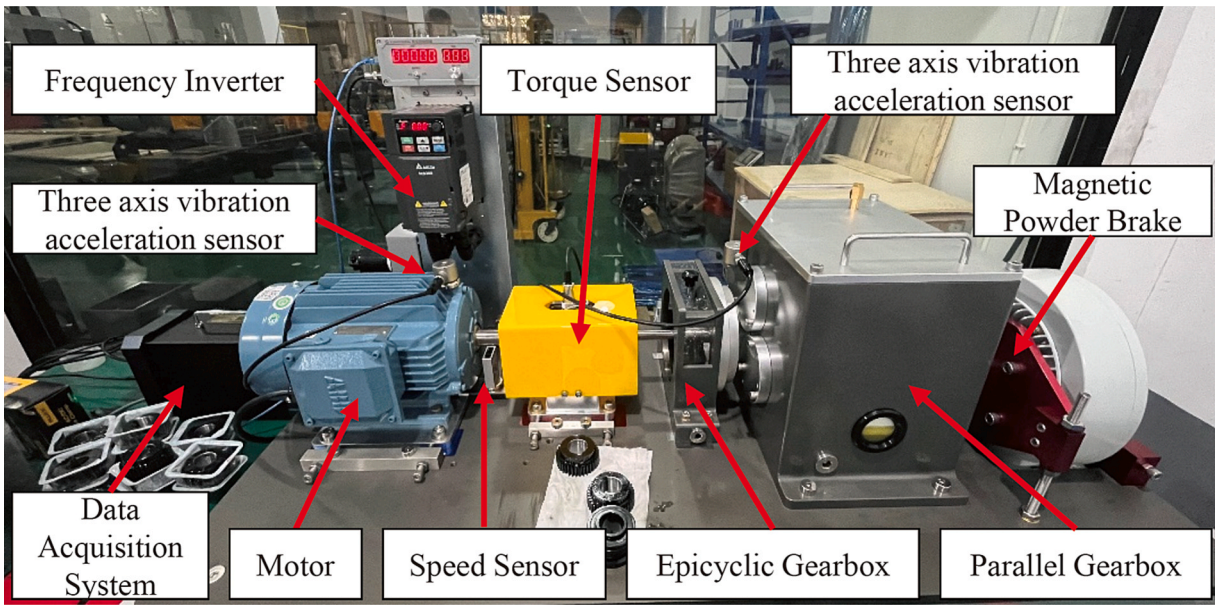


Fig. 3. Tsinghua gearbox test rig.

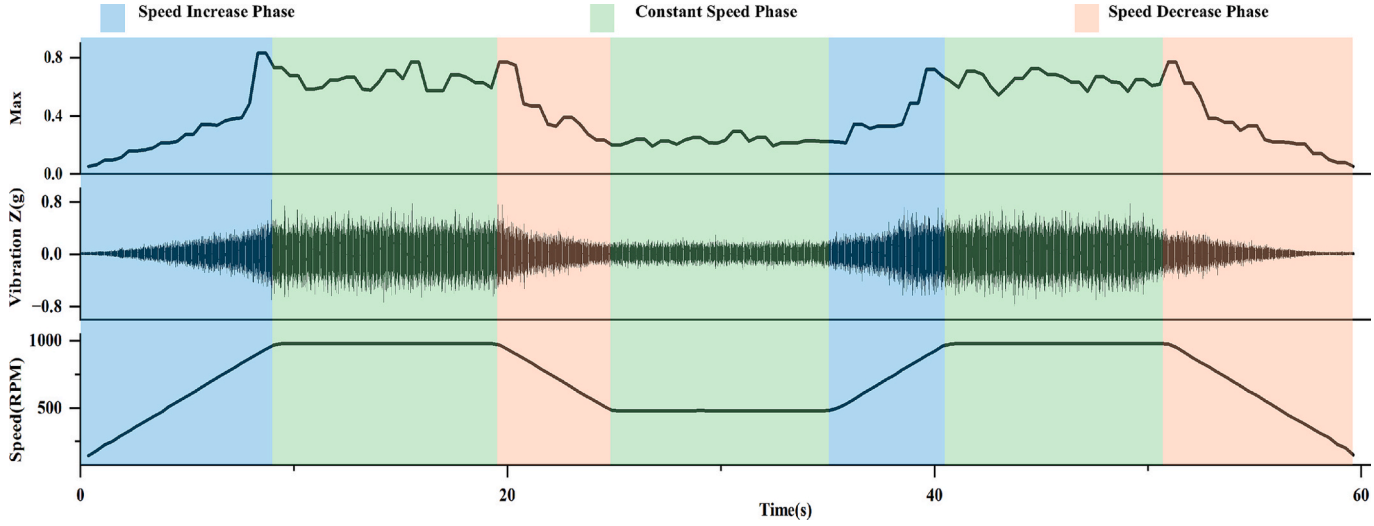
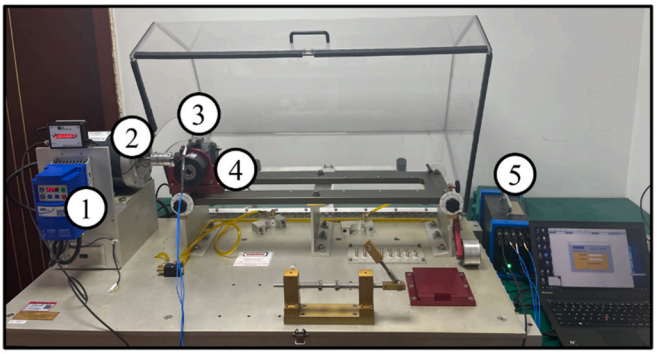


Fig. 4. Data visualization includes speed, vibration Z and max of vibration Z.

Table 2
Key information for the Tsinghua Gearbox Dataset.

Fault ID	Fault Type	Sampling frequency	Number of distinct load levels	Total sampling duration	Rotational speed range
F1	Tooth pitting	12.8 kHz	2	180 s	0–1000 rpm
F2	Tooth missing			60 s	
F3	Tooth broken + bearing inner broken			180 s	
F4	Tooth broken + bearing outer broken			180 s	



1.Speed control 2.Motor 3.Acceleration sensor
4.Gearbox 5.Data acquisition board

Fig. 5. HUST gearbox test rig.

