Abrasive Wear Characteristics and Failure Diagnosis Method of Mechanical Seal Faces

Jing Wang[1,*], Zhonghua Zhou[2]

[1] Chongqing Technology and Business Institute, Chongqing Open University, Chongqing 401520, China

[2] China Oilfield Services Limited, Sanhe 065200, Hebei, China

Corresponding Author: Jing Wang (1065560315@qq.com)

Funding: Scientific and Technological Research Project of Chongqing Municipal Education Commission (No. KJQN202204018): Research on Sealing Mechanism of Mechanical Equipment Lip Seal

Abstract: A failure diagnosis method based on multi-source input fusion was proposed to solve the problems of increased leakage, abnormal friction temperature rise and failure discrimination lag caused by abrasive wear of mechanical seal end face. The feature vectors of scratch density, wear area proportion, texture contrast, friction torque, temperature rise and vibration are constructed by taking the end face morphology image and the running state signal as input. The gated fusion module is used to complete the unified expression of image features and signal features, and the weighted cross-entropy loss is used to optimize the classification boundary of different wear levels. In the experiment, 1200 groups of effective samples were collected and divided into training set, validation set and test set according to 840/180/180. The test results show that the diagnostic accuracy of the proposed model reaches 95.3%, the F1-score reaches 94.6%, and the average inference time of a single sample is 12 ms, which is 3.9 percentage points higher than that of the CNN model, and it can realize the stable recognition of the end surface wear state.

Keywords: Mechanical seal; Abrasive wear; Feature fusion; Failure diagnosis

1. Introduction

The end face of mechanical seal carries axial load and friction in pumps, compressors and rotating equipment. It is easy to produce abrasive wear under long-term operation, resulting in end face roughness, increased leakage and abnormal friction temperature rise, which seriously affects the stability of equipment [1]. Traditional diagnosis methods based on empirical observation or single parameter threshold are lagging behind and cannot meet the requirements of high precision and real-time. To solve this problem, this study takes the abrasive wear characteristics of mechanical seal end face as the core, and proposes a failure diagnosis method based on multi-source feature fusion. Through the synchronous acquisition of wear end face images and running state signals, the morphology texture features, friction torque, temperature rise and vibration signals are encoded into feature vectors that can be input into the model. A fusion module is designed to realize the unified expression of image and signal features, and a wear state discrimination network is constructed to

realize end face failure classification.

2. Characterization and Input Modeling of end Surface Wear State

2.1 Definition of Wear State Variables and Diagnostic Labels

The wear state of mechanical seal end face should be transformed into a computable variable, and the diagnostic input is jointly characterized by the roughness of end face, scratch distribution, wear area, friction torque and temperature rise response [2]. In order to avoid the problem that a single threshold is not sensitive to local damage, a weighted score is used to describe the comprehensive degradation degree of the end face:

$$S=\omega_1 R+\omega_2 D+\omega_3 A+\omega_4 T+\omega_5 H \tag{1}$$

Where, S is the comprehensive score of end face wear; R is the roughness normalized value; D is the normalized value of scratch density; A is the proportion of wear area; T is the friction torque fluctuation coefficient; H is the normalized value of end face temperature rise; ω1, ω2, ω3, ω4, ω5 are the corresponding weight coefficients.

The diagnostic label is determined by the composite score interval:

$$Y=\begin{cases} 0, 0\le S<\tau_1 \\ 1, \tau_1\le S<\tau_2 \\ 2, \tau_2\le S<\tau_3 \\ 3, S\ge\tau_3 \end{cases} \tag{2}$$

Where, Y is the diagnostic category; The values τ1, τ2, τ3 are the wear level division thresholds, which are jointly determined by the end face detection results and the operating abnormal boundaries.

In order to ensure a clear correspondence between the diagnostic label and the wear state of the end face, it is necessary to uniformly classify the morphology characteristics, operation response and failure categories under different wear levels. The wear state classification of the mechanical seal end face is shown in Table 1.

Table 1: Classification Table of Wear State of Mechanical Seal end Face.

