

Article

Identification Method of Dynamometer Cards Based on Improved MobileNetV2 in Fault Identification of Sucker Rod Pumping System

Wen Sun , Lin Guo  and Wei Zeng

School of Mechanical Engineering, Xi'an Shiyou University, Xi'an 710065, China; 17739394030@163.com (L.G.); wzeng@xsyu.edu.cn (W.Z.)

* Correspondence: sunwen@xsyu.edu.cn

Featured Application

This method can provide a useful lightweight model reference for dynamometer-card-based fault identification in a sucker rod pumping system.

Abstract

Existing fault identification methods still face problems such as insufficient lightweight design, inadequate multi-scale feature representation capability, and difficulty in model training under small-sample conditions, making it difficult to meet the requirements of oilfield production sites. For the dynamometer-card-based fault identification task of the sucker rod pumping system, this paper makes targeted improvements to MobileNetV2 and a lightweight fault identification method based on transfer learning and improved MobileNetV2 is proposed. The method uses MobileNetV2 as its basic framework, improves the activation function, and introduces a multi-scale feature extraction module and an attention mechanism, thereby improving the model's identification capability while maintaining a lightweight design. Experiments were carried out using a constructed dynamometer card dataset based on competition data and measured oilfield data. The proposed method was compared with representative deep learning methods, and a cross-source external validation experiment was further conducted using independently measured oilfield samples. The results show that the improved MobileNetV2 achieved a mean accuracy of 95.01%, with a standard deviation of 0.09%, 2.28 M parameters, a model size of 9.01 MB, 536.72 M MACs, and a CPU inference latency of 29.62 ms per image. In terms of mean accuracy, the improved MobileNetV2 outperforms EfficientNetV2, ResNet34, DenseNet121, MobileNetV3, ShuffleNetV2, and the original MobileNetV2, while achieving a mean accuracy of 90.16% on the cross-source external validation set. This method achieves a relatively good balance between identification performance and model complexity on the dataset used in this study, and it can provide a useful lightweight model reference for dynamometer-card-based fault identification in a sucker rod pumping system.



Received: 10 July 2026

Revised: 21 September 2026

Accepted: 10 October 2026

Published: 14 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: sucker rod pumping system; fault identification; dynamometer card; transfer learning; multi-scale feature extraction; MobileNetV2

1. Introduction

The sucker rod pumping system is the core unit for onshore oilfield exploitation and has long been responsible for lifting downhole crude oil to the surface. The production

environment in which it operates is harsh, and its stable operation is influenced by many factors, with various faults prone to occur, particularly under extreme climates, corrosive environments, or high-load conditions. These faults not only reduce oilfield production but also create safety hazards. Therefore, accurate fault identification in the sucker rod pumping system is of great significance for achieving energy savings and production enhancement in oilfields.

In practical applications, researchers often use oil well production data, such as electrical parameters [1,2] and dynamometer cards [3] to carry out fault identification in the sucker rod pumping system. Among them, dynamometer cards are closed curves drawn based on the variation relationship between polished rod load and displacement within a complete reciprocating motion cycle [4]. The load–displacement relationship reflected by dynamometer cards is closely related to the dynamic response of the sucker rod pumping system. During the reciprocating operation of the pumping unit, the surface polished-rod load is not only related to the operating state of the downhole pump, but is also affected by factors such as the elastic deformation, longitudinal vibration, inertia, friction, and impact of the sucker rod string. When abnormal operating conditions such as insufficient fluid supply, gas impact, or upstroke pump bumping occur in the system, the operating process of the downhole pump and the stress state of the sucker rod string change accordingly, which are further reflected in differences in the contour, local morphology, and load distribution of the dynamometer cards. To address the vibration and dynamic load problems during the operation of sucker rod strings, previous studies have carried out relevant investigations from the perspective of mechanical structures. For example, slotted-shell friction dampers and shell-type shock absorbers can change the stiffness and vibration response characteristics of the sucker rod string suspension system through frictional contact, elastic deformation, and damping effects [5,6], providing a mechanical mechanism reference for understanding the dynamic stress of sucker rod strings and the load variations in dynamometer cards. The shape characteristics of dynamometer cards for different types of operating conditions are significantly different, which can directly reflect the operating state of the sucker rod pumping system. Therefore, the fault identification method based on dynamometer cards has become the most commonly used fault identification method for the sucker rod pumping system in oilfields at present.

Since Gibbs et al. [7] established the wave equation and laid the theoretical foundation, the fault identification method based on dynamometer cards has gradually evolved from manual experience judgment to computer-aided analysis. Early studies mainly used an expert system [8] to simulate the judgment rules of human experts to realize fault identification of the sucker rod pumping system. Such methods rely entirely on expert experience, have no learning ability, and possess poor generalization capability. With the development of artificial intelligence technology, image identification methods based on deep learning have provided new ideas for fault identification of dynamometer cards. Xu et al. [9] constructed a self-organizing competitive neural network to realize feature extraction and operating condition clustering of dynamometer cards. Zhang et al. [10] extracted dynamometer card features through fast discrete curvelet transform and used the sparse multi-graph regularized extreme learning machine algorithm as a classifier to conduct fault identification of the sucker rod pumping system. Cheng et al. [11] proposed a fault identification method based on transfer learning and AlexNet-SVM, in which AlexNet was used to extract dynamometer card features and support vector machine was combined to complete identification. Wu et al. [12] combined transfer learning with the DenseNet model to alleviate problems such as insufficient samples, imbalanced data distribution, and performance degradation of deep networks. Ye and Yi [13] improved the AlexNet structure and realized multiple types of fault identification by simplifying the

network structure and introducing BN layers. Li et al. [14] combined graphic features with Fourier descriptors and improved the model's representation capability for morphological differences in dynamometer cards through a dual-backbone network and an interactive fusion module, significantly improving the robustness and accuracy of fault identification for the sucker rod pumping system. Li et al. [15] proposed an oil well fault identification method based on improved Fourier descriptors and a fast density clustering RBF neural network. Zhang et al. [16] combined transfer learning with the Vision Transformer to improve the identification performance for similar fault categories. He et al. [17] proposed a working condition identification method for pumping wells based on a four-segment time-frequency feature matrix and convolutional neural network. By extracting time-frequency features and constructing feature matrices with multiple dynamometer cards, this method addressed the problems of large computational cost in traditional image-based methods and large differences in field data. Dong et al. [18] fused transfer learning with Swin Transformer to enhance the identification capability for dynamometer cards. Sreenivasan and Krishna [19] combined a pre-trained GoogLeNet with machine learning classifiers to construct a fault identification model for dynamometer cards. In addition, Qian et al. [20] proposed a single-positive multi-label fault identification method for concurrent multi-fault problems, while Wang et al. [21] alleviated the problems of insufficient dynamometer card samples and imbalanced category distribution based on the ParaDC data augmentation mechanism. Xu et al. [22] proposed VOC-LLM for mechanical fault diagnosis under varying operating conditions. By integrating multidimensional vibration features and performing lightweight adaptation of a large language model, the method achieves fault identification and severity assessment, demonstrating the application potential of large language models in fault diagnosis under complex operating conditions and providing a new research direction for intelligent and interpretable industrial fault diagnosis.

In summary, the above studies have extensively investigated feature extraction and operating-condition clustering, model training under limited and imbalanced sample conditions, discrimination of similar fault categories, simultaneous identification of multiple faults, as well as fault identification and severity assessment under varying operating conditions, and have achieved certain progress. However, for lightweight fault identification, challenges remain in distinguishing similar fault categories, inadequate representation of multi-scale morphological features, and difficulty in model training under limited sample conditions.

To address the above problems, this paper considers the subtle local morphological differences, significant contour-scale variations, and lightweight model requirements in the dynamometer-card-based fault identification task. Based on the MobileNetV2 network framework, improvements are made for the fault identification task of the sucker rod pumping system from three aspects: the attention mechanism, activation function, and multi-scale feature extraction structure. A lightweight fault identification method based on transfer learning and improved MobileNetV2 is proposed.

The proposed method designs the feature extraction strategy according to the morphological characteristics of dynamometer cards. The coordinate attention mechanism is embedded into the inverted residual modules of MobileNetV2 to enhance the model's representation capability for local morphological variations and spatial positional information. In terms of the activation function, the original ReLU6 in MobileNetV2 is replaced with LeakyReLU6 to preserve part of the negative feature responses and improve the representation of subtle fault features. Meanwhile, an improved Inception A multi-scale feature extraction module is introduced at the front end of the backbone network to enhance the network's early-stage extraction capability for contour variations and local detail features. In addition, ImageNet-1K pre-trained weights are adopted for transfer learning to alleviate

the training difficulty under limited dataset samples. Experimental results show that the proposed method achieves relatively good fault identification performance and maintains the lightweight model scale with only a slight increase in the number of parameters.

2. Problem Description

The sucker rod pumping system is mainly composed of three parts: first, is the surface driving equipment (pumping unit); second, is the oil well pump installed at the lower part of the tubing string; third, is the sucker rod string that transmits the motion and power of the surface equipment to the plunger of the downhole oil well pump to make it reciprocate up and down [23]. Its core function is to lift the oil in the oil well to the surface by using a special pumping device, so as to maintain the pressure difference in liquid flow between the bottom hole and the oil layer, and ensure that oil and gas continuously flow to the bottom hole.

Dynamometer cards are closed curve diagrams composed of curves representing load variation with displacement. Their shapes can directly reflect the operating state of the sucker rod pumping system and help operators understand the equipment condition of the sucker rod pumping system. The theoretical dynamometer cards of the sucker rod pumping system are closed curve diagrams formed with suspension point displacement as the abscissa and suspension point load as the ordinate under static load, presenting a regular parallelogram shape. Their four sides correspond to the four working stages of the sucker rod pumping system respectively: loading during upstroke, constant load during upstroke, unloading during downstroke, and constant load during downstroke [24].

In actual production, the sucker rod pumping system is easily disturbed by factors such as gas impact (GI), insufficient fluid supply (IFS), and upstroke pump bumping (UPB). These disturbances change the liquid suction and discharge process of the downhole oil well pump and the stress state of the sucker rod string, causing dynamometer cards to deviate from the regular parallelogram and show corresponding shape variations. Therefore, dynamometer cards serve as the core basis for intuitively judging the operating state of the sucker rod pumping system and achieving accurate fault identification. This paper considers 10 typical operating conditions: normal operation condition (NOC), insufficient fluid supply (IFS), gas impact (GI), air lock (AL), upstroke pump bumping (UPB), downstroke pump bumping (DPB), delayed closing of moving valve (DCMV), leakage of moving valve (LMV), sand influence (SI), and sucker rod parting (SRP). The dynamometer cards corresponding to various operating conditions are shown in Figure 1.

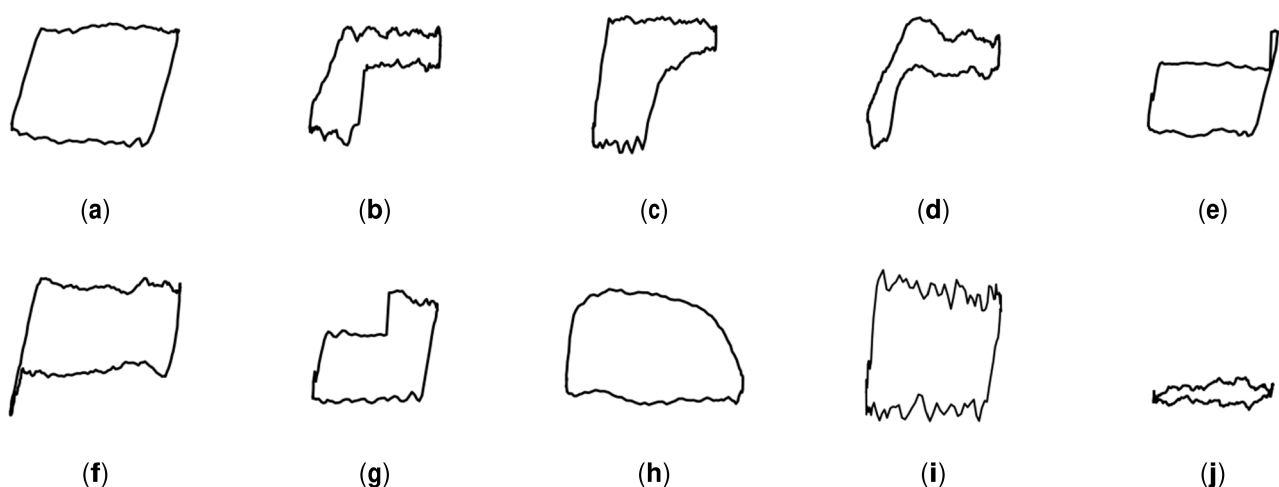


Figure 1. Dynamometer cards corresponding to ten typical operating conditions: (a) NOC; (b) IFS; (c) GI; (d) AL; (e) UPB; (f) DPB; (g) DCMV; (h) LMV; (i) SI; (j) SRP.

