IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 December 2023, accepted 2 January 2024, date of publication 5 January 2024, date of current version 12 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3350170

RESEARCH ARTICLE

A Novel Methodology for Microgrid Power Quality Disturbance Classification Using URPM-CWT and Multi-Channel Feature Fusion

JUNZHUO JIANG[1,2], HAO WU[1,2], (Member, IEEE), CHANGHUA ZHONG[1,2], YUAN CAI[1,2], AND HONG SONG[3]

[1]Department of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, Sichuan 643000, China
[2]Artificial Intelligence Key Laboratory of Sichuan Province, Zigong, Sichuan 643000, China
[3]School of Automation and Information Engineering, Aba Teachers College Aba, Sichuan 623002, China

Corresponding author: Hao Wu (11305076@qq.com)

This work was supported in part by the Project of Sichuan Provincial Science and Technology Department under Grant 2022YFS0518 and Grant 2022ZHCG0035; in part by the Artificial Intelligence Key Laboratory of Sichuan Province Foundation under Grant 2023RYY06; in part by the Enterprise Informatization and Internet of things Measurement and Control Technology Key Laboratory Project of Sichuan Provincial University under Grant 2022WYY04; in part by the Talent Introduction Project of Sichuan University of Science and Engineering under Grant 2021RC12; and in part by the Project of Zigong Science and Technology Bureau under Grant 2019YYJC13, Grant 2019YYJC02, and Grant 2020YGJC16.

ABSTRACT Addressing the limitations inherent in conventional Power Quality Disturbance (PQD) identification systems, particularly regarding the restricted information obtainable from single image features and the compromised noise immunity of single-channel networks, an innovative approach, integrating Uniform Relative Position Matrix-Continuous Wavelet Transform (URPM-CWT) and multi-channel feature fusion, is presented. This method capitalizes on the principle of feature fusion to enhance microgrid PQD identification. To begin with, each PQD signal undergoes processing through the URPM and CWT, followed by horizontal splicing to yield the URPM-CWT feature image. This is followed by the parallel deployment of three refined networks—MobileNetV2, ResNet50, and ShuffleNetV2—using the Self Fusion Module (SFM) to yield a multi-channel feature fusion classification model. The final stage involves feeding the URPM-CWT feature image into the multi-channel feature fusion classification model and applying a fully connected layer for training, leading to comprehensive perturbation recognition. Constructed using the PyTorch framework, the proposed model is evaluated on an exhaustive database of 28 distinct PQD types. In a 30db white noise environment, the method demonstrates an average classification accuracy of 99.35%, surpassing the performance of standalone deep learning recognition approaches. Simulation experiments corroborate the model's high classification accuracy, effective recognition, and robust resistance to noise when dealing with PQD signals. Thus, the model offers promising potential for practical applications in PQD identification and classification.

INDEX TERMS Power quality disturbances, deep learning, relative position matrix, wavelet transform, feature fusion.

I. INTRODUCTION

A microgrid is a self-contained, compact power system encompassing diverse distributed power sources, loads, energy storage, and control systems. It offers key advantages such as efficient resource use, swift setup, and broad applicability, making distributed power an appealing choice for energy generation and management. The widespread utilization of power electronic devices in microgrids leads to the injection of a substantial number of harmonic signals into the grid. This causes power quality deterioration manifested in voltage waveform distortion, fluctuation,

The associate editor coordinating the review of this manuscript and approving it for publication was Sinisa Djurovic.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

flickering, and three-phase unbalance. Furthermore, mini-grids, typically independent nodes operable in both connected and isolated modes, help optimize the use of distributed power. They enable local energy management, provide redundancy, reinforce grid resilience, and reduce transmission losses. By balancing local generation and load, mini-grids mitigate peak demand effects and voltage fluctuations on the larger grid, enhancing power supply reliability and overall grid efficiency. Given their substantial role, microgrids are at the forefront of global power system research and development. To fully leverage the advantages of microgrids, it is essential to address fundamental power system issues. One crucial technology for microgrids is the quick and accurate determination of power quality disturbances. Proper identification of the types of disturbances, such as voltage sags, swells, harmonics, and flicker, is vital, as it lays the foundation for effectively managing power quality issues in the future [1].

Researchers worldwide are diligently working on the accurate identification of power quality disturbance signals, and various methodologies have been proposed to tackle this omplex problem. One such proposal suggests utilizing Variational Mode Decomposition (VMD) combined with a Random Discriminative Projection Extreme Learning Machine (RDPEML) [2]. While this approach demonstrates commendable classification prowess, another introduces a Discrete Fourier Transform (DFT)-based method for harmonic estimation, which flaunts exceptional accuracy and adaptability [3]. Also notable are the practical applications of Empirical Mode Decomposition (EMD)-based noise reduction techniques that showcase the promise of real-time harmonic signal analysis [4]. Other works have combined the double-resolution S-transform with directed acyclic graph support vector machines to cater to both single and combined disturbances [5].

However, as smart grids evolve and data volume swells, the challenges escalate. Traditional methods struggle, necessitating the introduction of deeper machine learning models like the 1D-MR model, grounded on the Inception-ResNet framework [7]. Recognizing the limitations of traditional algorithms, this study embarks on refining the SSPQDD algorithm. By embracing a multiple sequence approach, it brings forth an enhanced detection mechanism and recognizes two concurrent disturbances within a singular window [8]. Additionally, with the rise of complex multilabel classification tasks, this research introduces LGAN, a deep learning marvel adept at extracting, directing, and predicting PQDs [9]. To ensure the interpretations of these classifiers remain lucid, an explainable artificial intelligence method is proposed, emphasizing comprehensible decision-making [10]. The burgeoning field also sees the introduction of an ensemble convolution neural network (ECNN) targeting the quality of electric vehicle charging [11]. The realm of power quality disturbance recognition is undergoing transformative changes. Recent research has shown the potential of transitioning from one-dimensional signals to two-dimensional visuals using the Markov transition field (MTF), coupled with the deep residual network (Resnet). This technique has reported an astounding recognition rate of 99.875% [12]. Moreover, the challenge of overfitting, a persistent issue in machine learning, has been addressed by another study through the dynamic adjustment of sub-decision tree weights, demonstrating enhanced accuracy, notably in renewable energy system disturbances [13]. In the quest for optimized classification, the DAE network has emerged, offering superior performance metrics compared to prominent techniques like SVM, SAE, and PCA [14]. Adding to this line-up of advancements, the adoption of fully convolutional networks (FCNs), bidirectional gated recurrent unit (BiGRU), and the squeeze-and-excitation network (SENet) presents a promising shift, holding an advantage over current deep learning paradigms [15].

Each of these approaches offers unique insights and advancements in the field of power quality disturbance identification, contributing to the development of more robust and accurate methods for power systems analysis.

The following two major problems commonly exist in the current microgrid Power Quality Disturbance recognition: 1. Insufficient consideration of the complementary nature of time-frequency analysis and image conversion methods, leading to inadequate consideration of the static structure and dynamic behavior of signals. 2. The limitations of single-channel network features present significant challenges, primarily due to the inadequate feature extraction which results in considerable feature loss. This consequently diminishes recognition accuracy, an issue that becomes increasingly prominent in high-noise environments. In an effort to confront and overcome these obstacles, this paper makes the following noteworthy contributions:

1) A novel URPM-CWT feature image: The traditional Relative Position Matrix (RPM) conversion method is streamlined and upgraded, and a more effective Symbolic Relative Position Matrix (URPM) conversion algorithm is introduced. This advanced algorithm not only accelerates the conversion time but also produces superior results. By fusing the URPM with the Continuous Wavelet Transform (CWT) on a level plane, we generate the URPM-CWT feature map. This innovative procedure heightens the consistency of the image features and bolsters its capability to resist noise interference.
2) A Self Fusion Module (SFM): This module is designed to adapt to the multi-channel feature fusion classification model, automatically adjusting the importance of different features, and accommodating various inputs and tasks. The adaptive feature fusion module enables the model to perform well in handling complex scenarios.
3) A multi-channel feature fusion classification model: Comprising improved MobileNetV2, ResNet50, ShuffleNetV2, and the adaptive feature fusion module, this model employs three parallel networks for feature

extraction, thereby enhancing recognition accuracy. The final recognition and classification are achieved through the fully connected layer, resulting in effective performance even in high-noise environments.

II. IMAGE PREPROCESSING

Image Preprocessing is the process of processing the original image to improve the quality of the image data or to extract some important characteristics about the image so that subsequent image processing or image analysis can be more effective. This section introduces the Relative Position Matrix (RPM) [16] and its optimized Uniform Relative Position Matrix (URPM) and Continuous Wavelet Transform (CWT) [17], and finally the two are horizontally stitched to obtain the URPM-CWT feature image.

A. URPM IMAGE GENERATION

1) TREND EXTRACTION

In this section, the focus is on trend extraction from original time series data. First, a standard normal distribution Z is obtained by the following z-score normalization method:

In this section, the focus is on trend extraction from original time series data. First, a standard normal distribution Z is obtained by the following z-score normalization method:

$$z_t = \frac{x_t - \mu}{\sigma}, \quad t = 1, 2, \ldots, N \tag{1}$$

where μ, denotes the mean value of X and σ denotes the standard deviation of X.