State Level	End-face Morphology Performance	Operating Response Performance	Diagnostic Category
Normal State	End face is intact, with slight local polishing marks	Torque is stable, temperature rise is low, and leakage shows no obvious change	Normal
Mild Wear	Local fine scratches, with discontinuous wear marks	Slight torque fluctuation and small temperature rise	Mild Wear
Moderate Wear	Continuous scratches increase, and shallow pitting appears locally	Vibration increases, and temperature rise and leakage continue to increase	Moderate Wear
Severe Failure	Deep scratches, peeling, and wear grooves are obvious	Leakage increases sharply, torque fluctuates violently, and operating stability decreases	Severe Failure

2.2 Construction of Feature Parameters of end Surface Topography

The end surface morphology features are used to describe the damage degree of abrasive wear in spatial distribution, and the feature construction is carried out around scratches, wear areas and texture changes [3]. After image graying, end region cropping and binary segmentation, the scratch distribution per unit area can be expressed as:

$$D_s = \frac{\sum_{i=1}^{n} l_i}{A_r} \tag{3}$$

Where, D_s is the scratch length density, li is the pixel length of the i_{th} scratch, n is the number of scratches, and A_r is the effective detection area of the end face.

The proportion of wear area is used to characterize the range of end face material removal:

$$P_w = \frac{A_w}{A_r} \times 100\% \tag{4}$$

Where, P_w is the proportion of wear area, A_w is the pixel area of wear mark, pitting erosion and spalling area, and A_r is the area of effective detection area of end face.

Texture dispersion is used to describe the degree of gray level mutation caused by abrasive wear:

$$G_c = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 p(i,j) \tag{5}$$

Where, G_c is the gray level co-occurrence matrix contrast, L is the gray level series, and $p(i,j)$ is the normalized probability of row i and column j in the gray level co-occurrence matrix.

In order to ensure that the end face image processing process is consistent with the subsequent feature vector input, it is necessary to connect the region clipping, damage identification and parameter calculation into a unified process. See Figure 1 for the end face topography feature extraction process.

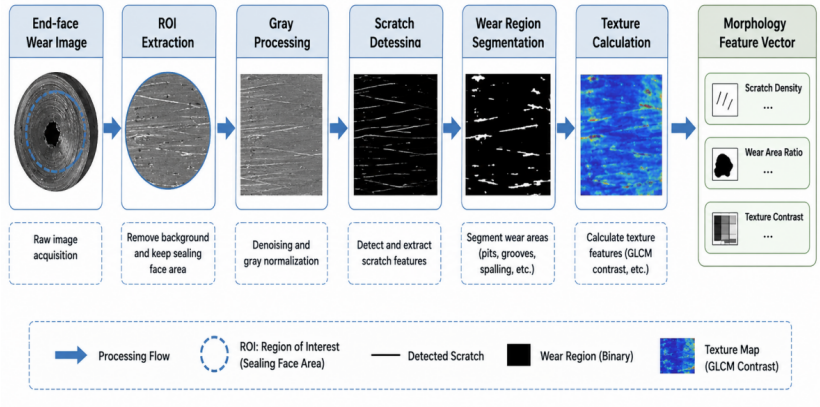

Figure 1: Flow Chart of Mechanical Seal end Surface Topography Feature Extraction.

2.3 Encoding the Feature Vector of the Running Signal

The abrasive wear of mechanical seal end face is not only reflected in the morphology change, but also reflected in the friction torque, end face temperature rise, vibration and acoustic emission and other operating signals. In order to unify the input model, different types of operational signals need to be encoded as feature vectors [4]. Each type of signal is segmented according to the time window, and the statistical characteristics such as mean, variance and peak value are calculated to form a preliminary vector representation:

$$X_s = [\overline{T}, \sigma_T, T_{max}, \overline{V}, \sigma_V, V_{max}] \tag{6}$$

Where, X_s is the feature vector of the end face operation signal; $\overline{T}, \sigma_T, T_{max}$ are the mean, standard deviation and peak value of the friction torque, respectively; $\overline{V}, \sigma_V, V_{max}$ are the mean, standard deviation and peak value of the vibration signal, respectively.

In order to take into account the comparability of signals of different dimensions, the vectors are normalized and the final input vector is constructed:

$$X=[X_s^{norm} \mid X_m^{norm}] \tag{7}$$

Where, X_s^{norm} is the normalized operating signal feature and X_m^{norm} is the normalized end face topography feature. The vector is used as the unified input of the multi-source feature fusion model to provide basic data for the diagnosis model.

3. Design of Failure Diagnosis Model for Multi-source Input Fusion

3.1 Design of Image and Signal Input Fusion Module

The scale, dimension and sensitivity interval of image topography features and operation signal features are different, and direct splicing will amplify the influence of high amplitude variables on diagnostic results. The fusion module adopts linear mapping and gated weighting structure, projects the end surface shape vector and operation signal vector into the same hidden space, and then assigns weights according to the characteristic response intensity, so that the model can simultaneously pay attention to scratches, wear area, texture mutation, torque, vibration and temperature rise changes [5]. The feature mapping process can be expressed as follows.

$$Z_m=\phi(W_mX_m+b_m),\ Z_s=\phi(W_sX_s+b_s) \tag{8}$$

Where, Z_m is the hidden layer feature of end surface topography; Zs is the hidden layer feature of operation signal; X_m is the shape feature input; X_s is the operation signal feature input; W_m, W_s is the weight matrix; b_m, b_s are bias terms; $\phi(\cdot)$ is a nonlinear activation function.

