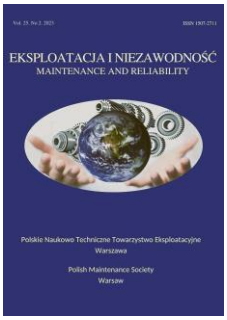


Eksploatacja i Niezawodnosc – Maintenance and Reliability
Volume 28 (2026), Issue 3
journal homepage: http://www.ein.org.pl

Article citation info:
Ma Z, Luo J, Li S, Zhang S, Li Y, AHC-SCLS-Driven Cluster Quality Improvement of Hierarchical Sample Entropy for Rotating Machinery Fault Feature Extraction, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(3) http://doi.org/10.17531/ein/217947

AHC-SCLS-Driven Cluster Quality Improvement of Hierarchical Sample Entropy for Rotating Machinery Fault Feature Extraction

Indexed by: Web of Science Group

Zhao Ma [a], Jiesi Luo [a,*], Shengbo Li [a], Shaohui Zhang [b], Yeni Li [a]

[a] School of Mechanical and Automotive Engineering, Xiamen University of Technology, China
[b] School of Mechanical Engineering, Dongguan University of Technology, China

Highlights

- Hierarchical clustering reduces redundancy among multi-scale entropy features.
- Silhouette-based clustering determines optimal feature groups automatically.
- Laplacian Score selects the most representative feature within each cluster.
- The optimized entropy features enhance discrimination of mechanical fault patterns.
- Cross-dataset experiments confirm strong robustness across multiple classifiers.

Abstract

To address insufficient fault discriminability from excessive feature redundancy in rotating machinery analysis, an AHC-SCLS-driven clustering quality enhancement method for hierarchical sample entropy (HSE) is proposed. Hierarchical entropy decomposition first decouples multi-scale entropy features across frequency bands. Agglomerative hierarchical clustering (AHC) then constructs a hierarchical feature tree and reduces dimensionality via redundant attribute merging. A dual-criterion SCLS framework is integrated, where the average Silhouette Coefficient (SC) selects optimal cluster number and the Laplacian Score (LS) screens the most discriminative feature per cluster to form a refined subset. Experiments on the Ottawa University bearing and laboratory gearbox datasets validate the superiority of AHC-SCLS-optimized features. The PSO-SVM classifier achieves 99.29% accuracy on both datasets, other classifiers maintain 95.04%–97.87% accuracy, and the method outperforms mRMR, PCA and other traditional approaches across all classifiers.

Keywords
fault diagnosis, laplacian score, hierarchical sample entropy, agglomerative hierarchical clustering, silhouette coefficient

1. Introduction

Fault Fault diagnosis methods based on machine learning have received widespread attention in vibration signal fault recognition. Such methods usually include three key steps: feature extraction, dimensionality reduction, and pattern recognition. With the development of nonlinear theory, a large number of signal processing methods based on statistical features and nonlinear dynamics principles have been introduced into the field of vibration signal analysis, aiming to more effectively reveal the potential fault information implied in equipment state changes [1]. Against this background, entropy-based features have become important tools for feature extraction due to their ability to quantify the intrinsic complexity and uncertainty of signals. Typical entropy algorithms include Approximate Entropy (AE) [2], Sample Entropy (SE) [3] and Permutation Entropy (PE) [4], etc.

In recent years, various entropy variants have emerged to address core challenges in mechanical fault diagnosis with limited samples, such as insufficient feature discrimination and poor generalization. Time-shift multi-scale phase entropy (RTSMPhE) adopts time-shift multi-scale technology to avoid

(*) Corresponding author.
E-mail addresses: Z. Ma, (ORCID: 0009-0009-0659-646X) mastarry@163.com, J. Luo (ORCID: 0000-0001-6385-9822) luojiesi_2004@163.com, S. Li (ORCID: 0009-0000-6950-0415) hit4057@xmut.edu.cn, S. Zhang (ORCID: 0000-0002-9784-4546) shaohui1985@hotmail.com, Y. Li (ORCID: 0009-0007-0665-8494) liyn@xmut.edu.cn

single-scale information incompleteness and invalid entropy interference, enabling cross-device generalized diagnosis [5]. Multi-modal multi-scale multi-level fusion quadrant entropy (MMMFQE) integrates vibration and acoustic multimodal data to enhance feature stability and complementarity [6]. Optimal weighted multi-scale entropy-energy ratio (OWMEER) employs WOA to optimize the weighted ratio of multi-scale permutation entropy and RMS, addressing traditional EER limitations and characterizing bearing degradation and early faults [7]. Multi-scale perception multi-level feature fusion image quadrant entropy (MPMFFIQE) transforms transient signals into temporal-information-preserving images via GASF to comprehensively extract dynamic features [8]. These variants promote the practical application of data-driven diagnosis. As a core foundational metric, Sample Entropy (SE) has algorithmic stability and sensitivity to temporal structural disturbances, it quantifies the probability of similar pattern repetitions to capture subtle early fault dynamics, avoiding the data length dependency of Approximate Entropy [9]. This solidifies its status as the cornerstone entropy indicator that underpins the innovation of various entropy variants in mechanical fault diagnosis.

However, traditional SE only focuses on features at a single scale and is difficult to effectively characterize the multi-scale dynamic characteristics of signals. It is especially prone to information loss when processing high-dimensional complex vibration signals [10]. To address the above problem, Costa [11] et al. proposed a method called Multiscale Sample Entropy (MSE), whose core idea is to perform coarse-graining of the original time series at multiple scales, and then calculate the sample entropy at each scale separately. This approach overcomes the limitation of traditional sample entropy, which may miss certain features when applied at a single scale.Multiscale sample entropy has also achieved success in feature extraction of mechanical vibration signals. For example, Wang [6] improved multiscale sample entropy and proposed the Generalized Refined Composite Multiscale Sample Entropy, which not only reduced the probability of invalid entropy values but also improved the stability and accuracy of entropy values, making the extracted features have good classification performance. However, multiscale sample entropy has two inherent limitations during the coarse-graining process: first, the coarse-graining process essentially functions as a low-pass filtering approach implemented via linear smoothing, primarily emphasizing low-frequency signal characteristics. However, it may overlook subtle fault information embedded in high-frequency components, which is often critical for early fault diagnosis. Such information is often of key value for early fault diagnosis. Second, the excessive smoothing effect at large scales leads to irreversible loss of dynamic details in the original signal. In 2011, Jiang [12] proposed Hierarchical Entropy (HE), which uses a tree-structured hierarchical mechanism to fundamentally overcome the information loss defect of linear coarse-graining, providing a more reliable entropy feature tool for complex system condition monitoring.

Based on the advantages of sample entropy and hierarchical decomposition, this paper adopts Hierarchical Sample Entropy (HSE) to extract fault features. Compared with MSE, HSE has stronger robustness and expressive power in retaining high-frequency features, and is especially suitable for analyzing high-dimensional vibration signals with complex structures and dynamic variations. If the number of decomposition layers in HSE is set too small, it may lead to the loss of high-frequency fault features and aggravation of low-frequency noise interference, thereby affecting the effective characterization of fault states by entropy values. In order to retain effective information at different scales as much as possible, this paper sets the maximum number of decomposition layers to 4 and extracts the HSE features corresponding to each layer to enhance the fine-grained expression ability of fault states [13].

However, the characterization ability of sample entropy at different layers for fault states varies, and some features may be redundant, highly correlated, or lack discriminative power. With the increase of decomposition layers, the dimensionality of extracted features increases significantly. If not processed, it may adversely affect the performance and computational efficiency of subsequent classification models. Common dimensionality reduction methods such as PCA and LDA are only applicable to linear or approximately linear high-dimensional datasets with Gaussian distributions [14,15]. And mRMR adopts a greedy strategy in the feature selection process, adding the currently optimal feature step by step without considering the global optimality of the feature subset, which is prone to fall into local optima and affects the final diagnosis

performance.