3. Materials and Methods

3.1. Relevant Theories

3.1.1. MobileNetV2

MobileNetV2 is a lightweight convolutional neural network model proposed by the Google team in 2018. Taking depthwise separable convolution as the basic architecture, this model has the advantages of high accuracy, fast computing speed and a small number of parameters [25]. The basic flow of depthwise separable convolution is shown in Figure 2.

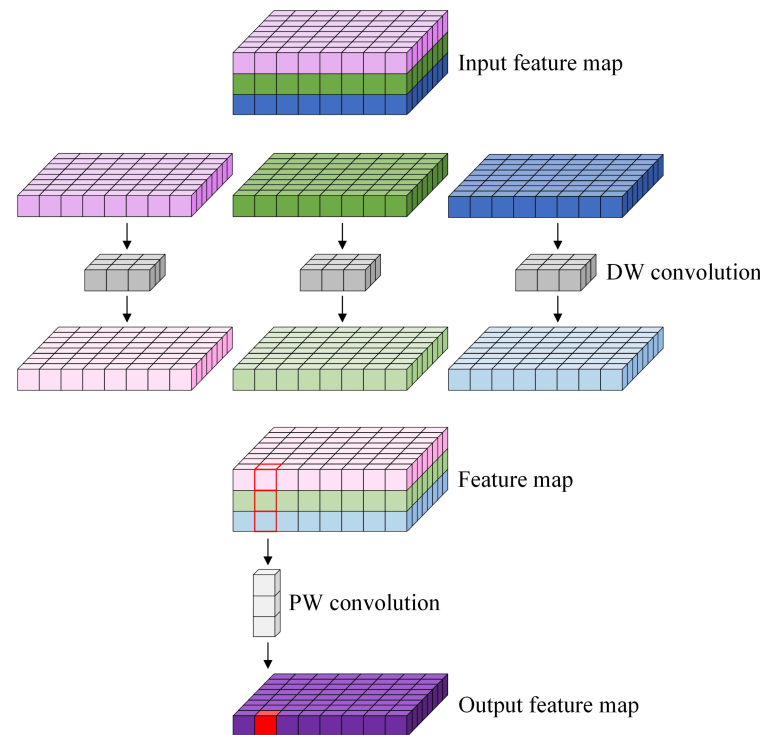


Figure 2. Basic procedure of depthwise separable convolution.

Compared with standard convolution, depthwise separable convolution splits the spatial feature extraction and channel feature fusion of standard convolution into depthwise convolution (DWC) and pointwise convolution (PWC). It reduces the number of parameters and computational complexity of the model without impairing the feature extraction ability. Depthwise convolution allocates an independent convolution kernel to each input channel of the input feature map for spatial feature extraction, without cross-channel information fusion. Pointwise convolution uses 1×1 convolution kernels to integrate cross-channel features, fusing spatial features from different channels into new feature dimensions.

The computational complexity of DWC is expressed in Equation (1):

$$P_D = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \quad (1)$$

The comparison of computational complexity with standard convolution is expressed in Equation (2):

$$\frac{P_D}{P_C} = \frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (2)$$

where P_D is the computational complexity of depthwise separable convolution, P_C is the computational complexity of standard convolution, D_K is the convolution kernel size, D_F is

the height and width of the input feature matrix, M is the number of channels of the input feature matrix, and N is the number of channels of the output feature matrix.

In addition to taking depthwise separable convolution as the basic architecture, MobileNetV2 also proposes an inverted residual structure. Unlike conventional residual connections, its residual shortcut directly connects the low-dimensional bottleneck layers of the input and output, instead of the high-dimensional feature layers in traditional residual connections. The inverted residual structure of MobileNetV2 first realizes dimension expansion via pointwise convolution. After capturing rich features through depthwise convolution, it performs dimension reduction on the feature map by pointwise convolution again. This structure enables the model to retain excellent feature extraction capability while achieving lightweight advantages. In MobileNetV2, the inverted residual structure has two strides: 1 and 2. A residual connection is applied only when the stride is 1 and the input and output feature matrices have the same size and depth. Furthermore, in the last pointwise convolution layer of the inverted residual module, a linear activation function replaces the nonlinear activation function to prevent the loss of low-dimensional feature information.

3.1.2. Coordinate Attention Mechanism

An attention mechanism dynamically allocates different weights according to the importance of different parts of the input content for the current task, enabling the model to focus on key parts and ignore task-irrelevant parts during information processing, thereby improving the processing efficiency and accuracy of the model [26]. Coordinate attention (CA) is a lightweight spatial-channel fusion attention mechanism proposed by Hou et al. [27] at CVPR 2021. It embeds spatial positional information into channel attention, endowing attention weights with both channel feature screening capability and spatial position perception capability. This compensates for the shortcoming of conventional attention mechanisms that emphasize channel feature mining but neglect spatial position correlation, making it more suitable for the feature extraction requirements of dynamometer-card-based fault identification and a suitable choice for the dynamometer-card-based fault identification task of the sucker rod pumping system.

In the fault identification task in the sucker rod pumping system, the core fault features of dynamometer cards are mainly reflected in the morphological changes in the two-dimensional space composed of load and displacement. For example, IFS leads to an advance of the unloading point, a missing lower-right corner of the image, and a sudden inclined drop of the unloading line from a nearly horizontal state; GI results in slow loading and unloading, a missing lower-right corner of the image, and an arc-shaped unloading line. Such similar fault types place forward higher requirements for the spatial perception capability of the attention mechanism. CA mainly consists of two core components: coordinate information extraction and coordinate attention weight generation, whose structure is shown in Figure 3.

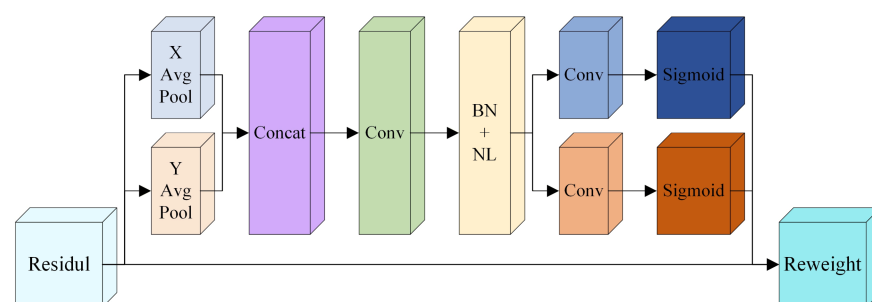


Figure 3. The structure of coordinate attention.

Coordinate information extraction performs vertical global average pooling and horizontal global average pooling on the input feature map, respectively; decomposes the two-dimensional spatial features of the feature map into two independent one-dimensional spatial dimensions along vertical and horizontal directions, and obtains spatial information and position-associated features for each direction separately. Vertical pooling performs global average pooling along the width W for each channel, only aggregates features of the same row, and retains coordinate information along the height H dimension, generating a feature map with the shape of $C \times H \times 1$. Horizontal pooling performs global average pooling on each channel along the height H direction, aggregates features of the same column and retains coordinate information along the width W dimension, generating a feature map with the shape of $C \times 1 \times W$. For an input feature map with C channels, height H and width W , the output representations of the c -th channel in the two directions are shown in Equations (3) and (4) respectively.

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (3)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (4)$$

Coordinate attention weight generation involves spatially concatenating vertical and horizontal direction feature maps, reducing the number of channels via a 1×1 convolutional layer to embed spatial and positional information into the channel features, and then applying a non-linear transformation using batch normalization and the ReLU activation function. Subsequently, the fused features are split along the spatial dimension into two independent features corresponding to the vertical and horizontal directions, respectively. These are restored to their original channel counts via 1×1 convolutional layers, and the sigmoid activation function is used to generate attention weights within the $[0, 1]$ interval. The input features are then subjected to channel-wise and position-wise weighted optimization. The final output is shown in Equation (5).

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (5)$$

3.1.3. LeakyReLU6

The choice of activation function directly influences the model's ability to capture fault features in dynamometer cards. The ReLU6 activation function used in MobileNetV2 restricts the output to the interval $[0, 6]$, thereby preventing the vanishing gradient problem caused by excessively large activation values, while reducing the model size and computational complexity to meet the requirements of lightweight deployment. However, when the input to the ReLU6 activation function is negative, its output gradient is always 0, which prevents the neuron from updating its parameters, leading to the phenomenon of neuron death and impairing the model's learning ability. Its mathematical expression is shown in Equation (6).

$$\text{ReLU6}(x) = \min[\max(0, x), 6] \quad (6)$$

The LeakyReLU activation function is an improved version of the ReLU activation function; it introduces a negative slope coefficient α in the negative region, thereby mitigating the problem of neuron death caused by gradient vanishing [28]. However, while this activation function preserves the negative input gradient, it has no upper bound on the

output, which can easily lead to feature divergence and is not conducive to the quantization and deployment of the model. Its mathematical expression is shown in Equation (7).

$$\text{LeakyReLU}(x) = \begin{cases} x & x \geq 0 \\ \alpha x & x < 0 \end{cases} \quad (7)$$

The LeakyReLU6 activation function incorporates the negative slope coefficient α from the LeakyReLU activation function into the ReLU6 activation function, thereby retaining the upper output clipping property of ReLU6 while preserving negative outputs and non-zero gradients for negative inputs through the negative slope coefficient α . Its mathematical expression is shown in Equation (8).

$$\text{LeakyReLU6}(x) = \begin{cases} 6 & x > 6 \\ x & 0 \leq x \leq 6 \\ \alpha x & x < 0 \end{cases} \quad (8)$$

During the experiments, α was typically set to a value between 0.01 and 0.2. As shown in Figure 4, the LeakyReLU6 activation function effectively mitigates the “neuron death” issue commonly observed with the ReLU6 activation function, while also preventing the divergence of feature values caused by the lack of an upper output limit in the LeakyReLU activation function. Consequently, this enhances the model’s ability to learn subtle fault features in dynamometer cards.

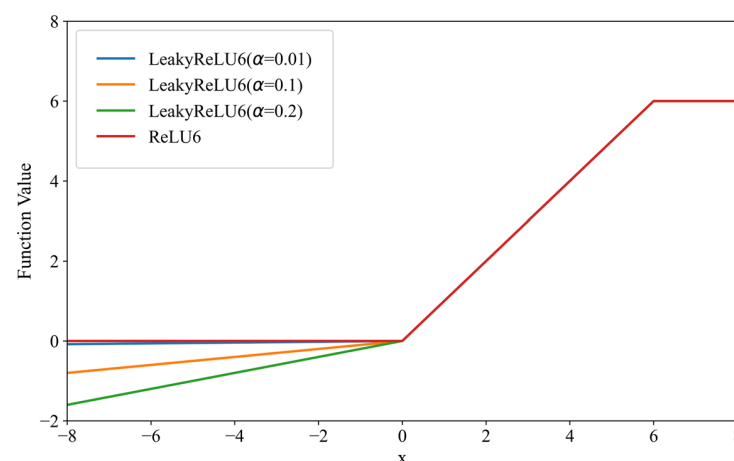


Figure 4. Activation function comparison diagram.

3.1.4. Multi-Scale Feature Extraction Module

Inception is a deep convolutional neural network architecture proposed by the Google team and serves as the core subnetwork structure within GoogLeNet [29]. The original Inception module structure is shown in Figure 5. Its core idea is to adaptively extract feature information at different scales by integrating multi-scale convolutional operations and pooling branches in parallel, thereby enhancing the model’s ability to capture multi-scale image features. The module consists of four branches: 1×1 convolutions provide the smallest receptive field to capture local point features; 3×3 convolutions extract spatial structures, local shapes, and edge textures at a medium scale; 5×5 convolutions offer a larger receptive field to better capture large-scale spatial correlations; and 3×3 max-pooling filters and retains the most prominent features within local windows, achieving spatial dimensionality reduction and noise suppression, thereby complementing the convolution branches. The feature maps output by each branch are fused through concatenation to form the module’s final output, thereby enabling the simultaneous capture of multi-scale

features. However, because the module incorporates large-scale 5×5 convolutions, while expanding the receptive field, it also introduces a significant number of parameters and computational complexity, increasing the network's complexity.

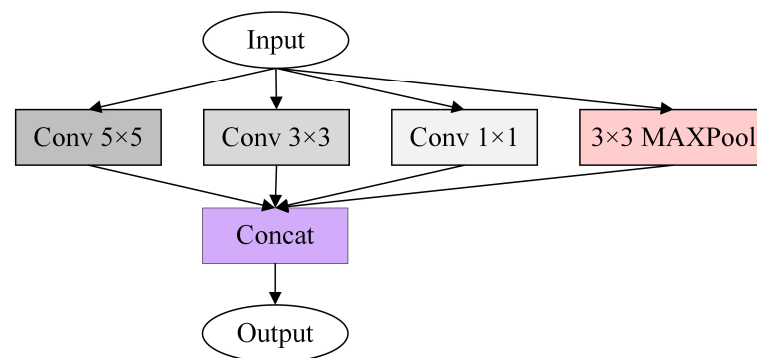


Figure 5. Original inception network structure.