Fig. 6. Images of different gear health conditions in the HUST Gearbox Dataset. (a) health gear, (b) gear with broken tooth, (c) gear with missing tooth.

Table 3
Key information for the HUST Gearbox Dataset.

Fault ID	Fault Type	Sampling frequency	Number of distinct load levels	Total sampling duration	Rotational speed range
F1	Tooth broken	25.6 kHz	2	10.2 s	0-3600RPM
F2	Tooth missing			10.2 s	

Table 4
Detailed configuration of model parameters.

Parameter	Setting	Description
window length	9126	Length of data segments for feature extraction
window overlap rate	50 %	Overlap between consecutive sliding windows
Number of histogram bins (B)	50	Number of bins for HBOS histogram construction
Anomaly ratio parameter (γ)	0.08	Percentile threshold for anomaly detection
Queue length in TQAS	10	Length of the queue used in alarm strategy
Sensitivity coefficient (θ)	0.7	Threshold for anomaly ratio within the queue

data from anomalies by learning a hyperplane in a high-dimensional feature space.

DeepSVDD [42]: a deep learning–based anomaly detection model that learns to map data into a hypersphere using a neural network and identifies outliers based on their distance from the centre.

VAE [43]: a probabilistic generative model that learns the latent representation and distribution from normal data. Anomalies are detected based on high reconstruction error or low likelihood under the learned distribution.

4.1.4. Evaluation metrics and computing infrastructure

To evaluate the anomaly detection performance of the models, we adopt three commonly used metrics: True Positive Rate (TPR), False Positive Rate (FPR), and F1-Score, as defined in the table below [44]. To ensure a fair comparison between AWHBOS and the baseline models, the TQAS is excluded from this evaluation. The evaluation indicators are shown in Table 5.

Table 5
Evaluation indicators of anomaly detection. True Positives (*TP*) and True Negatives (*TN*) signify correct predictions of positive and negative instances, while False Positives (*FP*) and False Negatives (*FN*) represent errors in predictions. *Pr* and *Re* denote precision and recall.

Indicator	Formula	Definition
TPR	$\frac{TP}{TP+FP}$	Probability of fault being correctly detected
FPR	$\frac{FP}{TP+FN}$	Probability of normal sample being identified as fault
F1-Score	$\frac{2 \bullet Pr \bullet Re}{Pr+Re}$	Harmonic mean of precision and recall

4.2. Comparison results

As demonstrated in Tables 6-8, the proposed AWHBOS method was evaluated across multiple fault types in both datasets. The results of the study yielded the following conclusions:

a) In most comparative experiments, the AWHBOS model achieved the highest TPR and F1-score, demonstrating its superior performance in gearbox anomaly detection tasks. These results validate that the introduced adaptive weighting mechanism effectively enhances the model's sensitivity to critical anomaly features, improving overall detection performance.
b) Although the LOF model achieved the lowest FPR in most experiments, its TPR was relatively low, resulting in a high miss rate and a significantly lower F1-score than the other models. The results indicate that while LOF effectively reduces false alarms, its overall detection accuracy is limited, undermining its practical applicability.
c) On the Tsinghua Gearbox Dataset, which features more complex data distributions and hierarchical fault types (mild, moderate, and severe), AWHBOS demonstrated robust performance. Except for slightly lower detection performance for the F2 fault, AWHBOS achieved the best overall results in the remaining three fault categories.

Receiver Operating Characteristic (ROC) curves for all models on both datasets are presented in Fig. 7. The proposed AWHBOS method consistently demonstrated superior anomaly detection performance across all test scenarios, achieving an average AUROC exceeding 0.91. It significantly outperformed traditional methods such as LOF, OCSVM, and MCD and even surpassed deep learning models, including VAE and DeepSVDD.

The ROC curves of AWHBOS are notably steep and exhibit the narrowest confidence intervals, indicating high detection performance and exceptional stability across different data samples. These results highlight the robustness and reliability of the proposed method. The performance advantage of AWHBOS is particularly evident on the more complex Tsinghua Gearbox Dataset. In contrast, the performance of VAE and DeepSVDD varies considerably across datasets. Although these models achieve high AUROC scores (with AUROC > 0.8) in some scenarios, their performance drops significantly—sometimes approaching or falling below 0.6—in others, revealing limitations in generalization and stability.

4.3. Fitting process analysis

To demonstrate the operation of the proposed algorithm, we use the Tsinghua Gearbox Dataset as an example and present the intermediate results obtained during the execution process. First, based on the correlation analysis method introduced in this study, features with a Pearson correlation coefficient greater than 0.95 to rotational speed are selected to construct a new feature matrix. As shown in Figs. 8, 11 features exhibit a correlation greater than 0.95 with rotational speed.

Features strongly correlated with rotational speed were fitted using a cubic polynomial via the least squares method. As shown in Fig. 9, the RMS and Absolute Energy features under health conditions display distinct nonlinear relationships with rotational speed.

As shown in Fig. 9(a) and Fig. 9(c), the cubic polynomial fitting outperforms both linear and quadratic models, achieving higher coefficient of determination R^2 values (0.985 for RMS and 0.981 for Absolute Energy), thereby more accurately capturing the intrinsic relationship between features and rotational speed. The comparison plots in Fig. 9(b) and Fig. 9(d) further confirm that the predicted values closely follow the actual values, especially during stable operational

Table 6
On the Tsinghua Gearbox Dataset, the anomaly detection performance of each model under 10 Nm load using residual features.

Indicator	TPR				FPR				F1-Score			
Fault ID	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
MCD	89.5	94.9	72.5	57.8	8.23	8.23	8.23	8.23	90.0	93.3	77.3	66.3
IForest	91.8	94.9	74.6	54.2	8.23	8.23	8.23	8.23	91.8	93.3	78.9	63.6
LOF	42.2	72.6	31.1	34.4	**5.06**	**5.06**	**5.06**	**5.06**	55.4	83.8	47.1	49.5
OCSVM	42.6	72.0	35.3	35.0	10.1	10.1	10.1	10.1	54.4	81.0	49.0	48.7
DeepSVDD	91.8	96.2	71.9	57.8	10.1	10.1	10.1	10.1	91.3	93.0	76.4	65.8
VAE	90.7	95.5	70.4	54.2	8.23	8.23	8.23	8.23	91.0	**93.7**	75.8	63.6
AWHBOS	**95.8**	83.4	**85.0**	**77.6**	8.23	8.23	8.23	8.23	**94.8**	87.6	**86.7**	**81.2**

Table 7
On the Tsinghua Gearbox Dataset, the anomaly detection performance of each model under 20 Nm load using residual features.