To address the above problems, this paper proposes a feature selection method based on HSE and AHC-SCLS cluster quality optimization. This method optimizes cluster quality to select representative low-dimensional features from the high-dimensional features extracted by HSE, in order to improve the recognition accuracy and generalization ability of the model. Specifically, this method introduces Agglomerative Hierarchical Clustering (AHC) [16] to construct a bottom-up hierarchical clustering tree of HSE features; during the feature clustering stage, the average Silhouette Coefficient (SC) [17] is introduced as a clustering quality evaluation index to automatically determine the optimal number of clusters and achieve iterative merging of redundant features; in each cluster, the Laplacian Score (LS) [18] is used to evaluate each feature dimension to quantify its discriminative ability between different classes, and then the most representative features with the strongest discrimination ability are selected.

In order to comprehensively evaluate the fault diagnosis performance of the proposed method, this study adopts five commonly used pattern recognition algorithms for performance evaluation, including Particle Swarm Optimization - Support Vector Machine (PSO-SVM) [19], Random Forest (RF) [20], Probabilistic Neural Network (PNN) [21], Extreme Learning Machine (ELM) [22], and Artificial Neural Network (ANN) [23]. These algorithms were applied to fault diagnosis experiments on both the laboratory self-collected wind turbine gearbox dataset and the public vibration dataset provided by the University of Ottawa, Canada. The experimental results show that the HSE features after dimensionality reduction by the proposed method can achieve good diagnosis performance under different classifier models.

In addition, to further verify the superiority of the proposed method, it was compared with commonly used feature selection/dimensionality reduction methods, including: Maximum Relevance Minimum Redundancy (mRMR) [24], Analysis of Variance (ANOVA) [25], Principal Component Analysis (PCA) [26], and Autoencoder (AE) [27]. The results indicate that the proposed method has significant advantages in fault diagnosis accuracy.The main contributions of this paper are:

1. It proposes an Agglomerative Hierarchical Clustering (AHC)-based method for hierarchical entropy feature redundancy reduction. This method organizes the multi-scale features extracted by Hierarchical Sample Entropy (HSE) into a hierarchical tree structure and employs an iterative merging strategy to efficiently compress feature dimensionality, significantly reducing feature redundancy.
2. It proposes a dual-criterion feature selection strategy (SCLS) based on the Silhouette Coefficient and Laplacian Score. This strategy uses the average Silhouette Coefficient to automatically determine the optimal number of clusters and selects the most discriminative feature from each cluster according to the Laplacian Score, ultimately constructing an optimized feature subset with low redundancy and high discriminative power.
3. It proposes a mechanical fault feature quality optimization method based on HSE and AHC-SCLS. Comprehensive validation has been conducted across multiple datasets and various classification models, demonstrating the effectiveness and superiority of the proposed method.

The structure of the paper is as follows: Section 2 overviews the basic theory of Hierarchical Sample Entropy and Agglomerative Hierarchical Clustering; Section 3 introduces the definition, calculation method, and the overall fault diagnosis framework of AHC-SCLS; Section 4 verifies the AHC-SCLS method using the self-tested wind turbine gearbox dataset and the public dataset from the University of Ottawa; finally, Section 5 provides the conclusions.

2. Basic Principles

2.1. Sample Entropy

Sample Entropy is a method used to evaluate the complexity of a time series. It assesses the unpredictability of the series by examining whether similar patterns remain consistent when the embedding dimension increases. The basic idea can be described as follows:

Assume a time series of length N: $\{X(n)\} = x(1), x(2), \ldots, x(N)$. The computation of Sample Entropy involves the following steps: First, construct a set of m-dimensional vectors: $X_i^m = \{x(i), x(i+1), \ldots, x(i+m-1)\}, 1 \le i \le N-m+1$.

Each vector represents mmm consecutive values starting from the i-th point. The distance between two vectors is defined as the maximum absolute difference among corresponding components:

$$d[X_m(i), X_m(j)] = max_{k=0,\dots,m-1}(|x(i+k) - x(j+k)|) \quad (1)$$

Count the number of vectors X_j^m whose distance from X_i^m is less than a threshold r, where $j \neq i$. Denote this count as $q_i^{m,r}$, and define the ratio of similar vectors as:

$$\psi_i^{m,r} = \frac{1}{N-m-1} q_i^{m,r} \quad (2)$$

Then, compute the average over all i:

$$\phi^{m,r} = \frac{1}{N-m} \sum_{i=1}^{N-m} \psi_i^{m,r} \quad (3)$$

Next, increase the embedding dimension to $m+1$, construct new vectors X_i^{m+1}, and repeat the same process to compute:

$$\psi_i^{m+1,r} = \frac{1}{N-m-1} q_i^{m+1,r} \quad (4)$$

$$\phi^{m+1,r} = \frac{1}{N-m} \sum_{i=1}^{N-m} \psi_i^{m+1,r} \quad (5)$$

Finally, Sample Entropy is defined as the negative natural logarithm of the ratio between the probabilities of matching at dimensions m and $m+1$:

$$SampEn(m,r,N) = \lim_{N\to\infty} \left\{ -ln\left[\frac{\phi^{m+1,r}}{\phi^{m,r}}\right] \right\} \quad (6)$$

When N is finite, it is typically estimated as:

$SampEn(m,r,N) = -ln\left[\frac{\phi^{m+1,r}}{\phi^{m,r}}\right]$, namely

$$SE(m,r) = -ln\left[\frac{\phi^{m+1,r}}{\phi^{m,r}}\right] \quad (7)$$

In this study, the embedding dimension m is set to 2 and the tolerance threshold r is set to 0.15 times the standard deviation of the signal [9].

2.2. Hierarchical Sample Entropy

The Hierarchical Sample Entropy (HSE) method extends traditional sample entropy to provide a more detailed characterization of time series complexity. By constructing a hierarchical structure, it calculates the entropy of sub-signals at different levels, thereby better capturing the complex features of the sequence.

To calculate it, a time series of length 2^n is defined as $X := (x_j : j \in Z_{2^{n-1}})$. Through successive decomposition, sub-signals of different time scales are obtained. The successive decomposition is performed by calculating the average and difference operators of the time series, so that sub-signals of different time scales are extracted. The average operator extracts the low-frequency part of the signal, while the difference operator extracts the high-frequency part. The average and difference operators are defined at each level as shown in Fig.1.

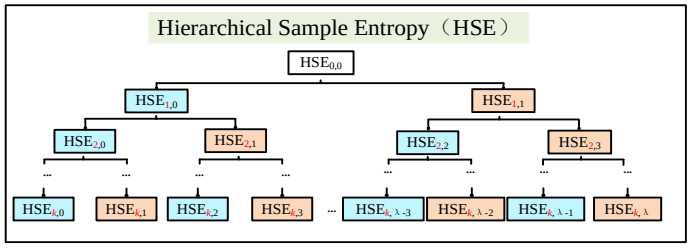


Figure 1. Hierarchical Sample Entropy Decomposition Flowchart.

The average operator $Q_0(X)$ and the difference operator $Q_1(X)$ are defined as in formulas (8) and (9):

$$Q_0(X) = \frac{x_{2j}+x_{2j+1}}{2}, j \in Z^{2n-1} \quad (8)$$

$$Q_1(X) = \frac{x_{2j+1}-x_{2j}}{2}, j \in Z^{2n-1} \quad (9)$$

The operator Q_j for $j \in Z_2$ is constructed in matrix form. When $j = 0$, it is an average operator; when $j = 1$, it is a difference operator. It is represented by a matrix as follows:

$$Q_j = \begin{bmatrix} \frac{1}{2} & \frac{(-1)^j}{2} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{(-1)^j}{2} & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{2} & \frac{(-1)^j}{2} \end{bmatrix}_{2^{n-1}*2^n} \quad (10)$$

To perform multiscale analysis on the time series X, the above operators need to be repeatedly applied. Let $N \in Z_+$ be a positive integer, and define $[\alpha_1, \alpha_2, \dots, \alpha_n] \in Z_2^n$. A non-negative integer e is then given by

$$e := \sum_{j=1}^{n} \alpha_j 2^{n-j} \quad (11)$$

When $n \in Z_N^+$ is fixed, each given non-negative integer corresponds to a unique vector $[\alpha_1, \alpha_2, \dots, \alpha_n] \in Z_2^n$. Based on this, the hierarchical decomposition of a time series $X \in R^{2^N}$ can be expressed as:

$$X_{n,e} := Q_{\alpha_n} \circ Q_{\alpha_{n-1}} \circ \dots \circ Q_{\alpha_1}(X) \quad (12)$$

For $k \in Z_{N+1}^+$, define the index set $J_k := \{(n,e) : n \in Z_k, e \in Z_{2^n}\}$. Given k, $X_{n,e}$ forms the k-level decomposition of X. The SE of each $X_{n,e}$ is then computed to obtain the HSE. The hierarchical decomposition process is shown in Fig. 1, where the root node

$HSE_{0,0}$represents the SE of the original time series X. At each decomposition level, the time series is transformed to a higher resolution, capturing details at finer scales or higher frequencies.