To address this issue, the Google team subsequently proposed InceptionV4 which optimized the network architecture and computational efficiency [30]. The Inception A module in InceptionV4 retains the core architecture of four branches, as shown in Figure 6, and introduces several key improvements over the original Inception module: First, Inception A reduces the number of model parameters while maintaining the same receptive field by equivalently replacing 5×5 convolutions with two consecutive layers of 3×3 convolutions. Second, 1×1 convolutions are introduced before the 3×3 and 5×5 convolution branches to reduce the dimensionality of the feature channels, and max pooling is replaced with average pooling to better extract global structural features; and finally, a 1×1 convolution is added after the pooling branch to transform features, and a batch normalization layer is incorporated after all convolutional branches, effectively improving the model's training stability and convergence speed.

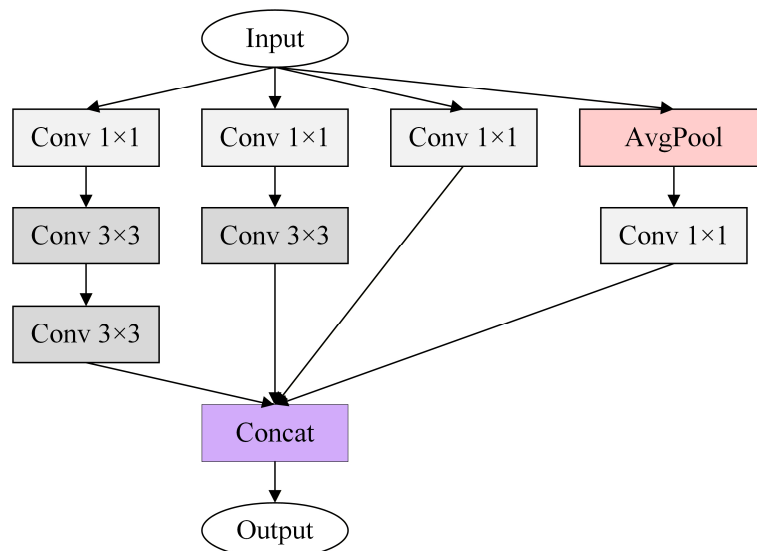


Figure 6. Network structure of Inception A.

3.1.5. Transfer Learning

Transfer learning is an efficient deep learning training strategy. Its core concept involves transferring knowledge and feature representations learned from one domain or task to another related but distinct domain or task with limited data and specific conditions, such as fault identification in the sucker rod pumping system. This approach addresses

issues such as scarce samples in the target domain, slow model convergence, and poor generalization capabilities, thereby enhancing the model's ability to identify target fault conditions. In practical engineering applications, dynamometer card datasets for the sucker rod pumping system often face challenges such as difficult sample collection, high manual annotation costs, and high feature similarity among similar faults. Training deep learning models from scratch is prone to overfitting and convergence difficulties, making it hard to ensure the accuracy and stability of fault identification; in contrast, transfer learning leverages the general feature extraction capabilities of pre-trained models to reduce the training difficulty of the sucker rod pumping system fault identification tasks, improve training stability and feature learning under limited samples, increase training efficiency, and better adapt to the rapid deployment requirements of lightweight models. The transfer learning process is illustrated in Figure 7.

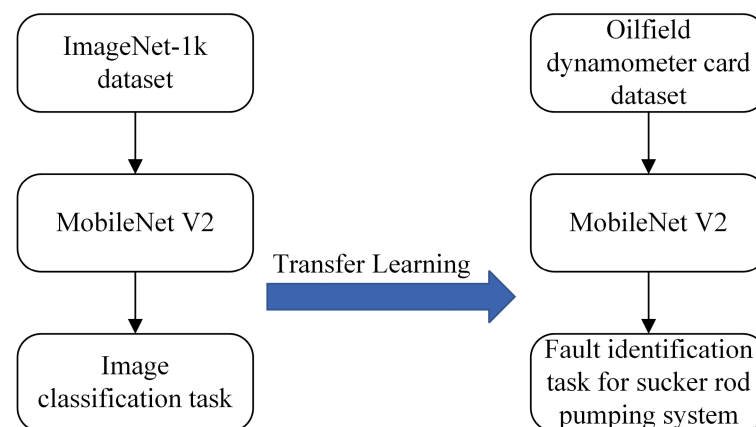


Figure 7. The process of transfer learning.

3.2. Model Design

MobileNetV2 centers on depthwise separable convolutions and an inverted residual structure, utilizing linear bottleneck layers to achieve lightweight feature extraction. While ensuring the model's fundamental feature representation capabilities, it significantly reduces the number of parameters and computational complexity, making it suitable for edge deployment. However, in the dynamometer-card-based fault identification task in the sucker rod pumping system, the original MobileNetV2 still has certain limitations. The fault features of dynamometer cards are usually reflected in morphological differences such as contour deformation, local missing regions, slope changes, and corner variations, and the differences between some similar fault categories are relatively subtle. For example, insufficient fluid supply and gas impact may exhibit similar overall contour deformation, but they differ in local morphological characteristics, such as the position of the unloading point, the presence of missing regions, and the shape of the unloading line. Therefore, effective fault identification requires the model to capture both global contour information and subtle local morphological variations. However, the original MobileNetV2 mainly focuses on lightweight feature extraction through depthwise separable convolution and inverted residual structures, which may limit its ability to fully capture subtle local morphological differences and location-sensitive fault features in dynamometer cards. The ReLU6 activation function truncates negative half-axis responses, which may lead to the loss of some weak feature information. In addition, the original network mainly adopts fixed-scale convolution kernels within each layer for feature extraction, which may limit its ability to simultaneously represent global contour variations and local detail differences in dynamometer cards.

To address these problems, this paper focuses on the morphological fault features of dynamometer cards and the requirements for lightweight identification. Targeted improvements are made to the feature extraction structure, attention representation method, and activation function of MobileNetV2. A lightweight fault identification method for the sucker rod pumping system based on transfer learning and an improved MobileNetV2 is proposed. On the basis of maintaining a small parameter scale, the proposed method improves the fault identification performance on the current dataset. The overall structure is shown in Figure 8.

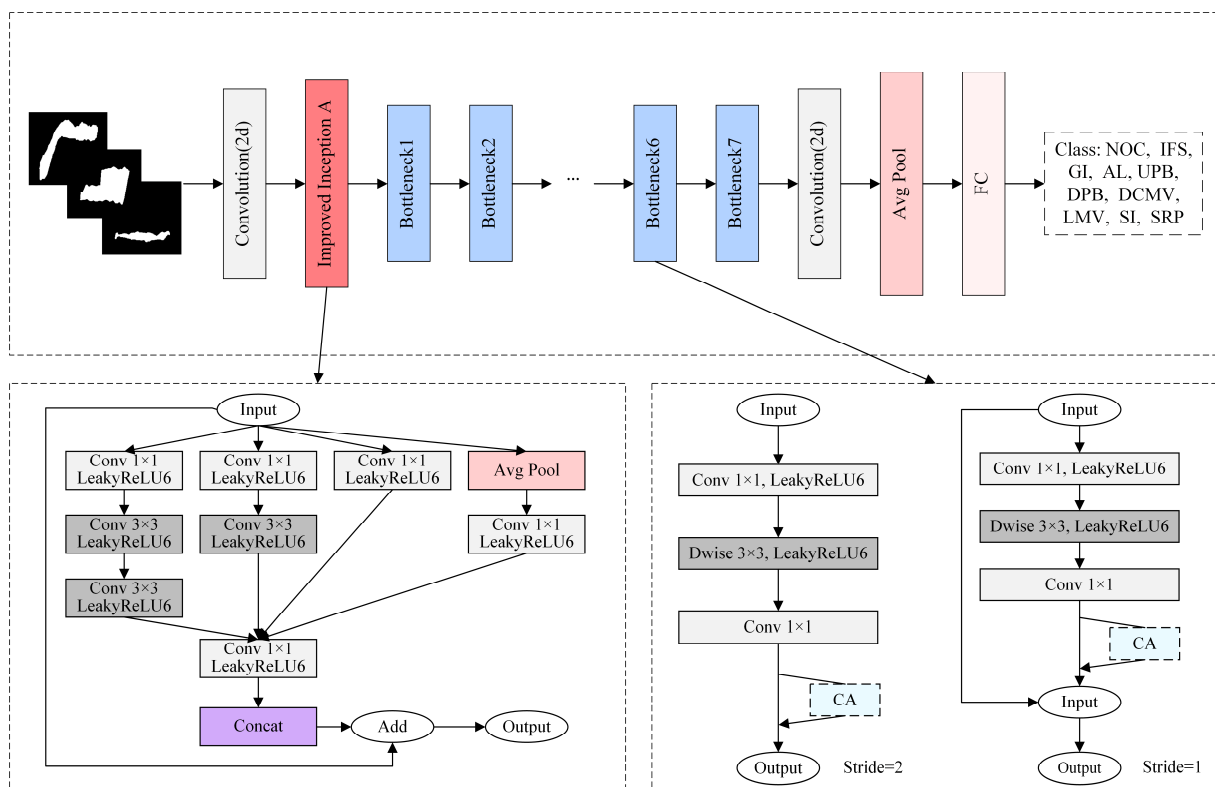


Figure 8. Overall structure of the model.

3.2.1. Attention Mechanism

CA is embedded into all inverted residual modules of MobileNetV2. The inverted residual module is the main feature extraction unit of MobileNetV2. After depthwise convolution and pointwise convolution, its output feature maps already contain certain spatial structural information and channel information. Introducing CA at this stage can use the coordinate information in the horizontal and vertical directions to reweight the extracted features. For the dynamometer-card-based fault identification task, fault information is mainly reflected in local morphological variations, such as local missing regions, corner deformation, and slope changes, and these local variations are closely related to their spatial positions in the load–displacement plane. Therefore, embedding CA into the inverted residual modules helps enhance the model’s representation capability for spatial positional information and local fault-sensitive regions. In addition, the introduction of CA only adds a small number of extra parameters, which is consistent with the lightweight design requirement of MobileNetV2.

3.2.2. Activation Function

The original ReLU6 activation function of the model is replaced with the LeakyReLU6 activation function. Based on ReLU6, this activation function introduces a non-zero slope

in the negative region. It retains the upper output clipping property of ReLU6 while preserving negative outputs. This design allows negative responses to continue participating in feature propagation and parameter updating. For the dynamometer-card-based fault identification task, the differences between some similar fault categories are mainly reflected in local contour variations, slope changes, or weak edge details. Preserving part of the weak negative responses helps the model represent these subtle differences on the current dataset.

3.2.3. Multi-Scale Feature Extraction

To address the lack of efficient multi-scale feature extraction in MobileNetV2, this paper introduces the Inception A module into the original model. While retaining the original Inception A's four-branch parallel structure and multi-scale feature extraction capabilities, we improve the Inception A module as follows: the ReLU activation function is replaced with LeakyReLU6 to maintain consistency with the activation function of the backbone network and preserve part of the negative half-axis responses. After the features from different branches are concatenated, a 1×1 convolution is used for feature fusion and channel adjustment, so that the output dimension can match the subsequent network structure. A residual connection is introduced to add the fused multi-scale features and the module input element-wise, thereby reducing information loss during feature transmission and improving the training stability of the model. The improved Inception A module is added after the first convolutional layer of the model and before the backbone network to enhance the multi-scale feature extraction capability of the model.

The selection of this position is mainly based on the morphological characteristics of dynamometer cards. The fault information in dynamometer cards is usually reflected in contour deformation, local missing regions, slope changes, and corner variations. These features are more obvious in shallow feature maps before multiple downsampling operations. Therefore, extracting multi-scale features at an early stage can provide richer contour and local detail information for the subsequent backbone network. At the same time, these early-extracted multi-scale features also provide a relatively sufficient feature basis for the CA modules in the backbone network to weight spatial positional information, enabling the model to further focus on local regions related to fault categories based on multi-scale morphological features.

4. Experimental Validation

4.1. Data Processing

The dataset used in this study consists of two parts. The first part comes from the competition dataset of the CNPC Second Dream Cloud Innovation Competition, which contains eight operating condition categories shown in Figure 1, including NOC, IFS, GI, AL, UPB, DPB, DCMV, and LMV. The competition dataset is published under a CC0 license. The dataset can be accessed at <https://aistudio.baidu.com/dataset/detail/57195/intro> (accessed on 24 October 2025). This competition was jointly organized by CNPC Beijing Richfit Information Technology Co., Ltd. (Beijing, China) and Baidu Online Network Technology Co., Ltd. (Beijing, China). Since the competition dataset of the CNPC Second Dream Cloud Innovation Competition does not include the SI and SRP operating condition data shown in Figure 1, the second part comes from measured dynamometer card data from a certain block of YanChang Oilfield Co., Ltd. (Yan'an, China), including the two operating conditions of SI and SRP. The collection period was from 2023 to 2025. The final dataset covers the ten typical operating condition types shown in Figure 1, and each sample corresponds to the displacement and load data within a complete stroke cycle.