The fusion weights are adaptively generated by the gating function:

$$F=g\odot Z_m+(1-g)\odot Z_s,\ g=\sigma(W_g[Z_m;Z_s]+b_g) \tag{9}$$

Where, F is the fused diagnostic feature; g is the gating weight; $\sigma(\cdot)$ is the Sigmoid function. W_g is the gating mapping matrix; b_g is the gating bias; $\odot$ is element-wise multiplication; $[Z_m; Z_s]$ represents the concatenation result of morphology features and signal features.

In order to highlight the connection relationship of multi-source inputs in the diagnosis model, it is necessary to put topography coding, signal coding, gated fusion and diagnostic output into the same structure. See Figure 2 for the structure of failure diagnosis model of multi-source input fusion.

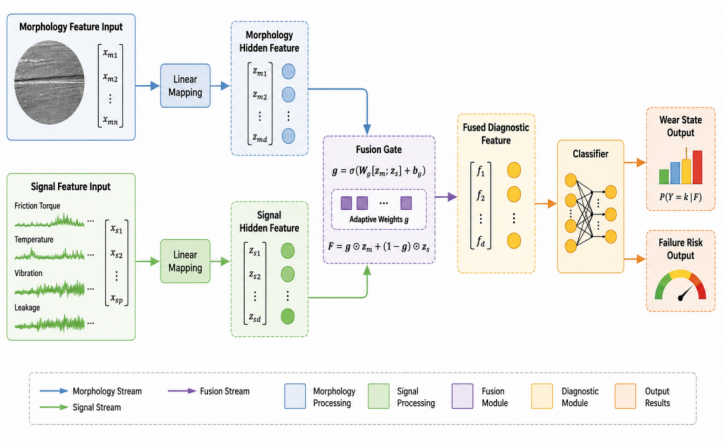

Figure 2: Structure Diagram of Failure Diagnosis Model for Multi-source Input Fusion.

3.2 Construction of Wear State Discrimination Network

The fused feature vector is used as the input of the wear state discrimination network, and the high-order features are extracted through the multi-layer fully connected structure, and the failure class probability of the end face is output. Linear mapping and nonlinear activation function are used for feature transformation between network layers:

$$H^{(l)}=\phi(W^{(l)}H^{(l-1)}+b^{(l)}) \tag{10}$$

Where, $H^{(l)}$ is the hidden feature of the LTH layer, $H^{(0)}=F$ is the fusion feature input, $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias of the LTH layer, and $\phi(\cdot)$ is the activation function.

The probability of the network output terminal failure class is transformed by Softmax:

$$P(Y=k|F)=\frac{\exp(H_k^{(L)})}{\sum_{i=0}^{C-1}\exp(H_i^{(L)})} \tag{11}$$

Where $H_k^{(L)}$ is the feature value corresponding to the class k output of the last layer of the network, C is the total number of classes, and $P(Y=k\mid F)$ is the class probability.

In order to constrain the network prediction to be consistent with the actual label, a weighted cross-entropy loss function is introduced:

$$L=-\sum_{k=0}^{C-1} w_k Y_k \log P(Y=k|F) \tag{12}$$

Where, Y_k is the true label, w_k is the category weight, which is used to balance the proportion of samples with different wear levels, and L is the optimization objective of network training. The network can accurately distinguish different wear stages and provide the core decision-making basis for end face failure diagnosis.

3.3 Loss Function and Model Training Strategy

In the model training phase, it is necessary to balance the accuracy of class discrimination and parameter stability. Since the number of severe failure samples is usually less than that of normal and mild wear samples, the training objective adopts a class-weighted method, so that the minority class samples get higher penalty weights in the back propagation, and reduces the case that the severe failure state is misjudged as moderate wear. The loss function is defined as follows.

$$L_{cls}=-\frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{C-1}\lambda_k y_{ik}\log(\hat{y}_{ik}) \tag{13}$$

Where, L_{cls} is category weighted cross-entropy loss; N is the number of training samples; C is the number of wear categories; λ_k is the weight of the KTH class. y_{ik} is the true class label. $\hat{y}_{ik}$ is the model prediction probability.

