ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on Acoustic Signal Feature Selection for Fault Diagnosis of Petroleum Drilling Conductive Slip Rings

Bin Lv[1], *Keqin Zhao*[2], *Xingxing Xu*[2], *Hongming Ni*[1], *Jian Zhou*[1] *and Fei Gao*[1,*]

[1] CSSC Jiujiang Jingda Technology Co., Ltd., Jiujiang 332000, Jiangxi, China
[2] Jiujiang Precision Measurement and Testing Technology Research Institute, Jiujiang 332000, Jiangxi, China

SUMMARY: *Under conditions of long-term operation of high-power conductive slip rings with high current and strong electromagnetic interference, and rotational contact coupling, the acoustic radiation characteristics of the housing may be altered due to bearing ball/raceway damage, spindle fatigue cracks, dynamic eccentricity, and brush jumping. Given the problems of high dimensionality, strong redundancy and a small sample size in the multi-scale candidate features obtained through empirical mode decomposition of the original audio signal, this paper enhances the Lasso feature selection method based on 60kW fault simulation platform data and introduces a segmented adaptive Lasso feature selection strategy. Lasso has sparse filtering properties, so a corresponding way was proposed to reduce the bias in coefficients resulting from uniform penalization; this method progressively performs filtering on groups of features to reduce the dimension of the single regression solution while retaining the traceability relationship among feature number, selected variables, and classification results. Five types of operational audio data were obtained from a 60kW high-power conductive slip ring fault simulation platform as the objects, and a 45-dimensional feature set of the original signal and IMF1-IMF8 was constructed; ten-fold cross-validation was performed on KNN, MLP and SVM classifiers. The results show that the segmented adaptive Lasso can provide good classification results in all five states, with an average accuracy of 0.873 when n=10; Compared with Lasso, its accuracy increased by 0.034, 0.010, 0.040 and 0.075 under normal, spindle fatigue crack, dynamic eccentricity and brush jumping conditions, respectively, and the average calculation time decreased by 38.61%. Improve the generalisation ability and real-time performance under high-dimensional small-sample conditions while keeping the interpretability of acoustic features.*

KEYWORDS: *High power conductive slip ring; sound signal; Fault diagnosis; Empirical Mode Decomposition; Distributed adaptive Lasso*

1 Introduction

Marine drilling, heavy-duty rotating platforms and high-power transmission equipment all have relatively high demands for continuous power supply capability and signal transmission stability of conductive slip rings. High-power conductive slip rings are used to transmit electrical energy and control signals between the rotating end and the fixed end, and at the contact pair, bearings, spindle, brush, etc., various mechanical loads, thermal effects from electrical contact, and environmental disturbances occur simultaneously. If the bearing

*LeonChang0226@163.com
https://doi.org/10.65102/is2026896
Published: 30 April, 2026

clearance of the slip ring increases, the main shaft will crack, dynamic eccentricity or electric brush jumping will occur; the contact pressure and local friction state will change, and consequently, fluctuations in contact resistance, abnormal heating, discharge and mechanical vibration will arise. For continuous-operation equipment such as offshore drilling platforms, such failures often do not have a large downtime maintenance window, and condition monitoring and early diagnosis directly impact unplanned downtime costs and operational safety. Research on the condition monitoring of mechanical equipment has gradually moved from a single threshold alarm to data-driven health assessment, and diagnostic models now need to consider fault sensitivity, online computing cost and interpretability of results simultaneously [1, 2].

Most studies on the fault diagnosis of conductive slip rings have focused on electrical parameters, temperature, vibration and contact status. Ma and others have investigated the different damage states of the surface of the wind power excitation slip ring using vibration signals, and determined that mechanical contact abnormalities can be observed as changes in vibration bandwidth [3]. Turel and others have studied the effect of rotational speed, current and temperature on the contact resistance and wear of metal-graphite brushes under dynamic excitation, and found that there is a certain electromechanical coupling at the micro-contact level between the brush and the slip ring [4]. The type of research above provides a foundation for locating slip ring contact faults; however, under conditions of large currents and severe electromagnetic interference or restricted installation spaces, it is difficult to install contact sensors densely. Only electrical and vibration quantities are used, so the differences between brush jumping, raceway impact and friction sound cannot be fully distinguished.

Sound can be used to observe whether the conductive slip ring is in good working order. Research on the fault diagnosis of rolling bearings has shown that crack propagation, ball impact and surface peeling can generate modulation components and impact pulses in the acoustic or vibration response [5]. Acoustic measurement can be used to obtain radiation information near the rotating component in a non-contact manner, and at the same time, a safe distance can be maintained between the sensor and the high-voltage and high-current circuit compared with contact vibration sensors installed on the housing. Hou et al. have enhanced weak impact features for acoustic diagnosis of bearings using the sparse guided MOMEDA method; Li et al. have extracted rolling bearing fault sound source features with a coprime circular microphone array; and Luo et al. have further improved diagnostic performance in the presence of noise through an acoustic feature learning network [6-8]. The above works show that acoustic signals can be used to diagnose faults in rotating machinery, and therefore, it is necessary to focus on the distribution of acoustic features at different time scales and frequencies in this article.

Based on the structural components of the conductive slip ring, the bearing, spindle and brush contact pair are not independent causes of failure. The bearing raceway will affect the rotor support condition; dynamic eccentricity will alter the contact pressure of the brush ring; and jumping of the electric brush may increase fluctuations in contact resistance and local friction noise. Therefore, acoustic diagnosis should not be limited to a single impact indicator; rather, the distribution of amplitude, frequency centroid, spectral diffusion and component energy proportion also need to be observed simultaneously. Although the above problems have been addressed by a multi-scale feature system, it is also prone to feature redundancy and variable correlation.

Although the intelligent diagnostic model construction has expanded applications for complex signals, problems at the feature-construction stage have also emerged. Deep models can learn representations from raw signals, but when the sample size is small, fault categories are few, and on-site noise is complicated [9], the training cost and interpretability of the model

are still relatively high. Traditional machine learning models, such as KNN, SVM and MLP, have relatively small sample size requirements and are still applicable when engineering experimental data is lacking. At this time, the quality of feature extraction and feature selection will directly affect whether the following classifier operates normally. Many kinds of features can be extracted from audio signals in the time domain, frequency domain, energy domain and decomposed components. As the number of candidate features increases, some may be redundant or weakly correlated, which can lead to an overfitted model and thus perform poorly in classification.

Reduce redundant dimensions and keep the fault information by feature selection. Guyon and Elisseeff have shown that, by using feature selection, it is possible to improve the predictive ability of models, reduce the computational overhead and increase their interpretability of the data-generating process [10]. Cai and others have summarized supervised, unsupervised, and semi-supervised feature selection methods based on machine learning, and pointed out that invalid variables need to be eliminated during high-dimensional data processing [11]. Chandrashekar and Sahin have divided the three categories of feature selection methods into filtering, wrapping and embedding. Among them, embedded methods perform variable selection simultaneously with model training and generally have good trade-offs in terms of computational cost and classification accuracy [12]. Lasso regression uses an L1 penalty to achieve sparsity by directly reducing the regression coefficients of less important features, and is thus suitable for acoustic candidate feature screening.

Fault sound signals of high-power conductive slip rings are also divided into different categories and have distinct coupling characteristics depending on the operating environment. Under normal circumstances, acoustic response consists mainly of stable rotation and environmental noise; rolling track damage and brush jumping may cause intermittent impact; there is a local overlap between the periodic modulation induced by dynamic eccentricity and the frequency shift due to spindle fatigue cracks. If the feature selection method only considers the sparsity of the training set, weak features that are suitable for a certain type of fault will be lost. The reason for dividing the candidate features into lower-dimensional subspaces and performing gradual judgments in the segmented adaptive strategy is to reduce the penalty intensity of important variables through weight coefficients; thus, it is more suitable for experimental environments that contain dispersed acoustic fault information and have a limited sample size.

For the acoustic diagnosis of high-power conductive slip rings, the candidate features usually come from the original waveform and several decomposed components, and there is a clear correspondence between feature numbers, component sources, and acoustic meanings. If only the final accuracy of the main text is reported, it is impossible to know what acoustic changes the model used to achieve this discrimination, and it is also unknown whether feature selection only removed low-scoring variables. Therefore, this paper will use the same set of evidence to treat the experimental platform, raw audio signal, EMD component, Spearman score, correlation coefficient matrix, and selected feature number as one chain, thereby providing an acoustic mechanism explanation for the feature selection results. Lasso bias, adaptive weighting and partial iteration mechanisms also need to be developed in the method section to ensure that the improvement sources correspond to subsequent classification accuracy, convergence and computation time.