Since the value ranges of displacement and load differ among different oil wells, the displacement and load data were first normalized to the interval [0, 1]. To solve the problem of the limited number of samples in some categories of the dataset, data augmentation techniques were used to expand the dataset before dividing it into the training set and test set, including random perturbation, curve translation, and scaling transformation. For operating condition categories with a relatively large number of original samples, such as IFS and GI, data augmentation was mainly used to increase the morphological diversity of samples within the same category. For the other operating condition categories with relatively small numbers of original samples, namely all operating conditions except IFS and GI, data augmentation was used to supplement the total number of samples. On the premise of not destroying the original closed-curve structure and main fault morphological features of the dynamometer cards, the above operations can improve the adaptability of the model to load fluctuations, acquisition errors, curve position offsets, and other situations. When dividing the augmented dataset into the training set and test set, the data were divided according to different oil wells to ensure that the training set and test set did not contain data from the same oil well. That is, data from a single oil well would not appear in both the training set and the test set, thereby reducing the risk of data leakage. The number of samples, the number of oil wells, and the distributions of the training and testing sets for each category are shown in Table 1.

Table 1. Composition and distribution of the constructed dynamometer card dataset.

Class	Original Dataset	Number of Oil Wells	Augmented Dataset	Final Dataset	Training Set	Testing Set
NOC	1495	24	2990	1300	1040	260
IFS	7555	10	7555	1300	1040	260
GI	4471	13	4471	1300	1040	260
AL	150	7	1256	1256	1005	251
UPB	711	6	1422	1300	1040	260
DPB	681	8	2043	1300	1040	260
DCMV	553	5	1659	1300	1040	260
LMV	297	8	1782	1300	1040	260
SI	364	19	1820	1300	1040	260
SRP	226	10	1356	1300	1040	260
Total	16,503	110	26,354	12,956	10,365	2591

As shown in Table 1, there are certain differences in the number of original samples among different categories. The original sample numbers of the IFS and GI categories have reached a certain scale, so data augmentation is mainly used to increase their intra-class morphological diversity, and 1300 samples are randomly selected from each category as the final sample number. The original sample numbers of the NOC, AL, UPB, DPB, DCMV, LMV, SI, and SRP categories are relatively small, so data augmentation is mainly used to supplement their sample numbers. This processing helps reduce the influence of category sample number differences on model training and avoids the model being overly biased toward categories with larger sample sizes. Among them, 1256 dynamometer card samples are finally retained for the AL category after augmentation, while 1300 dynamometer card samples are retained for each of the other categories. The constructed final dataset contains 12,956 dynamometer card samples, including 10,365 samples in the training set and 2591 samples in the test set.

In addition, to further verify the applicability of the model under different data sources, this paper constructed a cross-source external validation set using measured dynamometer cards from YanChang Oilfield Co., Ltd. (Yan'an, China). During the actual data collection process, the three operating conditions of UPB, DPB, and DCMV did not occur. Meanwhile, the two operating conditions of SI and SRP had already been included in the dataset shown in Table 1. Therefore, the cross-source external validation set only includes five operating

condition categories consistent with those in the competition dataset, namely NOC, IFS, GI, AL, and LMV. The collection period was from 2023 to 2025. This external validation set did not participate in model training, training set and test set partitioning, data augmentation, or hyperparameter tuning, and was only used to evaluate the identification performance of the trained model under different data sources. During the experiment, the model still performed identification according to the 10 operating condition categories. The number of samples and the number of oil wells for each category of the cross-source external validation set are shown in Table 2.

Table 2. Composition of the cross-source external validation set.

Class	NOC	IFS	GI	AL	LMV	Total
Quantity	142	316	39	107	97	701
Number of oil wells	94	80	32	68	43	317

Dynamometer cards are two-dimensional images composed of closed curves, with various faults represented by changes in the shape of these closed curves. However, such shape differences often manifest as abrupt changes in local lines or variations in slope, resulting in low visual distinguishability from normal regions. This makes it difficult for conventional lightweight models to effectively capture such poorly distinguishable fault information. Therefore, this paper applies a series of operations to the dynamometer cards, including grayscale, binarization, edge detection, and filling. While ensuring that the shape of the dynamometer cards remains unchanged, these operations amplify the shape feature differences corresponding to various faults and enhance the visual distinction between faulted and normal regions, thereby providing high-quality image inputs for subsequent feature extraction and fault identification by lightweight networks. The dynamometer card images before and after processing are shown in Figure 9.



Figure 9. Comparison of dynamometer cards before and after processing: (a) Unprocessed dynamometer cards; (b) Processed dynamometer cards.

4.2. Experimental Environment and Parameters

The Python version used in the experiment is 3.9, the deep learning framework is PyTorch, and its version is 2.5.1+cu118, the CUDA version is 11.8, the CPU model is Intel Core i9-13900H, the GPU model is NVIDIA GeForce RTX 4060, and the operating system is Windows 11.

During model training, the AdamW optimizer was used together with a cross-entropy loss function. The number of training epochs was set to 50, the batch size was set to 32, and the initial learning rate was set to 0.0001. A cosine annealing learning rate scheduling strategy was also adopted to improve the stability of the training process. The newly added modules were initialized using the default initialization strategy of the model. Transfer learning was performed using the official MobileNetV2 weights pre-trained on the ImageNet-1K

dataset. The input dynamometer card images were resized to 224×224 pixels and normalized using the ImageNet mean and standard deviation. During training, the feature extraction layers were frozen. To improve the stability of the experimental results, each model was trained three times using different random seeds under the same training strategy. The overall accuracy, precision, recall, F1-score, and specificity were reported as the mean values of the three repeated experiments, together with their corresponding standard deviations. Meanwhile, the confusion matrices corresponding to the three repeated experiments were also presented. However, to facilitate a detailed analysis of the misidentification relationships among specific categories from the perspective of the actual numbers of misclassified samples, this paper selected the experimental result whose accuracy was closest to the mean accuracy as the representative result for analysis, while the confusion matrices of the other two experiments were used to further evaluate the consistency of category confusion under different random seeds. To ensure the fairness of the experimental comparison, all models adopted the same hyperparameter settings and pre-training strategy.

4.3. Fault Identification Results of the Sucker Rod Pumping System

4.3.1. Overall Identification Accuracy and Corresponding Evaluation Metrics

Since the proposed model performs ten-class fault identification, the category-level evaluation metrics were calculated using the one-versus-rest (OvR) strategy. Each operating condition was regarded as an independent positive class, while the remaining nine operating conditions were regarded as negative classes. Based on this strategy, precision, recall, F1-score, and specificity were calculated separately for each category. For overall model comparison, recall and F1-score were calculated using the macro-average strategy. Overall identification accuracy reflects the proportion of correctly identified samples in all samples and is the most intuitive performance metric of the model. Precision reflects the proportion of samples that truly belong to a certain fault among the samples predicted as this fault and is a key metric for measuring model misjudgment. Recall reflects the proportion of actual samples of a certain fault category that are correctly identified by the model and is an important metric for measuring missed detections. The F1-score is the harmonic mean of precision and recall and can comprehensively evaluate the model's balance between false alarms and missed detections. Specificity reflects the model's ability to correctly identify non-target samples as non-target categories and is an important metric for evaluating the model's ability to exclude negative samples. The formulas are shown in Equations (9)–(13).

$$Accuracy = \frac{N_{correct}}{N} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

$$Specificity = \frac{TN}{TN + FP} \quad (13)$$

Here, $N_{correct}$ refers to the number of samples that are correctly classified by the model; N refers to the total number of test samples; TP refers to true positives, i.e., the number of samples that actually have the target fault and are correctly identified by the model; TN refers to true negatives, i.e., the number of samples that do not belong to the target category and are not predicted as the target category by the model; FP refers to false positives,

i.e., the number of samples that actually have other faults but are incorrectly identified as having the target fault by the model; *FN* refers to false negatives, i.e., the number of samples that actually have the target fault but are incorrectly identified as having other faults by the model.

The training results of the improved MobileNetV2 model are shown in Figure 10, the overall mean identification accuracy and standard deviation are shown in Table 3, and the identification performance metrics for each category, including precision, recall, F1-score, and specificity, are shown in Table 4.

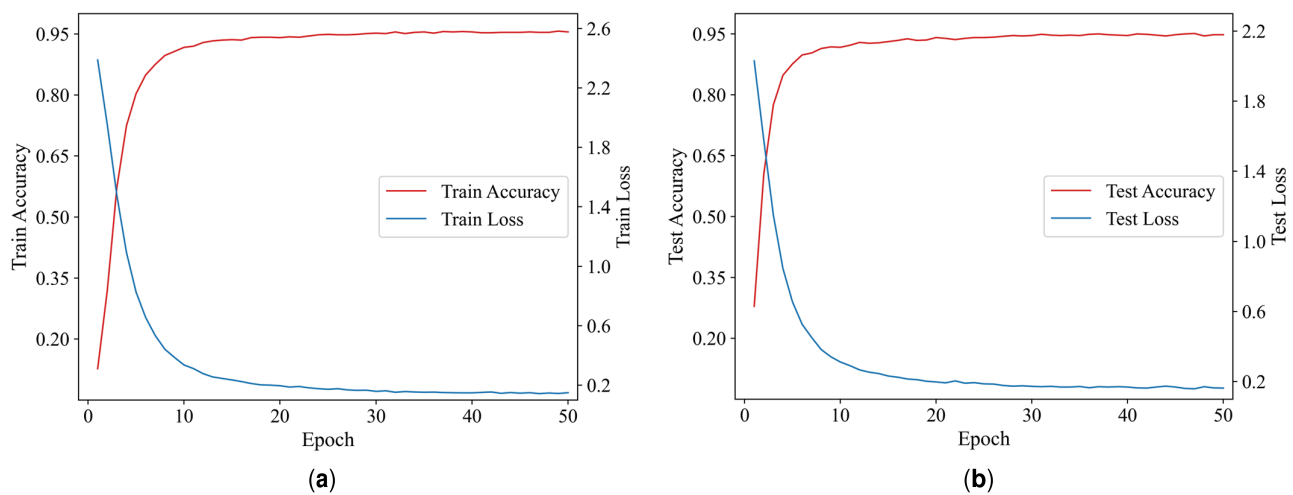


Figure 10. Training and testing accuracy and loss curves of the improved MobileNetV2 model: (a) Training process; (b) Testing process.

Table 3. Overall accuracy and stability of the improved MobileNetV2 model.

Model	Mean Accuracy (%)	Accuracy Std (%)
Improved MobileNetV2	95.01	0.09

Table 4. Identification performance of the improved MobileNetV2 model.

Class	Mean Precision (%)	Precision Std (%)	Mean Recall (%)	Recall Std (%)	Mean F1-Score (%)	F1-Score Std (%)	Mean Specificity (%)	Specificity Std (%)
NOC	83.98	0.50	95.38	2.34	89.31	0.76	97.97	0.12
IFS	91.85	1.00	90.90	2.32	91.36	1.16	99.10	0.13
GI	92.43	0.21	89.36	2.73	90.86	1.51	99.18	0.00
AL	97.04	1.07	99.73	0.46	98.37	0.48	99.67	0.12
UPB	98.81	0.80	95.90	1.46	97.33	1.09	99.87	0.09
DPB	93.85	1.02	92.18	2.93	93.00	1.98	99.33	0.10
DCMV	97.80	0.92	95.77	3.36	96.74	1.37	99.76	0.11
LMV	99.58	0.41	91.03	1.55	95.11	0.75	99.96	0.05
SI	96.67	1.80	100.00	0.00	98.31	0.93	99.61	0.22
SRP	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00

As shown in Figure 10, as the number of training epochs increased, the model's accuracy during both training and testing gradually rose, while the loss value continued to decrease and converge. This indicates that the model was able to learn fault-related features from dynamometer cards and perform fault identification for the sucker rod pumping system. After 50 epochs of iterative training, the improved MobileNetV2 model achieved a mean fault identification accuracy of 95.01% on the current dataset, suggesting that the model obtained relatively good identification performance under the current experimental setting.

As shown in Table 3, the improved MobileNetV2 model achieved a mean accuracy of 95.01% on the current dataset, with a standard deviation of 0.09%. This indicates that the overall identification results of the model fluctuate only slightly under different random seeds and show a certain degree of stability. From the mean values and standard deviations of precision, recall, F1-score, and specificity across three repeated experiments with different random seeds shown in Table 4, it can be seen that categories such as SRP, SI, AL, UPB, and DCMV all achieved relatively high mean identification performance. Among them, the mean precision, recall, and F1-score of SRP all reached 100.00%, with standard deviations of 0.00%. The mean recall of SI reached 100.00%, with a standard deviation of 0.00%, while its mean F1-score was 98.31%, with a standard deviation of 0.93%. The mean F1-scores of AL and UPB reached 98.37% and 97.33%, with standard deviations of 0.48% and 1.09%, respectively, indicating that the model has relatively good identification capability for these operating conditions with relatively obvious morphological features.