2.3. Agglomerative Hierarchical Clustering Algorithm

In Agglomerative Hierarchical Clustering (AHC), each data point starts as an independent cluster. The algorithm then progressively combines clusters based on their similarity, proceeding in a bottom-up manner until the entire dataset is unified into a single cluster or the desired number of clusters is formed. The AHC process for five data objects is shown in Fig. 2.

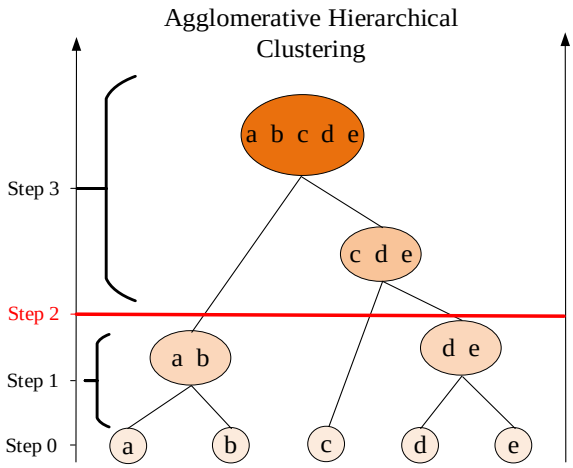


Figure 2. Flowchart of Hierarchical Sample Entropy Decomposition.

The basic steps of AHC are as follows:

Step 0: Initialization. Treat each feature as an individual cluster; there are n clusters in total. As shown in Fig. 2: the bottom nodes a, b, c, d, e form five clusters, each corresponding to one feature.

Step 1: Compute the distance matrix. Calculate the distance between all clusters. As shown in Fig. 2: if a and b are most similar in the initial stage, they are merged into ab; d and e are merged into de.

Step 2: Update the distance matrix. Update the distance between the new cluster and the other clusters according to the selected linkage method. As shown in Fig. 2, the red dashed lines indicate updates to the distance matrix. After each merge, the clustering structure changes and the distance matrix must be recalculated to reflect the new structure.

Step 3: Repeat merging until one of the following conditions is met: all samples are merged into a single cluster (complete linkage clustering), or the preset number of clusters (cutoff) is reached. As shown in Fig. 2: c and de are merged into cde, and finally ab and cde are merged into a single cluster.

3. Fault Diagnosis Method Based on AHC-SCLS

This section first describes the construction principle of the hierarchical sample entropy clustering tree driven by AHC, analyzes the distance measurement criterion of the vibration map; then proposes the SCLS dual-indicator evaluation framework: first, the average silhouette coefficient (SC) is used to determine the optimal number of clusters automatically, and then Laplacian Score (LS) is used to select the representative features of each cluster to form the optimal feature subset; finally, it explains how AHC-SCLS inputs the subset into the classifier for fault identification.

3.1. Construction Principle of the Hierarchical Sample Entropy Clustering Tree Driven by AHC

In order to effectively identify the correlation and structural differences between multi-dimensional features, this paper adopts AHC to divide features: Using feature similarity as the metric and following a bottom-up aggregation principle, a hierarchical clustering tree is iteratively constructed. The clustering distance measurement relationship is as follows.

Construction Method of Feature Distance Matrix Based on Pearson Correlation Coefficient

In order to evaluate the similarity among features, this study develops a correlation matrix utilizing the Pearson Correlation Coefficient (PCC). The Pearson coefficient quantifies the degree of linear dependence between pairs of feature variables. Given two featuresx_i andx_j (corresponding to the i-th and j-th column vectors of the feature set), the coefficient is computed as follows:

$$r_{ij} = \frac{\sum_{k=1}^{N}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{N}(x_{ki}-\bar{x}_i)^2}\cdot\sqrt{\sum_{k=1}^{N}(x_{kj}-\bar{x}_j)^2}} \tag{13}$$

where: x_{ki} is the value of the k-th sample on the i-th feature; $x_{kj}$is the value of the k-th sample on the j-th feature;N is the number of samples;

$\bar{x}_i$,$\bar{x}_j$are the means of the i -th and j -th features, respectively：

$$\bar{x}_i = \frac{1}{N}\sum_{k=1}^{N} x_{ki}, \quad \bar{x}_j = \frac{1}{N}\sum_{k=1}^{N} x_{kj} \tag{14}$$

The calculated correlation matrix is expressed as:

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,30} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,30} \\ \vdots & \vdots & \ddots & \vdots \\ r_{30,1} & r_{30,2} & \cdots & r_{30,30} \end{bmatrix} \quad (15)$$

The value range of the PCC is $[-1,1]$, and it is converted to a distance metric to meet clustering requirements.Therefore, this paper uses the following method to convert correlation into distance:

$$d_{ij} = 1 - |r_{ij}| \quad (16)$$

For a symmetric distance matrix D of size $n \times n$, where each D_{ij} represents the distance between feature i and feature j, the upper triangular elements (excluding diagonal elements) are extracted in row-major order and transformed into a vector d:

$$d = [D_{12}, D_{13}, \dots, D_{1n}, D_{23}, \dots, D_{(n-1)n}] \quad (17)$$

Principle of Hierarchical Clustering Based on Ward's Method

In this study, Ward's method is employed to perform hierarchical clustering. The algorithm operates by reducing the increase in the error sum of squares (ESS) when combining clusters, which serves as the merging strategy. Through a progressive merging process that consistently selects the pair of clusters causing the minimal rise in overall within-cluster variance, a well-defined hierarchical clustering tree is generated. The principle is:

If C_i and C_j are two clusters to be merged, the Ward distance is defined as:

$$d(C_i, C_j) = \sqrt{\frac{|C_i||C_j|}{|C_i|+|C_j|}} \cdot \left\|\bar{x}_i - \bar{x}_j\right\|_2 \quad (18)$$

where $|C_i|$ and $|C_j|$ are the sample counts of clusters C_i and C_j, and $\left\|\bar{x}_i - \bar{x}_j\right\|_2$ is the Euclidean distance between cluster centroids $\bar{x}_i$ and $\bar{x}_j$, respectively. The centroids are defined as:

$$\bar{x}_i = \frac{1}{|C_i|}\sum_{x \in C_i} x, \quad \bar{x}_j = \frac{1}{|C_j|}\sum_{x \in C_j} x \quad (19)$$

3.2. Determination of Optimal Clustering Structure Based on SC

To facilitate the automated selection of the optimal feature subset, this study employs the average Silhouette Coefficient (SC) as an evaluation metric during the feature clustering phase, guiding the determination of the most suitable number of clusters. SC is a widely used unsupervised clustering evaluation metric with two main advantages in complex fault diagnosis: it does not rely on label information, and it evaluates both inter-class separation and intra-class compactness, thereby improving clustering performance.