Then, we employed the Symbolic Aggregate approximation (SAX) [18] methodology to process the power quality disturbance signals. Initially, a Piecewise Aggregate Approximation (PAA) was utilized to represent the features of the signals, thereby facilitating a more robust analysis and interpretation of the data in question.

$$\bar{c}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} Z_j \tag{2}$$

In order to reduce the dimensionality of the n-dimensional original power quality disturbance signals to w dimensions, the original signals were segmented into w fragments. Let $\bar{c}_i$ denote the mean value of the i-th segment. This procedure reduces the dimensionality of the power quality disturbance signals from n to w, thereby segmenting the signal into w equal-sized "frames". It is imperative to ensure that the compression ratio $\frac{n}{w}$ is an integer.

Subsequent to the dimensionality reduction, the PAA feature representation was discretized into SAX, aligning the resulting symbols with the corresponding features of the time series, which share identical probabilities. The results demonstrate that the values of the z-score normalized time series adhere to a normal distribution.

Leveraging the properties of the methodology, it becomes feasible to employ a lookup table approach to ascertain the coordinates of a straight line under the normal curve, thereby partitioning the area under the Gaussian curve. In the context of SAX, these x-coordinates of the lines are referred to as breakpoints, and the breakpoints listed in the table partition the data in the N (0,1) distribution into a equal regions. By assigning a corresponding alphabet symbol *alpha* to each interval $[\beta_{j-1}, \beta_j)$, the vector's transformation PAA coefficients $\hat{c}$ to the string $\bar{c}$ is implemented as follows:

$$\hat{c} * i = alpha * j, iif, \bar{c} * i \in [\beta_{j-1}, \beta_j) \tag{3}$$

SAX introduces a new metric for measuring the distance between strings by extending the Euclidean distance and the PAA distance. This function returns the minimum distance between string features of two original time series $\hat{Q}$ and $\hat{C}$.

$$MINDIST(\hat{Q}, \hat{C}) \equiv \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^{w} (dist(\hat{q} * i, \hat{c} * i))^2} \tag{4}$$

2) IMAGE GENERATION

To perform relative position computation between two two minimum distances M, the computation follows the SAX algorithm. The preprocessed power quality disturbance signals are then converted into a two-dimensional matrix H, resulting in a Uniform Relative Position Matrix (URPM) image. This approach eliminates the need for the traditional relative position matrix's min-max normalization step, significantly improving image conversion efficiency and enhancing the characteristics of the resulting image.

$$H = \begin{bmatrix} M_1 - M_1 & M_2 - M_1 & \cdots & M_m - M_1 \\ M_1 - M_2 & M_2 - M_2 & \cdots & M_m - M_2 \\ \vdots & \vdots & \ddots & \vdots \\ M_1 - M_m & M_2 - M_m & \cdots & M_m - M_m \end{bmatrix} \tag{5}$$

Which eliminates the traditional relative position matrix in the use of min-max normalization to convert the resulting matrix to gray value matrix this step, greatly improving the image conversion efficiency as well as enhance the characteristics of the picture law. The obtained image is shown in Figure.1, and it can be clearly seen that the optimized image features are more obvious.

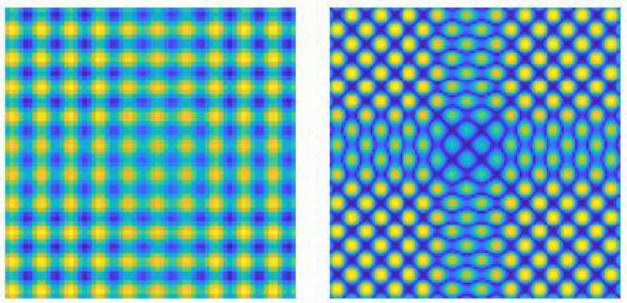

FIGURE 1. Comparison of URPM and RPM images.

B. CWT IMAGE GENERATION

Wavelet time-frequency image representation is a powerful tool capable of visualizing the characteristics of a signal in

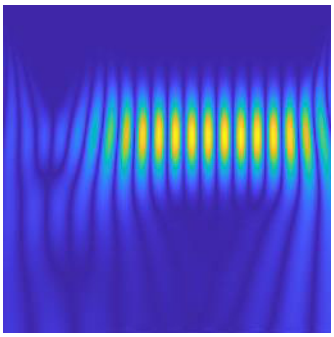

FIGURE 2. CWT images.

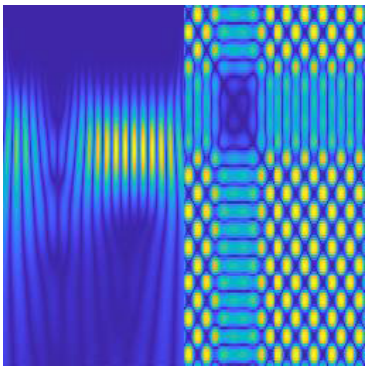

FIGURE 3. URPM-CWT images.

time and frequency simultaneously. The Continuous Wavelet Transform (CWT) [17] builds on the Fourier Transform (FT) [19] by introducing scale and translation factors, and by using finite-length and decaying wavelet basis functions instead of the sine and cosine basis functions used in the FT. The finite length and decaying nature of wavelet basis functions allow for localized frequency analysis, an advantage not present in traditional FT. This allows the window function to vary with frequency characteristics, enabling detailed analysis at different scales and locations.

The advantage of CWT lies in its adaptive window property, for high-frequency components, the window length is shorter to accurately obtain the location of the event, and for low-frequency components, the window length is longer to obtain more accurate frequency information. This flexibility makes CWT able to show the time domain and frequency domain characteristics of the signal and obtain better time-frequency analysis.

First, we select a mother wavelet (base wavelet), denoted as ψ. In this paper, we use the Morlet wavelet, known for providing an optimal balance between time and frequency resolution, thus facilitating the consideration of both time and frequency domain characteristics of the signal. The mother wavelet is a locally fluctuating, complex function with zero mean, which is mainly used to analyze the characteristics of signals at different time frequencies. The basic formula for the continuous wavelet transform is:

$$\mathrm{CWT}(s,f) = \frac{1}{2\pi\sqrt{s}} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} y(t)\psi^* \left(\frac{t-\tau}{s} \right) dt \right) e^{-j2\pi f\tau} d\tau \tag{6}$$

where s and τ denote the scale and translation factors, respectively, ψ^* denotes the complex conjugate of the scale and translation wavelet functions, and y(t) denotes the perturbation signal.

Then, a collection of scale and translation factors is systematically selected, often based on the specific time-frequency analysis needs of the given signal. Finally, a Figure 2 is obtained which represents the information of the original signal at different times and frequencies.

C. URPM-CWT IMAGE GENERATION

In our study, we combine the two unique image representations mentioned above, the symbol relative position matrix (URPM) and the wavelet time-frequency image (based on CWT), to reveal the structural, frequency, and time-domain properties of the original perturbed signals more comprehensively. Specifically, we generate a URPM image and a wavelet time-frequency image for each signal, and then horizontally stitch these two images to generate a combined image, as shown in Figure. 3.

In this combined image, the RPM image portion reflects the relative positional information of the elements in the signal, which captures and preserves some of the underlying structure and patterns of the original time series. The wavelet time-frequency image portion, on the other hand, demonstrates the behavior of the signal in time and frequency, which can reveal some of the variations and dynamic properties of the signal, including periodicity, abrupt changes, and trends.

By horizontally splicing these two images, we have fused the relative position information and time-frequency characteristics, two distinct perspectives on signal analysis, into one combined image. In this way, we can consider both the static structure (via URPM) and the dynamic behavior (via CWT) of the signal. This multi-perspective, multi-property representation not only preserves the original complexity of the signal, but also enhances the accuracy and reliability of our understanding and analysis of the signal by providing a more holistic view.

III. MULTI-CHANNEL FEATURE FUSION CLASSIFICATION MODEL

In deep learning, feature extraction and feature fusion are crucial aspects as they help in representing the data in a way that enhances the learning capability of models. In this section, we first explain the working principles of the Self Fusion Module and the Convolutional Block Attention Module. Then, we introduce the basic structures of MobileNetV2 [20], ResNet50 [22], and ShuffleNetV2 [23], along with the improvement measures. Finally, we present the classification model based on URPM-CWT and multi-channel feature fusion.

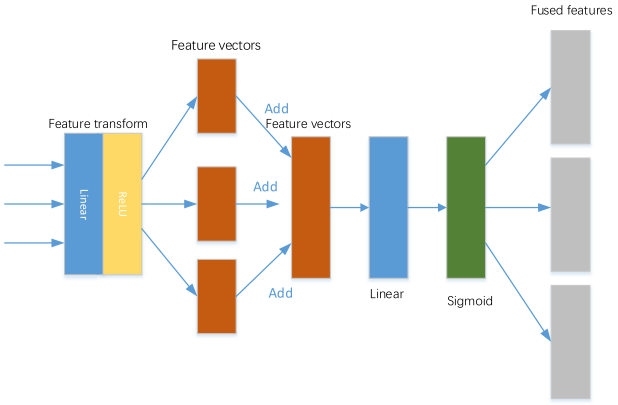

FIGURE 4. Self fusion module.