In addition, the identification performance of categories such as NOC, IFS, GI, and DPB still has certain room for improvement. The mean precision of NOC was only 83.98%, with a standard deviation of 0.50%, which was obviously lower than that of the other categories, while its mean recall reached 95.38%, with a standard deviation of 2.34%. This indicates that most normal operating condition samples could be correctly identified, but some samples from other fault categories were still misidentified as the normal operating condition, thereby reducing the precision of NOC. The mean recall of GI was 89.36%, with a standard deviation of 2.73%, which was relatively low among all categories, indicating that some GI samples were still prone to being misidentified as other operating conditions with similar morphological features. The mean recall of IFS was 90.90%, with a standard deviation of 2.32%, while that of DPB was 92.18%, with a standard deviation of 2.93%, also indicating that these categories still exhibited a certain degree of missed identification. Although LMV achieved a relatively high mean precision of 99.58%, with a standard deviation of 0.41%, its mean recall was 91.03%, with a standard deviation of 1.55%, indicating that samples predicted as LMV had relatively high reliability, but some actual LMV samples were still identified as other categories.

From the fluctuations observed in the repeated experiments, the standard deviations of precision and specificity for each category were generally small, whereas the standard deviations of recall were relatively large for some categories. For example, the standard deviations of recall for DCMV, DPB, and GI were 3.36%, 2.93%, and 2.73%, respectively, indicating that different random initializations mainly had a certain influence on the missed identification of some similar operating conditions. In contrast, the standard deviations of F1-score for all categories did not exceed 1.98%, indicating that the model generally maintained relatively stable performance in terms of the combined performance of precision and recall. The mean specificity of each category remained at a relatively high level. Among them, NOC had the lowest mean specificity of 97.97%, with a standard deviation of 0.12%, while the mean specificity values of the other categories were all higher than 99%, indicating that the model generally had a relatively good ability to exclude non-target categories.

From the above analysis results, it can be seen that the improved MobileNetV2 maintained relatively high overall identification performance and relatively small result fluctuations across three repeated experiments with different random seeds under the current dataset partition setting, indicating that the proposed method has certain application value. However, further optimization is still needed for similar fault categories with small morphological differences.

4.3.2. Confusion Matrix

As shown in Tables 3 and 4, the improved MobileNetV2 achieved relatively high overall identification accuracy across three repeated experiments with different random seeds, but certain differences in identification performance still existed among different operating condition categories. Among them, some precision, recall, or F1-score values for categories such as NOC, GI, and LMV were relatively low. To further analyze the specific sources of these performance differences among categories, the confusion matrices from the three repeated experiments were analyzed to observe the consistency of category confusion under different random seeds.

The confusion matrix is a crucial tool in machine learning for evaluating a model's identification performance. It provides an intuitive understanding of the model's performance across different categories, including the number of correctly and incorrectly identified samples. The horizontal axis represents the model's predicted class, the vertical axis represents the true class, and the values along the diagonal represent the number of correctly identified samples. The confusion matrices from three repeated experiments with the improved MobileNetV2 are shown in Figure 11.

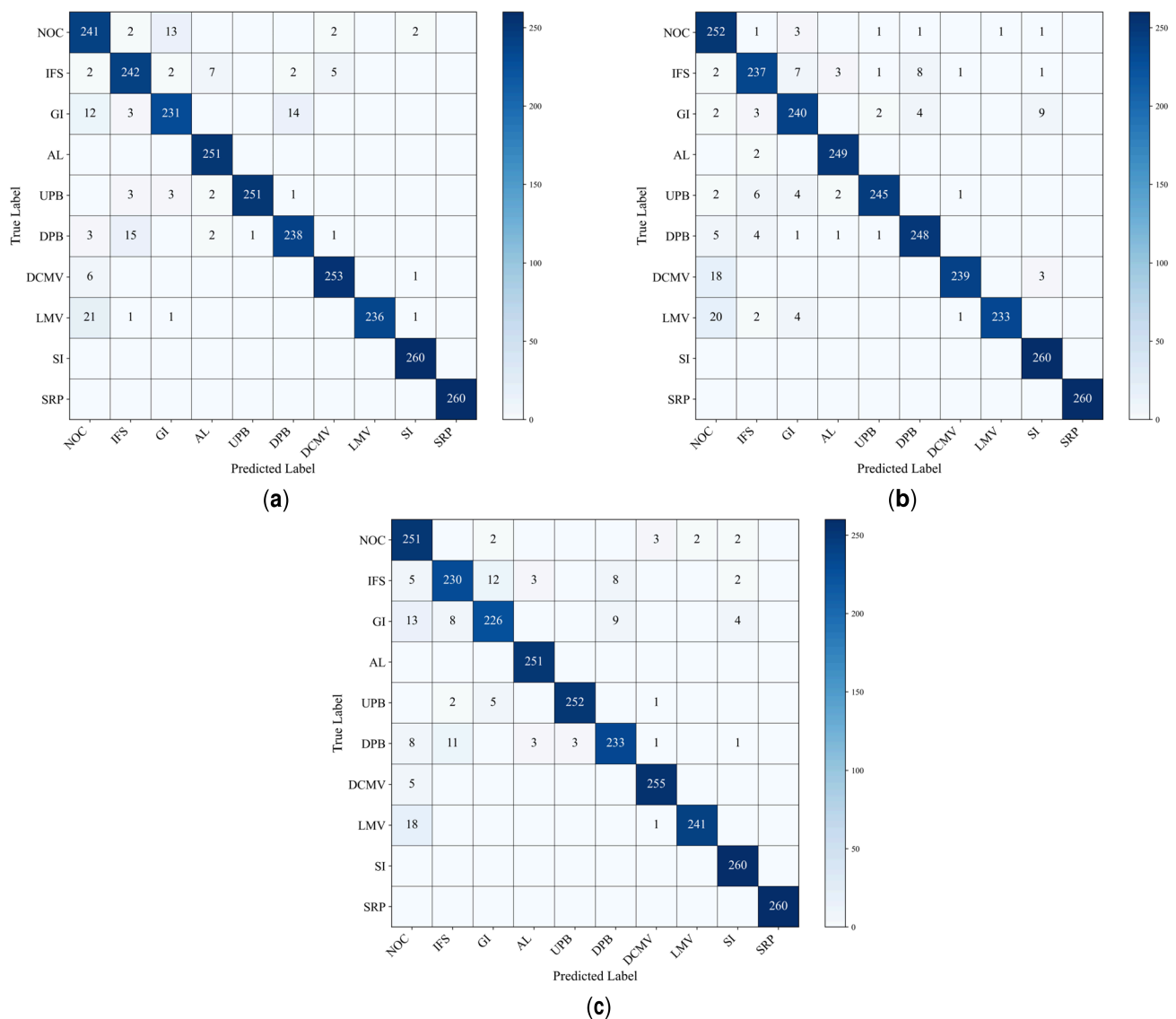


Figure 11. Confusion matrices of the improved MobileNetV2 across three repeated experiments: (a) the first experiment; (b) the second experiment; (c) the third experiment.

Figure 11 shows the confusion matrices obtained from three repeated experiments of the improved MobileNetV2 under different random seeds, with Figure 11a–c corresponding to the results of the first, second, and third experiments, respectively. To facilitate the analysis of the misidentification relationships among specific categories in combination with the actual numbers of misclassified samples, this paper mainly discusses the results in Figure 11a, while Figure 11b,c are presented as the results of the other two repeated experiments to observe the consistency of category confusion of the model under different random seeds.

As shown in the confusion matrix in Figure 11a, most samples are distributed along the main diagonal, indicating that the improved MobileNetV2 model shows relatively good identification performance for most fault categories. In the first experiment, the SI, AL, and SRP categories are all correctly identified, indicating that the model achieves relatively good identification performance for these fault categories in this run. However, some confusion still exists among several categories. Specifically, GI is mainly misidentified as NOC and DPB, DPB is mainly misidentified as IFS, and a relatively large number of LMV samples are misidentified as NOC. These major misidentification relationships are generally consistent with the class-specific results averaged over the three repeated experiments in Table 4, where relatively lower precision, recall, or F1-score values are observed for categories such as NOC, GI, and LMV.

As shown in Figure 11a, some LMV and GI fault samples are identified as NOC. This is because faults such as LMV and GI do not completely destroy the overall closed-curve shape of the dynamometer cards at the early stage or when the fault severity is relatively low. Their overall contours are still close to those of normal operating condition dynamometer cards, mainly showing slight local slope changes, local load differences, or local curve fluctuations. The morphological differences are relatively weak, causing the model to easily misidentify such slight fault samples as NOC.

The confusion among GI, DPB, and IFS observed in Figure 11a is mainly related to their similarity in local morphology. When the degree of gas influence is relatively strong, local corner deformation may appear in the dynamometer card. Such changes have certain similarities with the load mutation or sharp-corner features at the end of the downstroke in downstroke pump bumping dynamometer cards. Therefore, some GI samples are easily misidentified as DPB. Meanwhile, typical downstroke pump bumping dynamometer cards usually show sharp corners or load mutations at the end of the downstroke. However, in the dataset used in this study, some DPB fault samples not only contain local sharp corners but also show atypical morphological features such as lower-right missing regions, collapse, or abnormal unloading segments. These features have certain similarities with the local missing regions and abnormal unloading features of IFS or GI, thereby increasing the difficulty for the model to distinguish the three similar operating conditions of GI, DPB, and IFS. In response to the above phenomena, future research can further introduce oil well production parameters and finer-grained local feature constraints to improve the model's ability to distinguish similar fault categories and complex fault categories.

As shown in Figure 11a–c, there were certain differences in the specific numbers of correctly identified and misidentified samples for each category across the three repeated experiments with different random seeds, indicating that random initialization had a certain influence on the final classification results of some samples, but the overall category confusion did not show large deviations. This is basically consistent with the results reflected by the mean values and standard deviations of precision, recall, F1-score, and specificity for each category in Table 4. Therefore, from the above analysis, it can be seen that under the current dataset partition setting, the category identification results of the model generally maintained relatively good consistency across repeated experiments with

different random seeds, although the specific numbers of misclassified samples for some categories still showed certain fluctuations.

4.3.3. ROC Curve

The receiver operating characteristic (ROC) curve is an important tool for evaluating a model's identification performance. It plots the model's performance at different thresholds by plotting false positive rate (FPR) on the horizontal axis and the true positive rate (TPR) on the vertical axis. The TPR refers to the proportion of samples that the model correctly identifies as positive; a higher value is better, with a maximum of 1. The FPR refers to the proportion of samples that the model incorrectly identifies as positive among all samples that are actually negative; a lower value is better, with a minimum of 0. The calculation formulas are shown in Equations (14) and (15). The ROC curves for each category are shown in Figure 12.

$$TPR = \frac{TP}{TP + FN} \quad (14)$$

$$FPR = \frac{FP}{FP + TN} \quad (15)$$

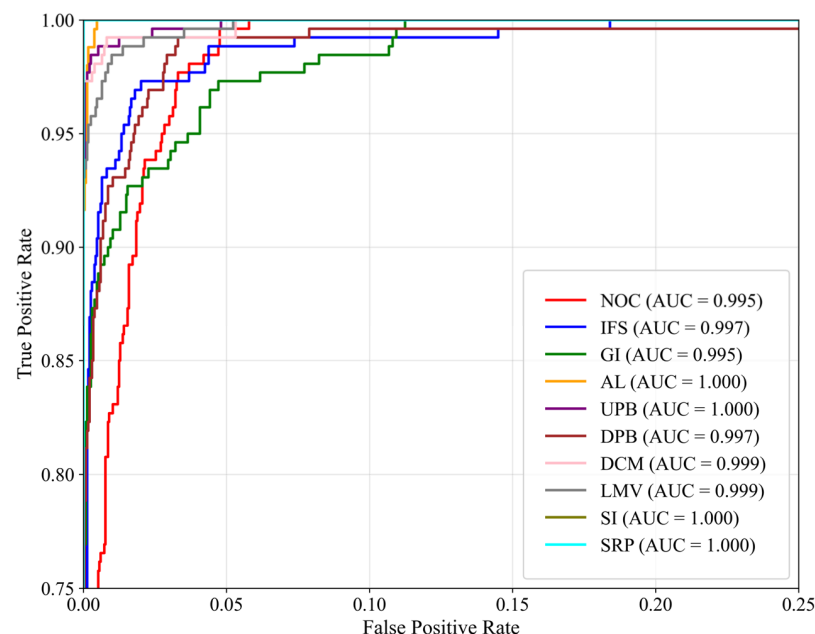


Figure 12. ROC curves for each category.