Indicator	TPR(%)				FPR(%)				F1-Score(%)			
Fault ID	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
MCD	65.8	94.3	50.7	62.1	8.3	8.3	8.3	8.3	78.0	**93.1**	66.1	75.4
IForest	91.1	70.7	68.6	66.4	8.3	8.3	8.3	8.3	93.9	79.0	80.0	78.5
LOF	67.2	52.9	42.7	59.1	**7.6**	**7.6**	**7.6**	**7.6**	79.1	65.9	58.8	73.2
OCSVM	46.8	36.9	32.7	54.3	10.2	10.2	10.2	10.2	62.2	50.2	48.0	68.8
DeepSVDD	90.8	89.2	66.9	64.5	10.2	10.2	10.2	10.2	93.4	89.5	78.6	76.8
VAE	89.2	**92.4**	66.7	64.7	8.3	8.3	8.3	8.3	92.8	92.1	78.7	77.3
AWHBOS	**94.0**	68.2	**80.0**	**88.5**	8.3	8.3	8.3	8.3	**95.5**	77.3	**87.6**	**92.5**

Table 8
On the HUST Gearbox Dataset, each model's average anomaly detection performance under two load conditions, L0 and L1, using residual features.

Load	L0						L1					
Indicator	TPR(%)		FPR(%)		F1-Score(%)		TPR(%)		FPR(%)		F1-Score(%)	
Fault ID	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
MCD	74.7	100.0	8.1	8.1	81.8	96.1	98.0	100.0	8.1	8.1	95.1	96.1
IForest	74.7	100.0	8.1	8.1	81.8	96.1	92.9	100.0	8.1	8.1	92.5	96.1
LOF	44.4	89.9	**7.1**	**7.1**	58.7	91.3	91.9	91.9	**7.1**	**7.1**	92.4	92.4
OCSVM	47.5	89.9	10.1	10.1	60.3	89.9	88.9	93.9	10.1	10.1	89.3	92.1
DeepSVDD	74.7	99.0	10.1	10.1	80.9	94.7	90.9	100.0	10.1	10.1	90.5	95.2
VAE	74.7	100.0	8.1	8.1	81.8	96.1	93.9	100.0	8.1	8.1	93.0	96.1
AWHBOS	**74.7**	**100.0**	8.1	8.1	**81.8**	**96.1**	**99.0**	**100.0**	8.1	8.1	**95.6**	**96.1**

phases. The residuals remain within 0.025 for RMS and 25 for Absolute Energy, indicating high fitting accuracy and robustness.

By subtracting the fitted values from the observed feature values, we obtain residuals largely insensitive to rotational speed fluctuations. These residuals provide a more reliable basis for anomaly detection, as they predominantly reflect variations caused by factors such as operational disturbances, gear backlash, or incipient faults, rather than regular speed-induced changes.

We performed residual analysis on standardized residual features extracted from health state data in the Tsinghua Gearbox Dataset. As shown in Fig. 10, the Q-Q plot indicates that the Peak to Peak Distance residual closely follows a normal distribution, with a narrow 95 % confidence interval that contains almost all samples. In contrast, the RMS residual appears approximately normal in the central region but deviates significantly in the tails, exhibiting a heavy-tail phenomenon. This observation is further supported by the Anderson-Darling test statistic and p-value, which confirm the non-normality of RMS residuals.

Nevertheless, regardless of the specific residual feature, when compared to the original feature distributions shown in Fig. 1, the residuals under varying rotational speeds exhibit a more concentrated and regular distribution, making them more suitable for statistical modelling and anomaly detection.

4.4. Ablation study of model components

In this section, ablation experiments are conducted on the Tsinghua Gearbox Dataset to evaluate the effectiveness of two key components in the proposed method: the AWM and the TQAS. In the experiments, the queue length l was set to 10, and the tolerance-aware sensitivity coefficient θ was set to 0.7. To assess the individual contributions of these components to gear fault detection performance, a series of controlled experiments are designed under different load conditions (10 Nm and 20 Nm) and across multiple fault types (F1–F4). Three model configurations are compared:

M1: the baseline model using the standard HBOS algorithm.

M2: AWHBOS, which incorporates the adaptive weighting mechanism.

M3: the complete model that further integrates TQAS on M2.

Under the 10 Nm load condition (Fig. 11(a)–(c)), M2 demonstrates overall performance improvements over M1, with particularly notable gains in fault types F3 and F4. For F4, the F1-score of M2 increases by approximately 9.8 %, from 0.838 to 0.921, and precision improves from 0.688 to 0.785. Under the 20 Nm load condition (Fig. 11(d)–(f)), M2 achieves a substantial improvement in F2, where the TPR increases from 0.605 to 0.675, corresponding to an 11.6 % gain. These ablation results demonstrate that incorporating the AWM into HBOS provides a consistent advantage in gearbox anomaly detection under varying load conditions.

Moreover, M3, which integrates the TQAS into AWHBOS, achieves consistent and significant improvements across all three evaluation metrics. It demonstrates superior performance under both load conditions and across all four fault types, further validating the effectiveness of TQAS in enhancing the generalization capability and robustness of the proposed anomaly detection framework.