Based on the above problems, this paper will focus on the research of feature selection for high-power conductive slip ring audio signals. First, a 60 kW high-power conductive slip ring fault simulation platform is used to obtain acoustic data in a normal state and four typical fault states, and the original signals and EMD decomposition diagrams of the bearing ball/raceway damage samples are provided. Second, construct 45-dimensional acoustic features for the

original signal and IMFs1-IMFs8, and then combine Spearman scores and correlation coefficient matrices to evaluate the feature contribution and redundancy among them. Finally, the full feature set, Lasso, and segmented adaptive Lasso are used to train KNN, MLP, and SVM classifiers, and their classification accuracy, convergence, computation time and diagnostic efficiency are compared. The three contributions of this paper are as follows: first, to build an interpretable multi-scale audio feature numbering system; second, to put forward a segmented adaptive Lasso filtering strategy for high-dimensional small-sample acoustic features; and third, to verify the feature selection results through accuracy, time and selected feature maps.

2 Methods

2.1 Experimental Platform and Acoustic Data Organization

Based on a 60 kW high-power conductive slip ring fault simulation experimental platform, the method was verified in this paper. The experimental objects are high-power conductive slip rings and their bearings, main shafts and brush contact parts. The experimental platform is as follows: Figure 1.

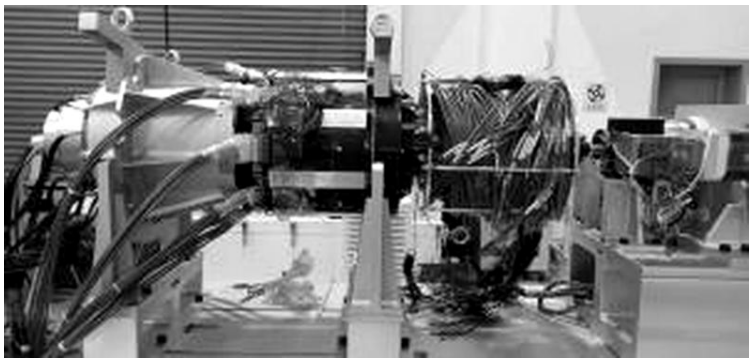

Figure 1: 60 kW High-power conductive slip ring fault simulation experimental platform

As shown in Figure 1, a closed-loop test environment for the slip ring body, drive end, load end and sensing acquisition unit can be established to simulate normal operation and typical mechanical contact faults in the presence of power frequency noise. A simple bench-vibration test cannot replicate the rotational contact, structural support and external noise conditions that occur during high-power operation of a conductive slip ring, so it is not suitable for engineering diagnostic applications.

The five types of equipment states for the experiment are: normal state, bearing ball/raceway damage, spindle fatigue crack, dynamic eccentricity, and brush jumping. Collect 100 sets of audio data per category, for a total of 500 sets, and each group is 30 seconds long. A fibre-optic acoustic emission sensor is used for the acoustic sensor, has a frequency response range of 20 kHz - 1 MHz, is non-contact type, and is mounted externally on the slip ring housing at a distance of no more than 50 mm for monitoring. The above arrangement can avoid a strong electromagnetic interference generated by a slip ring working current of 2000A and is also able to receive bearing impact, brush ring friction and structural transmission noise. The collection end has a 16-channel data acquisition system, samples at a rate of 30 kHz, and has 22,630 effective samples per segment. Power frequency noise in the experiment is approximately 87 dB.

The five types of fault mechanisms are different, and thus the effective information of audio response will not be concentrated in a single frequency band. Ball/raceway damage to bearings

usually presents as periodic impacts and short-term energy fluctuations; fatigue cracks in the spindle change local stiffness and frequency distribution; dynamic eccentricity results in an increase in periodic modulation components; and electric brush jumping is related to intermittent contact, micro-arc and friction state changes. To prevent the classification model from learning only a single amplitude scale or a local frequency band, each audio segment is treated as an independent sample, feature scale estimation is completed in the training set, and then the same scale transformation is applied to the test set. Standardised parameters do not use test set statistics to avoid premature leakage of test information.

Before gathering data for feature extraction, this paper does not add extra data augmentation to the different fault states to keep the acoustic differences from the original bench experiments. First determine if the number of sampling points in a sample meets the uniform-length requirement, and then set a sample index based on the state label. After completing the standardised parameter calculation in the training set, the test set only calls the mean and standard deviation of the training set for scale transformation. Processing that is relatively more sensitive to the regression penalty term and thus to the size of the coefficients is needed for Lasso-type methods. If the dimensions of the various features are significantly different, the weight of the penalty will be too small for the features with small value ranges. The distance measure of KNN and the interval estimation for SVM are also affected by the dimension, so standardisation needs to be performed before feature selection and classification training.

In the sample organisation step, the same quantity of data is kept for all states to prevent the classifier from being affected by the small number of samples in some regions due to imbalance. Each 30-second audio segment is first divided into effective segments, and then a fixed number of sampling points is used for feature extraction. Given that there is about 87 dB of power-frequency noise in the experimental environment, the original waveform is not directly input to the classifier in this paper. Acoustic information is divided into parts and statistical features for compression instead. Processing can reduce the influence of background noise on instantaneous waveforms and retain the interpretable variables of fault impact, spectral diffusion and energy allocation. All the training set, test set, and cross-validation folds are divided by the same rule for state partitioning, so the different feature selection methods are presented with the same data boundaries.

To ensure the comparability of audio features across different states, a fixed sampling length, decomposition order and numbering rule have been adopted at the feature level in this paper. The raw signal is the unprocessed audio before EMD decomposition, and IMF1 to IMF8 are the individual intrinsic mode functions sorted by increasing frequency. Each set of states obtains a 45-dimensional feature vector through the same process, and the classification labels only represent the device state and do not participate in feature normalisation parameter estimation. This way, we can avoid differences in data-processing metrics among the methods, and thus, the performance disparities of full features, Lasso, and segmented adaptive Lasso will be attributed solely to their feature-selection strategies.

2.2 EMD Decomposition and Construction of 45-Dimensional Candidate Features

The fault sound signal of a conductive slip ring is non-stationary, and the same signal may contain low-frequency structural vibration, high-frequency impact and broadband friction noise simultaneously. Empirical mode decomposition can adaptively separate several intrinsic mode functions according to the local extrema of a signal, and is suitable for handling non-linear and non-stationary time series [13]. The damage status of bearing balls/raceways is used as an example, and the original audio signal is shown in Figure 2.

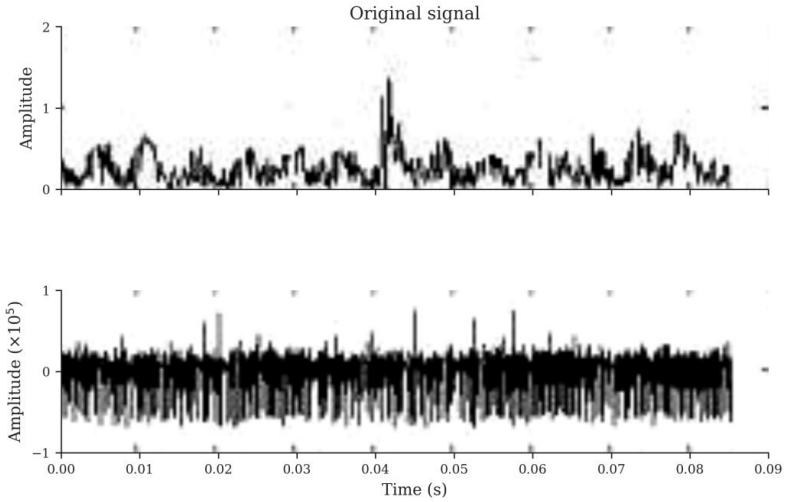

Figure 2: Raw Acoustic Signal with Bearing Ball/Raceway Damage

As shown in Figure 2, the upper wave shape is the result of the change in the original amplitude over time, and the lower wave shape is the densely distributed high-frequency disturbance at a sampling point. The maximum value of this kind of signal is not uniform; therefore, the combination of fault impact and background noise has a multi-scale distribution.

As shown in Figure 2, after EMD decomposition of the original audio signal, 8 IMF components can be obtained, as shown in Figure 3.