The area under the curve (AUC) is the area under the ROC curve and is used to measure the model's identification performance. The closer the AUC value is to 1, the stronger the model's identification performance; conversely, the lower the AUC value, the weaker the model's identification performance.

As shown in Figure 12, the ROC curves for all categories lie close to the top-left corner, and the AUC values for all categories are no less than 0.995. Notably, the AUC values for AL, UPB, SI, and SRP all reach 1, indicating that the improved MobileNetV2 model shows relatively strong discrimination ability on the current dataset for various types of fault dynamometer cards and demonstrates high identification accuracy for these four fault types: AL, UPB, SI, and SRP.

4.3.4. P-R Curve

The precision–recall (P-R) curve is a graph with recall on the x-axis and precision on the y-axis. The area under the curve, known as average precision (AP), serves as the

primary evaluation metric. Its values range from 0 to 1. A higher AP value that is closer to 1 means the model can maintain high recall while ensuring high precision. The P-R curves for each category are shown in Figure 13.

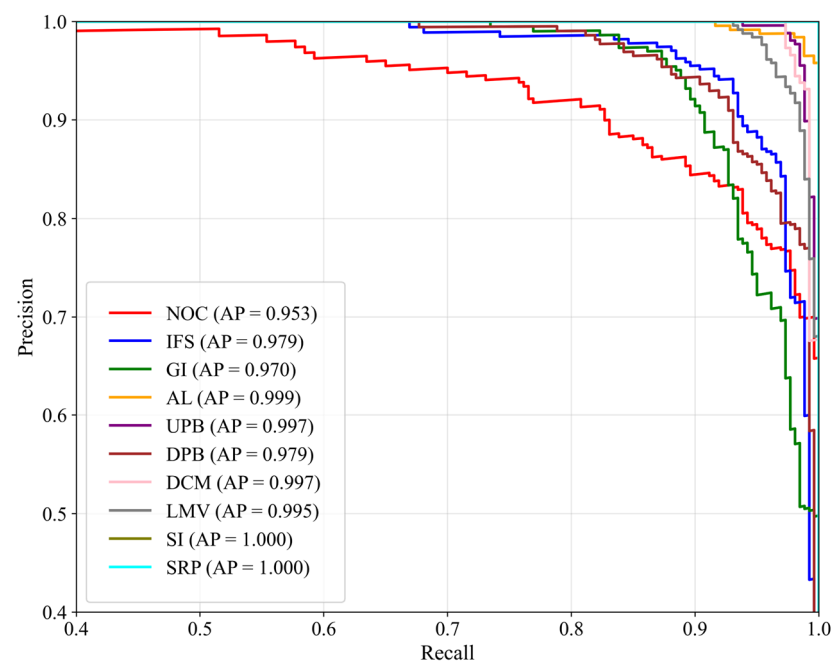


Figure 13. P-R curves for each category.

As shown by the distribution of P-R curves in Figure 13, the curves for all operating condition categories lie closely along the upper-right corner, with AP values exceeding 0.95 for each category. Most categories have AP values close to 1, indicating that the improved MobileNetV2 model achieves relatively good overall identification performance in identifying various fault dynamometer cards, maintaining relatively high precision while achieving high recall.

4.3.5. Interpretability Analysis

The Grad-CAM method [31] was employed to generate category activation heatmaps. The color gradient from blue to red in the heatmaps represents the importance of different regions in the dynamometer cards to the identification results. The red high-response regions indicate image regions that make relatively large contributions to the model's current classification decision, while the blue regions indicate regions that make relatively small contributions to the classification result. Grad-CAM reflects the discriminative regions that the model focuses on in the input image. Figure 14 displays the heatmaps of ten operating conditions generated by the Grad-CAM method.

The Grad-CAM results show that the model tends to focus on different local morphological regions under different operating conditions. For example, UPB mainly focuses on the sharp corner variation region in the upper-right corner, while IFS mainly focuses on the missing region in the lower-right corner. These high-response regions show a certain degree of spatial consistency with the diagnostically meaningful morphological differences in the dynamometer cards corresponding to the respective operating conditions, indicating that the model's classification decisions utilize morphological information related to the operating condition categories to a certain extent. However, this consistency cannot prove that the model directly learned the physical mechanisms of fault formation.

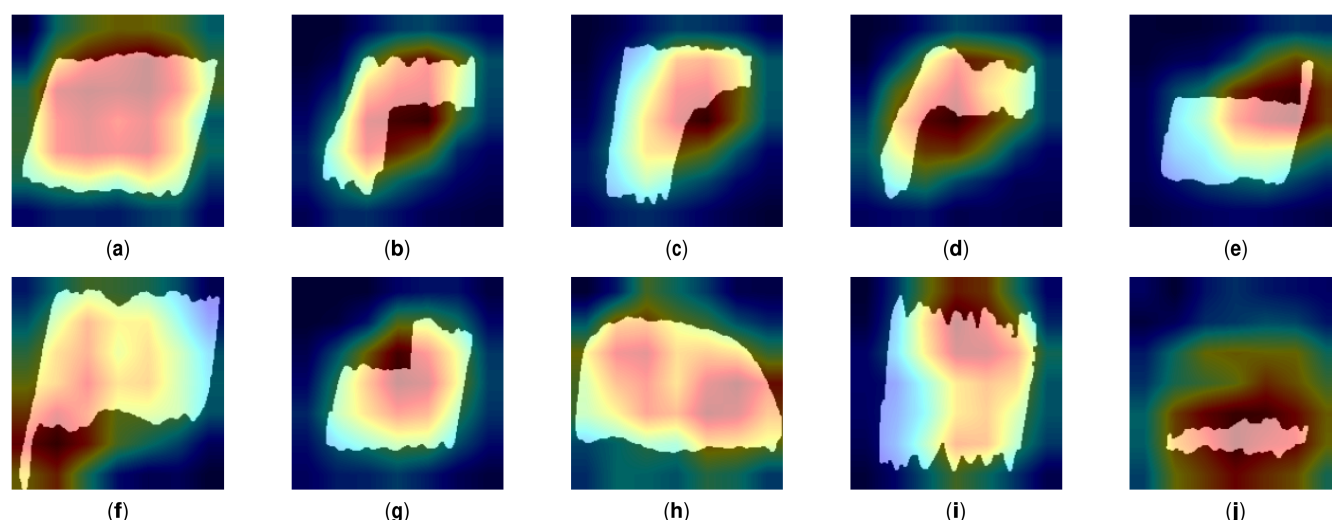


Figure 14. Heatmap analysis results for ten categories: (a) NOC; (b) IFS; (c) GI (d) AL; (e) UPB; (f) DPB; (g) DCMV; (h) LMV; (i) SI; (j) SRP.

4.3.6. The Impact of Different Negative Slope Values in LeakyReLU6 on Identification Performance

The LeakyReLU6 activation function mitigates the vanishing gradient problem of the traditional ReLU6 activation function for negative inputs by introducing a negative slope coefficient α . This coefficient serves as a key factor in controlling the gradient magnitude within the negative range. To determine the selected value of α , three commonly adopted values, including 0.01, 0.1 and 0.2 were selected for comparative experiments against the MobileNetV2 baseline model. Table 5 presents the identification accuracy of the baseline MobileNetV2 model and improved models with LeakyReLU6 under different α values. The corresponding confusion matrices of each model are illustrated in Figure 15, arranged in order as (a), (b), (c) and (d).

As shown in Table 5, compared with the original ReLU6 activation function, replacing it with LeakyReLU6 improves the mean accuracy of the model under different negative slope settings. Specifically, the mean accuracy corresponding to ReLU6 is 91.00%, with a standard deviation of 0.23%. When α is set to 0.01, 0.1, and 0.2, the mean accuracy of the model is 91.51%, 92.51%, and 92.35%, respectively, and the corresponding standard deviations are 0.28%, 0.24%, and 0.29%, respectively. The results show that the model achieves the highest mean accuracy when $\alpha = 0.1$, while its standard deviation is close to that of ReLU6, indicating that this negative slope setting has relatively good identification performance and stability on the current dataset.

Table 5. Accuracy of the model under different activation functions.

Activation Function	Mean Accuracy (%)	Accuracy Std (%)
ReLU6	91.00	0.23
LeakyReLU6 ($\alpha = 0.01$)	91.51	0.28
LeakyReLU6 ($\alpha = 0.1$)	92.51	0.24
LeakyReLU6 ($\alpha = 0.2$)	92.35	0.29

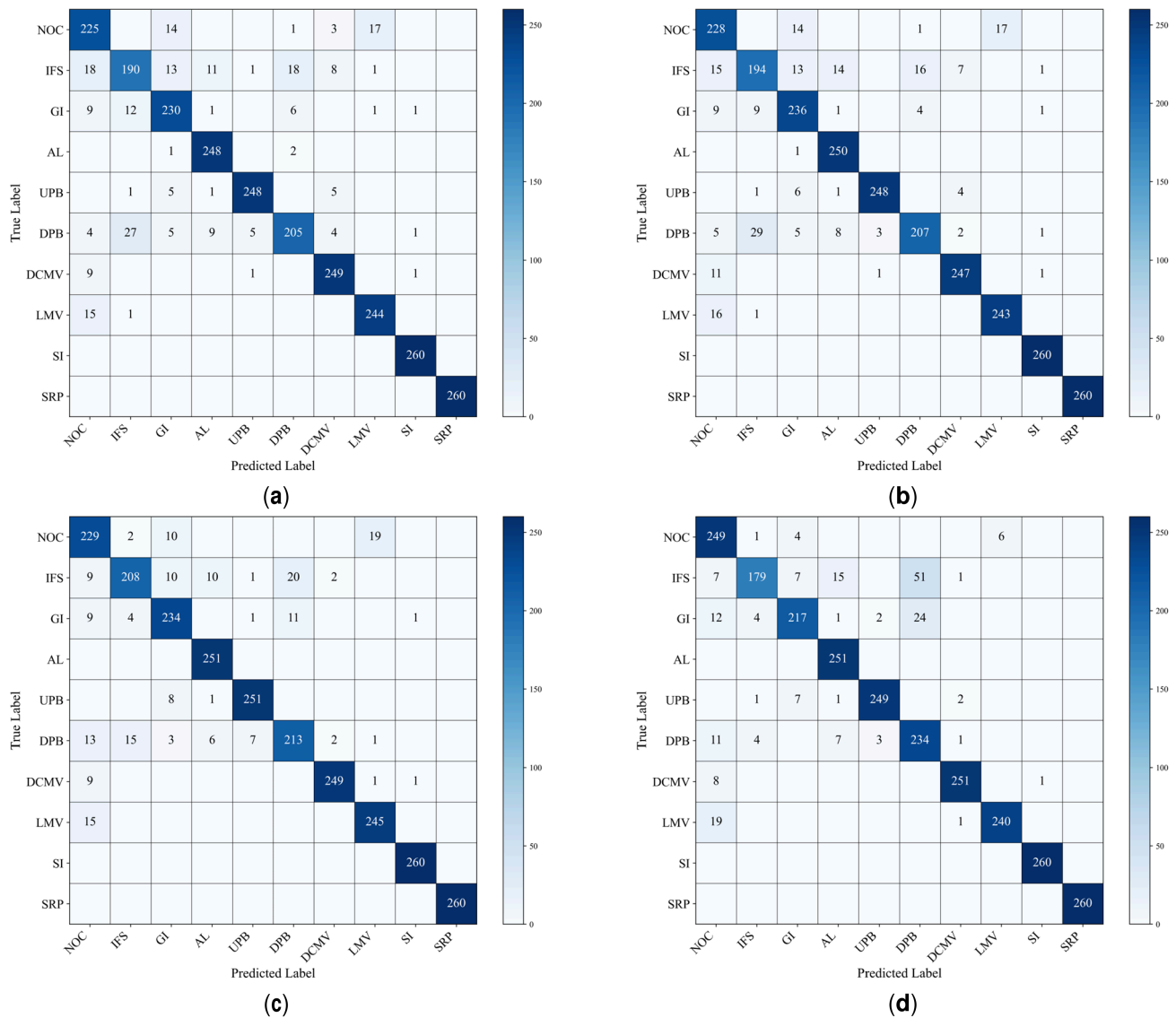


Figure 15. Confusion matrices under different activation functions: (a) ReLU6; (b) LeakyReLU6 ($\alpha = 0.01$); (c) LeakyReLU6 ($\alpha = 0.1$); (d) LeakyReLU6 ($\alpha = 0.2$).

Further combined with the confusion matrices in Figure 15, it can be seen that the identification performance of the model for each category differs under different negative slope settings. ReLU6 directly truncates the feature responses on the negative half-axis to 0, which may suppress some weak fault features during forward propagation. For categories with relatively similar morphologies, such as GI, IFS, DPB, LMV, and NOC, their differences are often reflected in local contour variations, slope changes, or edge details. Excessive truncation of weak responses may affect the model's ability to learn these subtle differences. When $\alpha = 0.01$, the negative slope is small, and the negative half-axis responses are retained to a limited extent. Therefore, the improvement in distinguishing some similar fault categories is relatively limited. When $\alpha = 0.2$, although the model retains more negative half-axis responses, it may also introduce some non-discriminative responses, making the feature boundaries between similar categories less clear and increasing the risk of misidentification. When $\alpha = 0.1$, the model achieves a more appropriate balance between retaining negative half-axis information and suppressing non-discriminative responses. Therefore, $\alpha = 0.1$ is selected as the negative slope parameter of LeakyReLU6 in the subsequent experiments.