To improve clustering effectiveness, this paper determines the optimal number of clusters by maximizing the average SC, making the final formed feature subsets more structurally separable. For a given data point x_i, SC is defined as:

$$s(x_i) = \frac{b(x_i) - a(x_i)}{max(a(x_i), b(x_i))} \quad (20)$$

Here, $a(x_i)$ denotes the average distance between sample x_i and all other points within the same cluster, representing the intra-cluster cohesion:

$$a(x_i) = \frac{1}{|C_k|-1}\sum_{j \neq i, x_j \in C_k} d(x_i, x_j) \quad (21)$$

C_k is the cluster to which x_i belongs. $b(x_i)$ is the minimum average distance from sample x_i to all points in other clusters:

$$b(x_i) = \min_{C_m \neq C_k} \frac{1}{|C_m|}\sum_{x_j \in C_m} d(x_i, x_j) \quad (22)$$

The average SC for a given number of clusters k is calculated by taking the mean silhouette coefficient of all samples:

$$S_k = \frac{1}{N}\sum_{i=1}^{N} s(x_i) \quad (23)$$

For the selection of the optimal number of clusters, if the number of clusters is too small (small k), it may lead to different fault categories being misclassified and merged, reducing diagnostic accuracy. If the number of clusters is too large (large k), it may result in the same type being divided into multiple classes, increasing computational complexity and affecting the classification accuracy of the model. Searching within the range of 2 to 10 can ensure that key fault categories are not missed while avoiding overfitting.

The maximum number of clusters is set to 10, and the S_k for $k \in [2,10]$ is calculated. The k^* corresponding to the maximum value is selected as the optimal number of clusters:

$$k^* = arg \max_{k \in \{2,\dots,10\}} S_k \quad (24)$$

3.3. Intra-cluster Representative Feature Selection Based on Laplacian Score (LS)

After determining the feature clustering structure and obtaining the optimal number of clusters, this section further selects the most discriminative representative feature within each cluster. For this purpose, the Laplacian Score (LS) is used to evaluate the features within each cluster and select the optimal features.

LS is an unsupervised feature evaluation criterion based on graph theory. It can measure the ability of features to maintain local neighborhood consistency and is suitable for filtering representative features in high-dimensional signals. Specifically, let the feature matrix be $X = [x_1, \dots, x_n]^T \in R^{n \times d}$, where$n$is the number of samples and d is the feature dimension. To construct the structural relationships between samples, the k-nearest neighbor (KNN) method is used to build a stable similarity matrix. For each sample x_i, its k nearest neighbors are selected, and based on similarity strength or distance matching, a neighborhood matrix $W \in R^{n \times n}$ is constructed.

$$W_{ij} = \begin{cases} cos(x_i, x_j) & \text{if } x_j \in KNN(x_i) \text{ or } x_i \in KNN(x_j) \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

The degree matrix D and the Laplacian matrix L are defined as:

$$D_{ii} = \sum_j W_{ij}, \quad L = D - W \quad (26)$$

For each feature $f_j \in R^n$, its Laplacian Score is defined as:

$$\text{Laplacian Score}(f_j) = \frac{f_j^T L f_j}{f_j^T D f_j} \quad (27)$$

The numerator reflects the variation of the feature in the neighborhood, and the denominator is its normalization term. A smaller score indicates that the feature values are smoother among local neighbors and can better preserve the local structure. For each cluster, the feature with the smallest score is selected as the representative feature. Denote:

$$f^* = arg \max_{j \in C_I} F_j, \quad \forall i \in \{1,2, \dots, k^*\} \quad (28)$$

By performing the above screening operation within each cluster, the final feature index set constituting the optimal feature subset is obtained： $F^* = \{f_1^*, f_2^*, \dots, f_{k^*}^*\}$。

3.4. Specific Procedure of the AHC-SCLS Based Diagnosis Method

The fault diagnosis method proposed in this paper is shown in Fig. 3.

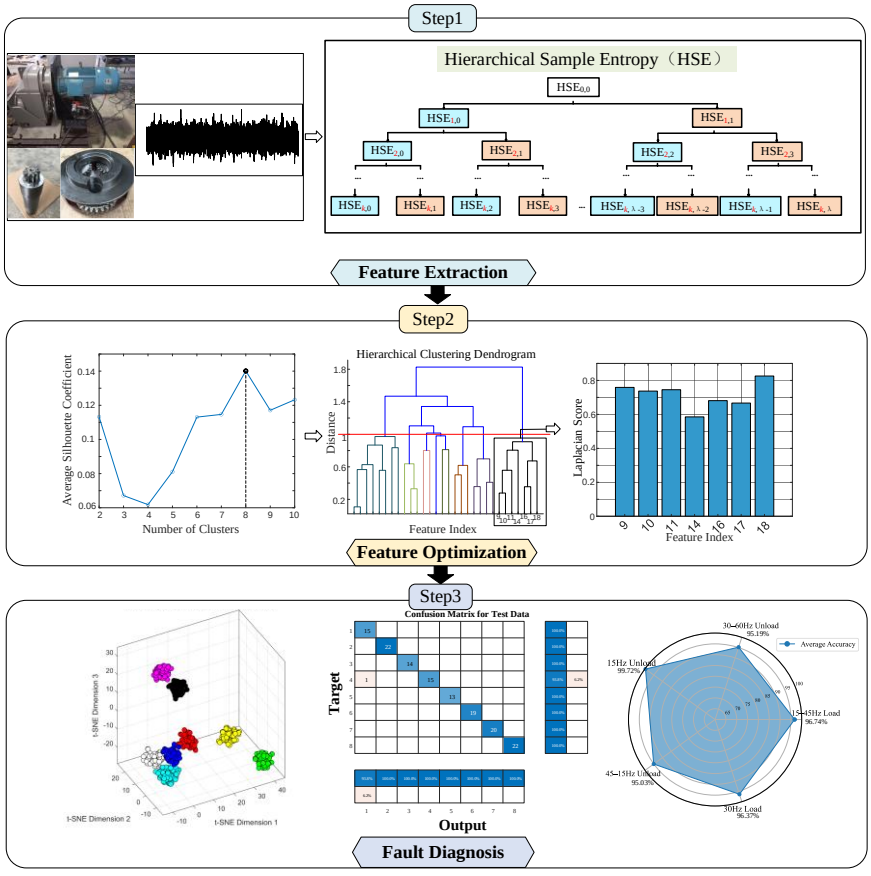


Figure 3. Overall Experimental Flowchart**.**

The specific process includes the following steps:

Step 1: Processing the signal and extracting its key features. First, the raw vibration data is ob-tained from the experimental equipment. Then, the HSE method is applied to perform multi-level analysis on the signal, and the entropy values of sub-frequency bands are extracted layer by layer to construct a high-dimensional feature set containing rich dynamic information.

Step 2: Optimal feature selection. Based on the high-

dimensional features obtained in step 1, the AHC method is used to construct the feature tree, and the average SC is calculated to evaluate the clustering effect, so as to determine the optimal number of clusters. On this basis, the LS is calculated in each cluster, and the optimal feature is selected. This process constitutes the SCLS feature selection method proposed in this paper.

Step 3: Pattern recognition. The features selected in Step 2 are used to construct the final optimal feature set. The dataset is split into 70% for training and 30% for testing to assess and validate the classification ability of the pattern recognition model. The confusion matrix derived from the test set is then used to evaluate diagnostic accuracy, confirming the effectiveness of the proposed method.

4. Experimental Verification and Analysis

To assess the validity and advantages of the proposed approach, experiments were performed using two datasets: the first is a motor fault vibration dataset from the University of Ottawa [28], and the second is a wind turbine gearbox fault dataset obtained in our laboratory [29].

4.1. Experiment 1: Motor Fault Vibration Dataset from the University of Ottawa, Canada Data Description

The experimental platform of the University of Ottawa is shown in Fig. 4. The sampling frequency was set to 42 kHz, and the acquisition duration was 10 s, yielding a total of 420,000 samples to ensure that speed variations could be distinguished.The dataset covers four motor speeds and multiple fault types. A total of eight motors of the same model were tested (one healthy and seven faulty), with a consistent sensor arrangement adopted to reduce acquisition noise.