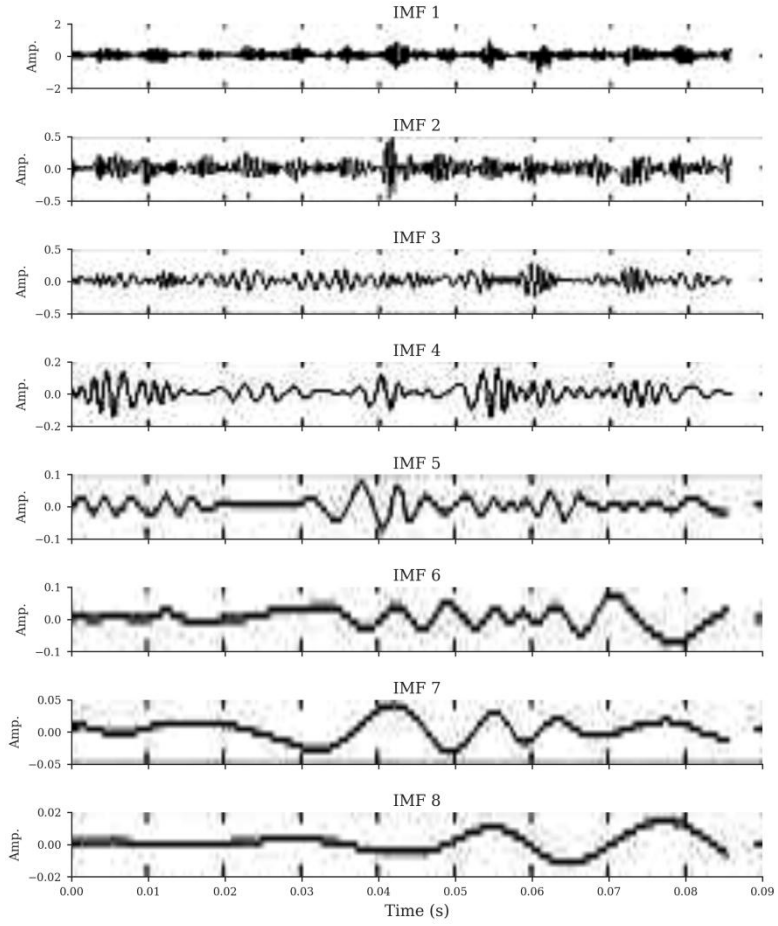

Figure 3: IMF Components Obtained by EMD Decomposition of the Raw Acoustic Signal

As shown in Figure 3, IMF1 and IMF2 have relatively strong high-frequency shocks and local oscillations; the amplitude fluctuations in IMF3 to IMF5 are relatively gentle, indicating more intermediate-frequency modulation and structural transmission components; IMF6 to IMF8 have smaller amplitudes and mainly carry low-frequency trends and residual components. The decomposed multi-scale structure provides component-level objects for the next feature extraction step, and thus the fault diagnosis is no longer limited to the overall statistics of the original signal.

Five types of acoustic features are extracted from the original signal and IMF1-IMF8 in this paper to form a 45-dimensional candidate feature set, as shown in Table 1. The five kinds of features are kurtosis, peak coefficient, centre frequency, frequency variance and energy coefficient. Kurtosis can be used to describe the effect of impact pulses on the tail of the amplitude distribution; damage to bearing balls/raceways and brush runout generally increases the response of this feature. The peak coefficient is the relative relationship between the extreme amplitude and the root mean square level, and it can differentiate between continuous friction noise and intermittent bouncing noise. The center frequency represents the centre of gravity of acoustic energy distribution and is suitable for observing frequency shifts due to crack propagation or eccentricity; frequency variance indicates the degree of spectral diffusion and is sensitive to broadband impact and modulation response; and the energy coefficient reflects the contribution of a specific component to the total acoustic energy. Apply the above characteristics to Raw and IMF1-IMF8 simultaneously to avoid detecting the fault state in a single time scale.

Table 1: Numbering of 45-Dimensional Features from the Raw Signal and IMF Components

Feature	Raw	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8
K	1	6	11	16	21	26	31	36	41
CF	2	7	12	17	22	27	32	37	42
FC	3	8	13	18	23	28	33	38	43
VF	4	9	14	19	24	29	34	39	44
E	5	10	15	20	25	30	35	40	45

Keep the feature numbers for the following interpretation. For example, feature 4 is the frequency variance of the original signal, feature 9 is the frequency variance of IMF1, and feature 28 is the centre frequency of IMF5. If a certain method repeatedly selects the same feature family and focuses on a few IMF components, it suggests that the model's discrimination is mainly due to specific frequency bands; if the selected features are distributed simultaneously in the original signal and multiple IMF components, it indicates that the fault information occurs at different time scales. Next, based on the above numbering system, a graphical analysis of the selected features in the damaged state of bearing balls/raceways will be performed in this article.

The five kinds of features are all based on the same signal segment or IMF component before calculation. Kurtosis and peak coefficient focus on describing the time-domain amplitude distribution, are more sensitive to local impacts and intermittent contacts; center frequency and frequency variance focus on describing the energy structure in the frequency domain, and are more sensitive to frequency band transitions caused by cracks, eccentricity and changes in friction states; the energy coefficient is used to describe the proportion of component energy and can reflect the distribution of fault excitation in different IMFs. This combination of features does not have too high a dimension, but it still retains statistical data that are directly related to mechanical contact faults of slip rings, so the features selected by Lasso-like methods can be traced back to specific acoustic meanings.

Feature organisationally, the 45-dimensional vector consists of two kinds of complementary information. The first type is multi-index information for the same signal object; that is, kurtosis, peak coefficient, centre frequency, frequency variance and energy coefficient are all available in the raw signal and describe different attributes of the object's amplitude and spectral distribution. The second type is the scale transfer information of the same indicator in different components; for example, the changes in frequency variance in Raw, IMF1, IMF3 and IMF8 can reflect the process of fault energy diffusion from high-frequency impact to low-frequency structural response. If the following selected features cross both types of information, it can be inferred that the model is using both local impact strength and energy distribution among components.

2.3 Partitioned Adaptive Lasso Feature Selection and Evaluation Protocol

Among the 45-dimensional candidate features, some are highly correlated with the fault category, and others only show environmental noise or repetitive information among different parts. If all features are directly input into the classifier, the model needs to determine the classification boundary in a high-dimensional space and is likely to overfit or have a large computational cost. Lasso regression obtains sparse coefficients using the L1 penalty and is a typical embedded feature selection method [14]. Assuming that a linear approximation relationship exists between the sample response and the candidate features, it can be expressed as:

$$y_i = \alpha + \sum_{j=1}^{p} x_{ij}\,\beta_j + \varepsilon_i, \quad i = 1,2,\ldots,n$$

In the formula, y_i is the class response encoding of the ith sample, x_{ij} is the jth dimensional candidate feature of the ith sample, β_j is the regression coefficient of the jth dimensional feature, α is the constant term, ε_i is the random error, n is the number of samples, and p is the number of candidate features. In this article, p=45. Lasso estimates sparse coefficients by minimizing the residual term and L1 penalty term, with the objective function being:

$$\widehat{\boldsymbol{\beta}}^{L} = \arg\min_{\alpha,\boldsymbol{\beta}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \alpha - \sum_{j=1}^{p} x_{ij}\,\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

In the equation, $\widehat{\boldsymbol{\beta}}^{L}$ is the estimated Lasso regression coefficient, $\boldsymbol{\beta}$ is the vector of coefficients to be estimated, and λ is the regularization parameter. As λ increases, more coefficients are compressed to 0 and corresponding features are removed. This mechanism can reduce the input dimension, but Lasso applies the same penalty intensity to all features, and important features will also be compressed to the same extent, resulting in biased coefficient estimation. For high-dimensional small sample acoustic features, this bias may cause weak fault sensitive features to be removed in advance.

Adaptive Lasso changes the strength of the penalty for different features by using weight coefficients, thus reducing the bias issue of uniform penalties in Lasso [15, 16]. Use the initial estimation coefficients to construct weights in this paper:

$$w_j = \frac{1}{\left(|\hat{\beta}_j^{(0)}| + \delta \right)^{\gamma}}, \quad \gamma > 0,\ \delta > 0$$

In the formula, w_j is the penalty weight corresponding to the jth dimensional feature, $\hat{\beta}_j^{(0)}$ is the initial estimation coefficient, γcontrols the strength of weight decay, and δ is a stable constant to prevent the denominator from being zero. If a certain feature shows a significant contribution in the initial estimation, then w_j is smaller and its subsequent penalty intensity is weakened; If the initial estimate approaches 0, w_j increases, making the feature easier to compress. The adaptive Lasso objective function is written as:

$$\widehat{\boldsymbol{\beta}}^A = \arg\min_{\alpha,\boldsymbol{\beta}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \alpha - \sum_{j=1}^{p} x_{ij}\, \beta_j \right)^2 + \lambda \sum_{j=1}^{p} w_j \, |\beta_j| \right\}$$

In the equation, $\widehat{\boldsymbol{\beta}}^A$ $is the adaptive Lasso coefficient estimation. This formula preserves the sparsity of L1 penalty, while setting differentiated penalties for different features through w_j, which helps to reduce the shrinkage bias of important feature coefficients. For conductive slip ring audio signals, this design can preserve the high contribution features of amplitude impact, frequency transfer, and energy redistribution, while reducing dependence on noise components.