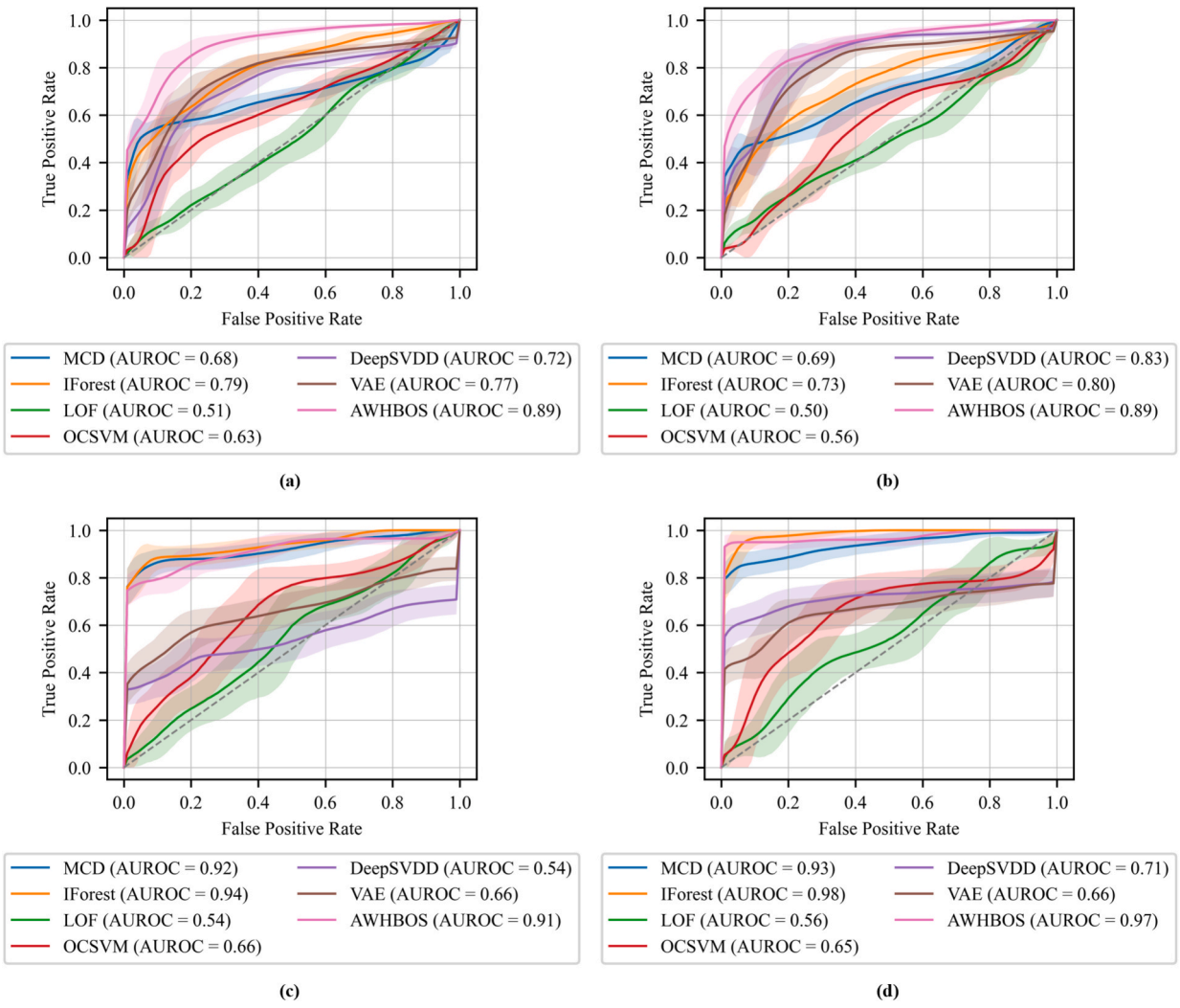


Fig. 7. ROC curves illustrate the anomaly detection performance of different models. Solid lines indicate the average results obtained from five-fold cross-validation, while shaded areas represent the variability across different data samplings, corresponding to the confidence intervals of the ROC curves. (a) Performance under 10 Nm load on the Tsinghua dataset; (b) Performance under 20 Nm load on the Tsinghua dataset; (c) Performance under L0 load on the HUST dataset; (d) Performance under L1 load on the HUST dataset.

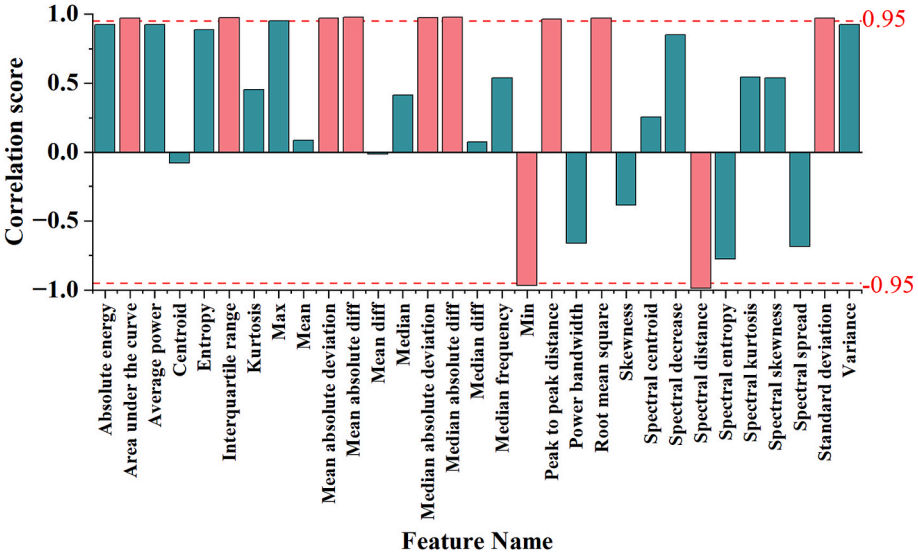


Fig. 8. Correlation analysis results between features and varying rotational speed.

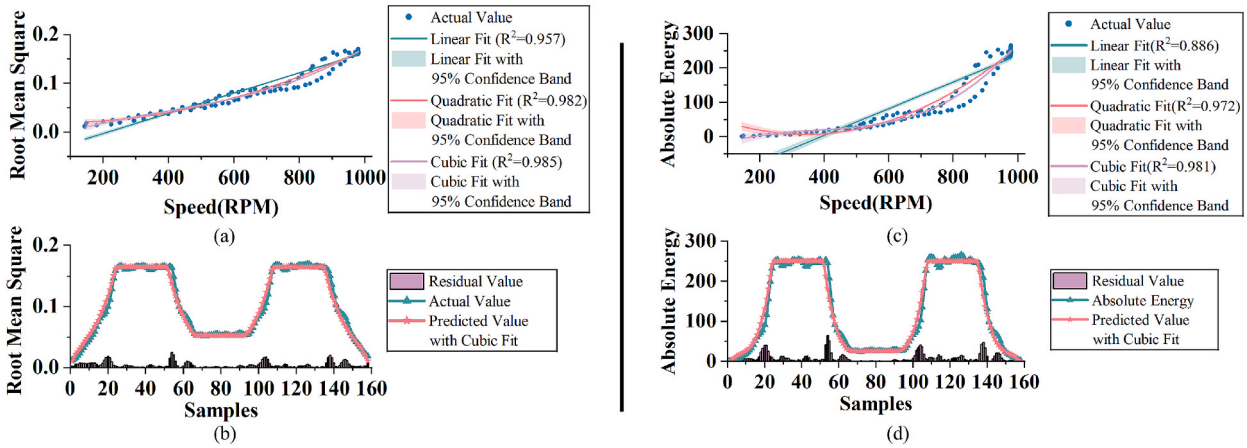


Fig. 9. Fitting results and comparison. (a) Cubic polynomial fitting of the vibration signal RMS values under varying rotational speeds. (b) Comparison between cubic fit predicted and actual values for RMS, along with the corresponding residuals. (c) Cubic polynomial fitting of the absolute energy values of the vibration signal under varying rotational speeds. (d) Comparison between cubic fit predicted and actual values for absolute energy, along with the corresponding residuals.

