ISSN: 2688-8645

# Research on a CNN-CBAM-Based Method for Seal Fault Diagnosis of Flow Control Valves

Hongbin Liu, Bangyao Tang, Yi Zhang, Daqin Zhang, Zhen Wang

School of Mechanical Engineering, Southwest Petroleum University, Chengdu, Sichuan, China

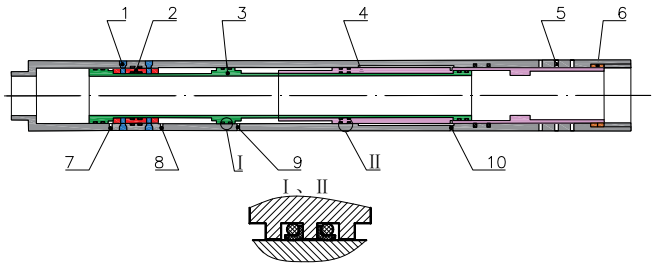
## Abstract

**The seal of a flow control valve is a critical component in hydraulic systems. Seal failure may cause leakage, reduced energy efficiency, and unplanned shutdowns, posing severe threats to safe system operation. Traditional seal fault diagnosis methods often suffer from strong hysteresis and excessive reliance on expert experience; hence an efficient, real-time online seal monitoring technique is urgently needed. This paper proposes a seal fault diagnosis method for flow control valves based on a Convolutional Neural Network with a Convolutional Block Attention Module (CNN-CBAM). First, a reciprocating seal test rig was established, and seal models with multiple fault modes—including abrasion, aging, scuffing, and poor lubrication—were constructed. Friction force–time series data were collected. Then, time–frequency analysis methods such as the S-transform, Short-Time Fourier Transform (STFT), and Continuous Wavelet Transform (CWT) were used to convert the signals into time–frequency maps, which served as input features to the CNN-CBAM network. On this basis, we analyzed how the three time–frequency methods affect the performance of the CNN-CBAM hybrid architecture. Experimental results show that the S-transform + CNN-CBAM model achieved a test accuracy of 99.33% on 3000 samples and better preserved the frequency-adaptive characteristics of the signal. The proposed method significantly improves diagnostic accuracy and robustness, demonstrating that the CNN-CBAM hybrid architecture can effectively identify reciprocating seal faults.**

## Keywords

**Reciprocating Seal; Fault Diagnosis; CBAM; S-transform; Convolutional Neural Network.**

## 1. Introduction




1—set screw; 2—baffle; 3—first-stage hydraulic cylinder; 4—second-stage hydraulic cylinder; 5—throttle orifice; 6—sealing device; 7—opening end of the first-stage cylinder; 8—closing end of the first-stage cylinder; 9—closing end of the second-stage cylinder; 10—opening end of the second-stage cylinder

**Figure 1.** Schematic diagram of the flow control valve structure

A flow control valve is the actuator of a downhole flow control system. It adjusts the opening of a throttle orifice through hydraulic signals to realize flow regulation for different producing zones. Proper sealing is crucial for accurate flow control. Downhole operating conditions are

ISSN: 2688-8645

harsh, placing stringent requirements on the sealing performance and service life of seals. Therefore, it is necessary to study fault diagnosis for non-metallic seals in the hydraulic system of flow control valves. The structure of a flow control valve is shown in Figure 1. By controlling the hydraulic oil flow direction through ports 7, 8, 9, and 10, the throttling area can be adjusted[1].

Foreign research on sealing technology began relatively early, originating in the United Kingdom. With the development of the petroleum industry and materials processing technology, seals have evolved from simple structures to diversified sealing structures with different types and materials. Studies on mechanical seal failures have also increased. Michalis, P.M.M. of CFR Total (France) reported a fault tree analysis (FTA) of mechanical seals for light hydrocarbon pumps in refineries, using a light hydrocarbon mechanical seal as an example to illustrate fault-tree modeling and related analyses [2]. EMayer et al. measured the friction torque on mechanical seal faces using a support-reaction method [3]. Min Zou et al. measured the clearance between seal faces using eddy-current methods [4]. Digard J. et al. also measured seal-face clearance via capacitance methods [5]. Clearance measurements can be used to determine whether seal faces contact and whether faults occur. Acoustic technologies are also widely used for seal condition monitoring: W.B. Anderson [6–7] and T. Reddyhoff [8] monitored the contact state of mechanical seals using acoustic monitoring and achieved good results.