In order to suppress the over-fitting problem caused by high-dimensional fusion features, a parameter regularization term is added to the training objective:

$$L=L_{cls}+\eta \sum_{r=1}^{M}\|\Theta_r\|_2^2 \quad (14)$$

Where, L is the final optimization objective; η is the regularization coefficient; M is the number of trainable parameter groups. Θ_r is the r_{th} group of network parameters. AdamW optimizer was used for training, and the learning rate was dynamically attenuated according to the change of the validation set loss. Batch normalization and Dropout were used to improve the convergence stability under small sample conditions.

In order to ensure the reproducibility of the model training process, the optimizer, learning rate, batch size, training rounds and regularization strategy should be set uniformly; see Table 2 for the failure diagnosis model training parameter configuration.

Table 2: Configuration of Training Parameters for Failure Diagnosis Model.

Parameter Category	Configuration Item	Configuration Value	Function Description
Optimizer	AdamW	weight decay 0.0001	Completes network parameter updating
Learning Setting	Initial Learning Rate	0.001	Controls the initial convergence step size
Training Batch	Batch Size	32	Balances GPU memory usage and gradient stability
Iteration	Epochs	100	Ensures sufficient model convergence
Regularization	Dropout	0.3	Reduces the risk of overfitting in fused features
Loss Strategy	Class Weight	inverse frequency	Increases the recognition weight of minority failure classes

4. Experimental Verification and Result Analysis

4.1 Experimental Environment, Data Set and Evaluation Index

The experimental data were collected by the abrasive wear bench of the mechanical seal end face. The test was set up in four states: normal, mild wear, moderate wear and severe failure, and the data of end face image, friction torque, end face temperature, vibration acceleration and leakage were collected synchronously in each state. The morphological features of the end face image are generated after region cropping, gray normalization and wear area labeling. The statistical features of the running signal are extracted according to a fixed time window. A total of 1200 valid samples were obtained, which were divided into training set, validation set and test set according to 7:1.5:1.5. The model training is completed in a unified hardware and software environment, the validation set is used for parameter selection, and the test set is only used for the final performance evaluation. In

order to ensure the reproducibility of the experimental process, data scale and evaluation method, the experimental environment, data set division and evaluation indicators are shown in Table 3.

Table 3: Experimental Environment, Dataset Division and Evaluation Indicators.

Category	Item	Configuration Content	Function Description
Hardware	CPU / GPU	Xeon Silver 4314 ×2 / RTX 4090 24GB	Supports feature calculation and model training
Software	Python / PyTorch	Python 3.10 / PyTorch 2.2	Builds the diagnostic network and training process
Image Data	End-face Images	1200 groups	Provides end-face morphology wear input
Signal Data	Torque / Temperature / Vibration / Leakage	1200 groups	Provides operating state signal input
Data Split	Train / Validation / Test	840 / 180 / 180	Completes training, parameter tuning, and independent testing
Metrics	Accuracy / Precision / Recall / F1-score	classification evaluation	Evaluates wear state diagnostic performance
Efficiency	Inference Time	millisecond level	Evaluates online diagnostic efficiency

4.2 Design of Comparison Experiment and Ablation Experiment

In order to verify the effectiveness of the proposed diagnostic model, a comparison experiment and an ablation experiment were designed. The comparison experiments include the traditional machine learning methods SVM, Random Forest, XGBoost, and the deep learning method CNN with the Proposed multi-source fusion diagnosis Model, to test the performance improvement of the fusion features and network structure for the discrimination of end face wear states. The ablation experiment carried out control experiments for the key modules of the model: removing the end face topography features, the operation signal features, the feature fusion gating module and the weighted loss strategy, respectively, to analyze the contribution of each module to the overall diagnostic accuracy. All experiments were carried out under the unified training set, validation set and test set, and the batch size, training rounds and optimizer parameters were kept consistent to ensure comparison fairness. The experimental protocol and the model Settings and ablation configurations of each group are shown in Table 4.

Table 4: Configuration Table for Comparison Experiments and Ablation Experiments.

Experiment Category	Model / Module	Configuration Description
Comparative Experiment	SVM	Single-feature input and fully connected classification
Comparative	Random Forest	100 trees, with the same feature input

Experiment Comparative Experiment	XGBoost	Tree-based model, with consistent feature input
Comparative Experiment	CNN	Single-channel image feature convolution, with signal feature concatenation
Comparative Experiment	Proposed Model	Multi-source fusion of image and signal, with gated weighting
Ablation Experiment	w/o Morphology Feature	Removes end-face morphology features
Ablation Experiment	w/o Signal Feature	Removes operating signal features
Ablation Experiment	w/o Fusion Gate	Removes the gated fusion module
Ablation Experiment	w/o Weighted Loss	Removes class weights from the loss function

4.3 Diagnostic Performance, Error Samples, and Efficiency Analysis

The performance of each model is evaluated on an independent test set, which contains 180 groups of samples covering four types of states: normal, mild wear, moderate wear and severe failure. The evaluation metrics are Accuracy, Precision, Recall, F1-score and average inference time per sample. To visually present the performance differences between different diagnostic models and ablation groups, the results of the comparison experiment and ablation experiment are shown in Figure 3.