Based on the research on Lasso solution path and grouping variable selection [17, 18], this paper introduces a distributed computing strategy. Divide the 45-dimensional candidate features into n feature groups in numerical order, denoted as G1, G2, ..., GN. Divisional filtering is not solved in the full 45-dimensional space at once, but in step k, the previously selected features are combined with the current feature set to form a candidate set:

$$S_k = A_{k-1} \cup G_k, \quad A_k = supp\left(\widehat{\boldsymbol{\beta}}_{S_k}^A\right), \quad k = 1,2,\dots,n$$

In the formula, S_k is the candidate set selected in the k-th step, A_{k-1} is the feature set retained in the previous round, G_k is the k-th feature set, A_k is the non-zero coefficient feature set output in the k-th step, and $supp(\cdot)$ represents the non-zero coefficient support set. After the iteration is completed,A_n is the final selected feature. Compared with solving all features at once, the distributed strategy reduces the feature dimension of a single round regression and maintains the previous effective information by gradually carrying the selected features. This mechanism is suitable for acoustic diagnostic scenarios such as this one, where the number of candidate features is not very large but the sample size is limited and the feature correlation is strong.

The particular execution sequence of the distributed adaptive Lasso is as follows. Divide the 45-dimensional candidate features into n groups sequentially based on the number. If division is not feasible, the remaining features will be included in the last group. Second, conduct adaptive Lasso on the first group to obtain candidate feature sets that correspond to non-zero coefficients. Again, merge the set with the second set of features and continue filtering, and pass the new non-zero feature set to the next set. The above process is repeated for the nth group. Finally, output the feature numbers of the last non-zero coefficients and build the classifier input matrix based on them. Thus, the feature dimensions at each regression round are reduced from the full 45-dimensional space, and the features that have contributed in previous rounds are not lost outright in the subsequent rounds.

As shown in Table 1, the group order is used for execution, and the Raw signal and each IMF component will enter the filtering process sequentially. A small partition parameter n has a larger number of candidate features per group and thus contains more comprehensive information, but it will also have a high computational cost in a single round of solving; on the other hand, a larger value of n will reduce the candidate dimension in each round, but the effect

of early-screened results on subsequent combinations will be weakened. Simultaneously run the above three experiments to show how changes in the parameter affect convergence and accuracy for n=5, n=10, and n=15. Since each group of screening will carry the selected features from the previous round, the segmented calculation does not fully separate the different IMF components but retains the cross-component competition relationship and reduces the dimensionality.

Regularisation parameters are selected by search in the training set and do not need to be tuned using the test set. If no non-zero features are generated in a certain round of partition screening, a small number of candidate variables with larger absolute coefficients in the group will be retained for the next round to avoid information interruption caused by noise disturbance in the early grouping. The final feature set is based only on the training data, and the test set does not participate in the confirmation of the selected features. Thus, it will be possible to ensure that the classification accuracy reflects the generalisation ability of the method to unseen samples rather than post-adaptation to the distribution of the test set.

Three types of input schemes for stability tests of feature selection methods are used in this paper: full-feature input, Lasso filtering input, and segmented adaptive Lasso filtering input. Segmented parameters n are set to 5, 10 and 15, and the effect of different numbers of feature group partitions on diagnostic accuracy is studied. The proportion of the training set and the test set is 70:30. The training set is used to estimate the parameters for feature normalisation and selection, and to train a classification model; only the test set is reserved for the final performance evaluation. The evaluation indicators are classification accuracy, calculation time and diagnostic efficiency (accuracy/calculation time). Accuracy is employed to assess the distinction capability of fault categories, computation time is used to show the real-time performance of the algorithm, and diagnostic efficiency can be used to observe performance and cost at the same time. For the results that have not converged, this paper keeps the "Not converged" mark in the table and does not convert them to 0 for inclusion in the average; otherwise, there may be confusion between low accuracy and optimization failure, and inconsistency will arise between the caliber of the result and that in the subsequent graph. The three selected classifiers are KNN, MLP and SVM, and the corresponding types of discrimination mechanisms are based on neighborhood distance, nonlinear mapping and maximum interval classification [20-22]. The algorithm is realised in a Python environment and Scikit-learn toolkit [23], and classification accuracy is verified via ten-fold cross-validation with five repetitions per class dataset. The final report of each feature scheme shows the mean accuracy and mean computation time of 15 tests for KNN, MLP and SVM. Ten-fold cross-validation is employed to reduce the effect of random partitioning of the sample on accuracy estimation [24].

3 Results and Discussion

3.1 Correlation of Acoustic Features and Structure of Selected Features

First, determine whether the 45-dimensional candidate features have different task contributions and are not redundant before comparing classification accuracy. If all the features contribute to the classification target in a similar manner, then the advantage of feature selection will be relatively small; if there is a strong correlation among the high-scoring features, directly selecting the set of features with the highest scores may also result in repeated expressions. Therefore, this paper takes the damage status of bearing balls/raceways as a case study and uses the Spearman score and correlation coefficient matrix of 45-dimensional original features as the basis for feature quality discrimination. Spearman's coefficient of monotonicity is used to

measure how strongly features are correlated with the target state; the correlation matrix of all candidate features is employed to determine the linear redundancy among them. The purpose is to increase the correlation and reduce the redundancy of the features, thus making the high-quality features category-sensitive and less interdependent [19].

The Spearman coefficients of the 45-dimensional acoustic features are as follows:

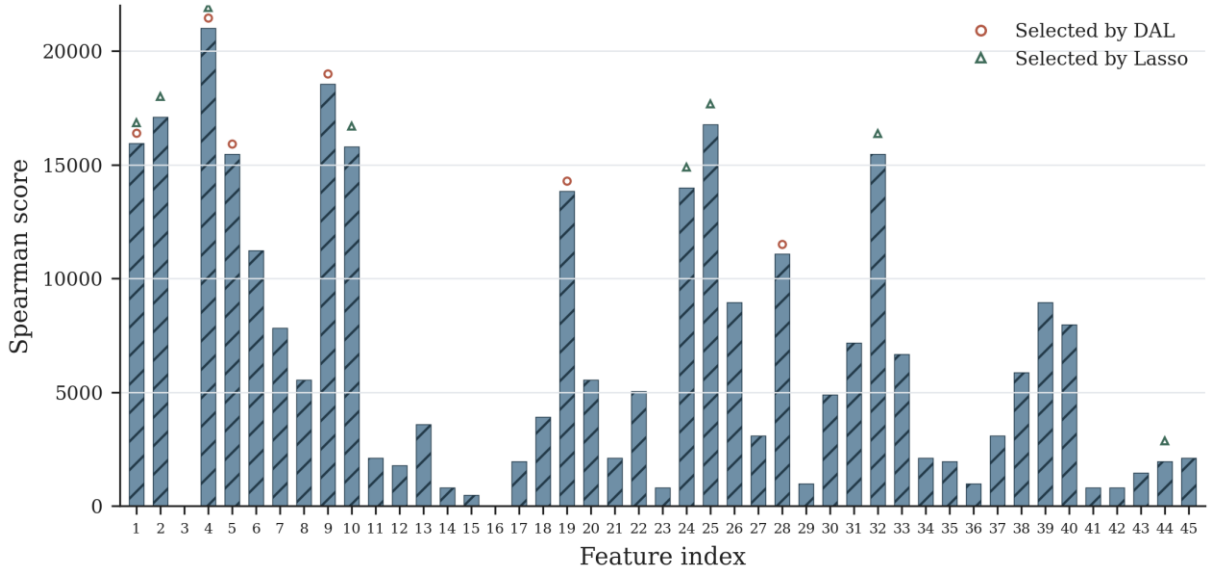

Figure 4: Spearman Scores of the 45-Dimensional Acoustic Features

As shown in Figure 4, the feature scores of the original signal and some low-order IMF components are relatively high; therefore, it can be concluded that the damage impact information caused by bearing ball/raceway damage is first manifested in the original waveform and high-frequency decomposition components. Features 1, 4, 5, 9, 19 and 28 are all included in the selected set of the distributed adaptive Lasso, and they correspond to the kurtosis of the original signal (1), the frequency variance of the original signal (4), the energy coefficient of the original signal (5), the frequency variance of IMF1 (9), the frequency variance of IMF3 (19) and the center frequency of IMF5 (28). The four types of information in the above numbers are amplitude shock, spectral diffusion, energy contribution and intermediate frequency migration, and they do not focus on a single feature family according to the results of the segmented adaptive Lasso.