China's sealing technology originated in the 1950s. Initially, it was mainly used to manufacture sealing parts for Soviet centrifugal pumps; complete sealing equipment began to be manufactured in Shanghai, Shenyang, and other places in the 1960s [9]. As the usage of sealing devices increased, failures during operation became evident. Research on seal fault diagnosis has attracted extensive attention. Gu Yongquan analyzed failures of centrifugal pumps used in petrochemical industries [10]. By comparing 12 wear images, he inferred causes of faults, but this approach requires professionals and provides only qualitative estimates. He also proposed a phase-state-informed fault analysis method because phase changes were not considered in mechanical seal design; however, this method only considers phase-state factors and cannot comprehensively diagnose and evaluate seals. Zhou Jianfeng et al. treated the seal as a series system and computed reliability of individual sealing elements based on Monte Carlo principles, then calculated overall seal reliability [11]. Wei Long et al. analyzed mechanical seal failure modes and determined ranges or values of parameters in the Weibull expression, enabling effective application of the method [12]. Qing Pan developed an experimental setup for measuring friction force in hydraulic cylinders under different working conditions based on displacement and chamber pressure measurements, and studied the effects of speed, external load, seal diameter, and seal profile on steady-state friction force [13]. Shen Min et al. used finite-element software to simulate pneumatic combined seals for shafts and analyzed how friction factors, reciprocating speed, compression ratio, and medium pressure affect sealing performance [14]. Chong Xiang et al. built an elastohydrodynamic lubrication simulation model for reciprocating rod seals with textured rods and analyzed the effects of different texture types and parameters on sealing performance [15]. XIANG et al. established transient dynamic equilibrium conditions for reciprocating rod seals with textured rods based on the viscoelasticity of rubber materials and computed lubricant film pressure under different textures [16].

In summary, fault diagnosis methods are diverse and widely applied in many fields. However, seal faults are difficult to detect and their fault features are often subtle; thus, relatively few studies focus on seal fault diagnosis for flow control valves, especially small-sample studies on reciprocating seal faults. Therefore, this paper investigates the leakage mechanism of reciprocating seals from a flow-field analysis perspective and proposes a CNN-CBAM hybrid architecture for fault diagnosis to identify different fault types. Moreover, to explore how different time–frequency methods affect the constructed network's performance and to

ISSN: 2688-8645

determine the best model for seal fault diagnosis, three stable and widely used methods—STFT、CWT and S-transform—are selected for comparative discussion.

## 2. Seal Signal Acquisition System

A dual-cylinder reciprocating seal test rig is used as an example. The rig and its schematic are shown in Figures 2. The test rig consists of a measurement unit, a control unit, and a hydraulic unit. The measurement unit includes a displacement sensor, a tensile (pull) sensor, and pressure gauges. The control unit includes a computer (PC), an electric motor, a lead-screw slide table, and a data acquisition system. The hydraulic unit consists of a fixed cylinder, a test cylinder, and a high-pressure manual pump. The motor drives the lead-screw slide and the piston in the test cylinder to move reciprocally. The measurement unit transmits measured pressure, tensile force, and displacement to the control unit. The left chamber of the fixed cylinder is connected to the piston oil chamber of the test cylinder. The high-pressure manual pump fills hydraulic oil into the left and right chambers of the fixed cylinder and the piston oil chamber of the test cylinder so that the initial pressures are equal and the piston remains stationary. When a defect exists in the piston seal of the test cylinder, leakage occurs and the pressure in the piston oil chamber of the test cylinder and the left side of the fixed cylinder decreases. The piston in the fixed cylinder shifts to one side. By collecting the tensile-force signal of the test cylinder, the friction-force signal can be obtained.

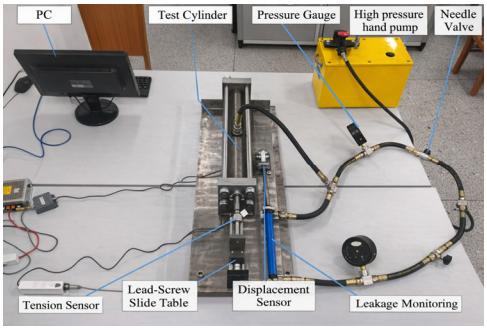



**Figure 2.** Schematic of the dual-cylinder reciprocating seal test rig

The sealing element adopts a slipper-ring reciprocating seal, as shown in Figure 3. The slipper-ring composite seal consists of a slipper ring and an O-ring. The slipper ring is made of polytetrafluoroethylene (PTFE) and directly contacts the piston rod, protecting the O-ring; the O-ring is made of nitrile rubber. PTFE is widely used in the sealing industry as a polymer material, especially in reciprocating seals. PTFE slipper-ring composite seals feature good wear resistance and a low friction coefficient, along with good thermal conductivity and corrosion resistance. They also have a wide service temperature range, meeting harsh downhole operating conditions and sealing requirements.

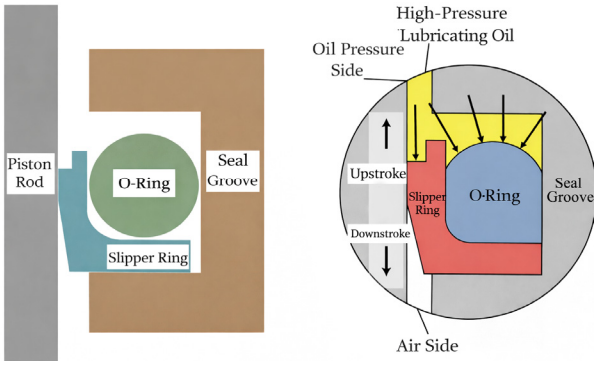



**Figure 3.** Structure of the slipper-ring composite seal

ISSN: 2688-8645

The operating conditions for seal signal acquisition are listed in Table 1. All conditions were tested under 10 MPa pressure. Four structural fault conditions and the normal condition were included. 5.