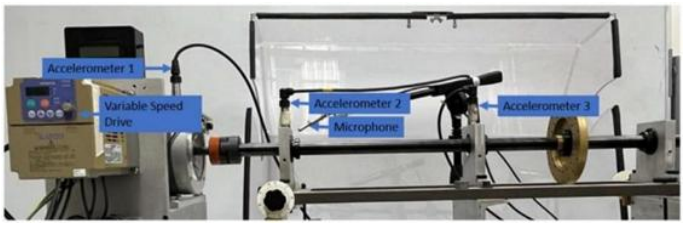


Figure 4. University of Ottawa Laboratory Platform.

Table 1. Motor operating conditions and health states.

Motor health condition	Motor operating conditions and load conditions				
	15-45Hz Load	30-60Hz Unload	15Hz Unload	30-60Hz UnLoad	30Hz Load
Healthy	HL-5-1	HL-6-0	HL-1-0	HL-6-1	HL-2-1
Rotor Unbalance	RU-5-1	RU-6-0	RU-1-0	RU-6-1	RU-2-1
Rotor Misalignment	RM-5-1	RM-6-0	RM-1-0	RM-6-1	RM-2-1
Stator Winding Fault	SW-5-1	SW-6-0	SW-1-0	SW-6-1	SW-2-1
Voltage Unbalance and Single Phasing	VU-5-1	VU-6-0	VU-1-0	VU-6-1	VU-2-1
Bowed Rotor	BR-5-1	BR-6-0	BR-1-0	BR-6-1	BR-2-1
Broken Rotor Bars	KA-5-1	KA-6-0	KA-1-0	KA-6-1	KA-2-1
Faulty Bearings	FB-5-1	FB-6-0	FB-1-0	FB-6-1	FB-2-1

Feature Extraction

A total of 464 motor vibration signals were collected in the experiment, covering the healthy state and 7 typical fault modes under different loads and variable speed conditions. Detailed working condition information is shown in Table 1. Each signal contains 2048 data points, and all samples are divided into 8 categories with 58 signals per category. Based on Hierarchical Sample Entropy (HSE), features are extracted by calculating entropy values from level 1 to 4 for each signal, resulting in a 30-dimensional feature vector. Finally, a 464×30 feature matrix is constructed.

Taking the 15–45 Hz Load dataset in Table 1 as an example, feature selection experiments are conducted. The hierarchical entropy clustering tree generated by Agglomerative Hierarchical Clustering (AHC) is shown in Fig. 5. The horizontal axis represents feature indices composed of 30

extracted HSE features, and the vertical axis represents clustering distance calculated using the Ward distance introduced in Section 3.1.

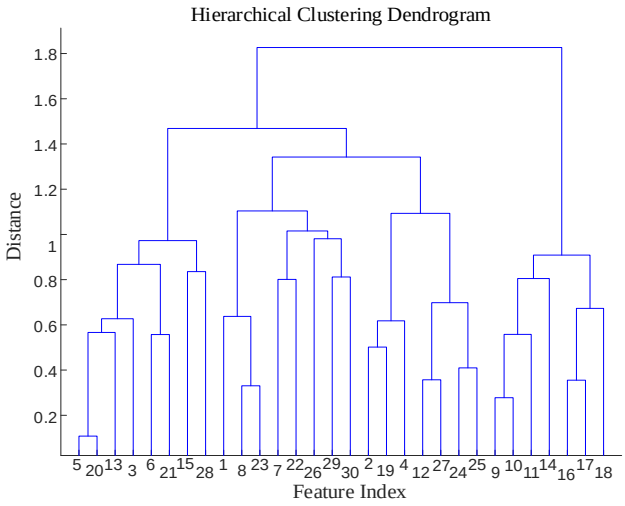


Figure 5. Hierarchical Sample Entropy Clustering Tree Based on AHC.

Feature Selection

After constructing the HSE clustering tree, the optimal cluster count is identified based on the average SC. The SC ranges from −1 to 1, with larger values representing better clustering performance. Fig. 6 shows the variation curve of average SC for cluster numbers ranging from 2 to 10. The peak value corresponds to the number of clusters being 8, so the dendrogram is finally divided into 8 clusters.

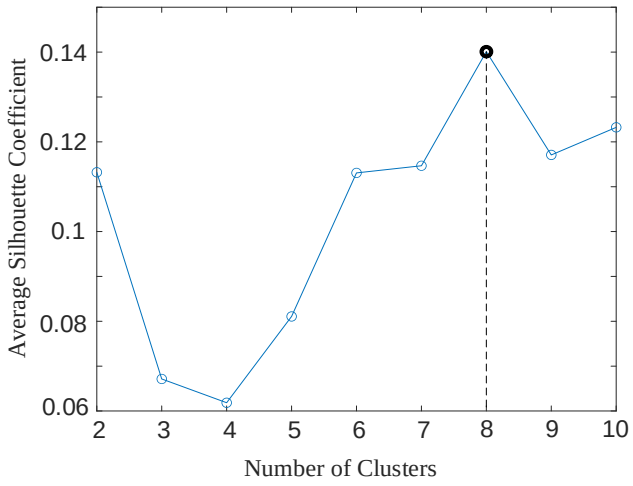


Figure 6. Average SC variation curve for number of clusters from 2 to 10.

Figure 7 shows the specific process of intra-cluster feature selection using LS in the AHC-SCLS feature selection method. First, according to the average SC mentioned above, the optimal number of clusters is determined as k=8, and the dendrogram is divided into 8 sub-clusters. Then, for the features within each sub-cluster, LS is used to score them, and the two features with the lowest scores are selected as the representative features of that cluster to construct the optimal feature subset. As shown in Figure 7, features 14 and 17 are finally selected as the optimal features in the first sub-cluster.

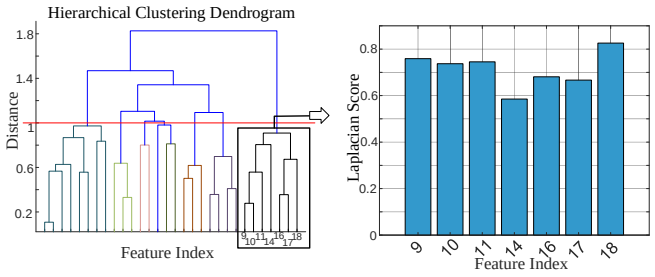


Figure 7. Specific process of selecting optimal features from a sub-cluster using the AHC-SCLS method.

Parameter settings of comparison methods

To ensure a fair comparison, all comparison methods were implemented with explicitly specified parameters and a unified feature dimension. For mRMR, the mutual information quotient (MIQ) criterion was adopted to select features based on maximum relevance and minimum redundancy, while ANOVA ranked features using the F-score derived from one-way analysis of variance.

For PCA, the first 6 principal components with the highest explained variance were retained. For the autoencoder (AE), a single-hidden-layer network with 6 hidden neurons was used for nonlinear feature reduction and trained in an unsupervised manner with 300 epochs, where L2 and sparsity regularization (sparsity proportion of 0.05) were applied to improve generalization.