The correlation coefficient matrix of the 45-dimensional acoustic features is as follows: Figure 5.

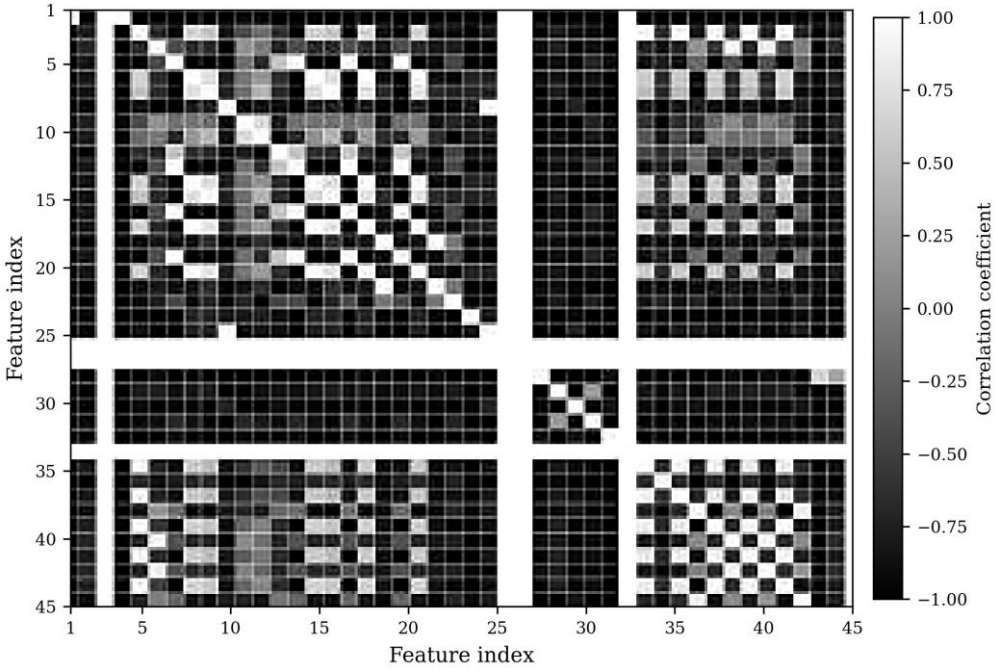

Figure 5: Correlation Coefficient Matrix of the 45-Dimensional Acoustic Features

As shown in Figure 5, there are several local block high-correlation regions in the matrix, which indicate that the frequency-domain features and energy features among adjacent IMF components, as well as some original signal features within the same signal object, are redundant. If only based on the univariate score, many variables with the same information in a correlation block may be selected; thus, the classifier can fit the training set better but will be more unstable on the test set. Partial adaptive lasso reduces such redundancy by means of a weighted penalty and grouping iteration mechanism, and the selected features are distributed in raw, imf1, imf3 and imf5, avoiding repeated selection in the same related block.

As shown in Figure 4 and Figure 5, the high-score features are not necessarily independent of each other, and low-correlation features may have weak task contributions. Therefore, the results of the feature selection need to be placed back in the score graph and correlation matrix for interpretation at the same time. If the selected features are concentrated in the same matrix block, it indicates that the model may be reusing the same information; if the selected features span different IMF components and cover different feature families, then more multi-scale fault evidence has been used by the model. This determination is particularly important for the fault diagnosis of a conductive slip ring, as bearing impact, spindle crack and brush runout may all generate high-frequency disturbance noises, but their frequency band distribution and energy distribution are not the same.

3.2 Classification Accuracy under Five Operating States and the Effect of Partition Parameter

Determine whether the candidate features can contribute to the differential distribution, and then compare the classification accuracy and convergence behaviour under different feature-input schemes. Table 2 shows the average classification accuracy of full feature, lasso and segmented adaptive lasso for the five types of equipment. As shown in Table 2, "non-convergence" is kept as the original experimental result and is not converted to 0 for the average accuracy rate to prevent mixing convergence failure of the algorithm with a low accuracy rate.

Table 2: Comparison of Classification Accuracy under Three Feature Schemes

Equipment Status	All Features	Lasso	Adaptive Lasso n=5	Adaptive Lasso n=10	Adaptive Lasso n=15
Normal	0.693	0.872	0.906	0.906	0.901
Bearing ball/raceway damage	Not converged	Not converged	0.751	0.773	0.776
Spindle fatigue crack	0.679	0.819	0.826	0.829	0.822
Dynamic eccentricity	Not converged	0.825	0.867	0.865	0.871
Brush bounce	0.732	0.917	0.992	0.992	0.991

As shown in Table 2, the full feature scheme has achieved accuracy rates of 0.693, 0.679 and 0.732 under normal, spindle fatigue crack and brush runout conditions, respectively; however, it failed to converge in the presence of bearing ball/raceway damage and dynamic eccentricity. According to the above experiments, the direct input of 45-dimensional features will result in severe redundancy and noise interference for the classifier, making it difficult to obtain a stable boundary in some fault states. After Lasso feature selection, the accuracy of the normal state increased to 0.872; the fatigue crack in the main shaft increased to 0.819; the dynamic eccentricity increased to 0.825; and the brush runout increased to 0.917, but the damage of bearing ball/raceway has not converged yet. Lasso can reduce some irrelevant features, but the coefficient compression induced by unified punishment may harm the effective feature retention under weak impact faults.

Partial adaptive lasso can perform the five kinds of state diagnosis in three settings with n=5, n=10 and n=15. When n=10, the accuracy rates of normal, bearing ball/raceway damage, spindle fatigue crack, dynamic eccentricity and brush runout were 0.906, 0.773, 0.829, 0.865 and 0.992 respectively, and the average accuracy rate of the five categories was 0.873. Compared with Lasso in the four states of common convergence, the piecewise adaptive lasso increased by 0.034, 0.010, 0.040 and 0.075 respectively for the normal, spindle fatigue crack, dynamic eccentricity and brush runout states, with an average increase of 0.0398. Compared with the three kinds of states in which all features converge together, the accuracy is increased by 0.2077 on average under the setting of n=10.

The connection among the accuracy of classification and the time required for calculation is shown in Figure 6.

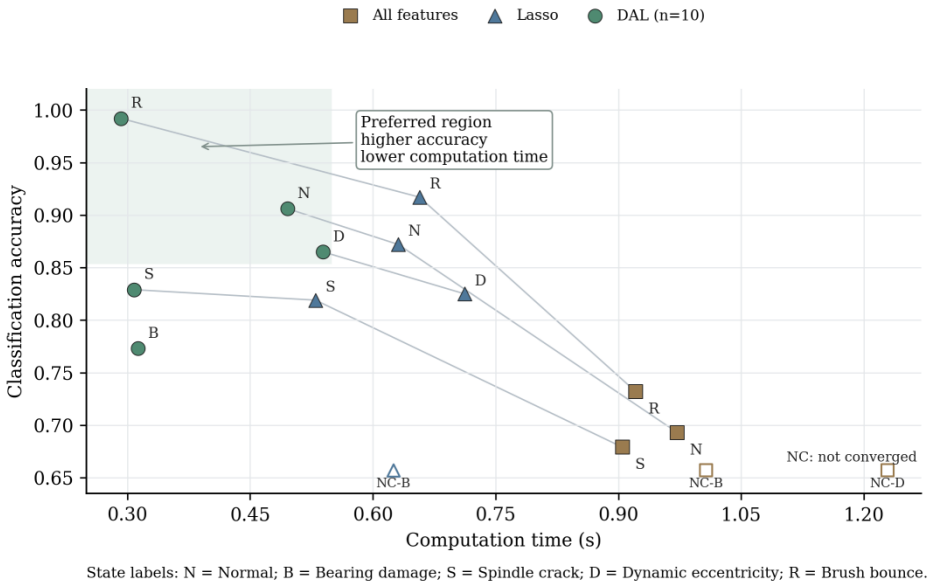

Figure 6: Coupling Relationship of Classification Accuracy and Computation Time

As shown in Figure 6, the closer it is to the upper left corner, the higher the accuracy of the model is at a lower calculation cost. The complete feature scheme is mainly located in the lower right or non-convergent area, indicating that the high-dimensional input does not bring stable gain; the Lasso scheme moves to the left and up as a whole, but the bearing ball/raceway damage state still lacks effective points; the segmented adaptive lasso forms effective points in all five types of states, and most of these points are distributed in the high-accuracy and low-time-interval region.