4.3.7. Comparative Experiments

To better verify the performance of the improved MobileNetV2 model in the sucker rod pumping system fault identification task, comparative tests were conducted under the same training strategy and dataset. The improved MobileNetV2 model was compared with EfficientNetV2 [32], ResNet34 [33], DenseNet121 [34], MobileNetV3 [35], ShuffleNetV2 [36] and MobileNetV2 [25]. In addition to evaluating the model based on fault identification performance, the lightweight characteristics of the models were also evaluated from three aspects: the number of parameters (Params), the number of multiply–accumulate operations (MACs), and the model weight file size (Model size). The results are shown in Table 6.

Table 6. Comparison results of various models.

Model	Accuracy (%)	Accuracy Std (%)	Recall (%)	Recall Std (%)	F1-Score (%)	F1-Score Std (%)	Params (M)	MACs (M)	Model Size (MB)
EfficientNetV2	91.04	0.62	91.07	0.62	90.96	0.64	20.19	2897.33	77.8
ResNet34	89.80	0.99	89.81	1.00	89.83	0.97	21.29	3678.23	81.3
DenseNet121	90.37	0.30	90.41	0.29	90.35	0.29	6.96	2895.99	27.1
MobileNetV3	86.64	0.40	86.68	0.40	86.70	0.40	1.53	61.18	5.95
ShuffleNetV2	81.36	0.29	81.34	0.29	80.92	0.40	1.26	151.70	4.98
MobileNetV2	91.00	0.23	91.02	0.23	90.09	0.26	2.24	326.28	8.76
Improved MobileNetV2	95.01	0.09	95.02	0.09	95.04	0.10	2.28	536.72	9.01

In Table 6, Recall denotes Macro Recall, and F1-score denotes Macro F1-score. Accuracy, Macro Recall, and Macro F1-score are reported as the mean values of three repeated experiments using different random seeds. Accuracy std, Recall std, and F1-score std denote the standard deviations of Accuracy, Macro Recall, and Macro F1-score across the three repeated experiments, respectively. Params denotes the total number of model parameters, MACs denotes the number of multiply–accumulate operations under an input size of 224×224 , and Model size denotes the size of the model weight file.

As shown in Table 6, the improved MobileNetV2 model achieves a mean accuracy of 95.01%, with Macro Recall and Macro F1-score of 95.02% and 95.04%, respectively, which are higher than those of the other comparison models. In addition, the standard deviations of Accuracy, Macro Recall, and Macro F1-score are 0.09%, 0.09%, and 0.10%, respectively, which are lower than those of EfficientNetV2, ResNet34, DenseNet121, MobileNetV3, ShuffleNetV2, and the original MobileNetV2. This indicates that the proposed model not only achieves relatively good overall identification performance on the current dataset but also shows smaller performance fluctuations under different random seeds.

From the perspective of model complexity, the improved MobileNetV2 retains the depthwise separable convolution and inverted residual structures of the original MobileNetV2 as the lightweight backbone network, while introducing additional modules to improve fault identification capability. Finally, the proposed model contains 2.28 million parameters and has a model size of 9.01 MB. Compared with the original MobileNetV2, which contains 2.24 million parameters and has a model size of 8.76 MB, only slight increases are observed, indicating that the proposed model still maintains relatively good lightweight characteristics. Compared with EfficientNetV2, ResNet34, and DenseNet121, the improved MobileNetV2 achieves higher Accuracy, Macro Recall, and Macro F1-score while maintaining a substantially smaller model scale and lower computational complexity. Compared with MobileNetV3 and ShuffleNetV2, the improved MobileNetV2 introduces slightly higher computational cost but achieves higher Accuracy, Macro Recall, and Macro F1-score.

The performance differences among different models suggest that the advantages of the improved model are closely related to the feature extraction strategy designed according to the morphological characteristics of dynamometer cards. The fault information in

dynamometer cards is mainly reflected in overall curve contour variations and subtle morphological differences, such as local missing regions, sharp corner deformations, and slope variations. However, some fault categories have similar overall morphologies, and their subtle differences may not be sufficiently captured by fixed-scale feature extraction alone. Therefore, this paper enhances the extraction capability of multi-scale contour information and local detail information through the improved Inception A module, strengthens the representation of fault-related spatial regions through the coordinate attention mechanism, and reduces the loss of weak response information using LeakyReLU6, enabling the model to more sufficiently learn the morphological differences among different fault categories. Therefore, the improved MobileNetV2 achieves better identification performance while retaining a compact parameter scale and model size on the current dataset.

4.3.8. Ablation Studies

To verify the impact of the proposed improvements on the identification performance of the base MobileNetV2 model, ablation studies were conducted. Under the same training strategy and dataset, the CA module, the LeakyReLU6 activation function, and the multi-scale feature extraction module were progressively introduced and combined in various pairs to construct multiple control models for performance comparison. In addition to Params, MACs, and model size, which are used to evaluate the lightweight characteristics of the model, inference latency (Latency) is further introduced to assess the practical inference efficiency of different model variants. The results are shown in Table 7.

Table 7. Ablation results and computational costs of different improvements.

Model	Improvements			Accuracy (%)	Std (%)	Params (M)	Size (MB)	MACs (M)	Latency (ms)
	CA	LR6	InA						
MobileNetV2				91.00	0.23	2.24	8.76	326.28	17.80
	✓			91.65	0.97	2.24	8.97	327.99	23.29
		✓		92.51	0.24	2.24	8.76	326.28	19.15
			✓	92.17	0.24	2.24	8.80	535.02	21.61
	✓	✓		93.05	0.68	2.28	8.97	327.99	25.73
		✓	✓	93.27	0.53	2.24	8.80	535.02	23.35
	✓		✓	94.03	0.08	2.28	9.01	536.72	28.82
	✓	✓	✓	95.01	0.09	2.28	9.01	536.72	29.62

In Table 7, LR6 denotes the LeakyReLU6 activation function, and InA denotes the improved Inception A module. Accuracy is the mean value obtained from three repeated experiments using different random seeds, std denotes the standard deviation of accuracy, size denotes the storage size of the model weight file, and latency denotes the inference latency under CPU-only inference.

As shown in Table 7, under the same training strategy and dataset conditions, different improvement components have different effects on the identification performance and computational cost of the model. The original MobileNetV2 achieves a mean accuracy of 91.00%, with a standard deviation of 0.23%, 2.24 M parameters, 326.28 M MACs, and a CPU inference latency of 17.80 ms.

After introducing CA, LeakyReLU6, and the improved Inception A alone, the mean accuracy of the model increases to 91.65%, 92.51%, and 92.17%, respectively, indicating that all three improvements can improve the identification performance of the model to different degrees. Among them, when CA is used alone, the accuracy improvement is limited, the standard deviation is 0.97%, and the MACs and CPU inference latency increase to some extent. This indicates that its results fluctuate relatively more under different random seeds, and the attention computation has a certain impact on inference efficiency.

LeakyReLU6 changes the feature response pattern by introducing a non-zero slope in the negative region, thereby preserving part of the negative responses. The additional computational cost introduced by this operation is relatively small compared with that of the entire network. At the current level of statistical precision, the number of parameters, model size, and MACs remain essentially unchanged, while the CPU inference latency for a single image increases only from 17.80 ms to 19.15 ms, corresponding to an increase of 1.35 ms. Therefore, considering the trade-off between identification performance and inference efficiency, these results suggest that the performance improvement achieved by LeakyReLU6 can reasonably compensate for the additional computational overhead it introduces. The improved Inception A has a small impact on the number of parameters, but because it is located at the shallow feature stage, MACs increase significantly to 535.02 M.

From the combination results, the mean accuracies of CA + LeakyReLU6, LeakyReLU6 + improved Inception A, and CA + improved Inception A reach 93.05%, 93.27%, and 94.03%, respectively, all of which are higher than those of the corresponding single-module models. This indicates that the three improvements do not merely work independently, but provide complementary effects from different aspects of feature representation. Among them, the combination of CA and the improved Inception A achieves the highest accuracy among the two-module combinations, with a standard deviation of 0.08%. This result suggests that the improved Inception A module and CA have a relatively clear complementary relationship. The improved Inception A module is placed in the shallow stage of the network and is mainly used to extract multi-scale contour and local detail features before repeated downsampling operations. The coordinate attention (CA) mechanism reweights fault-sensitive regions by introducing spatial positional information, but its effectiveness depends on the richness of the input features. Therefore, the shallow multi-scale features extracted by the improved Inception A module can provide a more complete feature basis for the subsequent spatial positional weighting of CA, while CA further highlights local regions related to fault morphology. This complementary mechanism enables the model to better represent subtle morphological differences in dynamometer cards, such as local missing regions and corner deformation.

After simultaneously introducing CA, LeakyReLU6, and the improved Inception A, the mean accuracy of the complete model reaches 95.01%, with a standard deviation of 0.09%, which is the highest among all ablation combinations. Compared with MobileNetV2, the mean accuracy of the complete model increases by 4.01%, the number of parameters increases from 2.24 M to 2.28 M, the model size increases from 8.76 MB to 9.01 MB, the MACs increase to 536.72 M, and the CPU inference latency increases to 29.62 ms. This indicates that the improved structure also introduces a certain computational cost while improving identification performance. The experimental results show that these three optimization methods improve the model's representation capability for dynamometer-card fault features from different dimensions. This performance improvement is accompanied by increased computational cost, while the parameter count and model size remain relatively compact compared with the baseline MobileNetV2.

4.3.9. Cross-Source External Validation

To further evaluate the applicability of the improved MobileNetV2 model to dynamometer card data from different sources, the model was tested using the cross-source external validation set, and the data in this validation set were not processed by data augmentation. The confusion matrix of the validation results is shown in Figure 16, and the corresponding evaluation metrics are shown in Table 8.

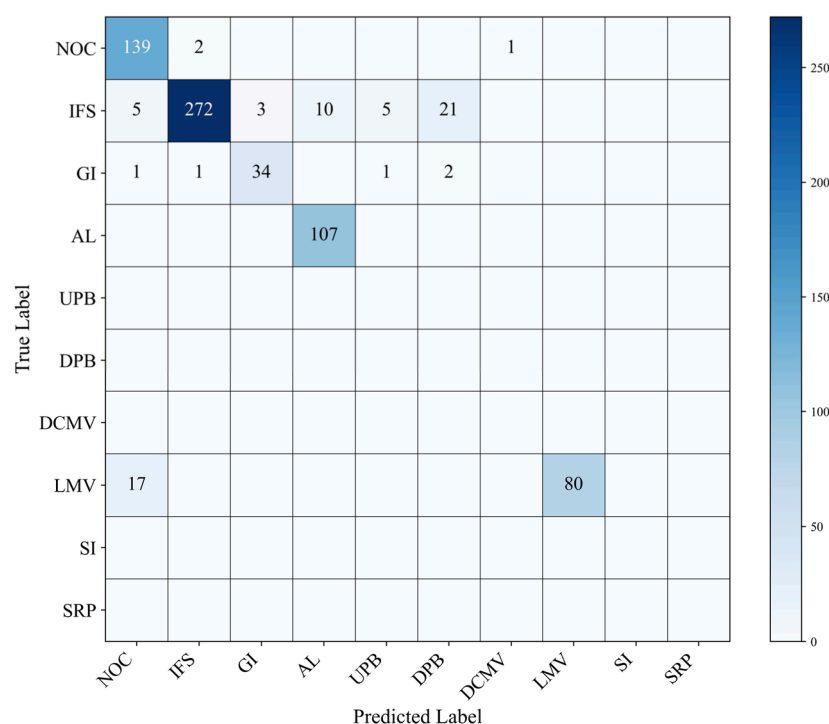


Figure 16. Confusion matrix of the improved MobileNetV2 model on the cross-source external validation set.

Table 8. Performance of the improved MobileNetV2 on the cross-source external validation set.

Class	Value (%)
Accuracy	90.16
Accuracy std	0.29
Macro Recall	90.63
Macro Recall std	0.22
Macro F1-score	91.53
Macro F1-score std	0.31

In Table 8, Accuracy, Macro Recall, and Macro F1-score are reported as the mean values of three repeated experiments using different random seeds. Accuracy std, Recall std, and F1-score std denote the standard deviations of Accuracy, Macro Recall, and Macro F1-score across the three repeated experiments, respectively.