**Table 1.** Test conditions for reciprocating seal data acquisition

| Category | Test Specimen | Number of Samples | Pressure | Label |
|---|---|---|---|---|
| Normal |  | 600 | 10 MPa | 0 |
| Poor lubrication |  | 600 | 10 MPa | 1 |
| Abrasion |  | 600 | 10 MPa | 2 |
| Aging |  | 600 | 10 MPa | 3 |
| Scuffing |  | 600 | 10 MPa | 4 |

# 3. Fault Diagnosis Method

## 3.1. Time–Frequency Analysis Methods

### 3.1.1. Short-Time Fourier Transform

STFT is a time–frequency analysis method for non-stationary signals [17,18]. It divides a long non-stationary signal into multiple shorter segments (which may overlap). Each segment can be approximately considered stationary, and a Fourier transform is performed on each segment to obtain the frequency components within each time interval. The principle is to use a sliding window to segment the signal into short, equal-length portions. Because the window is small, the signal within it can be assumed stationary. A Fourier transform is then applied to each portion to obtain spectral information. The frequency-domain signal within a time window is approximated as the instantaneous frequency-domain feature, and concatenating all segments yields the time–frequency representation.

Given a continuous-time signal x(t), its STFT is defined as:

$$STFT\{x(t)\}(\tau, f) = \int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-j2\pi ft}dt \quad (1)$$

where w(t−τ) is a window function centered at τ used to extract a short segment of the signal; τ is the time shift indicating the time point under analysis; and f is the frequency.

ISSN: 2688-8645

In STFT, the window used to segment the signal is called the window function. The Hamming window is particularly suitable for small impact signals; it has smaller sidelobe interference and yields clearer impact features, enabling distinction between adjacent fault characteristic frequencies. Therefore, a Hamming window is used in this paper.

The Hamming window is expressed as:

$$\omega[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1 \tag{2}$$

where N is the total window length and n is the current sample point.

### 3.1.2. Wavelet Transform

Traditional Fourier analysis is suitable for stationary signals but has the drawback that the analysis window does not change with frequency in localized analysis. The wavelet transform is a newer transform analysis method derived from STFT. It can perform localized time (or space)–frequency analysis by multi-scale refinement through scaling and translation operations: it refines the high-frequency part in time and the low-frequency part in frequency. As a result, signal details are concentrated on the time and frequency axes, with the time–frequency window size changing with frequency [19–21]. As a typical time–frequency analysis method for non-stationary signals, wavelet transforms are widely used in mechanical fault signal processing.

The continuous wavelet transform is given by:

$$CWT(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\sim} f(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{3}$$

where b is the translation factor, a is the scale factor, the normalization factor is included, ψ* denotes the complex conjugate of the wavelet function, and f(t) is the original signal.

The mother wavelet function is:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{4}$$

Because the extracted friction force originates from reciprocating motion and resembles a periodic signal, the Morlet wavelet is more appropriate. It can be expressed as:

$$\psi(t) = \frac{1}{\sqrt{\pi f_b}} e^{-t^2/f_b} e^{j2\pi f_c t} \quad \psi(t) = \pi^{-1/4} e^{-t^2/2} e^{i\omega_0 t} \tag{5}$$

,

where $f_b$ is the bandwidth parameter, $f_c$ is the center frequency, and j is the imaginary unit.

### 3.1.3. S-Transform

The S-transform is a time–frequency analysis method [22] that combines advantages of STFT and the Fourier transform, providing joint time and frequency information. By applying a variable-width window function in the frequency domain, it uses a smaller window for high frequencies and a larger window for low frequencies, thereby improving time–frequency resolution for non-stationary signals. A Hamming window is used.

The S-transform expression is:

$$S(t,f) = \int_{-\infty}^{\sim} x(\tau)\cdot\omega(t,f,\tau)\cdot e^{-j2\pi f\tau}d\tau \tag{6}$$

where x(τ) is the signal, τ is the time variable, and w(t,f,τ) is the time–frequency window function.

ISSN: 2688-8645

## 3.2. Convolutional Neural Network

Two-dimensional convolution (2D Convolution) is the core operation for processing image data in CNNs. It is designed to extract local features in two-dimensional spatial data. In reciprocating seal fault diagnosis, time–frequency maps are treated as 2D signals, in which each pixel contains energy information in the time–frequency domain. 2D convolution slides a kernel over the input feature map and performs a local weighted summation.

The mathematical expression is:

$$S(i,j)=\sum_{m=1}^{M}\sum_{n=1}^{N} I(i+m-1,j+n-1)\cdot K(m,n)+b \tag{7}$$

Here: S(i,j) represents the value of the output feature map at position (i,j), I is the input feature map with dimensions H×W, K is the convolution kernel (filter) with dimensions M×N, b is the bias term, and M×N is the receptive field of the convolution kernel.