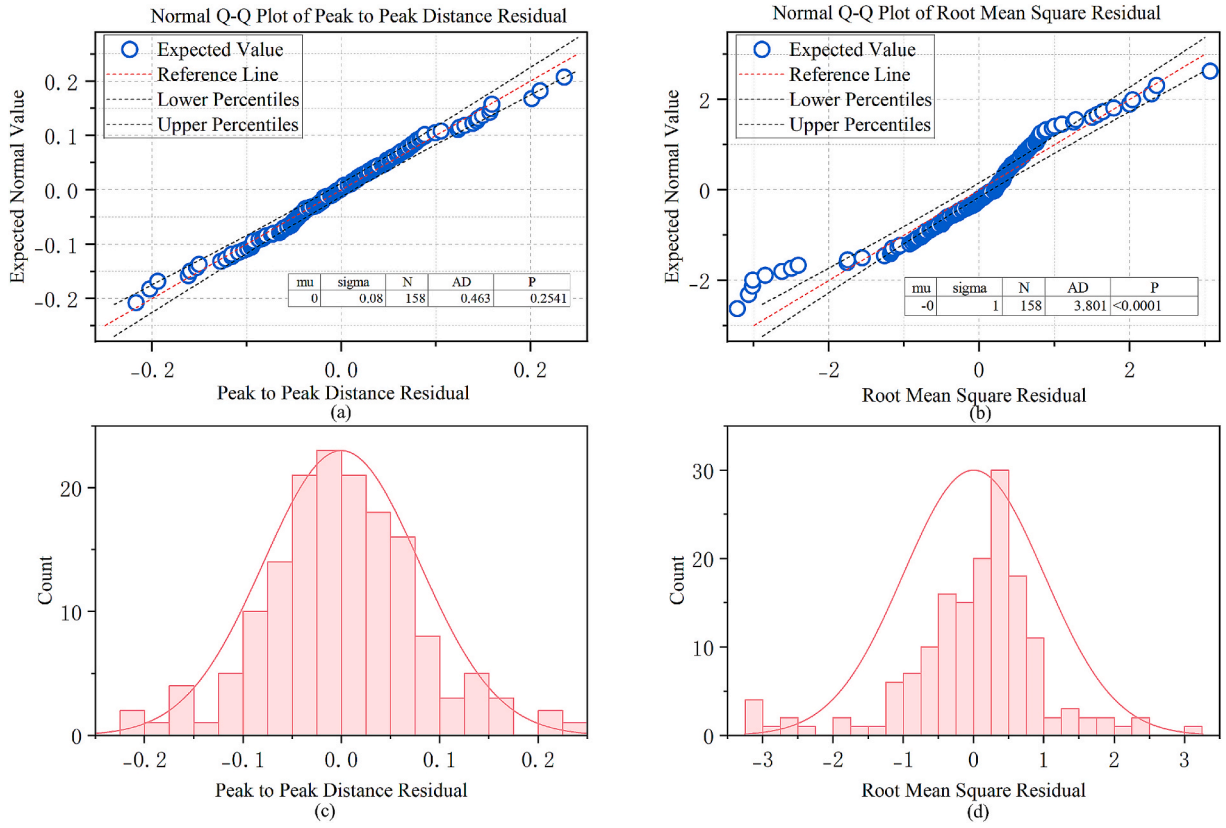


Fig. 10. Residual feature analysis under health state data. (a)Normal Q-Q plot of Peak to Peak Distance residual. (b) Normal Q-Q plot of RMS residual. (c) Peak to Peak Distance residual Histogram. (d) RMS residual Histogram.

4.5. Impact of the sensitivity coefficient

To evaluate the impact of the key parameter in the proposed TQAS, we investigated how different values of the tolerance-aware sensitivity coefficient θ (ranging from 0.1 to 1.0) affect model performance. As shown in Fig. 12, increasing the value of θ means that a higher proportion of anomalies is required within the queue to trigger an alarm. Consequently, the TPR gradually decreases, indicating a higher likelihood of missed detections. Meanwhile, the FPR also decreases, reflecting fewer false alarms. The TPR gradually decreases, indicating a higher likelihood of missed detections. Meanwhile, the FPR also decreases, reflecting fewer false alarms. The F1-score first increases and then declines, suggesting a trade-off between detection sensitivity and reliability. Within the recommended range of (0.4–0.7), the model achieves a desirable balance and demonstrates optimal overall performance. Additionally, it can be observed that although the TPR and F1-score follow a similar trend with varying θ across different fault types, the magnitude of change differs. For faults that exhibit more pronounced performance fluctuations with changing θ—such as F2—greater caution is required when determining the appropriate θ value to ensure reliable detection performance.