The effect of the partial parameter n on the classification accuracy is shown in Figure 7.

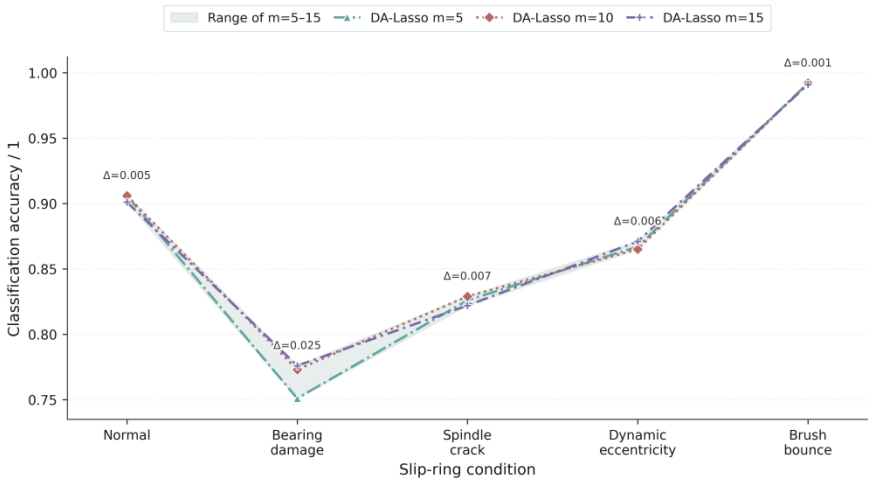

Figure 7: Effect of Partition Parameter n on the Accuracy of Partitioned Adaptive Lasso

As shown in Figure 7, the average accuracy of the five categories for n=5, n=10 and n=15 are 0.868, 0.873 and 0.872 respectively, and they are close. The normal states are 0.906 at $n=5$ and $n=10$, and 0.901 at $n=15$; the damage to the bearing ball/raceway increases from 0.751 to 0.776 with an increase in n; the fatigue crack of the spindle reached its maximum value of 0.829 at $n=10$, and the highest value of 0.871 was obtained when the dynamic eccentricity was $n=15$; and the brush jumping exceeds 0.991 under all three conditions. The above results

show that the sub-parameters mainly affect the order of feature inclusion in the iterative screening and the candidate dimension for a single round; the overall diagnostic performance is not sensitive to n. A total of 10 was selected reasonably because this setting is suitable for achieving good accuracy, convergence of bearing faults and computational efficiency.

When the bearing ball/raceway is damaged, Lasso selects feature numbers 1, 2, 4, 10, 24, 25, 32 and 44; The feature numbers selected by the segmented adaptive Lasso for n=10 are 1, 4, 5, 9, 19 and 28, as shown in Figure 8.

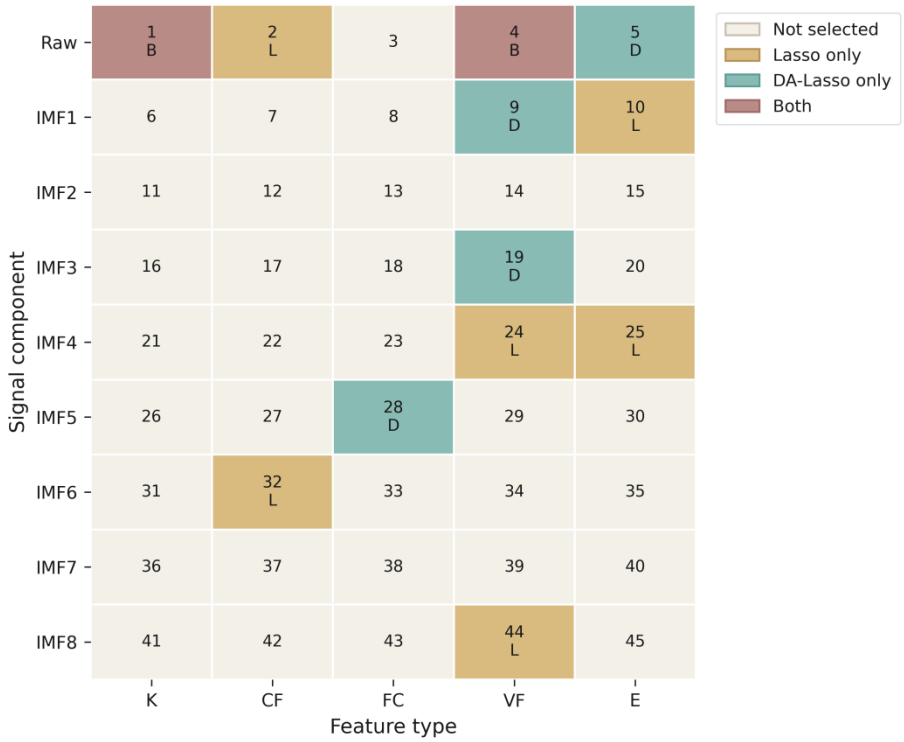

Figure 8: Distribution of Selected Features Under Bearing Ball/Raceway Damage

Two methods jointly selected features 1 and 4, indicating that the kurtosis and frequency variance of the original signal are stable features of this type of fault. The unique features of Lasso are 2, 10, 24, 25, 32 and 44, which have multiple energy or frequency variance variables and are located in areas of strong correlation; the unique features 5, 9, 19 and 28 of the segmented adaptive Lasso are distributed among the different IMF components and exhibit multi-scale information retention capabilities.

To better observe the three-dimensional relationships among feature schemes, fault states and accuracy intuitively, this paper will plot the data in Table 2 as a three-dimensional response graph, as shown in Figure 9.

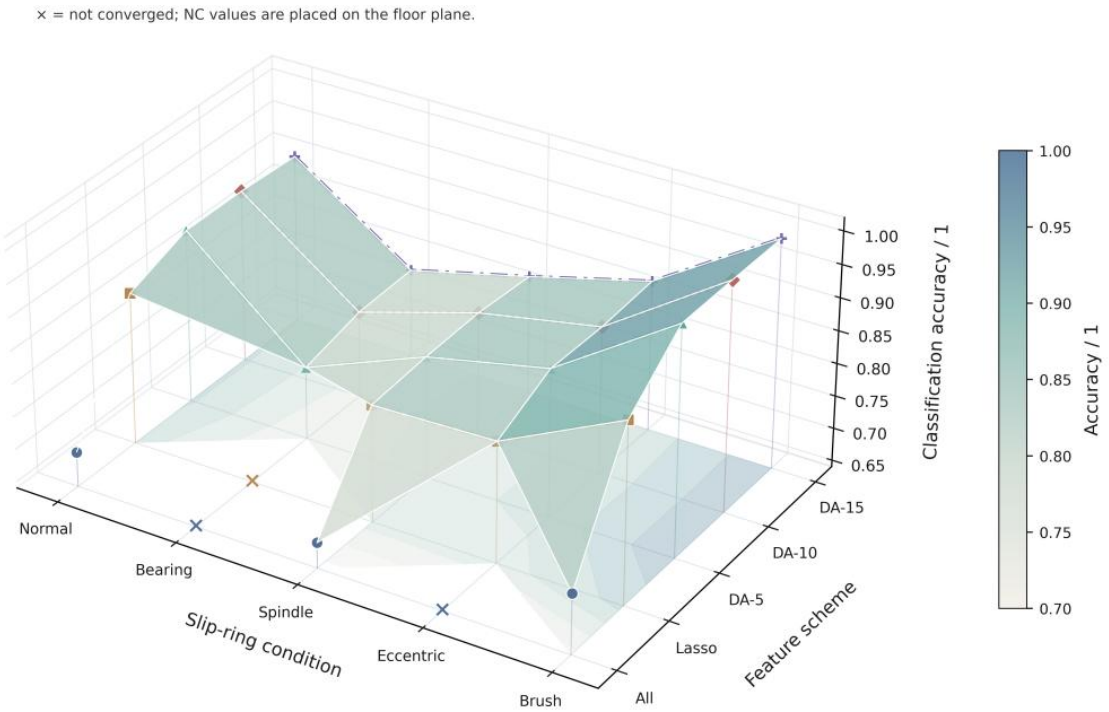

Figure 9: Three-Dimensional Response Relationship of Feature Scheme, Fault State and Classification Accuracy

As shown in Figure 9, the full feature scheme has convergence gaps at the locations of bearing ball/raceway damage and dynamic eccentricity, but the Lasso scheme also has gaps at the locations of bearing ball/raceway damage. The three-parameter surfaces of the segmented adaptive Lasso include five kinds of states. The electric brush jumping form has the highest response zone under the segmented adaptive Lasso, with an accuracy of approximately 0.992; bearing ball/raceway damage is the lowest response zone, but still has a reasonable identification level of 0.751-0.776. The three-dimensional diagram shows that the reasons for the advantages of this method are mainly two: first, it addresses the deficiency of full features and Lasso in some cases; second, it maintains high stability of the classification boundary for brush jumping, dynamic eccentricity, and normal states.