A. SELF FUSION MODULE AND CONVOLUTIONAL BLOCK ATTENTION MODULE

1) SELF FUSION MODULE

In order to fuse features from different layers more efficiently, this paper proposes a Self Fusion Module (SFM). This module consists of two linear layers (Linear), ReLU (Rectified Linear Unit) [24] activation function and Sigmoid [25] function. It is shown in Figure. 4.

The main part of the module can be divided into the following steps:

Feature Transformation: Each input feature undergoes a transformation through a linear layer and a ReLU activation function, enhancing the non-linearity of the model and facilitating the extraction of higher-level features. Given that the dimensions of the feature vectors output by the three networks (MobileNetV2, ResNet50, ShuffleNetV2) are different, the linear layer adjusts the dimension of each feature vector to a uniform size, enabling the fusion of features from all three models. For instance, in our study, the feature vector size of the ResNet50 output is 2048, while the output feature vector sizes for the other two models are 1024. Using the SFM module, we adjust the feature vector size to 2048 for all of them.

Attention Weight Calculation: The attention mechanism automatically calculates the weight for each feature, enabling adaptive fusion of different features. Specifically, suppose we have three feature vectors $x1$, $x2$, and $x3$ to be fused. First, these three feature vectors undergo non-linear transformations through a linear layer and a ReLU activation function, after which they are summed. This summed feature vector is then fed into a second linear layer to calculate the attention weights. These weights, representing the relative importance of each feature vector, are normalized using a Sigmoid function, ensuring their values lie between 0 and 1. This process allows the model to adaptively learn and focus on the most important features during fusion.

The attentional weights are computationally defined as:

$$\begin{aligned} \mathrm{f1} &= \mathrm{ReLU}\left(W_1 * x1 + \mathrm{b1}\right) \\ \mathrm{f2} &= \mathrm{ReLU}\left(W_1 * x2 + \mathrm{b1}\right) \\ \mathrm{f3} &= \mathrm{ReLU}\left(W_1 * x3 + \mathrm{b1}\right) \\ \mathrm{attn\_weights} &= \mathrm{Sigmoid}\left(W_2 * (f1 + f2 + f3) + b2\right) \end{aligned} \tag{7}$$

In the context of the process described, $x1, x2, x3$ represent the input features from the different networks. W_1, W_2 are the weight matrices of the two linear layers, b1, b2 are the corresponding bias terms, ReLU and Sigmoid are the nonlinear activation functions.

Feature fusion: We apply the attention weights, corresponding to each of the three input feature vectors, to obtain the fused features. Note that the fusion strategy here is not simply a weighted average of all the features, but rather a more sophisticated approach: for the first feature $x1$, we use the derived attention weight as its weight, whereas for the remaining two features $x2$ and $x3$, we use (1 minus the derived attention weight) as their weights.

This approach ensures a dynamic balance between the feature vectors. As the importance of the first feature $x1$ increases (i.e., the attention weight approaches 1), the importance of the remaining two features $x2$ and $x3$ decreases, and vice versa. This adaptive adjustment allows the model to focus on the most relevant features at any given moment, thereby enhancing the efficacy of the feature fusion process.

$$\begin{aligned} \text{fused features} = {} & \text{attn\_weights} * x1 \\ & + (1 - \text{attn\_weights}) * x2 \\ & + (1 - \text{attn\_weights}) * x3 \end{aligned} \tag{8}$$

where "fused features" is feature fusion, "attn_weights" is the weights of feature $x1$, and 1-"attn_weights" is the weights of features $x2$ and $x3$.

This adaptive feature fusion strategy not only automatically adjusts the importance of different features, but also adapts to different inputs and tasks, enabling our model to achieve good performance in dealing with various complex situations.

2) CONVOLUTIONAL BLOCK ATTENTION MODULE

CBAM (Convolutional Block Attention Module) [26] is an attention module that can effectively enhance the performance of convolutional neural networks. CBAM module introduces an attention mechanism in each of the two dimensions, channel and space, by seamlessly inserting it into an existing convolutional network. This design enables the model to pay more attention to the important channel and spatial regions, thus optimizing the model performance.

The CBAM module consists of two sub-modules: the Channel Attention Module and the Spatial Attention Module. As shown in Figure.5.

Channel Attention Module: for the input feature map $\boldsymbol{F} \in \mathbf{R}^{\boldsymbol{C}\times\boldsymbol{H}\times\boldsymbol{W}}$, where C is the number of channels, and H and W are the height and width, respectively, first, global average pooling and global maximum pooling are carried out, and the obtained results are denoted as $\boldsymbol{F}_{\mathrm{avg}} \in \mathbf{R}^{\boldsymbol{C}}$ and $\boldsymbol{F}_{\max} \in \mathbf{R}^{\boldsymbol{C}}$.Then, the two vectors are respectively passed through a

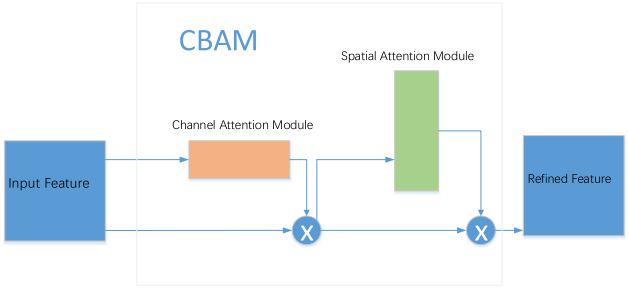

FIGURE 5. CBAM module.

multilayer perceptual machine (MLP) to be processed and the obtained results are denoted as $\mathbf{MLP}(\boldsymbol{F}_{\text{avg}})$ and $\mathbf{MLP}(\boldsymbol{F}_{\max})$. The two results are then summed and passed through a Sigmoid function to obtain the channel attention weights:

$$M_c(F) = \sigma(MLP(F_{avg}) + MLP(F_{max})) \tag{9}$$

where σ denotes the Sigmoid function. Finally, the obtained attention weights are multiplied with the original input feature map to obtain the output of the channel attention module:

$$F' = M_c \cdot F \tag{10}$$

Spatial Attention Module: first, the weighted feature map F' is subjected to the average and maximum operations in the channel dimension, and the obtained results are denoted as $F'_{\text{avg}} \in \mathrm{R}^{H\times W}$ and $F'_{\max} \in \mathrm{R}^{H\times W}$. Then, these two results are concatenated in the channel dimension and then passed through a convolutional layer of size 7×7 and a Sigmoid function to obtain the spatial attention weights:

$$M_s = \sigma\left(Conv\left(\left[F'_{avg}, F'_{max}\right]\right)\right) \tag{11}$$

Finally, the obtained spatial attention weights are subjected to an element-by-element multiplication operation with the output of the channel attention module to obtain the final output of the CBAM module:

$$F'' = M_s \cdot F' \tag{12}$$

Overall, the CBAM module allows the model to adaptively reshape the channel properties and spatial distribution of the feature map by introducing channel and spatial attention, thus optimizing the feature representation ability of the model.

B. INTRODUCTION AND IMPROVEMENT OF MOBILENETV2 MODEL

The MobileNet convolutional neural network, initially proposed by Google in 2017, uses Depthwise Separate Convolution (DSC) instead of standard convolution. This change makes the network more lightweight and reduces computation and parameters. MobileNetV2, introduced in 2018, further improved the model's performance and efficiency by introducing a linear bottleneck and an inverted residual structure [20].

The core of MobileNetV2 is the Depthwise Separable Convolution (DSC), which combines Depthwise Convolution (DC) and Pointwise Convolution (PC). DC applies a convolution kernel independently to each input channel, reducing computational load and preserving feature richness. PC, a 1×1 convolution operation, is used for upscaling and downscaling of feature channels.

TABLE 1. Detailed structure of improved MobileNetV2.

Input	Operator	t	c	n	s
$224^2\times3$	conv2d	-	32	1	2
$112^2\times32$	bottleneck	1	16	1	1
$112^2\times16$	bottleneck	6	24	2	2
$56^2\times24$	bottleneck	6	32	3	2
$28^2\times32$	bottleneck	6	64	4	2
$14^2\times64$	bottleneck	6	96	3	1
$14^2\times96$	bottleneck	6	160	3	2
$7^2\times160$	bottleneck	6	320	1	1
$7^2\times320$	bottleneck	6	320	1	1
$7^2\times320$	conv2d 1x1	-	1280	1	1
$7^2\times1280$	avgpool 7x7	-	-	1	-
$1\times1\times1280$	conv2d	-	k	-	

A key innovation in MobileNetV2 is the Linear Bottleneck structure, where a linear activation function is used in the last convolutional layer, instead of the ReLU activation function. Additionally, the Inverted Residual structure improves network efficiency by reducing computational complexity [20].

In this paper, we augment the original MobileNetV2 network with additional depth-separable convolutional blocks to capture more feature information. We replace the ReLU6 activation function with the SiLU [27] activation function, also known as the Swish function. This new activation function enables the network to learn more complex feature representations. We also include a Dropout [21] layer before the fully connected layer to prevent overfitting and add the CBAM Attention Mechanism module to help the model focus more on essential information during feature extraction.