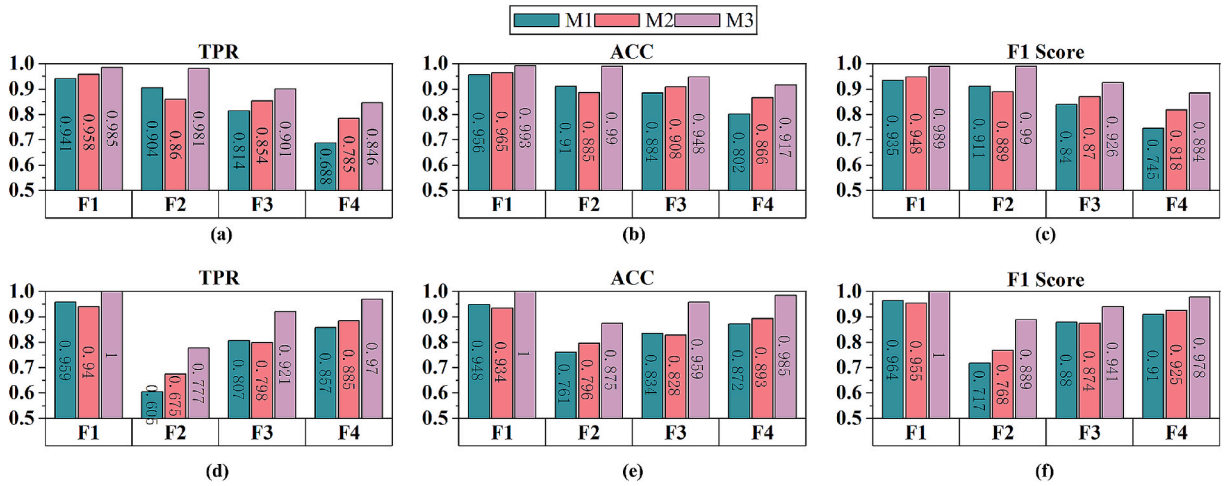


Fig. 11. Ablation experiment results on the Tsinghua Gearbox dataset. (a), (b) and (c) show the results under the 10 Nm load condition, while (d), (e), and (f) correspond to the 20 Nm load condition.

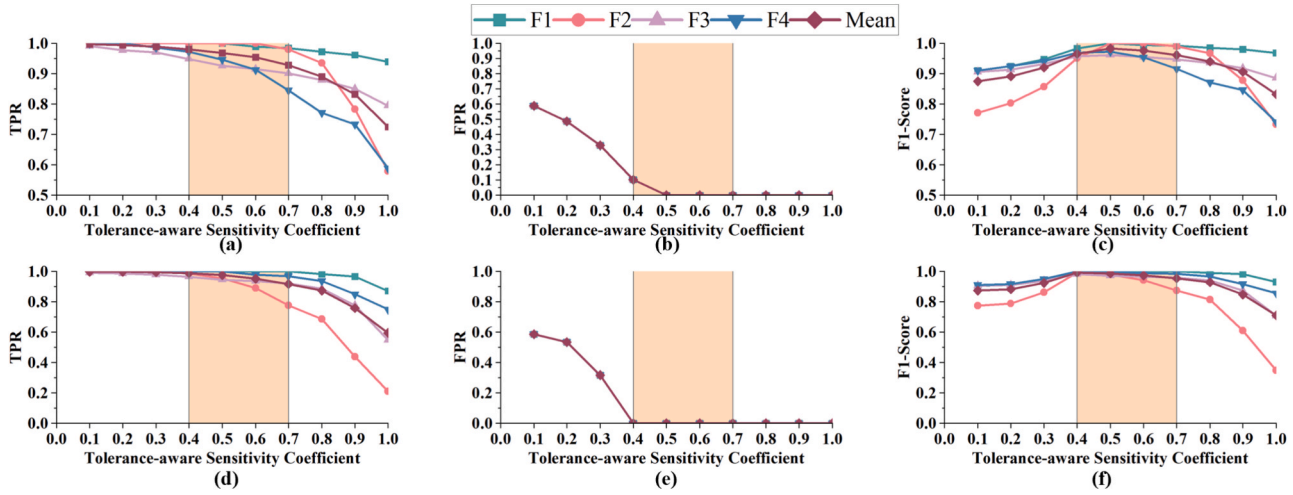


Fig. 12. Influence of the different tolerance-aware sensitivity coefficient θ on the anomaly detection performance. (a)–(c) 10 Nm load condition on the Tsinghua Gearbox Dataset; (d)–(f) 20 Nm load condition on the Tsinghua Gearbox Dataset. The yellow-shaded region indicates the recommended range of θ used in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Conclusion

In this study, we proposed a novel residual feature-based semi-supervised learning method for anomaly detection in gearboxes. By analysing 31 time-domain and frequency-domain features, we selected features highly correlated with rotational speed and derived residual features through cubic polynomial fitting. This study demonstrates that the newly developed residual features effectively mitigate the influence of time-varying rotational speed on anomaly detection model performance. Evaluations on two publicly available datasets involving time-varying rotational speeds and multiple gearbox fault types demonstrated that the proposed AWHBOS method, compared to the standard HBOS approach, dynamically emphasizes critical residual features and substantially improves anomaly detection performance. Experimental results confirm that the proposed model outperforms six commonly used unsupervised anomaly detection methods relying solely on health state data. Additionally, we introduced TQAS. Ablation studies demonstrate that an appropriately selected tolerance-aware sensitivity coefficient can effectively enhance the TPR while concurrently reducing the false alarm rate. This underscores the substantial practical value of the proposed approach.

Considering that our method depends on a certain amount of health state data to establish a reliable baseline, future work will explore deep neural network-based techniques to synthesize additional healthy-state data, thus reducing the reliance on initial healthy datasets. Furthermore, we will quantitatively investigate the impact of the TQAS on alarm delays, aiming to minimize implementation risks and promote the practical industrial deployment of this strategy.

CRediT authorship contribution statement

Hongliang Song: Writing – original draft, Software, Methodology, Formal analysis. **Yi Sun:** Visualization, Methodology, Data curation. **Hongli Gao:** Validation, Methodology, Funding acquisition. **Zhichao You:** Software, Funding acquisition. **Hao Xu:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Sichuan Science and Technology Program, No. 2024ZYD0271.

Data availability

Data will be made available on request.