A convolutional layer comprises multiple kernels. Convolution types include valid, same, and full convolution; this paper uses same convolution.

Pooling layers mimic the human visual system by reducing dimensionality and abstracting features. Two common pooling methods are max pooling and average pooling [23,24]. Max pooling retains the maximum value in each pooling region, while average pooling retains the mean value.

Fully connected layers connect features extracted by convolution and pooling and feed them into a classifier. For a c-class classification task, features are flattened into a 1D vector and then classified by a Softmax classifier.

The expression is:

$$z=Wx+b \tag{8}$$

where x is the input feature vector, W is the weight matrix, b is the bias vector, and z is the output vector.

Dropout regularization randomly drops neurons during training with probability p:

$$\overline{x}_i=\begin{cases} x_i/(1-p), & \text{with probability 1-p} \\ 0, & \text{with probability p} \end{cases} \tag{9}$$

The loss function measures the error between the actual and expected outputs to evaluate classification performance. A smaller loss indicates better classification. Cross-entropy loss is commonly used in CNNs and is expressed as:

$$\ell_{CE}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{C} y_{n,i}\log\left(\hat{y}_{n,i}\right) \tag{10}$$

Here: N represents the number of samples in the batch, C represents the total number of categories, $y_{n,i}$ is the true label of the n sample belonging to the i category, and $\hat{y}_{n,i}$ is the probability predicted by the model that the n sample belongs to the i category.

Activation functions introduce nonlinearity and are essential for model expressiveness. They enable the network to learn complex fault patterns from time–frequency maps and extract discriminative feature representations. For computational efficiency, mitigation of vanishing gradients, and faster feature extraction, the ReLU activation function is used in this paper [25], expressed as:

$$ReLU(x)=max(0,x) \tag{11}$$

ISSN: 2688-8645

### 3.3. CBAM Attention Mechanism

Attention mechanisms originate from the human visual system and simulate selective focus on important information. In deep learning, attention allows a model to dynamically focus on the most relevant parts of inputs. The Convolutional Block Attention Module (CBAM) is a lightweight and efficient attention mechanism designed for CNN. By sequentially applying channel attention and spatial attention submodules, CBAM enables the network to adaptively calibrate feature representations, enhancing important features and suppressing irrelevant ones.

The outputs of the channel and spatial attention modules effectively enhance the feature map. Channel attention selects important channels, and spatial attention focuses on important spatial regions. After two rounds of re-weighting, the resulting feature map emphasizes key information relevant to the task.The principle is shown in Figure 4.