The structure of the improved MobileNetV2 model is shown in Figure. 6. And the detailed structure of the model is shown in the Table 1.

C. INTRODUCTION AND IMPROVEMENT OF RESNET50 MODEL

The ResNet50 model, part of the ResNet (residual network) family, was designed to overcome challenges faced by deep learning networks, such as vanishing or exploding gradients. ResNet introduces "residual learning," where each layer learns a residual mapping between inputs and outputs, rather than directly learning the original mapping. This approach eases learning when the ideal mapping is close to a constant [22].

ResNet50 is a deep neural network consisting of 50 layers (including convolutional layers, activation layers, pooling layers, etc.). In these 50 layers, the main building blocks are multiple Residual Blocks. Each residual block usually contains three convolutional layers (1×1, 3×3, 1×1 convolutional kernel), one Identity Block and each convolutional layer is followed by a batch normalization (Batch Normalization) and ReLU activation function. In addition, each residual

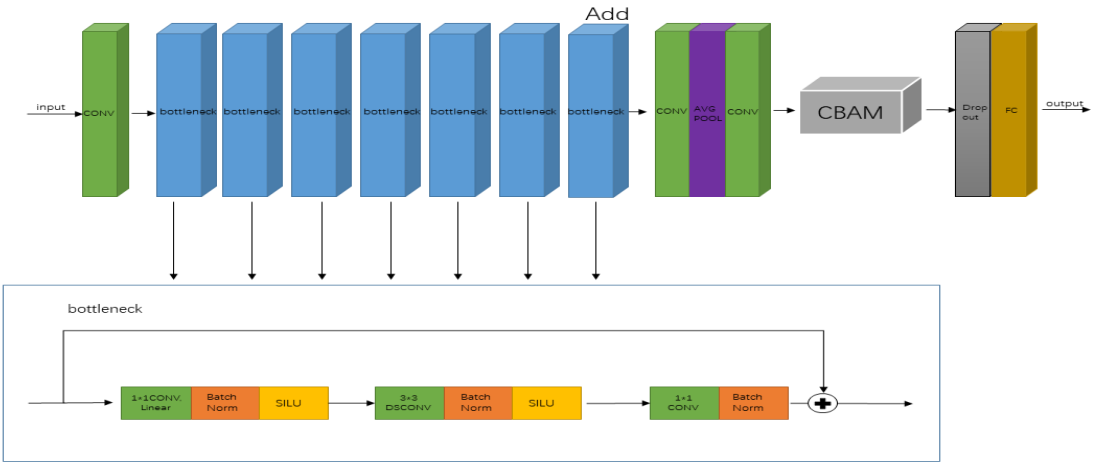

FIGURE 6. Improved MobileNetV2 model structure.

block contains a Shortcut Connection, which is the core innovation of ResNet. This shortcut connection allows the inputs of the network to be directly connected to its outputs, forming a constant mapping [22].

The input layer of the ResNet50 model accepts an RGB image of size 224 × 224, the first layer is a 7 × 7 convolution with a step size of 2, followed by a maximum pooling layer, also with a step size of 2. There are then four stages, each with a number of residual blocks, where the first residual block of each stage is used to reduce the size of the feature maps and to increase the number of channels, and the others maintain the size and the number of channels unchanged. The other residual locks maintain the size and number of channels unchanged. Stage 1 has three residual blocks, stage 2 has four, stage 3 has six, and stage 4 has three. Finally, a global average pooling layer, and a fully connected layer are connected, and the output layer is the number of categories for the classification task [22].

In our implementation, we replaced ResNet50's standard convolutional layers with depth-separable convolutional layers. We also added a CBAM attention mechanism module before the fully connected layer, helping the model focus on crucial information during feature extraction. This results in a more lightweight model, suitable for devices with limited computational resources.

The structure of the improved ResNet50 model is shown in Figure. 7. And the detailed structure of the model is shown in the Table 2.

TABLE 2. Detailed structure of improved ResNet50.

Layer Type	Output Size	Filter Size	Stride	Number of Filters
Input	224x224x3	-	-	-
DSConv1	112x112x64	7x7	2	64
MaxPool	56x56x64	3x3	2	-
Conv2 block	56x56x64	1x1	1	64
	56x56x64	3x3	1	64
	56x56x256	1x1	1	256
ID block	56x56x256	-	-	256
Conv2 block	28x28x128	1x1	2	128
	28x28x128	3x3	1	128
	28x28x512	1x1	1	512
ID block	28x28x512	-	-	512
Conv2 block	14x14x256	1x1	2	256
	14x14x256	3x3	1	256
	14x14x1024	1x1	1	1024
ID block	14x14x1024	-	-	1024
Conv2 block	7x7x512	1x1	2	512
	7x7x512	3x3	1	512
	7x7x2048	1x1	1	2048
ID block	7x7x2048	-	-	2048
Average Pool	1x1x2048	7x7	1	-

D. INTRODUCTION AND IMPROVEMENT OF SHUFFLENETV2 MODEL

ShuffleNetV2 is a highly efficient deep neural network architecture, designed for mobile and resource-constrained devices. The model aims to maintain or enhance performance while minimizing computational complexity, characterizing it with exceptional computational efficiency and fewer parameters [23].

The design of ShuffleNetV2 is based on four main factors: uniformity of input/output channels, equivalent width/height for balanced information flow, efficiency of matrix operations for computational speed, and reduction of network traffic to minimize data transfers. This reduction is achieved via a "channel shuffle" operation that rearranges feature channels

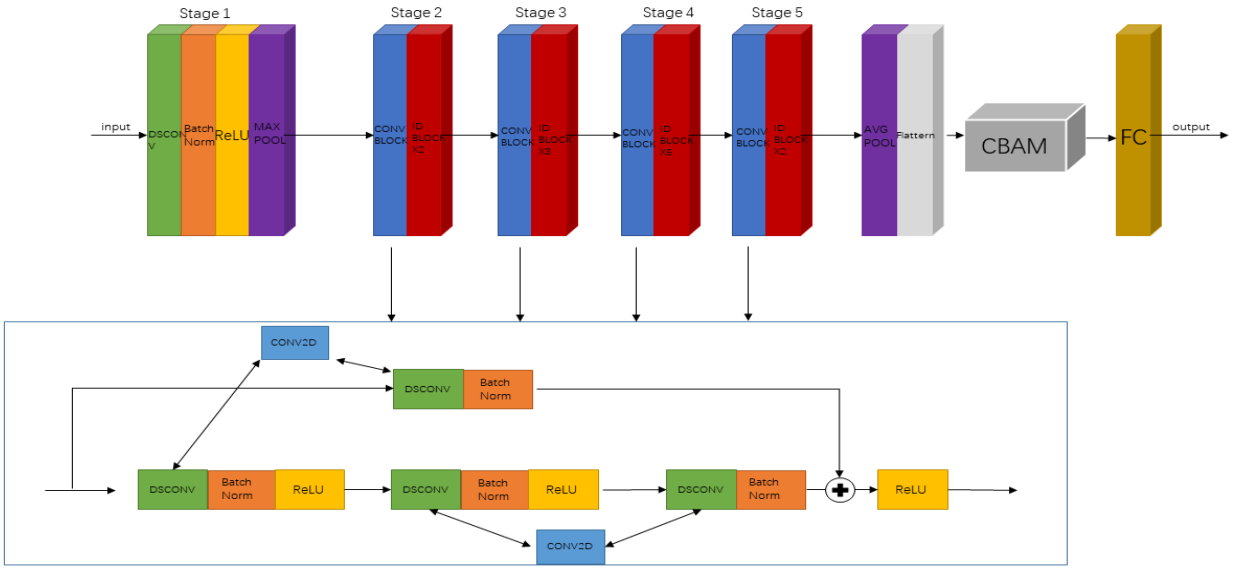

FIGURE 7. Improved ResNet50 model structure.

after each convolutional layer, allowing different features to be learned by various grouped convolutional layers.

The basic building block of ShuffleNetV2, called ShuffleUnit, includes grouped convolution, channel shuffling, and depth-separable convolution. Grouped convolution reduces the model's computational complexity, while the channel shuffle operation ensures adequate information exchange between individual groups. Depth-separable convolution further optimizes computational efficiency [23].

In our implementation, we added a CBAM attention mechanism module before the fully connected layer, enabling the model to focus on crucial information during the feature extraction process. This module can self-learn and automatically assign attention weights in the spatial and channel dimensions.

The structure of the improved ShuffleNetV2 model is shown in Figure. 8. And the detailed structure of the model is shown in the Table 3.

TABLE 3. Detailed structure of improved ShuffleNetV2.

Layer Type	Output Size	Filter Size	Stride	Channels	Operation
Input	224x224x3	-	-	-	-
Conv1	112x112x24	3x3	2	24	Con
MaxPool	56x56x24	3x3	2	-	-
ShuffleUnit (Stage 2, x4)	56x56x58	-	1	58	DSC + Conv
ShuffleUnit (Stage 3, x8)	28x28x116	-	2	116	DSC + Conv
ShuffleUnit (Stage 4, x4)	14x14x232	-	2	232	DSC + Conv
Conv5	14x14x1024	1x1	1	1024	Conv
Global Pool	1x1x1024	-	-	-	-

E. CLASSIFICATION MODEL BASED ON URPM-CWT AND MULTI-CHANNEL FEATURE FUSION

In this paper, we propose a classification model based on URPM-CWT and multi-channel feature fusion, the structure is shown in Figure. 9.