References

[1] A.A. Dubaish, A.A. Jaber, State-of-the-art review into signal processing and artificial intelligence-based approaches applied in gearbox defect diagnosis, Eng. Technol. J. (2023) 1–16, https://doi.org/10.30684/etj.2023.142462.1535.
[2] X. Zhao, J. Yao, W. Deng, P. Ding, Y. Ding, M. Jia, Z. Liu, Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intraclass and interclass convolutional neural network, IEEE Trans. Neural Networks Learn. Syst. 34 (2023) 6339–6353, https://doi.org/10.1109/TNNLS.2021.3135877.
[3] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: a review, Measurement 48 (2014) 292–305, https://doi.org/10.1016/j.measurement.2013.11.012.
[4] Y. Lin, M. Xiao, H. Liu, Z. Li, S. Zhou, X. Xu, D. Wang, Gear fault diagnosis based on CS-improved variational mode decomposition and probabilistic neural network, Measurement 192 (2022) 110913, https://doi.org/10.1016/j.measurement.2022.110913.
[5] D. Liu, L. Cui, W. Cheng, A review on deep learning in planetary gearbox health state recognition: methods, applications, and dataset publication, Meas. Sci. Technol. 35 (2023) 012002, https://doi.org/10.1088/1361-6501/acf390.
[6] J. Zhang, B. Zeng, W. Shen, L. Gao, A one-class Shapelet dictionary learning method for wind turbine bearing anomaly detection, Measurement 197 (2022) 111318, https://doi.org/10.1016/j.measurement.2022.111318.
[7] K. Hendrickx, W. Meert, Y. Mollet, J. Gyselinck, B. Cornelis, K. Gryllias, J. Davis, A general anomaly detection framework for fleet-based condition monitoring of machines, Mech. Syst. Sig. Process. 139 (2020) 106585, https://doi.org/10.1016/j.ymssp.2019.106585.
[8] C. Hu, J. Wu, C. Sun, X. Chen, R. Yan, Mutual information-based feature disentangled network for anomaly detection under variable working conditions, Mech. Syst. Sig. Process. 204 (2023) 110804, https://doi.org/10.1016/j.ymssp.2023.110804.
[9] H. Liu, H. Zhang, Y. Tang, Y. Yao, A unified detection approach for point and subsequence anomaly data from train axle temperature sensors, IEEE Sens. J. 23 (2023) 24772–24786, https://doi.org/10.1109/JSEN.2023.3307623.
[10] T. Kim, C.H. Park, Anomaly pattern detection in streaming data based on the transformation to multiple binary-valued data streams, J. Artif. Intell. Soft Comput. Res. 12 (2022) 19–27, https://doi.org/10.2478/jaiscr-2022-0002.
[11] K. Vos, Z. Peng, C. Jenkins, M.R. Shahriar, P. Borghesani, W. Wang, Vibration-based anomaly detection using LSTM/SVM approaches, Mech. Syst. Sig. Process. 169 (2022) 108752, https://doi.org/10.1016/j.ymssp.2021.108752.
[12] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, Automatic bearing fault diagnosis based on one-class ν-SVM, Comput. Ind. Eng. 64 (2013) 357–365, https://doi.org/10.1016/j.cie.2012.10.013.
[13] C. McKinnon, J. Carroll, A. McDonald, S. Koukoura, D. Infield, C. Soraghan, Comparison of new anomaly detection technique for wind turbine condition monitoring using gearbox SCADA data, Energies 13 (2020) 5152, https://doi.org/10.3390/en13195152.
[14] L. Puggini, S. McLoone, 'An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data, Eng. Appl. Artif. Intell. 67 (2018) 126–135, https://doi.org/10.1016/j.engappai.2017.09.021.
[15] Y. Lv, X. Guo, S. Shirmohammadi, L. Qian, Y. Gong, X. Hu, Intelligent cross-working condition fault detection and diagnosis using isolation forest and adversarial discriminant domain adaptation, IEEE Trans. Instrum. Meas. 73 (2024) 3531915, https://doi.org/10.1109/TIM.2024.3457923.
[16] F. Mateo, J. Vila-Frances, E. Soria-Olivas, M. Martinez-Sober, J. Gomez-Sanchis, A. J. Serrano-Lopez, Dynamic classifier auditing by unsupervised anomaly detection methods: an application in packaging industry predictive maintenance, Appl. Sci.-Basel 15 (2025) 882, https://doi.org/10.3390/app15020882.
[17] H. Tao, J. Qiu, Y. Chen, V. Stojanovic, L. Cheng, Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion, J. Franklin Inst. 360 (2023) 1454–1477, https://doi.org/10.1016/j.jfranklin.2022.11.004.
[18] A. Purarjomandlangrudi, A.H. Ghapanchi, M. Esmalifalak, A data mining approach for fault diagnosis: an application of anomaly detection algorithm, Measurement 55 (2014) 343–352, https://doi.org/10.1016/j.measurement.2014.05.029.
[19] B. Sheng, Q. Li, W. Mao, W. Jin, Outlier detection in sensor networks, in: MobiHoc '07: Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing, n.d.: pp. 219–228. doi: 10.1145/1288107.1288137.
[20] I. Aguilera-Martos, J. Luengo, F. Herrera, Revisiting Histogram Based Outlier Scores: Strengths and Weaknesses, in: Hybrid Artificial Intelligent Systems, HAIS, Springer International Publishing Ag, Cham, 2023, pp. 39–48, https://doi.org/10.1007/978-3-031-40725-3_4.
[21] R. Zhao, R. Yan, J. Wang, K. Mao, Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks, Sensors 17 (2017) 273, https://doi.org/10.3390/s17020273.
[22] W. Huang, B. Zhang, K. Zhang, H. Gao, R. Wan, Improved Auto encoder with LSTM module and KL divergence for anomaly detection, IEEE Trans. Instrum. Meas. 73 (2024) 3534011, https://doi.org/10.1109/TIM.2024.3460931.
[23] U. Yokkampon, S. Chumkamon, A. Mowshowitz, E. Hayashi, Anomaly detection using variational autoencoder with spectrum analysis for time series data, in: 2020 Joint 9th International Conference on Informatics, Electronics & Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 2020, pp. 1–6, https://doi.org/10.1109/ICIEVicIVPR48672.2020.9306570.
[24] X. Yu, K. Zhang, Y. Liu, B. Zou, J. Wang, W. Wang, R. Qian, Adversarial transformer-based anomaly detection for multivariate time series, IEEE Trans. Ind. Inf. 21 (2025) 2471–2480, https://doi.org/10.1109/TII.2024.3507211.
[25] Braei M., Wagner S., Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art, (2020). http://arxiv.org/abs/2004.00433 (accessed July 11, 2023).
[26] G. Pang, C. Shen, L. Cao, A.V.D. Hengel, Deep learning for anomaly detection: a review, ACM Comput. Surv. 54 (2022) 1–38, https://doi.org/10.1145/3439950.
[27] J. Yu, X. Gao, T. Wang, H. Lu, B. Li, F. Zhai, B. Xue, Z. Meng, A feature matching-based method for few-shot multivariate time series anomaly detection with symmetric patch mask Siam Transformer, Eng. Appl. Artif. Intell. 154 (2025) 110894, https://doi.org/10.1016/j.engappai.2025.110894.
[28] Y. Xi, Z. Lei, G. Wen, Z. Liu, Y. Su, Z. Zhang, X. Chen, Unsupervised fault detection method via time-series segmentation and contrastive masking learning, IEEE Trans. Instrum. Meas. (2025) 1, https://doi.org/10.1109/TIM.2025.3568963.
[29] J. Chen, W. Hu, D. Cao, Z. Zhang, Z. Chen, F. Blaabjerg, A meta-learning method for electric machine bearing fault diagnosis under varying working conditions with limited data, IEEE Trans. Ind. Inf. 19 (2023) 2552–2564, https://doi.org/10.1109/TII.2022.3165027.
[30] L. Kang, B. Xu, P. Li, K. Wang, J. Chen, H. Du, Q. Liu, L. Zhang, X. Lian, Controllable preparation of low-cost coal gangue-based SAPO-5 molecular sieve and its adsorption performance for heavy metal ions, Nanomaterials 15 (2025) 366, https://doi.org/10.3390/nano15050366.
[31] N. Cao, H. Du, J. Lu, Z. Li, Q. Qiang, H. Lu, Designing ionic liquid electrolytes for a rigid and Li+-conductive solid electrolyte interface in high performance lithium metal batteries, Chem. Phys. Lett. 866 (2025) 141959, https://doi.org/10.1016/j.cplett.2025.141959.
[32] H. Huang, N. Baddour, Bearing vibration data collected under time-varying rotational speed conditions, Data Brief 21 (2018) 1745–1749, https://doi.org/10.1016/j.dib.2018.11.019.
[33] R.G. Vinson, P.S. Heyns, T. Heyns, Rotating machine diagnosis using smart feature selection under non-stationary operating conditions, Insight - Non-Destructive Testing and Condition Monitoring 58 (2016) 417–422.
[34] Q. Xu, T. Xie, C. Jiang, Q. Cheng, X. Wang, Adaptive working condition recognition with clustering-based contrastive learning for unsupervised anomaly detection, IEEE Trans. Ind. Inf. 20 (2024) 12103–12113, https://doi.org/10.1109/TII.2024.3413952.
[35] Q. Li, C. Shen, L. Chen, Z. Zhu, Knowledge mapping-based adversarial domain adaptation: a novel fault diagnosis method with high generalizability under variable working conditions, Mech. Syst. Sig. Process. 147 (2021) 107095, https://doi.org/10.1016/j.ymssp.2020.107095.
[36] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, H. Gamboa, TSFEL: time series feature extraction library, SoftwareX 11 (2020) 100456, https://doi.org/10.1016/j.softx.2020.100456.
[37] D. Edelmann, T.F. Mori, G.J. Szekely, On relationships between the Pearson and the distance correlation coefficients, Stat. Probab. Lett. 169 (2021) 108960, https://doi.org/10.1016/j.spl.2020.108960.
[38] N. Paulauskas, A. Baskys, Application of histogram-based outlier scores to detect computer network anomalies, Electronics 8 (2019) 1251, https://doi.org/10.3390/electronics8111251.
[39] U. Binzat, E. Yıldıztepe, The adjusted histogram-based outlier score - AHBOS, MJST 9 (2023) 92–100, https://doi.org/10.22531/muglajsci.1252876.
[40] S. Chen, Z. Liu, X. He, D. Zou, D. Zhou, Multi-mode fault diagnosis datasets of gearbox under variable working conditions, Data Brief 54 (2024) 110453, https://doi.org/10.1016/j.dib.2024.110453.
[41] C. Zhao, E. Zio, W. Shen, Domain generalization for cross-domain fault diagnosis: an application-oriented perspective and a benchmark study, Reliab. Eng. Syst. Saf. 245 (2024) 109964, https://doi.org/10.1016/j.ress.2024.109964.
[42] D. Peng, W. Desmet, K. Gryllias, Reconstruction-based deep unsupervised adaptive threshold support vector data description for wind turbine anomaly detection, Reliab. Eng. Syst. Saf. 260 (2025) 110995, https://doi.org/10.1016/j.ress.2025.110995.
[43] Y. Chen, Y. Liu, S. Han, Y. Qiao, Multi-component condition monitoring method for wind turbine gearbox based on adaptive noise reduction, IET Renew. Power Gener. 17 (2023) 2613–2624, https://doi.org/10.1049/rpg2.12772.
[44] G. Forman, Counting positives accurately despite inaccurate classification, in: Machine Learning: ECML 2005, Proceedings, Springer-Verlag Berlin, Berlin, 2005, pp. 564–575.