Experiments

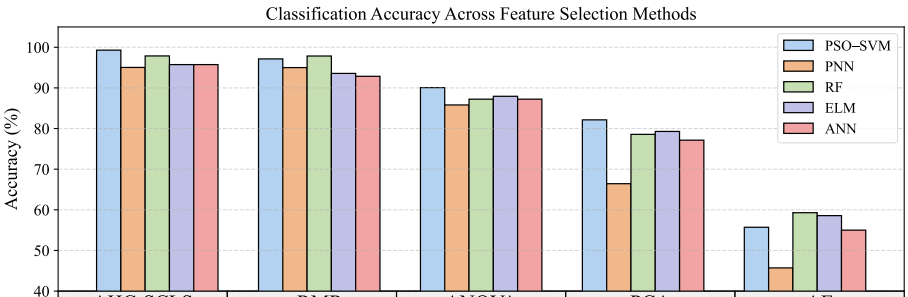
To verify the advantage of AHC-SCLS in HSE feature selection, under the 15–45Hz load condition, this paper uses HSE features as input, and systematically compares AHC-SCLS with four typical dimensionality reduction/selection methods. The experimental results are shown in Fig. 10. As shown in the classification accuracy comparison results in Fig. 10, the AHC-SCLS feature selection method proposed in this paper shows significant advantages across all classifiers. Specifically:

(1) In the PSO-SVM classifier, AHC-SCLS achieved the highest accuracy of 99.29%, an improvement of 2.15 percentage points over the next best method, mRMR

(97.14%);

(2) For classifiers such as PNN, RF, ELM, and ANN, the accuracy of the AHC-SCLS method (ranging from 95.04% to 97.87%) is consistently higher than that of other feature selection methods, with accuracy improvements over traditional dimensionality reduction methods (PCA, AE) reaching 20%–40%;

(3) Among the different feature selection methods, mRMR and ANOVA perform relatively well, while PCA and AE perform significantly worse, especially the AE method, which has a minimum accuracy of only 45.71%.These results fully validate the effectiveness of the AHC-SCLS method in discriminative feature extraction. Through hierarchical clustering and a dual-indicator optimization mechanism, it can significantly enhance the performance of various classification models in mechanical fault diagnosis tasks.



	AHC-SCLS	mRMR	ANOVA	PCA	AE
PSO–SVM	99.29	97.14	90.07	82.14	55.71
PNN	95.04	95.00	85.81	66.43	45.71
RF	97.87	97.85	87.23	78.57	59.29
ELM	95.74	93.57	87.94	79.29	58.57
ANN	95.74	92.86	87.23	77.14	55.00

Figure 8. Comparison results between AHC-SCLS and other feature reduction/selection methods.

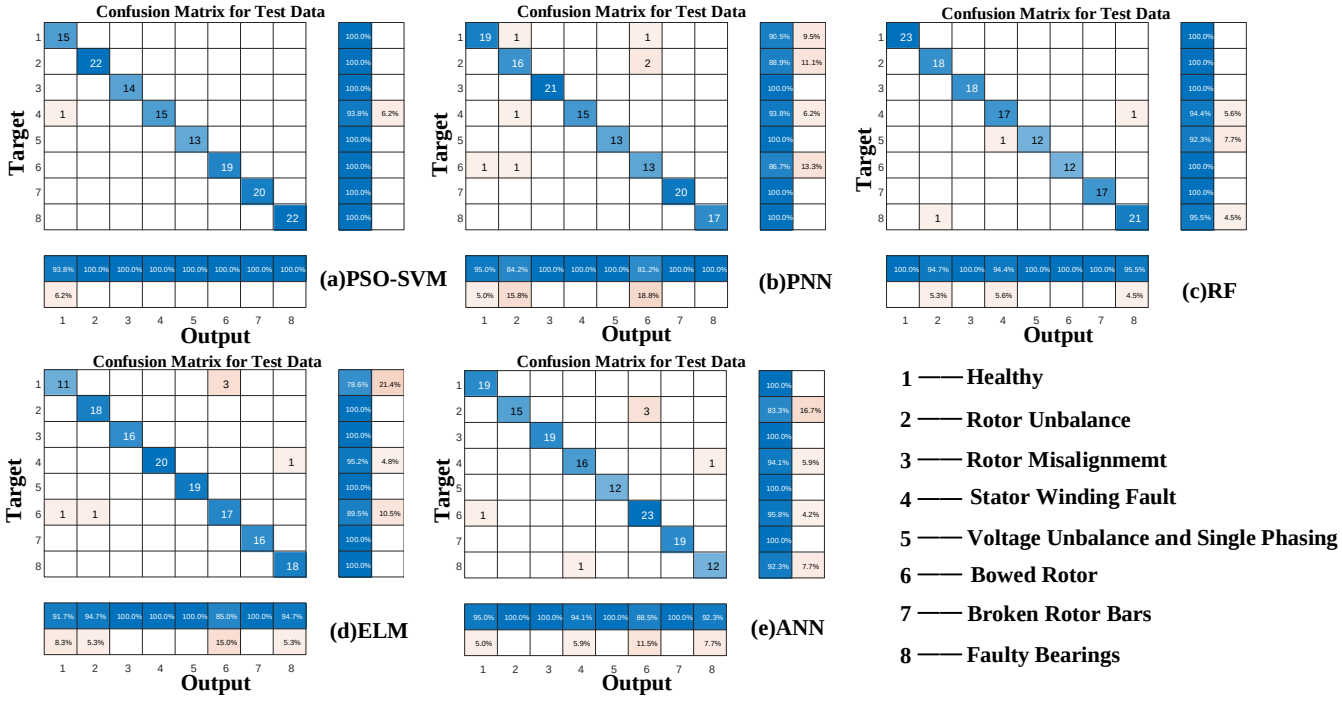


Figure 9. Confusion matrix of model prediction results.

Figure 9 shows the confusion matrices of the proposed method under different classifiers.It can be seen that one sample was misclassified in PSO-SVM; three samples were misclassified in RF; the number of misclassifications in the remaining cases did not exceed 7. This indicates that AHC-SCLS achieves high accuracy in overall pattern recognition classification.

Ablation Study

To verify the necessity and effectiveness of each module in the proposed framework, systematic ablation experiments were conducted based on the multi-level HSE features. Several feature processing schemes with different module combinations were constructed and evaluated using multiple classifiers.

As shown in Table 2, when the Laplacian Score (LS) is directly used for the selection of HSE features, the classification performance remains relatively low and unstable, with accuracies fluctuating roughly between 52% and 75%. This indicates that directly selecting features based solely on local manifold structure is insufficient to address the strong redundancy among multi-scale entropy features.

Table 2. Ablation results under different module combinations.

Method	PSO-SVM	PNN	RF	ELM	ANN
LS	73.57	52.14	75.00	70.71	61.43
AHC+LS	70.21	58.16	67.38	58.87	58.16
AHC+SC	97.16	90.07	96.45	90.07	92.91
AHC+SCLS	99.29	95.04	97.87	95.74	95.74

When hierarchical clustering is introduced without adaptive cluster number optimization (AHC+LS), the performance does not improve and remains at a similarly low level, demonstrating that clustering without a proper global structure constraint cannot form a reliable feature grouping.

By contrast, when the Silhouette Coefficient is used to guide cluster number determination (AHC+SC), the classification performance improves significantly, with accuracies rising to around 90%–97% across classifiers. This shows that a more reasonable clustering structure can effectively enhance feature organization.

When AHC, SC, and LS are jointly integrated into the complete framework, the proposed method achieves the highest and most consistent performance, with accuracies further increasing to approximately 95%–99%. Compared with all incomplete variants, the full framework yields both higher peak performance and reduced fluctuation, indicating that removing any component leads to performance degradation, while the integrated framework provides a more robust and discriminative feature representation.

Validation under other conditions

To verify the effectiveness of the AHC-SCLS-based feature selection method under other conditions, experiments were conducted on eight-class classification tasks under five working conditions as shown in Table 1, namely: 15–45Hz load, 30Hz load, 30–60Hz unload, 15Hz unload, and 45–15Hz unload. The experimental results are shown in Table 3, and the average accuracy is shown in Fig. 10. As can be seen from the figure, the proposed AHC-SCLS method exhibits high classification accuracy under different motor operating and loading conditions, with the lowest average accuracy exceeding 95%. Among them, the best performance is achieved under the 15Hz unload condition, with an average accuracy of 99.72%. This result indicates that the method has good stability and robustness under both constant-speed and variable-speed conditions, and also maintains good classification accuracy under both load and no-load states of the motor.