Firstly, the original signals are converted into two kinds of feature pictures, URPM and CWT, respectively, and then both of them are horizontally spliced to complete the preprocessing of the pictures to get more featured URPM-CWT feature pictures.

Secondly, we propose to unify the feature vectors of each channel to reach a suitable splicing dimension by a Self Fusion Module (SFM) module, and then let SFM perform adaptive weight calculation, which makes the model able to automatically adjust the importance of each channel feature, and finally realize the effective fusion of the three feature vectors.

Then this paper proposes a new Multi-channel feature fusion classification model (MCFFN). The model is mainly composed of three models, the improved MobileNetV2, ResNet50 and ShuffleNetV2, in parallel to realize multi-channel feature extraction. With the parallel model structure, each model can focus on learning different features of the input data, thus improving the overall model performance. And the advantages of each model can be integrated into the MCFFN model, MobileNetV2 can enhance the lightweight, high efficiency and structure optimization

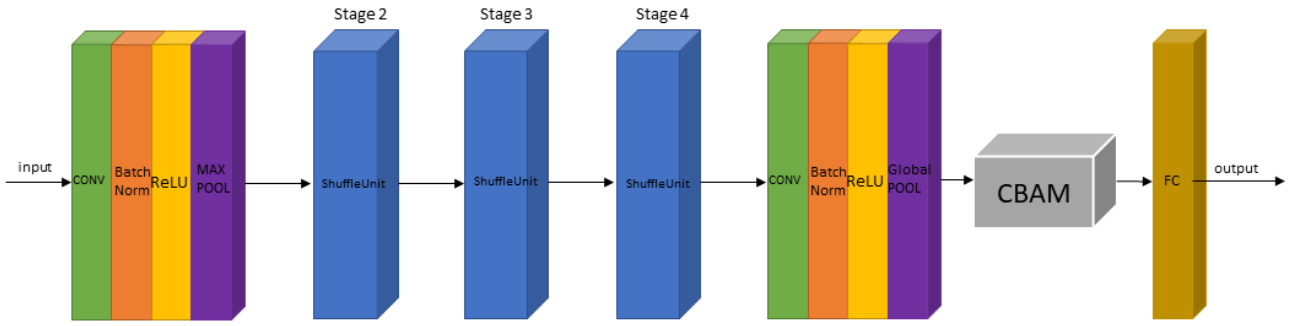

FIGURE 8. Improved ShuffleNetV2 model structure.

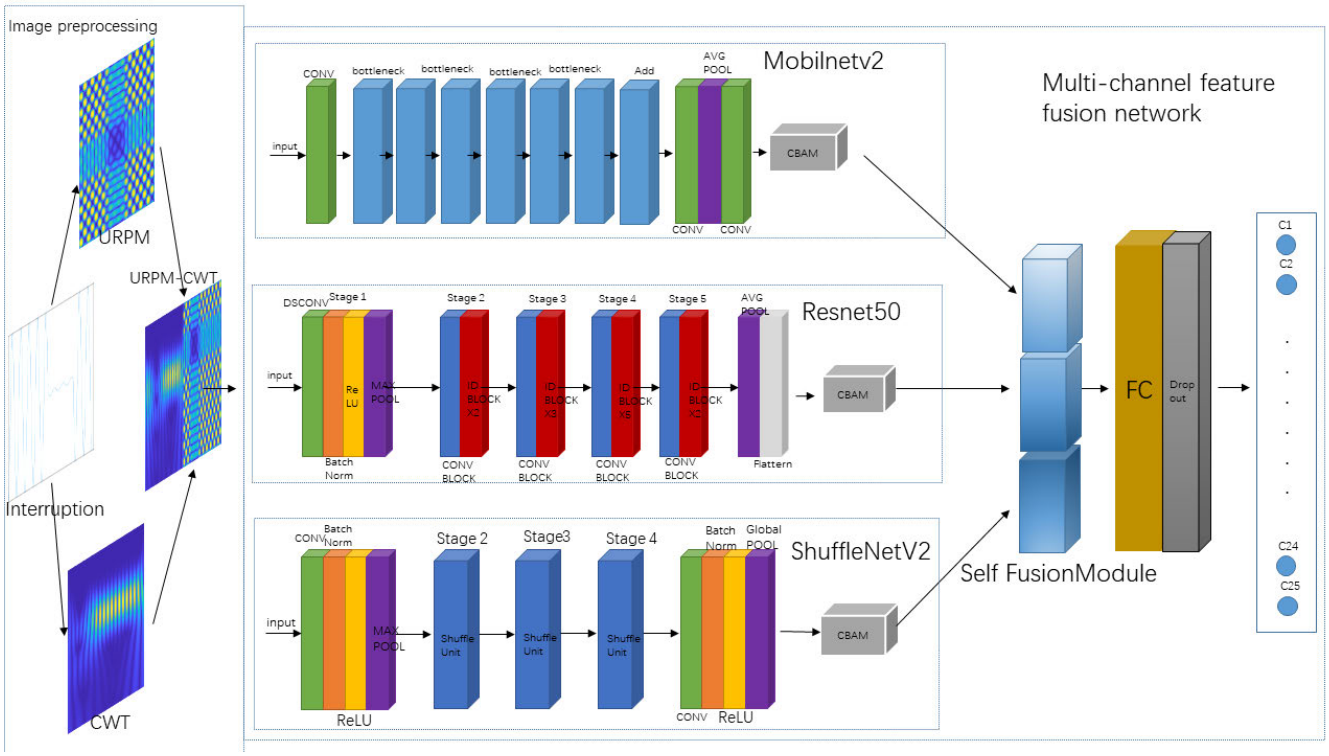

FIGURE 9. Classification model based on URPM-CWT and multi-channel feature fusion.

of the model, ResNet50 can solve the problem of gradient vanishing during the training of deep neural networks, and ShuffleNetV2 provides high efficiency and good performance, and finally the MCFFN model has strong robustness and generalization ability, which can be adapted to more kinds of perturbations and higher noise environments.

To cap off the process, we classify the features using a Fully Connected Layer (FC). This layer has the capacity to integrate all high-level features learned and facilitate deep feature interactions. The Fully Connected Layer is also designed to directly output the probability of target categories, enabling the model to easily adapt to complex classification tasks. At last, a Dropout layer is added behind the FC layer to prevent overfitting.

IV. RESULTS AND DISCUSSION

In order to verify the validity of the proposed model, this section describes the composition of the database and the comparison between the different approaches.

A. DISTURBANCE SIGNAL SAMPLE LIBRARY COMPOSITION

According to the IEEE 1159 [27] standard and in conjunction with related literature, 28 PQD signals are simulated in MATLAB given the power quality disturbance mathematical model. These signals consist of 8 single types, 15 double types and 5 triple types as shown in Table 4. where the fundamental frequency is set to 50Hz, the sampling frequency is set to 3.2kHz and the number of sampling points is set to

TABLE 4. Categories of PQDS.

Label	Category	Label	Category
C1	Nominal	C15	Sag + Oscillatory
C2	Sag	C16	Flicker + Oscillatory
C3	Swell	C17	Pulse + Flicker
C4	Interruption	C18	Interruption + Flicker
C5	Pulse	C19	Interruption + Pulse
C6	Flicker	C20	Pulse + Harmonics
C7	Harmonics	C21	Pulse + Oscillatory
C8	Oscillatory	C22	Pulse + Harmonics + Swell
C9	Swell + Harmonics	C23	Swell + Harmonics + Flicker
C10	Sag + Harmonics	C24	Flicker + Harmonics + Sag
C11	Interruption + Harmonics	C25	Pulse + Harmonics + Interruption
C12	Flicker + Harmonics	C26	Sag + Flicker
C13	Swell +	C27	Interruption + Oscillatory
C14	Oscillatory Swell + Pulse	C28	Interruption + Oscillatory + Pulse

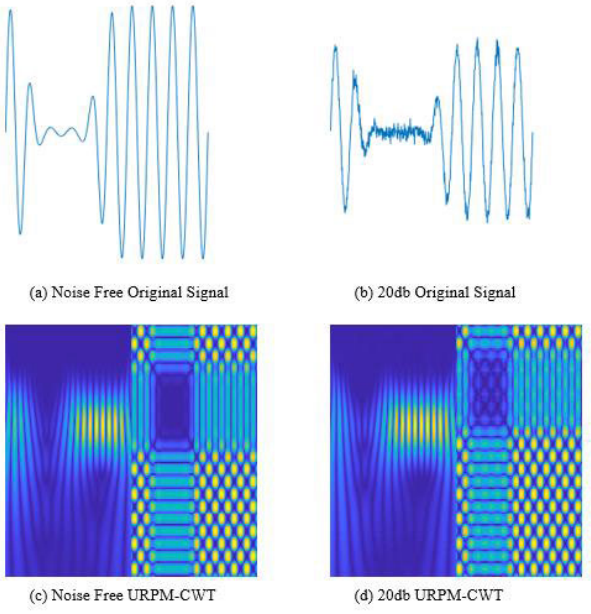

FIGURE 10. Plot of C4 class original signals with URPM-CWT features at different signal-to-noise ratios.