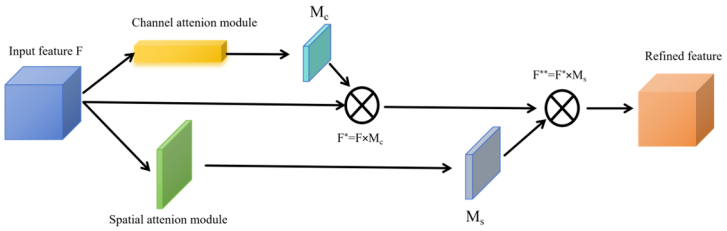



**Figure 4.** Principle of the CBAM module

## 4. Construction of the CNN-CBAM Hybrid Diagnostic Model

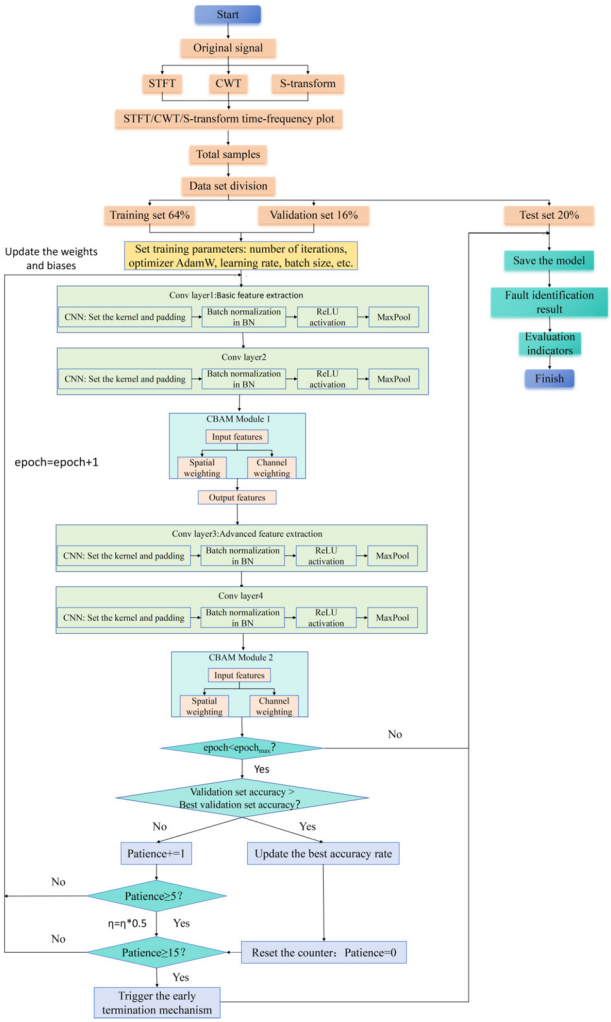



**Figure 5.** Training process of the CNN-CBAM model

ISSN: 2688-8645

The CNN-CBAM hybrid architecture is a deep learning model that organically combines a CNN with CBAM. On top of a conventional CNN, lightweight attention modules are introduced so that the network can adaptively focus on the most discriminative parts of input features, improving feature extraction efficiency and classification performance.

The implementation consists of four steps:

1)Acquire the raw friction-force time-series signal through experiments.

2)Perform simple preprocessing and convert the signal into time–frequency maps using the S-transform, STFT, and CWT; use the maps as inputs for the classification model.

3) Feed the time–frequency maps into the CNN-CBAM architecture for training and validation, and then pass features to fully connected layers for classification.

4) Evaluate the trained model on the test set to obtain accuracy.

AdamW is used to train the CNN-CBAM model. It decouples weight decay from gradient updates, providing effective regularization and reducing overfitting. AdamW adjusts learning rates for parameters using first- and second-moment estimates of gradients, enabling fast convergence early in training and stability later. Weight decay performs L2 regularization to prevent excessive model complexity.

A ReduceLROnPlateau learning-rate scheduler is used to dynamically adjust the learning rate during training. The scheduler monitors validation accuracy; when accuracy does not improve for several consecutive epochs, the learning rate is reduced by half. Early stopping is also used to optimize training: this paper sets patience = 15, and training stops when validation accuracy does not improve for 15 consecutive epochs. This prevents overfitting and saves training time. The training process is shown in Figure 5.

## 5. Results and Analysis

### 5.1. Parameter Settings of the Hybrid Model

**Table 2.** Network architecture parameters

| Module | Layer | Output Size | Kernel | Stride | Padding | Activation |
|---|---|---|---|---|---|---|
| Input | Input | 224×224×3 | - | - | - | - |
| Conv Block 1 | Conv1 | 224×224×32 | 3×3 | 1 | 1 | ReLU |
| Conv Block 2 | Conv2 | 112×112×64 | 3×3 | 1 | 1 | ReLU |
| CBAM 1 | CBAM1 | 56×56×64 | - | - | - | - |
| Conv Block 3 | Conv3 | 56×56×128 | 3×3 | 1 | 1 | ReLU |
| Conv Block 4 | Conv4 | 28×28×256 | 3×3 | 1 | 1 | ReLU |
| CBAM 2 | CBAM2 | 14×14×256 | - | - | - | - |
| Classifier Head | AdaptivePool | 1×1×256 | - | - | - | - |
| | Flatten | 256 | - | - | - | - |
| | Dropout1 | 256 | - | - | - | - |
| | FC1 | 128 | - | - | - | ReLU |
| | Dropout2 | 128 | - | - | - | - |
| | FC2 | 5 | - | - | - | - |
| Output | Output | 5 | - | - | - | Softmax |

The proposed CNN-CBAM hybrid diagnostic model is systematically designed in terms of parameters and inter-layer structure to ensure strong feature extraction ability while optimizing computational efficiency and generalization.

All convolution layers use small 3×3 kernels. They provide sufficient receptive fields while significantly reducing the number of parameters; stacking multiple layers builds deep nonlinear representation ability and helps learn complex fault feature patterns.

ISSN: 2688-8645

All convolution layers set padding = 1 to keep spatial dimensions unchanged before and after convolution, preserving boundary information and avoiding information loss due to early shrinkage in shallow layers.

Each convolution block is followed by a 2×2 max-pooling layer with stride 2. It retains the most salient responses, enhances spatial invariance, and provides robustness to small displacements and deformations. Pooling with stride 2 halves the feature-map size, reducing computational complexity while gradually expanding the receptive field of later convolution layers to capture more global fault patterns.

For CBAM parameters, channel attention uses a shared MLP to compute weights, and spatial attention uses a 7×7 kernel with padding = 3. A larger kernel allows the module to fuse contextual information over a broader neighborhood when generating the spatial weight map, enabling more accurate assessment of the importance of each spatial position and avoiding focus on isolated noise points. The network structure parameters are shown in Table 2.

Model training is controlled by carefully tuned hyperparameters to achieve fast convergence, strong performance, and overfitting prevention. This forms an adaptive and robust training loop: AdamW drives the model quickly toward an optimum; ReduceLROnPlateau fine-tunes learning rates in later stages; and early stopping terminates training once performance saturates, ensuring the best-generalizing model weights. Training hyperparameters are listed in Table 3.