Based on the type of fault, the identification results show that brush jumping is the most prominent; that is, intermittent contact and changes in the friction state exhibit clearer amplitude and energy characteristics in the audio signal. The accuracy of bearing ball/raceway damage is lower than that of other faults because rolling impact is easily affected by background noise and structural transmission paths after propagating through the shell, and some of the impact energy will be dispersed into multiple IMF components. Fatigue cracks and dynamic eccentricity of the spindle are at an intermediate level; the former is caused by frequency transfer and local stiffness changes, and the latter by periodic modulation characteristics. The segmented adaptive Lasso can maintain stable output in these different acoustic modes, and thus its feature selection results are not limited to a single fault mechanism.

3.3 Computation Time, Diagnostic Efficiency, and Engineering Applicability

Fault diagnosis methods are now applied to online monitoring and need to be more accurate. The conductive slip ring is in a long-term rotating state with strong current coupling, and the

diagnostic algorithm needs to complete feature screening and classification output within a limited sampling window. Table 3 shows the average computation time for full feature, Lasso and partial adaptive Lasso at n=10. At this time, we have averaged the test results of KNN, MLP and SVM classifiers on multiple state datasets.

Table 3: Classification Computation Time for the Three Feature Schemes

Dataset	All Features / s	Lasso / s	Adaptive Lasso n=10 / s
Normal	0.972	0.631	0.496
Bearing ball/raceway damage	1.007	0.625	0.313
Spindle fatigue crack	0.905	0.530	0.308
Dynamic eccentricity	1.229	0.712	0.539
Brush bounce	0.921	0.657	0.292

As shown in Table 3, the full feature scheme has a computation time of 0.905-1.229 seconds in five states, and the dynamic eccentricity state has the longest computation time at 1.229 seconds. Lasso has reduced the computation time to 0.530-0.712 seconds and can achieve sparse filtering to reduce the input dimension of the classifier. Segmented adaptive Lasso reduces the time to 0.292-0.539 seconds further, and the brush runout, spindle fatigue crack and bearing ball/raceway damage are as low as 0.292, 0.308 and 0.313 seconds, respectively. Segmented adaptive Lasso reduced the time by 48.97%, 68.92%, 65.97%, 56.14% and 68.30% compared to the full feature for normal, bearing ball/raceway damage, spindle fatigue crack, dynamic eccentricity and brush runout states, respectively, with an average reduction of 61.66%. Compared with Lasso, the five types of state times were reduced by 21.39%, 49.92%, 41.89%, 24.30%, and 55.56%, with an average reduction of 38.61%.

Both accuracy and computation cost need to be considered, so we set diagnostic efficiency as the ratio of classification accuracy to computation time, and a diagnostic efficiency heatmap is presented in Figure 10.

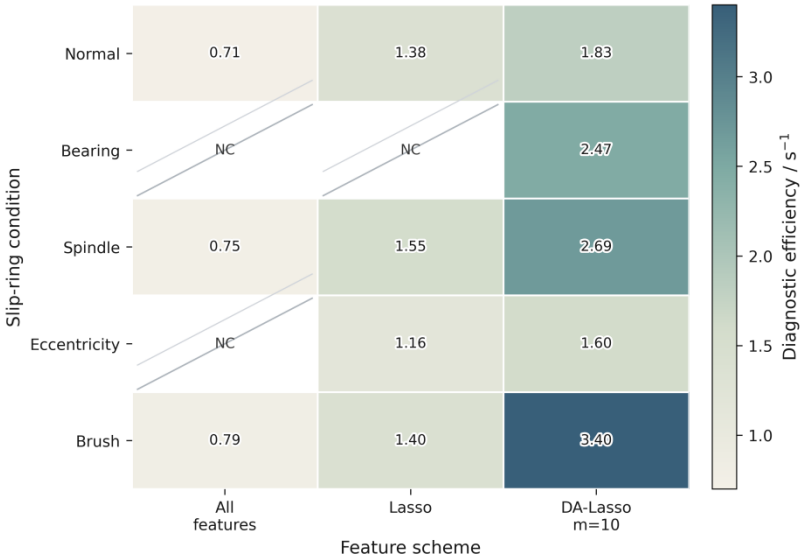

Figure 10: Diagnostic Efficiency Heatmap of Three Feature Schemes

As shown in Figure 10, the segmented adaptive Lasso is a high-efficiency region in all five states; among them, the diagnostic efficiency of the brush-jumping state is 3.397, and the fatigue cracks in the spindle and the damage to the bearing ball/raceway are 2.692 and 2.470,

respectively. Lasso's diagnostic efficiency under normal, spindle fatigue crack, dynamic eccentricity and brush jumping conditions is as follows: 1.382, 1.545, 1.159 and 1.396. The complete feature scheme is less than 0.80 at convergence. According to the efficiency heatmap, the improved accuracy of segmented adaptive Lasso is not due to any of the factors listed above; rather, it results from a reduction in the number of features, enhanced convergence stability and simplified classification inputs.

The accuracy and time-to-result of the segmented adaptive Lasso compared to the baseline are shown in Figure 11.

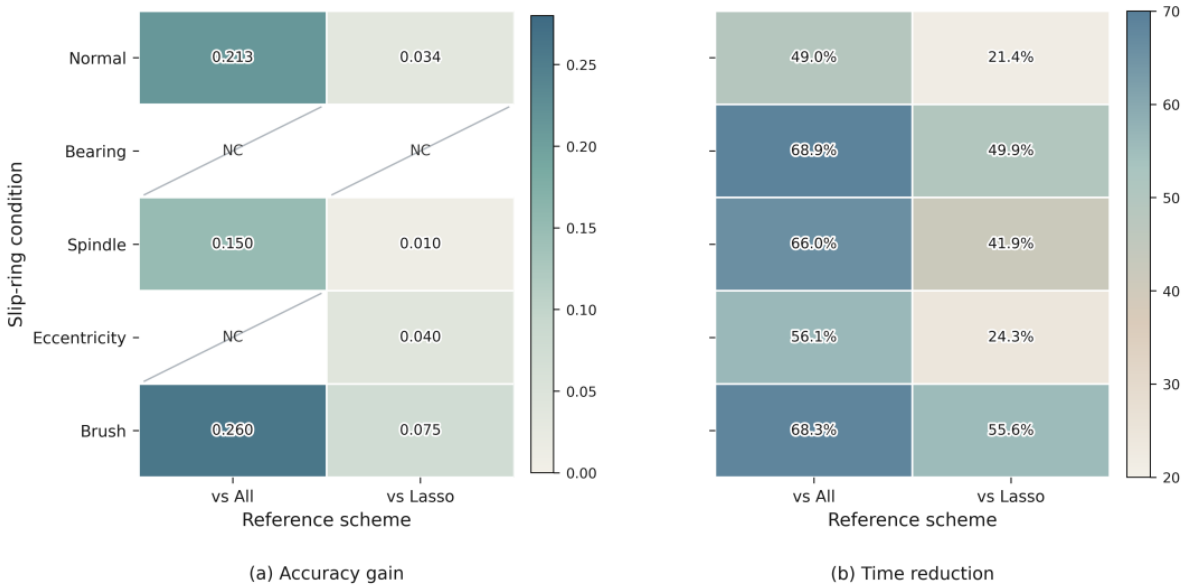

Figure 11: Accuracy and Time Gain Matrix of Partitioned Adaptive Lasso Versus Baseline Methods

As shown in Figure 11, when $n=10$ is used instead of the full feature set, an improvement in accuracy of 0.213, 0.150 and 0.260 was achieved under normal operation, spindle fatigue crack, and brush jumping conditions, respectively, with a significant reduction in computation time; for bearing ball/raceway damage and dynamic eccentricity where the full features have not converged, the segmented adaptive Lasso achieved the accuracy of 0.773 and 0.865, demonstrating that it is more trainable for complex acoustic faults. Compared with Lasso, the accuracy gain of brush jumping is 0.075; the dynamic eccentricity is 0.040; the normal value is 0.034; and the spindle fatigue crack is 0.010. In the state of bearing ball/raceway damage, Lasso did not converge, and the segmented adaptive Lasso still maintained an accuracy of 0.773.