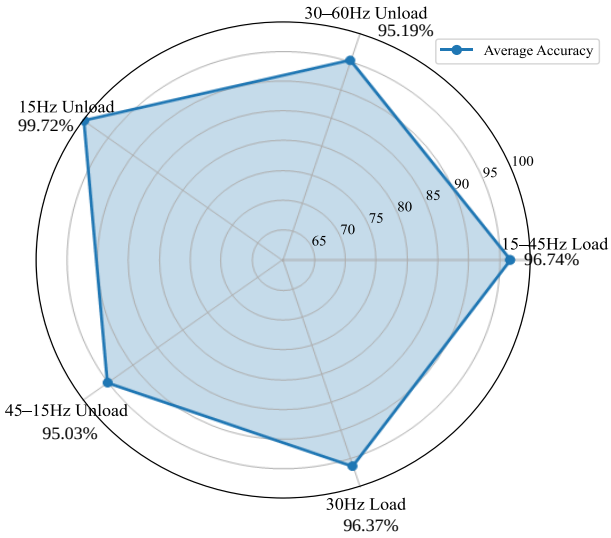


Figure 10. Average classification accuracy under different working conditions.

Table 3. Experimental results of AHC-SCLS method under different working conditions on Ottawa experimental dataset.

Algorithm	Motor operating conditions and load conditions				
	15-45Hz Load	30-60Hz Unload	15Hz Unload	45-15Hz Unload	30Hz Load
PSO-SVM	99.29	95.03	100	95.04	97.87
PNN	95.04	94.47	98.58	94.13	94.63
RF	97.87	98.58	100	97.16	97.16
ELM	95.74	93.23	100	94.61	97.16
ANN	95.74	94.62	100	94.20	95.04

Visualization results

To further validate the effectiveness of the proposed AHC-SCLS feature optimization method, t-SNE(t-Distributed Stochastic Neighbor Embedding) is used to perform three-dimensional visualization of feature data at different stages, as shown in Fig. 11.

The original vibration signals (A) show severe category overlap with no clear clustering structure. After applying the proposed AHC-SCLS method (B), the feature distribution becomes highly compact with clear class boundaries, presenting the best separability. In contrast, PCA (C) yields only partial separation, and the mRMR results (D) show limited improvement, with several fault categories still overlapping.

Overall, the AHC-SCLS method more effectively preserves key fault information, enhances inter-class separability and intra-class consistency, and demonstrates clear advantages in high-dimensional fault diagnosis scenarios.

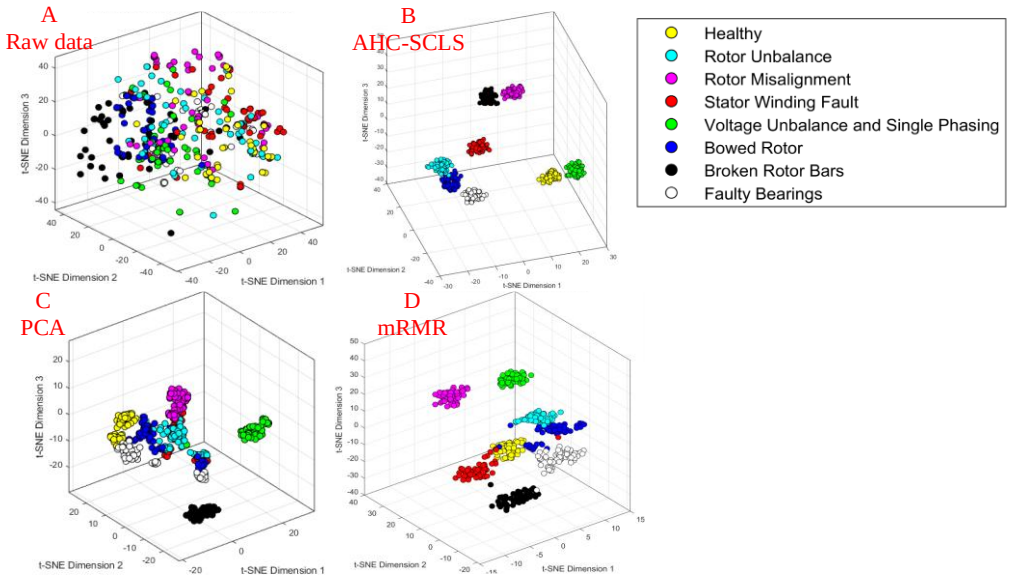


Figure 11. The results of t-SNE.

4.2. Experiment 2: Self-tested wind turbine gearbox vibration dataset analysis

Data description

The experimental data were collected from a self-constructed wind turbine gearbox test platform in the wind power gearbox laboratory at Xiamen. The configuration of the experimental platform is shown in Fig. 12. During the experiments, the driving motor was operated at a constant rotational speed of 300 r/min under an electrical load of 0.1 Ω.

Four different health conditions were considered in the dataset, including a normal condition, a cracked planetary gear, a broken planetary gear tooth, and a broken sun gear tooth. The fault conditions were artificially introduced by machining or pre-damaging the corresponding gear components prior to the experiments.

Vibration signals were collected using accelerometers mounted on the gearbox housing near the planetary gear stage, as shown in Fig. 12. Only vibration signals were used for subsequent analysis, and all vibration data were sampled at a frequency of 20 kHz.

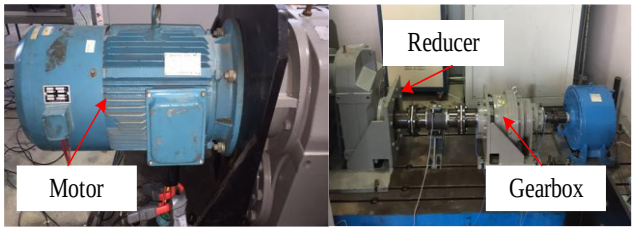


Figure 12. Self testing wind turbine gearbox Laboratory Platform.

Table 4. Experimental condition patterns for the wind turbine gearbox.

Pattern No	Fault location	Degree of failure
1	Normal	Normal
2	Planetary wheel	Fracture (~0.50 mm wide, ~0.30 mm deep)
3	Planetary wheel	Tooth entirely fractured
4	Sun gear	Tooth entirely fractured

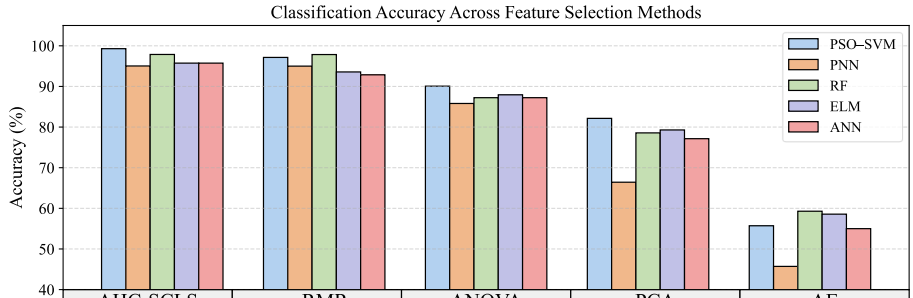
Experiments

To validate the effectiveness of the AHC-SCLS method, the same experimental scheme as in Section 4.1 was adopted. Comparative tests were conducted on the wind turbine gearbox vibration dataset, and the experimental results are shown in Fig. 13. According to the results, the AHC-SCLS method achieved 100% accuracy under PSO-SVM, PNN, ELM, and ANN classifiers, and also performed excellently under RF classifier

with 98.59%, outperforming all comparative methods.

In contrast, AE and PCA, as dimensionality reduction methods, performed relatively poorly with generally lower classification accuracy. Although mRMR and ANOVA achieved good results under some classifiers, their accuracy dropped to 88.57% and 94.29% respectively under RF, which was still inferior to the stability and robustness of the AHC-SCLS method.

The overall comparison shows that the AHC-SCLS method can effectively retain discriminative and low-redundancy key features during the feature selection phase, significantly improving the diagnostic performance of various classification models, highlighting its superiority in complex signal analysis and fault identification tasks.



	AHC-SCLS	mRMR	ANOVA	PCA	AE
PSO–SVM	99.29	97.14	90.07	82.14	55.71
PNN	95.04	95.00	85.81	66.43	45.71
RF	97.87	97.85	87.23	78.57	59.29
ELM	95.74	93.57	87.94	79.29	58.57
ANN	95.74	92.86	87.23	77.14	55.00

Figure 13. Experimental results of wind turbine gearbox data.