640 points (10 weekly waves). Among them, the initial phase of various disturbances and other indicators are completely randomized.

To emulate signals obtained in practical engineering scenarios, each group of perturbations is subjected to five different noise conditions: no noise, and signal-to-noise ratios of 20dB, 30dB, 40dB, and 50dB, in Figure 10 shows the original signal for C4 class in different signal-to-noise ratio with URPM-CWT feature map. Each class of signal-to-noise ratio generates 1800 samples, totaling 28 × 1800 samples. On average, each model is fed 28 × 600 samples. The dataset is then randomly split into training and test sets in an 8:2 ratio.

The experiments were conducted on a system equipped with an Intel Core i5-12600Kf 3.7GHz CPU, a GeForce RTX 3060Ti GPU, and 32 GB of RAM, running a 64-bit Windows operating system. The software environment utilized was Python 3.10, with the PyTorch 2.0.0 library, and CUDA version 11.7.

In this paper, the models are evaluated using recognition accuracy (Acc), precision of positive class prediction (Pre), sensitivity (Sen), and F1-score performance metrics [29], which are defined:

$$\begin{aligned} \text{Acc}(\%) &= \frac{TP+TN}{TP+TN+FP+FN}\times 100\% \\ \text{Pre}(\%) &= \frac{TP}{TP+FP}\times 100\% \\ \text{Sen}(\%) &= \frac{TP}{TP+FN}\times 100\% \\ F_1\text{−score}(\%) &= 2\times\frac{\text{Pre}\times\text{Sen}}{\text{Pre}+\text{Sen}}\times 100\% \end{aligned} \tag{13}$$

Here, TP (True Positives) refers to the number of correctly identified positive instances; TN (True Negatives) indicates the number of correctly identified negative instances; FP (False Positives) denotes the instances incorrectly identified as positive; and FN (False Negatives) refers to the positive instances that were incorrectly identified as negative.

B. COMPARATIVE ANALYSIS OF DIFFERENT SIGNAL PREPROCESSING METHODS

In Figure 11, the efficacy of the URPM-CWT feature image is substantiated through comparative experiments utilizing four distinct signal preprocessing methodologies within the ambit of an enhanced MobileNetV2 model framework. The optimized URPM preprocessing technique exhibits a notable augmentation in accuracy relative to the RPM method. Noteworthy is the supremacy of the URPM-CWT's recognition accuracy, which prevails as the highest across all examined noise environments, thereby evidencing its superior performance over the competing three methods.

The incisive analysis of these experimental outcomes provides profound insights into the robust nature of the URPM-CWT technique, especially under the duress of noise-afflicted environments, a prevalent complication in signal processing tasks. The resilience of URPM-CWT is ascribed to its advanced noise-resistant features, which adeptly counteract the adverse impacts of acoustic disturbances on signal categorization. Furthermore, the unwavering dominance of URPM-CWT, evident across a noise range extending from 30db to 50db, reinforces its credentials as a dependable preprocessing instrument across a diversity of operational contexts.

These empirical findings corroborate the initial assertion that the integration of Universal Recurrence Plot Mapping with Continuous Wavelet Transform not only encapsulates the inherent complexities within datasets with heightened fidelity but also amplifies the discriminative prowess of the model. Thereby, the URPM-CWT framework is affirmed as a markedly beneficial methodology for the classification of power quality disturbances in microgrids, heralding

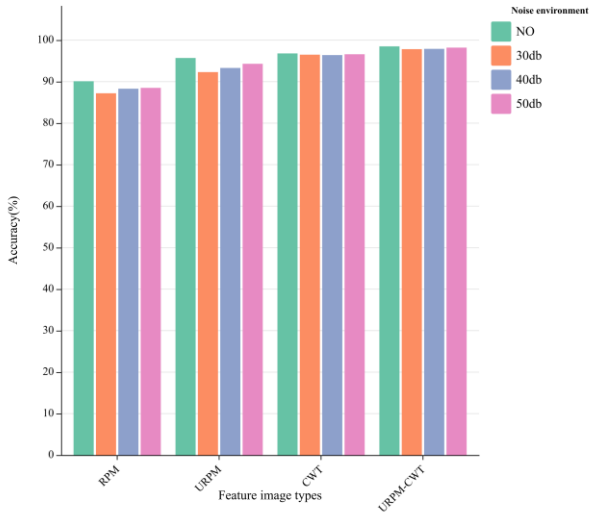

FIGURE 11. Comparison results of four preprocessing methods.

significant enhancements in scenarios where the dual imperatives of noise immunity and precision are paramount.

C. COMPARATIVE ANALYSIS OF DIFFERENT RECOGNITION NETWORK MODELS

1) COMPARATIVE ANALYSIS OF DIFFERENT NETWORKS UNDER THE SAME SIGNAL PREPROCESSING

In this study, we conducted a comparative analysis of five network models – Improved MobileNetV2, Improved ResNet50, Improved ShuffleNetV2, CNN and Proposed - under the same signal preprocessing conditions. The accuracy, precision, recall, and F1 scores of these models were evaluated under five conditions corresponding to different noise levels (20db, 30db, 40db, 50db, and no noise), as shown in Table 5.

In addition to the performance metrics, we also present the training and testing curves for the proposed model under the 30db noise condition in Figure. 12. These curves offer insights into the model's convergence behavior, potential overfitting, and generalization capabilities. Notably, it can be observed that the curves tend to stabilize after the 15th iteration, indicating a plateau in the model's learning progress. By examining these curves, we gain a comprehensive understanding of the model's training dynamics and can make informed decisions on potential improvements or modifications.

The results show that our proposed model performs optimally in all noisy environments, with its accuracy consistently above 99.35%, implying that this model outperforms the other four models in processing signals in these noisy environments. The performance of the conventional CNN in all noise conditions is significantly lower than the other models, which may indicate the lack of robustness of the conventional CNN in dealing with this type of data or noise scenarios compared to these improved and newly proposed models. The accuracy of the current model can be clearly found to be much higher than several other models at 20db noise level, reflecting the best noise robustness of the newly proposed model. As the noise level increases from 20db to

TABLE 5. Comparison of different network models.

Metrics	Improved MobileNetV2	Improved ResNet50	Improved ShuffleNetV2	CNN	**Proposed**
		20db			
Accuracy	0.9286	0.9432	0.9345	0.8632	**0.9765**
Precision	0.9384	0.9421	0.9326	0.8278	**0.9772**
Recall	0.9286	0.9433	0.9378	0.8323	**0.9765**
F1-score	0.9251	0.9423	0.9386	0.8246	**0.98**
		30db			
Accuracy	0.9785	0.9826	0.9751	0.9423	**0.9935**
Precision	0.9779	0.9828	0.9745	0.8522	**0.9936**
Recall	0.9787	0.9826	0.9746	0.8413	**0.9935**
F1-score	0.9783	0.9826	0.9753	0.8445	**0.9934**
		40db			
Accuracy	0.9808	0.9834	0.9763	0.9434	**0.9940**
Precision	0.9808	0.9823	0.9763	0.8525	**0.9942**
Recall	0.9807	0.9831	0.9765	0.8412	**0.9940**
F1-score	0.9806	0.9832	0.9753	0.8455	**0.9944**
		50db			
Accuracy	0.9825	0.9844	0.9801	0.9434	**0.9943**
Precision	0.9815	0.9844	0.9811	0.8545	**0.9941**
Recall	0.9823	0.983	0.9801	0.8482	**0.9943**
F1-score	0.9827	0.9842	0.9803	0.8435	**0.9945**
		Noise-free			
Accuracy	0.9856	0.9845	0.9813	0.9454	**0.9946**
Precision	0.9843	0.985	0.9807	0.8525	**0.9948**
Recall	0.9823	0.9849	0.9823	0.8481	**0.9946**
F1-score	0.9845	0.9852	0.9812	0.8439	**0.9948**

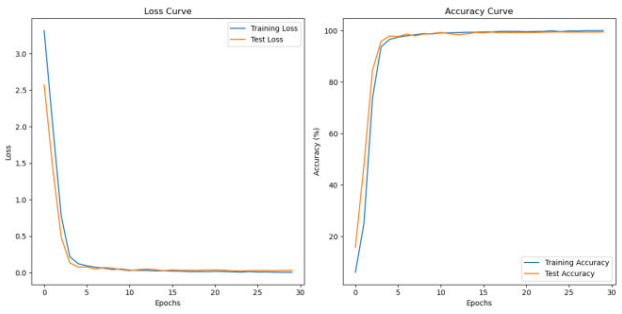

FIGURE 12. Training and testing curves.

TABLE 6. Comparative analysis with other methods.