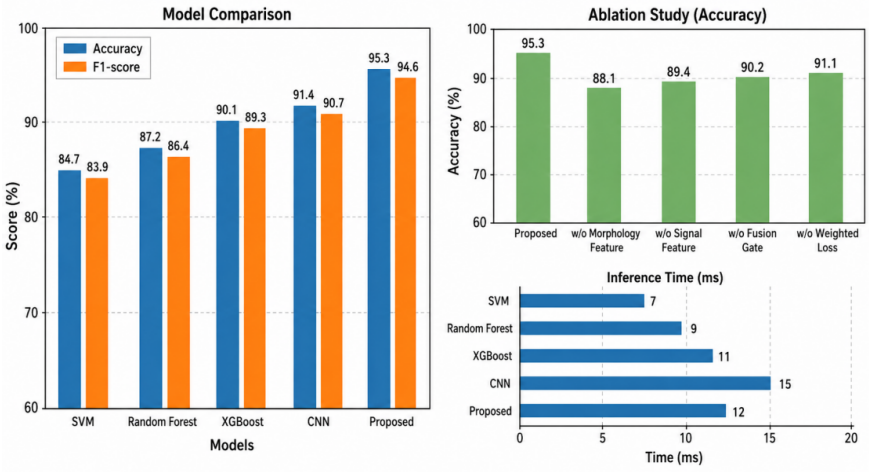

Figure 3: Comparison of Diagnostic Performance and Ablation Results of Different Models.

The Proposed Model has the best performance in four types of wear state recognition, with Accuracy of 95.3%, Precision, Recall and F1-score of 94.8%, 94.5%and 94.6%respectively, and the average inference time of a single sample is 12 ms. The Accuracy of SVM, Random Forest, XGBoost, and CNN is 84.7%, 87.2%, 90.1%, and 91.4%, respectively, which indicates that a single feature or shallow classification structure is insufficient to distinguish mild wear boundaries. The ablation results show that the Accuracy is reduced to 88.1% after removing the end surface topography features, 89.4% after removing the running signal features, and 90.2% after removing the fusion

gating module. The misjudgment samples mainly concentrated in the boundary area of mild wear and moderate wear, and the number of such misjudgments decreased from 18 groups to 7 groups after feature fusion. The model maintained good stability in diagnostic accuracy and online inference efficiency.

5. Conclusion

Focusing on the identification problem of wear state of mechanical seal end face, this paper completed the construction of end face morphology parameters, operation signal coding and multi-source fusion diagnosis network. The experimental results show that after fusing the end face image and the running signal, the model's ability to distinguish the mild wear and moderate wear boundary area is improved, and the misjudgment samples are reduced from 18 groups to 7 groups. After removing the morphology feature, the operation signal or the gating module, the accuracy was reduced to 88.1%, 89.4% and 90.2%, respectively, indicating that the spatial information of end surface damage, the dynamic response of operation and the allocation of fusion weights had a direct impact on the diagnosis results. The model maintains the inference time of 12 ms with the accuracy of 95.3%, which has a good basis for online discrimination. In the future, samples under high pressure, high speed and different media conditions can be further added to improve the generalization ability of the model in complex sealing environment.

References

[1] BADYKOV R, FALALEEV S, BENEDYUK M, et al. Dynamic models of mechanical seals for turbomachinery application. Lubricants, 2024, 12(10): 355.

[2] FRANCISCO A, BRUNETIÈRE N. Full and hybrid multiscale lubrication modeling. Lubricants, 2022, 10(12): 329.

[3] DESHPANDE P, WASMER K, IMWINKELRIED T, et al. Classification of progressive wear on a multi-directional pin-on-disc tribometer simulating conditions in human joints-UHMWPE against CoCrMo using acoustic emission and machine learning. Lubricants, 2024, 12(2): 47.

[4] STRABLEGG C, SUMMER F, RENHART P, et al. Prediction of friction power via machine learning of acoustic emissions from a ring-on-disc rotary tribometer. Lubricants, 2023, 11(2): 37.

[5] GUTIERREZ R, FANG T, MAINWARING R, et al. Predicting the coefficient of friction in a sliding contact by applying machine learning to acoustic emission data. Friction, 2024, 12(6): 1299-1321.