5. Conclusion

This paper proposes a mechanical fault feature selection method based on HSE and AHC-SCLS clustering quality optimization. This method utilizes HSE to capture the nonlinear dynamic differences of vibration signals over a wide frequency band, and combines AHC to construct a feature clustering tree that hierarchically merges redundant features. With the SCLS dual-constraint feature selection mechanism, the approach achieves collaborative optimization of global clustering quality and local feature importance, providing high-robustness and low-redundancy input features for subsequent fault diagnosis.

Experimental validations on the public dataset from the University of Ottawa and the self-tested wind turbine gearbox dataset show that the proposed method exhibits superior classification performance across multiple classifiers such as PSO-SVM and Random Forest, with recognition accuracy significantly outperforming traditional feature selection methods. Especially under complex working conditions such as variable load and variable speed, the method can still maintain stable diagnostic performance.

Although the proposed method has achieved remarkable results on experimental datasets, further research will be conducted to enhance its engineering applicability: 1)Integrate transfer learning techniques to address the issue of feature transfer across devices and working conditions; 2)Develop early fault warning capabilities to build a complete predictive maintenance system.

Acknowledgements

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Natural Science Foundation of Fujian Province，China (Grant no. 2025J01131029)，Foreign cooperation project of Fujian Province's science and technology plan, China (Grant no. 2024I0044)，Nation Natural Science Foundation of China (Grant Nos. 52375087).

References

1. Song Y, Cao J, Hu Y. In-process feature extraction of milling chatter based on second-order synchroextracting transform and fast kurtogram. Mechanical Systems and Signal Processing. 2024;208:111018. https://doi.org/10.1016/j.ymssp.2023.111018.

2. Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences. 1991;88(6):2297–2301. https://doi.org/10.1073/pnas.88.6.2297.
3. Richman JS, Moorman JR. Physiological time series analysis using approximate entropy and sample entropy. American Journal of Physiology, Heart and Circulatory Physiology. 2000;278(6):H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.H2039.
4. Cao Y, Tung WW, Gao JB, Protopopescu VA, Hively LM. Detecting dynamical changes in time series using the permutation entropy. Physical Review E. 2004;70(4):046217. https://doi.org/10.1103/PhysRevE.70.046217.
5. Wang Z, Zhang M, Chen H, Li J, Li G, Zhao J, Yao L, Zhang J, Chu F. A generalized fault diagnosis framework for rotating machinery based on phase entropy. Reliability Engineering & System Safety. 2025;256:110745. 10.1016/j.ress.2024.110745.
6. Wang Z, Liu Y, Bai R, Chen H, Li J, Chen X, Yao L, Zhao J, Chu F. Multi-modal multi-scale multi-level fusion quadrant entropy for mechanical fault diagnosis. Expert Systems with Applications. 2025;281:127715. https://doi.org/10.1016/j.eswa.2025.127715.
7. Shen C, Jiang P, Luo J, Lin G, Zhang S. Optimal weighted multi-scale entropy-energy ratio feature for machine fault diagnosis. Measurement. 2025;242:115782. https://doi.org/10.1016/j.measurement.2024.115782.
8. Wang Z, Liang P, Bai R, Liu Y, Zhao J, Yao L, Zhang J, Chu F. Few-shot fault diagnosis for machinery using multi-scale perception multi-level feature fusion image quadrant entropy. Advanced Engineering Informatics. 2025;63:102972. https://doi.org/10.1016/j.aei.2024.102972.
9. Lu N, Zhou TX, Wei JF, Yuan WL, Li RQ, Li ML. Application of a whale optimized variational mode decomposition method based on envelope sample entropy in the fault diagnosis of rotating machinery. Measurement Science and Technology. 2021;33(1):015014. 10.1088/1361-6501/ac3470.
10. Wang Z, Yao L, Cai Y. Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine. Measurement. 2020;156:107574. https://doi.org/10.1016/j.measurement.2020.107574.
11. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters. 2002;89(6):068102. https://doi.org/10.1103/PhysRevLett.89.068102.
12. Jiang Y, Peng CK, Xu Y. Hierarchical entropy analysis for biological signals. Journal of Computational and Applied Mathematics. 2011;236(5):728–742. https://doi.org/10.1016/j.cam.2011.06.007.
13. Zhou X, Yuan R, Lv Y, Li B, Fu S, Li H. Hierarchical multiscale fluctuation-based symbolic fuzzy entropy: A novel tensor health indicator for mechanical fault diagnosis. IEEE Sensors Journal. 2025;25(3):5013–5029. https://doi.org/10.1109/JSEN.2024.3512654.
14. Su Z, Tang B, Liu Z, Qin Y. Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. Neurocomputing. 2015;157:208–222. https://doi.org/10.1016/j.neucom.2015.01.016.
15. Zhao K, Xiao J, Li C, Xu Z, Yue M. Fault diagnosis of rolling bearing using CNN and PCA fractal based feature extraction. Measurement. 2023;223:113754. https://doi.org/10.1016/j.measurement.2023.113754.
16. Huang J, Yan X. Related and independent variable fault detection based on KPCA and SVDD. Journal of Process Control. 2016;39:88–99. https://doi.org/10.1016/j.jprocont.2016.01.001.
17. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987;20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
18. Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed. New York: Wiley; 2012.
19. Zhao S, Liang X, Wang L, Zhang H, Li G, Chen J. A fault diagnosis method for analog circuits based on EEMD–PSO–SVM. Heliyon. 2024;10(18):e21601. 10.1016/j.heliyon.2024.e38064.
20. Li Z, Wang Q, Xiong J, Cen J, Dai Q, Liang Q, Lu T. A building electrical system fault diagnosis method based on random forest optimized by improved sparrow search algorithm. Measurement Science and Technology. 2024;35(5):055110. 10.1088/1361-6501/ad2255.
21. Li M., Yan C., Liu W., Liu X., Zhang M., Xue J., Fault diagnosis model of rolling bearing based on parameter adaptive AVMD algorithm, Applied Intelligence, 2023;53(3):3150–3165, doi:10.17531/ein/163547.
22. Zhao Y, Yang X, Huang J, Gao J, Cui J. Improved weighted extreme learning machine with adaptive cost-sensitive strategy for imbalanced fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing. 2024;217:111526. https://doi.org/10.1016/j.ymssp.2024.111526.
23. Melnik R., Koziak S., Dižo J., Kuźmierowski T., Piotrowska E., Feasibility study of a rail vehicle damper fault detection by artificial neural

networks. Eksploatacja i Niezawodność – Maintenance and Reliability. 2023;25(1). doi:10.17531/ein.2023.1.5.

24. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;27(8):1226–1238. https://doi.org/10.1109/tpami.2005.159.
25. Miller GR. Beyond ANOVA – Basics of Applied Statistics. Boca Raton: Chapman & Hall/CRC; 1997.
26. Patange A, Soman R, Pardeshi S, Kuntoglu M, Ostachowicz W. Milling cutter fault diagnosis using unsupervised learning on small data: A robust and autonomous framework. Eksploatacja i Niezawodność – Maintenance and Reliability. 2024;26(1),doi:10.17531/ein/178274.
27. Niu M, Jiang H, Wu Z, Shao H. An enhanced sparse autoencoder for machinery interpretable fault diagnosis. Measurement Science and Technology. 2024;35(5):055108. 10.1088/1361-6501/ad24ba.
28. Sehri M, Dumond P. University of Ottawa constant and variable speed electric motor vibration and acoustic fault signature dataset. Data in Brief. 2024;53:110144. https://doi.org/10.1016/j.dib.2024.110144.
29. Luo J, Chen Y, Huang Q, Zhang X, Li H, Wei Z, Deng L. Joint application of VMD and IWOA–PNN for gearbox fault classification via current signal. IEEE Sensors Journal. 2023;23:13155–13164. https://doi.org/10.1109/JSEN.2023.3269594.