**Table 3.** Training hyperparameters

| Parameter | Value / Range | Description |
|---|---|---|
| Loss function | CrossEntropyLoss | Multiclass cross-entropy loss |
| Optimizer | AdamW | Adam with weight decay |
| Initial learning rate η | 0.004 | Initial learning rate of the optimizer |
| Weight decay | $1\times10^{-4}$ | L2 regularization coefficient |
| Batch size | 16 | Number of samples per batch |
| Max epochs | 80 | Maximum training iterations |
| LR scheduler | ReduceLROnPlateau | Adjust LR based on validation accuracy |
| Scheduler metric | Validation accuracy | Metric monitored by scheduler |
| Scheduler patience | 5 | Reduce LR after 5 epochs without improvement |
| Scheduler factor | 0.5 | LR reduction factor |
| Scheduler mode | max | Higher metric is better |
| Early stopping patience | 15 | Stop after 15 epochs without improvement |
| Early stopping metric | Validation accuracy | Metric monitored by early stopping |
| Min improvement | 0.001 | Minimum change regarded as improvement |

## 5.2. Influence of Different Time–Frequency Methods on Recognition Performance

With a sample size of 3,000, the dataset was split into training, validation, and test sets, with 1,920 and 480 samples for training and validation respectively, and the remaining 600 samples for testing. For the three input representations (S-transform, CWT, and STFT), CNN-CBAM diagnostic models were constructed and compared in terms of training process, convergence characteristics, and classification performance. The relationships between accuracy and epoch for training and validation sets under different inputs are shown in Figure 6.

From the trends of training and validation accuracy, the three input methods differ significantly in convergence speed, stability, and generalization. The S-transform + CNN-CBAM model shows rapid accuracy improvement in early training: training accuracy stabilizes around epoch 11,

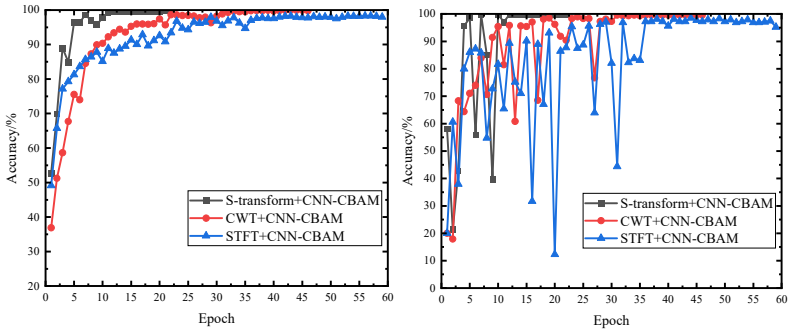
ISSN: 2688-8645

and validation accuracy stabilizes around epoch 10. The model converges in only about 22 epochs. Training and validation curves show similar trends with small fluctuations, indicating fast convergence and good training–validation balance with limited overfitting.

By contrast, the CWT + CNN-CBAM model exhibits more pronounced fluctuations. Training accuracy gradually converges around epoch 32, while validation accuracy fluctuates strongly within the first 28 epochs, sometimes dropping to about 60%. The total number of epochs reaches about 46. This suggests that although CWT can capture certain time–frequency characteristics, under limited samples it may introduce redundant information, making validation generalization less stable.

The STFT + CNN-CBAM model converges the slowest. Training and validation accuracies stabilize only after about 36 epochs, with the total number of epochs reaching about 59. Validation accuracy fluctuates greatly during training and may drop to as low as 12%, indicating sensitivity to sample distribution and susceptibility to noise and non-stationary components under STFT input, with a certain risk of overfitting.

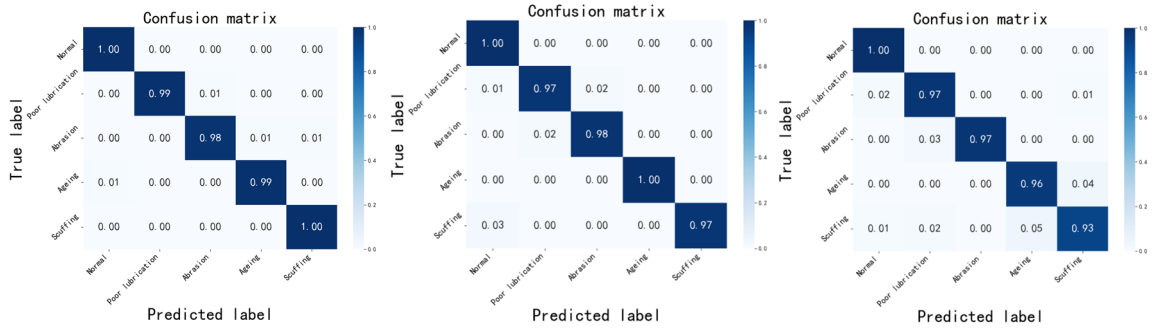
Overall, S-transform + CNN-CBAM clearly outperforms CWT and STFT in convergence speed, stability, and generalization, and is more suitable for feature extraction and intelligent diagnosis of reciprocating seal friction states.