Methods	Feature Extraction	No. of PQD	Accuracy (%) 20db	30db	40db	50db
FDST+DT[30]	Manual	13	-	97.49	98.8	99.28
DWT+PNN[31]	Manual	16	93.6	95.2	98.6	-
Image Combination +Resnet18[32]	Automatically	14	-	99.48	99.63	-
GCNN+AFEN[33]	Automatically	12	96.62	98.98	99.26	-
MTF+Resnet18 [34]	Automatically	20	-	-	-	97.58
1D VGG[35]	Automatically	17	96.74	98.49	99.36	-
DAE[14]	Automatically	15	97.99	98.69	98.93	-
DL-WMV[36]	Automatically	19	-	98.05	98.58	99.26
URPM-CWT +MCFFN	**Automatically**	**28**	**97.65**	**99.35**	**99.40**	**99.43**

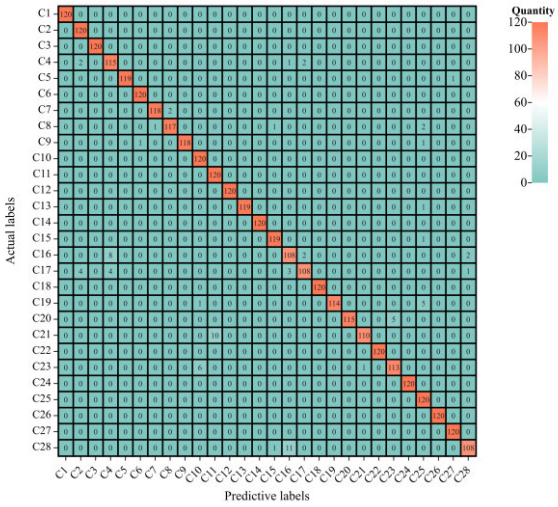

FIGURE 13. 20db confusion matrix.

50db, all four metrics improve for all models, indicating that these models may be more effective in recognizing signals in higher noise environments.

The Confusion Matrix [29], also known as the error matrix, is a specific matrix used to characterize the prediction performance of a model. The confusion matrix, shown in Figure 13, can provide information about the correct as well as incorrect classification of the model predictions. It can be observed that the correct recognition rate of C15, C16, C19, C20 fluctuates under high noise, while some of the perturbation types such as C9, C10, etc. can still maintain 100% recognition rate under 20db noise, which demonstrates the model's good ability to withstand noise as well as robustness, as shown in Figures. 13 and 14.

2) K-FOLD CROSS-VALIDATION

Within the domain of machine learning, the application of K-fold cross-validation [37], is a methodological cornerstone

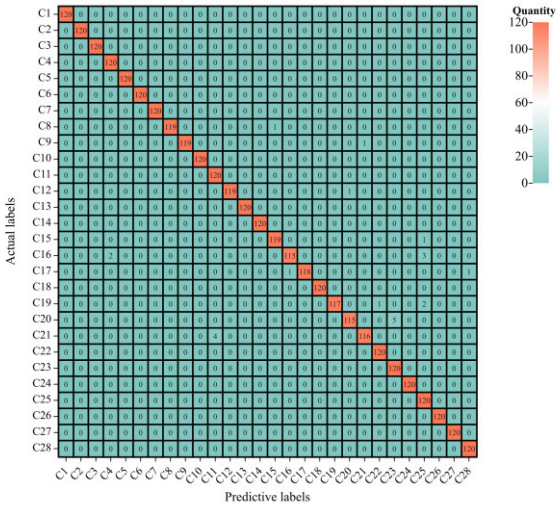

FIGURE 14. 30db confusion matrix.

for validating predictive models. This technique partitions the data into K segments of equal size, designating each in succession as the validation set while the remainder form the training set. Such a strategy not only facilitates a thorough evaluation of model efficacy but also fortifies against the potential for overfitting, thus enhancing model resilience and transferability.

In the expanded experimental discourse of our study, we have adopted an exhaustive K-fold cross-validation schema to scrutinize the robustness and transferability of our proposed model. Given the voluminous nature of our dataset, a quintuple-fold (K=5) structure was chosen to harmonize computational manageability with statistical robustness. The dataset was subjected to a noise level of 30db to closely simulate the intricate and noise-laden environments prevalent in actual power quality analysis.

As illustrated in Figure 15, the fivefold cross-validation yielded remarkably uniform accuracy rates: 99.3%, 99.27%, 99.4%, 99.35%, and 99.38%, with a mean accuracy of 99.34%. Figure 16 further delineates an average loss rate of 0.039 across the folds. This consistency in high performance over varied data subsets not only corroborates the model's efficacy in categorizing power quality disturbances but also its robustness amidst substantial noise levels.

The implications of these results are multifaceted. Primarily, the consistently high accuracy across disparate data partitions signifies an advanced degree of model generalization, vital for real-world applications across heterogeneous and dynamic environments. Additionally, the model's stalwart performance in a 30db noise milieu endorses its inherent resilience to sensory noise, thereby augmenting its suitability for field scenarios where data purity is not assured.

In conclusion, the extended analytical scrutiny presented here validates the proposed methodological framework, endorsing it as a dependable and adaptable instrument for the classification of power quality disturbances. This work also

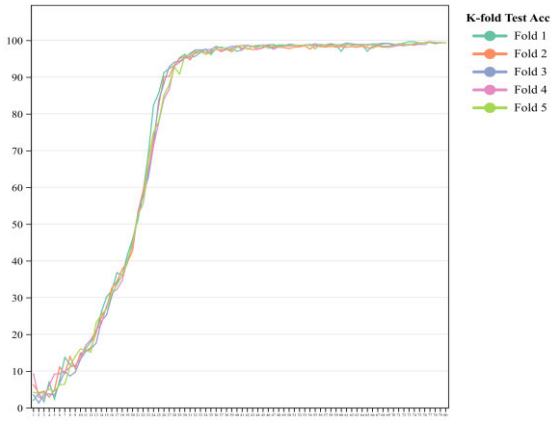

FIGURE 15. K-fold test accuracy.

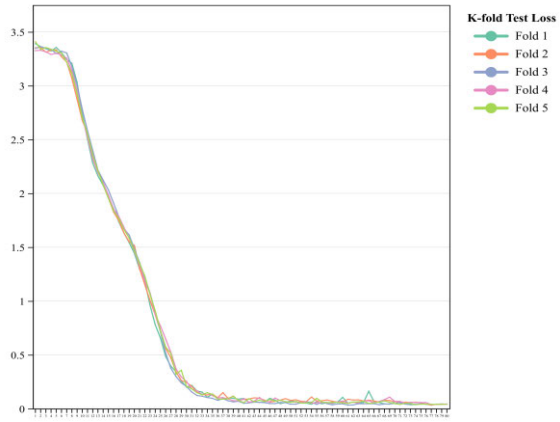

FIGURE 16. K-fold test loss.

lays the groundwork for further inquiry, inviting assessments of model robustness against escalated noise levels or within broader and more varied datasets, thus contributing to the progressive enrichment of power quality analysis research.

3) COMPARATIVE ANALYSIS WITH OTHER METHODS

In Table 3, a rigorous comparative analysis is conducted to assess the Multi-Channel Feature Fusion Network (MCFFN) against existing methodologies in terms of feature extraction capabilities, the diversity of Power Quality Disturbances (PQDs) that can be identified, and overall classification accuracy. The URPM-CWT+MCFFN framework stands out for its capacity to accurately classify an expanded set of 28 PQDs, thereby markedly broadening the scope of detectable disturbances compared to traditional single-label classification schemes.

Delving into the comparative statistics, it is evident that methodologies leveraging deep learning, as cited in [30] and [31], substantially outperform the conventional techniques such as Decision Trees (DT) and Probabilistic Neural Networks (PNN), detailed in [32] and [33]. When benchmarked against state-of-the-art deep learning models referenced in [34] and [36], the MCFFN demonstrates a notable advancement, both in terms of the spectrum of PQDs it can recognize and the fidelity of its classification accuracies.

The integration of Universal Recurrence Plot Mapping (URPM) with Continuous Wavelet Transform (CWT) within the URPM-CWT+MCFFN paradigm is instrumental in achieving this enhanced performance. This synergetic confluence of feature extraction and fusion methods is critical for the nuanced detection of a more comprehensive array of PQDs, exemplifying a significant leap in the domain of power quality analysis.

V. CONCLUSION

This research was conducted to address the problems of incomplete feature presentation in single-feature pictures and a significant reduction in recognition accuracy of single-channel networks in high-noise situations. This paper proposes a microgrid power quality disturbance identification method based on URPM-CWT and multi-channel feature fusion. The following conclusions were obtained through simulation experiments in MATLAB and PyTorch.

1) In terms of image preprocessing, this paper is based on the idea of feature complementarity, URPM-CWT can fully extract the original signal perturbation features, and has a certain anti-jamming ability, which can be better recognized after inputting into the model, and extends the previous image generation method of PQD signal.
2) In terms of the recognition model, the MCFFN model is composed of three improved network models and an adaptive feature fusion module. The whole model has excellent noise robustness. In the simulation validation, the average correct rate of classification for perturbation recognition under 30db noise condition can reach 99.35%.
3) Multi-channel feature fusion can make up for the shortcomings of single-channel recognition, and the performance advantages of each network can be complementary to each other, which can enable the network to maintain a good state of recognition under higher noise conditions.

Automatic feature extraction will be a more important concern in future work and is a prerequisite for the correct recognition of complex disturbed signals. Similarly, the multi-channel feature fusion recognition model will have an increased role.