HongliangSong obtained M.S. degree in Mechanical and Electronic Engineering from Southwest Jiaotong University, located in Chengdu, Sichuan, China, in 2020. He is currently pursuing Ph.D. in the same department at Southwest Jiaotong University. Throughout his academic journey, he has been deeply involved in research related to the fault diagnosis and intelligent maintenance technology of electromechanical equipment.

YiSun received the M.S. degree in mechanical engineering from Southwest Jiaotong University, Sichuan, P. R. China, in 2020, where he is currently pursuing the Ph.D. degree in mechatronic engineering. His research interests include machinery condition monitoring and structural load identification.

Hongli Gao received the Ph. D degree in mechanical engineering from Southwest Jiaotong University, Sichuan, P. R. China in 2005. He is currently a Professor at Southwest Jiaotong University. His research interests focus on designing and the reliability analysis of complex electromechanical equipment.

ZhichaoYou received the B.S. and Ph.D. degrees in mechanical engineering from Southwest Jiaotong University, Chengdu, P. R. China, in 2017 and 2022, respectively. He is currently an Assistant Professor at Southwest Jiaotong University. Prior to joining Southwest Jiaotong University in 2024, he was a Postdoctoral Research Fellow with the Shanghai Jiao Tong University, Shanghai, P. R. China. His research interests focus on the mechatronic prognostics and health management, smart manufacturing.

HaoXu received the B.S. degree in mechanical engineering from Anhui University of Technology, Ma'anshan, China in 2020. He is currently working toward the Ph.D degree in mechanical engineering with Southwest Jiaotong University. His research interests focus on machinery condition monitoring and health management.