(1) training set (2) validation set

**Figure 6.** Accuracy vs. epochs under different inputs

Based on this, the 600-sample test set was used to validate model performance. Test results indicate that the test accuracies of S-transform + CNN-CBAM, CWT + CNN-CBAM, and STFT + CNN-CBAM are 99.33%, 98.64%, and 96.67%, respectively. The corresponding confusion matrices are shown in Figure 7.




(a)S-transform + CNN-CBAM (b)CWT + CNN-CBAM (c)STFT + CNN-CBAM

**Figure 7.** Confusion matrices under different time–frequency inputs

From the confusion matrices, the S-transform + CNN-CBAM model achieves high recognition precision across all five operating-condition categories, with predictions concentrated along the main diagonal and very few misclassifications, demonstrating strong discriminative

ISSN: 2688-8645

capability. The CWT + CNN-CBAM model shows slight confusion between a few adjacent conditions, but overall classification remains good. In contrast, the STFT + CNN-CBAM model produces more misclassifications between states such as “Aging” and “Scuffing,” indicating limited ability of its time–frequency representation to characterize some complex friction features.

Further quantitative evaluation of the three models is provided in Table 4. Under 3,000 samples, S-transform + CNN-CBAM achieves 99.33% on accuracy, precision, recall, and F1 score, yielding the best overall performance; CWT + CNN-CBAM ranks second; and STFT + CNN-CBAM performs the worst.

**Table 4.** Performance metrics for models with different inputs

| Input | Samples | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) |
|---|---|---|---|---|---|
| S-transform + CNN-CBAM | 3000 | 99.33 | 99.33 | 99.33 | 99.33 |
| CWT + CNN-CBAM | 3000 | 98.67 | 98.69 | 98.67 | 98.67 |
| STFT + CNN-CBAM | 3000 | 96.67 | 96.68 | 96.67 | 96.66 |

In summary, the time–frequency features generated by the S-transform preserve frequency adaptivity while more effectively representing the non-stationary characteristics of friction signals. They match the CNN-CBAM network structure better, yielding clear advantages in convergence speed, stability, and classification accuracy.

## 5.3. Performance Comparison with Other Recognition Methods

To verify the effectiveness of the proposed model, comparative experiments were conducted among the proposed CNN-CBAM hybrid network and several typical fault recognition methods, including a plain CNN, Support Vector Machine (SVM), Random Forest (RF), and Extreme Learning Machine (ELM). Test accuracies on the same dataset are shown in Table 5.

As shown, deep-learning-based CNN methods significantly outperform traditional machine learning methods. The CNN-CBAM architecture achieves the highest test accuracy of 99.33%, demonstrating superior fault recognition. The plain CNN also achieves high accuracy with S-transform inputs but is slightly lower than the proposed method for the following reasons: (1) Attention enhances key feature representations. CBAM strengthens discriminative time–frequency features such as energy mutations in relevant frequency bands and frequency drifts, suppressing irrelevant background information. (2) Stronger discrimination among multiple fault modes. Reciprocating seal faults often show similar frequency distributions and subtle energy changes in S-transform maps; CBAM improves separability among highly similar categories. (3) Stable performance gain even in high-accuracy regimes. When accuracy is already around 98%, further improvement is typically difficult, yet CNN-CBAM still yields stable gains, indicating stronger generalization and robustness.

Compared with traditional machine learning methods (SVM, RF), CNN-based methods show clear advantages mainly because: (1) Feature acquisition differs fundamentally. SVM and RF rely on manually extracted wavelet features with limited representational power, which cannot fully capture complex non-stationary time–frequency characteristics; CNN-CBAM uses S-transform maps directly and learns multi-scale features automatically via multi-layer convolution, enabling end-to-end feature extraction and classification. (2) Traditional methods adapt poorly to complex operating conditions. RF can yield unstable decision boundaries when sample distributions are complex and class boundaries are blurred, and may underperform on minority classes; deep networks can learn more robust boundaries when sample sizes are sufficient.

ISSN: 2688-8645

**Table 5.** Comparison of recognition methods

| Method | Sample Type | Test Accuracy |
|---|---|---|
| CNN-CBAM architecture | S-transform time–frequency map | 99.33% |
| CNN | S-transform time–frequency map | 98.17% |
| SVM (hyperplane) | Wavelet feature values | 91.67% |
| Random Forest (RF) | Wavelet feature values | 87.64% |

## 6. Conclusion

This paper proposes a CNN-CBAM-based fault diagnosis method for reciprocating seals in flow control valves, aiming to improve the accuracy and robustness of seal fault diagnosis in hydraulic systems. By combining attention mechanisms with time–frequency analysis, the diagnostic performance is enhanced.

(1) A reciprocating seal test rig was successfully built and friction-force data were collected for different fault modes (e.g., abrasion, aging, scuffing, poor lubrication). Using S-transform, CWT, and STFT, the signals were converted into time–frequency maps as input features for the CNN-CBAM network, enabling effective classification of different fault states.