The decline in diagnostic accuracy is also due to the reduction in the scope of the classifier's distance and boundary after feature selection. KNN is highly sensitive to redundant dimensions, and irrelevant features reduce the distance to neighbours; SVM may form a more complex support vector structure under high-dimensional small-sample conditions; and MLP is prone to learning noise related to local patterns in the training set. Distributed adaptive Lasso compresses the candidate features first, and then all three classifiers perform discrimination in a lower-dimensional space to reduce computation time and improve test set performance simultaneously. The above mechanism explains why the accuracy and the time-return occur simultaneously in Figures 6, 10 and 11.

Compared with recent studies on acoustic vibration fusion diagnosis, this way avoids using complex deep networks and multi-sensor fusion structures, and instead reduces the deployment bar by using interpretable statistical features and sparse screening mechanisms. In cases of high-

power slip-rings where test samples are scarce and fault simulation costs are high, this light-weight method is reasonably applicable. It has a relatively short algorithm chain; the feature number can be directly mapped to the Raw or IMF component, and thus the diagnostic results can be more easily linked with the fault mechanism trace.

The above results can be applied directly to the problem of online diagnosis for high-power conductive slip rings. First, after EMD decomposition and numbering of acoustic features, different fault mechanisms can be associated with amplitude shock, spectral diffusion, frequency transfer and energy distribution changes, and the model output has interpretable entry points. Second, the distributed adaptive Lasso avoids convergence failure of all feature inputs in the high-dimensional and small-sample regime, and also reduces feature bias under the Lasso uniform penalty. Thirdly, the algorithm has a relatively short computation time compared with the full feature and Lasso schemes, and is thus more suitable for embedding in slip ring state monitoring systems. It should be pointed out that the verification in this paper is still based on a fixed platform and five typical states, and has not yet covered the conditions of continuous changes in speed, current, temperature and load. In the future, multiple conditions of acoustic acquisition and acoustic vibration fusion data will be combined to further verify the adaptive determination of sub-parameters, feature stability drift, and threshold alarm strategies [25].

4 Conclusion

This paper studies the problem of high-dimensional feature selection for fault diagnosis in high-power conductive slip ring audio signals. Based on the experimental platform, the original acoustic lines and result data, the Lasso feature selection method is adaptively improved in a segmented manner, and classification accuracy, calculation time and feature interpretation analysis have been carried out for five operating states.

(1) Taking the 60 kW high-power conductive slip ring fault simulation platform as the object, five types of state data are retained: normal, bearing ball/raceway damage, spindle fatigue crack, dynamic eccentricity, and brush jumping. Construct 45 dimensional acoustic features using the original audio signal and IMF1-IMF8 components, and illustrate the fault contribution and redundancy relationship of different features using Spearman scores, correlation coefficient matrices, and selected feature maps.

(2) Distributed adaptive Lasso reduces the coefficient bias caused by uniform penalty through adaptive weights and iteratively reduces the single round solution dimension through feature grouping. In the experiment, the average accuracy of the five categories was 0.873 when n=10; Compared with the four states that converge together with Lasso, the average accuracy is improved by 0.0398 and the average computation time is reduced by 38.61%.

(3) The effectiveness of this method has been verified under fixed bench and typical fault conditions, but it still does not cover the drift caused by continuous changes in speed, current, load, and temperature. In the future, multi condition data should be expanded, combined with acoustic vibration fusion and online update mechanisms, to further verify the stability of the selected features under long-term operating conditions.

Funds

This work was supported by the "Ganpo Juncai" Support Program for Leading Academic and Technological Leaders in Key Disciplines of Jiangxi Province (Project Number: 20232BCJ23064).

About the Author

Bin Lv was born in Jiujiang, Jiangxi, China in 1988. He graduated with a master's degree from Nanchang Hongqiao University in China. He works at CSSC Jiujiang Jingda Technology Co., Ltd. now. The first direction of his research is electric rotary transmission technology.

Qingqin Zhao was born in Cangzhou, Hebei Province, China, in 1984. He graduated from Dalian University of Technology in China with a bachelor's degree. He works at Jiujiang Precision Measurement and Testing Technology Research Institute now. His first research direction is automatic control technology.

Xingxing Xu was born in Jiujiang, Jiangxi Province, China, in 1988. Jiangxi Agricultural University in China granted him a college degree for graduation. He works at Jiujiang Precision Measurement and Testing Technology Research Institute now. The first research direction is electric rotary transmission technology.

Hongming Ni was born in Jiujiang, Jiangxi, China in 1994. He earned a Master's degree from Nanchang Hongqiao University in China. He works at CSSC Jiujiang Jingda Technology Co., Ltd. now. The head of this study is the technology of electric rotary reducers.

Jian Zhou was born in Jiujiang, Jiangxi, China in 1995. He graduated with a master's degree from East China Jiaotong University in China. He works at CSSC Jiujiang Jingda Technology Co., Ltd. now. The first direction of his research is electric rotary transmission technology.

Fei Gao was born in Jiujiang, Jiangxi, China in 1989. He received a master's degree from Beijing Institute of Technology in China. He works at CSSC Jiujiang Jingda Technology Co., Ltd. now. The main research directions of him are electric rotary transmission technology.

References

[1] Jardine, A. K. S., Lin, D., Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483–1510.

[2] Lei, Y., Yang, B., Jiang, X., et al. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing, 138, 106587.

[3] Ma, H., Chen, T., Zhang, Y., et al. (2017). Research on the fault diagnosis method for slip ring device in doubly-fed induction generators based on vibration. IET Renewable Power Generation, 11(2), 289–295.

[4] Turel, A., Slavič, J., Boltežar, M. (2017). Electrical contact resistance and wear of a dynamically excited metal-graphite brush. Advances in Mechanical Engineering, 9(3), 1–8.

[5] Randall, R. B., Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. Mechanical Systems and Signal Processing, 25(2), 485–520.

[6] Hou, Y., Zhou, C., Tian, C., et al. (2022). Acoustic feature enhancement in rolling bearing fault diagnosis using sparsity-oriented multipoint optimal minimum entropy deconvolution adjusted method. Applied Acoustics, 195, 109105.

[7] Li, C., Chen, C., Gu, X. (2023). Acoustic-based rolling bearing fault diagnosis using a

co-prime circular microphone array. Sensors, 23(6), 3050.

[8] Luo, Y., Lu, W., Kang, S., et al. (2023). Enhanced feature extraction network based on acoustic signal feature learning for bearing fault diagnosis. Sensors, 23(21), 8703.

[9] Zhao, R., Yan, R., Chen, Z., et al. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213 – 237.

[10] Guyon, I., Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157 – 1182.

[11] Cai, J., Luo, J., Wang, S., et al. (2018). Feature selection in machine learning: A new perspective. Neurocomputing, 300, 70 – 79.

[12] Chandrashekar, G., Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16 – 28.

[13] Huang, N. E., Shen, Z., Long, S. R., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A, 454(1971), 903 – 995.

[14] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267 – 288.

[15] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418 – 1429.

[16] Hastie, T., Tibshirani, R., Wainwright, M. (2015). Statistical Learning with Sparsity: The Lasso and Generalizations. Boca Raton: CRC Press.

[17] Efron, B., Hastie, T., Johnstone, I., et al. (2004). Least angle regression. The Annals of Statistics, 32(2), 407 – 499.

[18] Yuan, M., Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B, 68(1), 49 – 67.

[19] Peng, H., Long, F., Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226 – 1238.

[20] Cover, T., Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21 – 27.

[21] Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533 – 536.

[22] Cortes, C., Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20, 273-297.

[23] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825 – 2830.

[24] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 14(2), 1137 – 1145.

[25] Jiang, M., Luo, M., Zhang, C., et al. (2024). Rolling bearing fault diagnosis based on acoustic-vibration data fusion and mode decomposition combined with the crested porcupine optimization algorithm. Heliyon, 10(22), e40351.