REFERENCES

[1] T. Chakravorti, R. K. Patnaik, and P. K. Dash, "A morphological filter based disturbance detection and classification technique for DFIG wind farm based microgrid," in *Proc. IEEE Power, Commun. Inf. Technol. Conf. (PCITC)*, Bhubaneswar, India, Oct. 2015, pp. 979–985.

[2] C. Zhao, K. Li, Y. Li, L. Wang, Y. Luo, X. Xu, X. Ding, and Q. Meng, "Novel method based on variational mode decomposition and a random discriminative projection extreme learning machine for multiple power quality disturbance recognition," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2915–2926, May 2019.

[3] C. M. Orallo, I. Carugati, S. Maestri, P. G. Donato, D. Carrica, and M. Benedetti, "Harmonics measurement with a modulated sliding discrete Fourier transform algorithm," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 781–793, Apr. 2014.

[4] S. Shukla, S. Mishra, and B. Singh, "Power quality event classification under noisy conditions using EMD-based de-noising techniques," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1044–1054, May 2014.
[5] J. Li, Z. Teng, Q. Tang, and J. Song, "Detection and classification of power quality disturbances using double resolution S-transform and DAG-SVMs," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 10, pp. 2302–2312, Oct. 2016.
[6] W. Gao and J. Ning, "Wavelet-based disturbance analysis for power system wide-area monitoring," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 121–130, Mar. 2011.
[7] J. Ma, J. Zhang, L. Xiao, K. Chen, and J. Wu, "Classification of power quality disturbances via deep learning," *IETE Tech. Rev.*, vol. 34, no. 4, pp. 408–415, Jul. 2016.
[8] G. Patrizi, C. Iturrino-García, A. Bartolini, F. Ermini, L. Paolucci, L. Ciani, F. Grasso, and M. Catelani, "Improving power quality measurements using deep learning for disturbance classification," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Kuala Lumpur, Malaysia, May 2023, pp. 1–6.
[9] D. Gu, Y. Gao, Y. Li, Y. Zhu, and C. Wu, "A novel label-guided attention method for multilabel classification of multiple power quality disturbances," *IEEE Trans. Ind. Informat.*, vol. 18, no. 7, pp. 4698–4706, Jul. 2022.
[10] R. Machlev, M. Perl, J. Belikov, K. Y. Levy, and Y. Levron, "Measuring explainability and trustworthiness of power quality disturbances classifiers using XAI—Explainable artificial intelligence," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5127–5137, Aug. 2022.
[11] M. Wang, Z. Deng, Y. Zhang, and Z. Zhu, "An automatic identification framework for complex power quality disturbances based on ensemble CNN," *IEEE Access*, vol. 11, pp. 56550–56560, 2023.
[12] H. Zhang, X. Wu, Q. Wang, Y. You, W. Ding, K. Dong, and J. Zhao, "Disturbance identification of power quality based on Markov transition field and deep residual network," in *Proc. 3rd Int. Conf. Energy Eng. Power Syst. (EEPS)*, Dali, China, Jul. 2023, pp. 539–543.
[13] C. Zhou, Z. Shao, F. Chen, and Y. Zhang, "Classification of power quality disturbance signals based on weighting random forest," in *Proc. 7th Asia Conf. Power Electr. Eng. (ACPEE)*, Hangzhou, China, Apr. 2022, pp. 2052–2056.
[14] P. Khetarpal, N. Nagpal, M. S. Al-Numay, P. Siano, Y. Arya, and N. Kassarwani, "Power quality disturbances detection and classification based on deep convolution auto-encoder networks," *IEEE Access*, vol. 11, pp. 46026–46038, 2023.
[15] Y. Liu, D. Yuan, H. Fan, T. Jin, and M. A. Mohamed, "A multidimensional feature-driven ensemble model for accurate classification of complex power quality disturbance," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.
[16] W. Chen and K. Shi, "A deep learning framework for time series classification using relative position matrix and convolutional neural network," *Neurocomputing*, vol. 359, pp. 384–394, Sep. 2019.
[17] H. Khorrami and M. Moavenian, "A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5751–5757, Aug. 2010.
[18] Y. Sun, J. Li, J. Liu, B. Sun, and C. Chow, "An improvement of symbolic aggregate approximation distance measure for time series," *Neurocomputing*, vol. 138, pp. 189–198, Aug. 2014.
[19] S.-C. Pei, M.-H. Yeh, and T.-L. Luo, "Fractional Fourier series expansion for finite signals and dual extension to discrete-time fractional Fourier transform," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2883–2888, Oct. 1999.
[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[21] P. Baldi and P. J. Sadowski, "Understanding dropout," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2013, pp. 2814–2822.
[22] I. Z. Mukti and D. Biswas, "Transfer learning based plant diseases detection using ResNet50," in *Proc. 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Khulna, Bangladesh, Dec. 2019, pp. 1–6.
[23] N. Ma, X. Zhang, and H. T. Zheng, "ShuffleNet v2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 116–131.
[24] Y. Chen, X. Dai, and M. Liu, "Dynamic ReLU," in *Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2020, pp. 351–367.
[25] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *Proc. Int. Workshop Artif. Neural Netw. (IWANN)*, Jun. 1995, pp. 195–201.
[26] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
[27] S. Zhou and A. Huang, "On design of SILU algorithm to enable our new ABPM system for stroke risk early-warning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[28] *IEEE Recommended Practice for Monitoring Electric Power Quality*, Standard IEEE1159-2009, 2009.
[29] J. T. Townsend, "Theoretical analysis of an alphabetic confusion matrix," *Perception Psychophys.*, vol. 9, no. 1, pp. 40–50, Jan. 1971.
[30] M. Biswal and P. K. Dash, "Detection and characterization of multiple power quality disturbances with a fast S-transform and decision tree based classifier," *Digit. Signal Process.*, vol. 23, no. 4, pp. 1071–1083, Jul. 2013.
[31] S. Khokhar, A. A. M. Zin, A. P. Memon, and A. S. Mokhtar, "A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network," *Measurement*, vol. 95, pp. 246–259, Jan. 2017.
[32] Y. Zhang, J. Ou, T. Jin, and G. Bi, "Power quality disturbance recognition method based on feature image combination and modified ResNet-18," *Proc. CSEE*, pp. 1–15, Dec. 2022.
[33] R. Zhang, C. Zhang, H. Gao, and Z. Cheng, "Power quality disturbances classification based on grouping convolutional network with adaptive feature enhanced network," *Proc. CSEE*, pp. 1–10, Jul. 2022.
[34] Y. Luo, K. Li, X. Xiao, Y. Chen, B. Li, and X. Li, "Multi-label classification of power quality composite disturbances based on Markov transfer field and Resnet," *Proc. CSEE*, pp. 1–11, Jul. 2023.
[35] Z. Wang, T. Deng, H. Wang, J. Tao, H. Zhang, and Q. Wang, "Power quality disturbance recognition method in park distribution network based on one-dimensional VGGNet and multi-label classification," in *Proc. 5th Asia Energy Electr. Eng. Symp. (AEEES)*, Chengdu, China, Mar. 2023, pp. 764–770.
[36] G. Bayrak, A. Küçüker, and A. Yılmaz, "Deep learning-based multi-model ensemble method for classification of PQDs in a hydrogen energy-based microgrid using modified weighted majority algorithm," *Int. J. Hydrogen Energy*, vol. 48, no. 18, pp. 6824–6836, Feb. 2023.
[37] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.

JUNZHUO JIANG received the B.S. degree from Chengdu Technological University, Chengdu, China, in 2022. He is currently pursuing the M.S. degree with the Sichuan University of Science and Engineering, Yibin, China. His research interests include signal processing and power quality disturbances.

HAO WU (Member, IEEE) received the B.S. degree in electrical engineering and automation from Southwest Jiaotong University, Chengdu, China, in 2003, the M.S. degree in pattern recognition and intelligent systems from the Sichuan University of Science and Engineering, and the Ph.D. degree in power systems and automation from Southwest Jiaotong University. His research interests include big data, artificial intelligence, deep learning, image processing, modern signal processing and artificial intelligence technology in power systems, transmission (distribution) grid fault diagnosis and fault positioning technology, intelligent distribution grid protection, control and microgrid technology, and conditional intelligent monitoring technology for electrical equipment.

CHANGHUA ZHONG received the B.S. degree from the Binjiang College, Nanjing University of Information Science and Technology, Nanjing, China, in 2021. He is currently pursuing the M.S. degree with the Sichuan University of Science and Engineering, Yibin, China. His research interest includes image processing.

YUAN CAI received the B.S. degree in electrical engineering and automation from the Sichuan University of Science and Engineering, Yibin, China, in 2021, where he is currently pursuing the M.S. degree. His research interests include intelligent information processing and photovoltaic power generation prediction.

HONG SONG is currently a Professor and a Reserve Candidate for Academic Leader in Sichuan. His research interests include big data, artificial intelligence, deep learning, image processing, modern signal processing and artificial intelligence technology in power systems, transmission (distribution) grid fault diagnosis and fault positioning technology, intelligent distribution grid protection, control and microgrid technology, and conditional intelligent monitoring technology for electrical equipment.