(2) After comparative analysis among the S-transform, CWT, and STFT, results show that the S-transform combined with CNN-CBAM performs best. It achieves faster convergence and higher accuracy, and shows strong stability and generalization, clearly outperforming CWT and STFT inputs.

(3) Compared with other recognition algorithms, the CNN-CBAM network achieves slightly higher accuracy than a plain CNN and markedly higher accuracy than traditional machine learning methods such as SVM and RF. This verifies the effectiveness of deep learning for non-stationary friction-signal analysis and shows that the CNN-CBAM architecture can enhance feature extraction capability.

## References

[1] Guo Dong. Parametric Design and Analysis of Downhole Flow Controller for Intelligent Wells [D]. Southwest Petroleum University, 2018.

[2] Xia Xiong. Study on Condition Monitoring and Fault Diagnosis Technology of Gas Seal Face Operating State [D]. Beijing University of Chemical Technology, 2012.

[3] Mayer E. Mechanical Seals [M]. London: Newnes-Butterworth, 1977.

[4] Min Zou, I. Green. Clearance control of a mechanical face seal [J]. Tribology Transactions, 1999, 42 (3): 535–540.

[5] Digard J., Gentile M. Experimental study on lubrication condition of low-pressure mechanical seals [C]. Proceedings of the International Gas Seal Conference, Beijing: China Machine Press, 1991.

[6] W. Anderson, R. Salant, J. Jarzynski. Ultrasonic detection of lubricating film collapse in mechanical seals [J]. Tribology Transactions, 1999, 42(4): 801–806.

[7] W. Anderson, J. Jarzynski, R. Salant. Condition monitoring for liquid-lubricated mechanical seals [J]. Tribology Transactions, 2001, 44(3): 479–483.

[8] T. Reddyhoff, R.S. Dwyer-Joyce, P. Harper. A new approach for the measurement of film thickness in liquid face seals [J]. Tribology Transactions, 2008, 51(2): 140–149.

[9] Li Jihe (ed.). Mechanical Seal Technology [M]. Chemical Industry Press, 1988.

[10] Gu Yongquan. Practical Technology of Mechanical Seals [M]. China Machine Press, 2001.

[11] Zhou Jianfeng; Gu Boqin. Reliability evaluation method of mechanical seals based on Monte Carlo [J]. Lubrication Engineering & Sealing, 2006(02): 108–110+141.

[12] Wei Long, Sun Jianjun. Reliability analysis of mechanical seals [J]. Chemical Industry and Engineering Technology, 2002(03): 20–22+3.

[13] Qing Pan, Yunlong Zeng, Yibo Li, Xuepeng Jiang, Minghui Huang. Experimental investigation of friction behaviors for double-acting hydraulic actuators with different reciprocating seals [J]. Tribology International, Volume 153.

[14] Shen Min, Song Meili, Zhang Hua. Simulation analysis of sealing performance of ZHM pneumatic combined seals for reciprocating shafts [J]. Machinery Manufacturing & Automation, 2021, 50(4): 104–108.

[15] Chong Xiang. Elastohydrodynamic lubrication simulation of reciprocating rod seal with textured rod [J]. Tribology International, Volume 158.

[16] XIANG C., GUO F., LIU X., et al. Thermo-elastohydrodynamic lubrication simulation of reciprocating rod seals under transient condition [J]. Tribology International, 2021, 153: 106603.

[17] Nawab S., Quatieri T., Lim J. Signal reconstruction from short-time Fourier transform magnitude [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1983, 31(4): 986–998.

[18] Griffin D., Lim J. Signal estimation from modified short-time Fourier transform [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(2): 236–243.

[19] Wang Jiangping, Sun Wenli. Analysis and fault diagnosis of gear vibration signals based on wavelet packet energy spectrum [J]. Journal of Mechanical Transmission, 2011, 35(01): 55–58.

[20] Wang Ling. Pattern recognition of surface EMG signals based on wavelet packet transform [J]. Modern Electronics Technique, 2011, 34(17): 122–124+128.

[21] Xu Pan, Su Guangwei. Steganalysis based on wavelet coefficient correlation [J]. Computer Engineering and Applications, 2012, 48(28): 178–182+213.

[22] Stockwell R.G., Mansinha L., Lowe R.P. Localization of the complex spectrum: The S-transform [J]. IEEE Transactions on Signal Processing, 1996, 44(4): 998–1001. DOI: 10.1109/78.492555.

[23] Boureau Y.-L., Roux N., Bach F., et al. Ask the locals: Multi-way local pooling for image recognition. Proceedings of ICCV 2011. Barcelona, Spain, 2011: 2651–2658.

[24] Zeiler M., Fergus R. Stochastic pooling for regularization of deep convolutional neural networks. arXiv, 2013:1301.3557v1.

[25] Nair V., Hinton G.E. Rectified linear units improve restricted Boltzmann machines [C] // Proceedings of ICML'10. Haifa: International Machine Learning Society, 2010: 807–814.