As shown in Figure 16 and Table 8, the improved MobileNetV2 model achieved a mean accuracy of 90.16%, a mean Macro Recall of 90.63%, and a mean Macro F1-score of 91.53% on the external validation set. Most samples are still distributed along the main diagonal of the confusion matrix. This indicates that the model maintains a certain identification capability when applied to dynamometer card data collected from YanChang Oilfield Co., Ltd. (Yan'an, China). However, compared with the original test set, the identification performance on the cross-source external validation set decreases to some extent. This is because differences in the geological conditions and acquisition processes of different oilfields can affect the shape of dynamometer cards, thereby affecting the identification performance of the model.

Among them, the SI and SRP samples in the training dataset were collected from YanChang Oilfield Co., Ltd. (Yan'an, China), while the external validation dataset was also obtained from YanChang Oilfield Co., Ltd. (Yan'an, China) but contained five operating conditions, including NOC, IFS, GI, AL, and LMV. Therefore, the external validation

results also provide supplementary evidence for analyzing whether the model relies on source-specific characteristics rather than operating-condition-related features.

As shown in Figure 16, although the external validation samples and the SI/SRP samples originated from the same oilfield source, these samples were not systematically misclassified as SI or SRP. Instead, the model maintained the ability to distinguish different operating conditions within the same data source. This suggests that the model does not rely solely on the data source information as a shortcut for identifying SI and SRP, and that operating-condition-related morphological information also contributes to its classification decisions.

The current cross-source external validation set only includes five operating conditions and does not contain independent SI and SRP samples collected from other oilfields. Therefore, although the above results provide supplementary evidence that the model does not simply rely on the Yanchang Oilfield data source as the classification basis for SI and SRP, the influence of data-source differences on these two categories cannot be completely excluded. Furthermore, the current experiment is still insufficient to fully demonstrate the broad engineering generalization ability of the proposed method. In future work, more dynamometer card data, especially SI and SRP samples collected from different oilfields, will be introduced to further verify the robustness and cross-oilfield generalization capability of the model.

5. Conclusions

This paper addresses the limitations in the dynamometer-card-based fault identification task of the sucker rod pumping system, including the limited ability of lightweight models to distinguish similar fault categories, insufficient representation of multi-scale morphological features, and the difficulty of model training under limited sample conditions. For the dynamometer-card-based fault identification task of the sucker rod pumping system, MobileNetV2 is used as the framework, and improvements are made to the attention mechanism, activation function, and multi-scale feature extraction structure, and a lightweight fault identification method based on transfer learning and improved MobileNetV2 is proposed. Through collaborative optimization of multiple modules, the proposed method improves the fault identification capability on the current dataset of the model while maintaining its lightweight characteristics.

1. The introduction of the coordinate attention mechanism enhances the model's representation capability for local fault regions and spatial positional information in dynamometer cards. By considering the spatial distribution characteristics of fault-related morphological features, this mechanism enables the model to focus more effectively on local variations related to fault categories, such as local missing regions, corner deformation, and slope changes.
2. The introduction of transfer learning alleviates the difficulty of model training under limited sample conditions, while the LeakyReLU6 activation function improves the representation of subtle fault features by preserving part of the negative feature responses. Specifically, the transfer learning strategy based on ImageNet-1K pre-trained weights provides effective initial feature representations for the model and reduces the model's dependence on a large number of labeled samples. The LeakyReLU6 activation function alleviates the information loss caused by the truncation of negative responses in ReLU6 by preserving responses in the negative region, thereby providing more complete feature representations for subsequent network layers.
3. The improved Inception A multi-scale feature extraction module enhances the model's representation capability for overall contour variations and local detail differences in dynamometer cards. By introducing parallel feature extraction branches with different

receptive fields, this module provides richer multi-scale morphological information for the network and improves the model's identification capability for fault categories with complex local morphological differences.

The experimental results and Grad-CAM class activation heatmaps show that the improved model can focus on different classification-relevant morphological regions according to different fault types, thereby better performing the fault identification task of the sucker rod pumping system. Compared with the baseline MobileNetV2 model, the mean accuracy is improved by 4.01 percentage points and reaches 95.01% with only 2.28 M parameters. Meanwhile, its Accuracy, Macro Recall, and Macro F1-score are all higher than those of EfficientNetV2, ResNet34, DenseNet121, MobileNetV3, ShuffleNetV2, and the original MobileNetV2. In addition, the cross-source external validation results show that the improved MobileNetV2 model achieves a mean accuracy of 90.16%, a mean Macro Recall of 90.63%, and a mean Macro F1-score of 91.53% on measured dynamometer card samples from YanChang Oilfield Co., Ltd. (Yan'an, China), indicating that the proposed method has a certain applicability to data from different sources. However, this external validation set only covers part of the operating condition categories, and the current results are still insufficient to fully demonstrate broad engineering generalization ability. Considering both identification performance and model complexity, the improved MobileNetV2 achieved a good balance between identification performance and model size on the evaluation dataset. However, the current validation work is mainly based on the existing dynamometer card samples, and the external validation dataset only covers some operating conditions. Therefore, before practical deployment, further validation using large-scale field data from different production environments and containing more diverse operating conditions is still required. Future research will focus on expanding multi-source datasets, improving the model's adaptability to different field environments, and integrating the proposed identification method into practical monitoring systems.

Author Contributions: Conceptualization, W.S. and L.G.; methodology, W.S. and L.G.; software, W.S. and L.G.; validation, W.S., L.G. and W.Z.; formal analysis, W.S., L.G. and W.Z.; resources, W.S. and L.G.; data curation, W.S. and L.G.; writing—original draft preparation, W.S. and L.G.; writing—review and editing, W.S., L.G. and W.Z.; visualization, W.S. and L.G.; project administration, W.S.; funding acquisition, W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the research projects of the Shaanxi Province Science and Technology Department in China, grant number 2023-YBGY-068.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: The authors gratefully acknowledge the financial support from the research projects of the Shaanxi Province Science and Technology Department in China under grant No. 2023-YBGY-068.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yuan, C.; Liao, Z.; Li, X. A fault diagnosis model for pumping units based on small sample electric parameters. *Sci. Rep.* **2025**, *15*, 22997. [[CrossRef](#)]
2. Feng, Z.; Wang, X.; Huang, J.; Tan, C.; Zhao, Y. Intelligent diagnostic method for beam pumping units based on torque indicator cards. *Measurement* **2025**, *242*, 115903. [[CrossRef](#)]

3. Zhu, Z.; Han, D.; Liu, Z.; Li, B.; Ge, Z.; Yin, X.; Wei, X. Research progress and prospects in fault diagnosis of pumping units using dynamometer cards. *Eng. Res. Express* **2025**, *7*, 042201. [\[CrossRef\]](#)
4. Zhao, H.; Wang, J.; Liu, T.; Yu, Y.; Hu, D.; Cai, C. Research on the soft-sensing method of indicator diagram of beam pumping unit. *Sensors* **2024**, *24*, 1794. [\[CrossRef\]](#)
5. Velychkovych, A.S.; Shatskyi, I.P.; Shopa, V.M.; Bedzir, O.O. Slotted-shell friction damper for sucker-rod string suspension: Analytical model and parametric study. *Arch. Mater. Sci. Eng.* **2025**, *134*, 49–60. [\[CrossRef\]](#)
6. Velichkovich, A.S. Shock absorber for oil-well sucker-rod pumping unit. *Chem. Pet. Eng.* **2005**, *41*, 544–546. [\[CrossRef\]](#)
7. Gibbs, S.G.; Neely, A.B. Computer diagnosis of down-hole conditions in sucker rod pumping wells. *J. Pet. Technol.* **1966**, *18*, 91–98. [\[CrossRef\]](#)
8. Foley, W.L.; Svinos, J.G. Expert advisor program for rod pumping. *J. Pet. Technol.* **1989**, *41*, 394–400. [\[CrossRef\]](#)
9. Xu, P.; Xu, S.; Yin, H.W. Application of self-organizing competitive neural network in fault diagnosis of suck rod pumping system. *J. Pet. Sci. Eng.* **2007**, *58*, 43–48. [\[CrossRef\]](#)
10. Zhang, A.; Gao, X. Fault diagnosis of sucker rod pumping systems based on Curvelet transform and sparse multi-graph regularized extreme learning machine. *Int. J. Comput. Intell. Syst.* **2018**, *11*, 428–437. [\[CrossRef\]](#)
11. Cheng, H.; Yu, H.; Zeng, P.; Osipov, E.; Li, S.; Vyatkin, V. Automatic recognition of sucker-rod pumping system working conditions using dynamometer cards with transfer learning and SVM. *Sensors* **2020**, *20*, 5659. [\[CrossRef\]](#)
12. Wu, Y.; Feng, Z.; Liang, J.; Liu, Q.; Sun, D. Fault diagnosis algorithm of beam pumping unit based on transfer learning and DenseNet model. *Appl. Sci.* **2022**, *12*, 11091. [\[CrossRef\]](#)
13. Ye, Z.; Yi, Q. Working-condition diagnosis of a beam pumping unit based on a deep-learning convolutional neural network. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2022**, *236*, 2559–2573. [\[CrossRef\]](#)
14. Li, J.; Shao, J.; Wang, W.; Xie, W. An evolutionary deep learning method based on multi-feature fusion for fault diagnosis in sucker rod pumping system. *Alex. Eng. J.* **2023**, *66*, 343–355. [\[CrossRef\]](#)
15. Li, B.; Raja, S.S.; Li, J.; Yao, Z.; Song, W.; Li, H. Pumping machine fault diagnosis based on fused RDC-RBF. *PLoS ONE* **2023**, *18*, e0291777. [\[CrossRef\]](#)
16. Zhang, L.; Wu, J.; Zhang, K.; Wang, Z.; Yan, X.; Liu, P.; Wang, Q.; Fan, L.; Yao, J.; Yang, Y.; et al. Diagnosis of pumping machine working conditions based on transfer learning and ViT model. *Geoenergy Sci. Eng.* **2023**, *226*, 211729. [\[CrossRef\]](#)
17. He, Y.-P.; Cheng, H.-B.; Zeng, P.; Zang, C.-Z.; Dong, Q.-W.; Wan, G.-X.; Dong, X.-T. Working condition recognition of sucker rod pumping system based on 4-segment time-frequency signature matrix and deep learning. *Pet. Sci.* **2024**, *21*, 641–653. [\[CrossRef\]](#)
18. Dong, G.; Li, W.; Dong, Z.; Wang, C.; Qian, S.; Zhang, T.; Ma, X.; Zou, L.; Lin, K.; Liu, Z. Enhancing dynagraph card classification in pumping systems using transfer learning and the Swin Transformer model. *Appl. Sci.* **2024**, *14*, 1657. [\[CrossRef\]](#)
19. Sreenivasan, H.; Krishna, S. Automatic sucker rod pump fault diagnostics by transfer learning using GoogLeNet integrated machine learning classifiers. *Process Saf. Environ. Prot.* **2024**, *191*, 14–26. [\[CrossRef\]](#)
20. Qian, K.; Tang, J.; Zhao, Q.; Zhao, S.; Min, F. Multi-label learning for fault diagnosis of pumping units with one positive label. *Appl. Soft Comput.* **2025**, *174*, 113014. [\[CrossRef\]](#)
21. Wang, X.; Liu, Y.; Cheng, X.; Wang, F.-Y. ParaDC: Parallel-learning-based dynamometer cards augmentation with diffusion models in sucker rod pump systems. *Neurocomputing* **2025**, *617*, 128973. [\[CrossRef\]](#)
22. Xu, C.; Wang, Z.; Jin, Y.; Nong, W. An adaptive industrial large language model for mechanical fault diagnosis under variable operating conditions. *Adv. Eng. Inform.* **2026**, *74*, 104821. [\[CrossRef\]](#)
23. Gao, Z.-W.; Jia, S. Modeling and control for beam pumping units: An overview. *Processes* **2024**, *12*, 1468. [\[CrossRef\]](#)
24. Gao, Z.-W.; Zhu, X.; Liu, Y. A review of fault diagnosis and prognosis for oil pumping machines. *Next Energy* **2026**, *11*, 100522. [\[CrossRef\]](#)
25. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [\[CrossRef\]](#)
26. Lv, H.; Chen, J.; Pan, T.; Zhang, T.; Feng, Y.; Liu, S. Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application. *Measurement* **2022**, *199*, 111594. [\[CrossRef\]](#)
27. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722. [\[CrossRef\]](#)
28. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv* **2015**, arXiv:1505.00853. [\[CrossRef\]](#)
29. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9. [\[CrossRef\]](#)

30. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; Volume 31, pp. 4278–4284. [[CrossRef](#)]
31. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929. [[CrossRef](#)]
32. Tan, M.; Le, Q.V. EfficientNetV2: Smaller models and faster training. *arXiv* **2021**, arXiv:2104.00298. [[CrossRef](#)]
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
34. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. [[CrossRef](#)]
35. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324. [[CrossRef](#)]
36. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In *Computer Vision—ECCV 2018, Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Cham, Switzerland, 2018; pp. 122–138. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.