

RESEARCH ARTICLE

Cable Incipient Fault Identification Method Using Power Disturbance Waveform Feature Learning

HONG LU, WEN-HAI ZHANG¹, (Member, IEEE), YING WANG, (Senior Member, IEEE), AND XIAN-YONG XIAO, (Senior Member, IEEE)

College of Electrical Engineering, Sichuan University, Chengdu 610065, China

Corresponding author: Wen-Hai Zhang (zhangwh1989@163.com)

This work was supported by the National Key R&D Program of China under Grant 2020YFF0305800.

ABSTRACT Cable incipient fault (CIF) is a potential fault which may recur over time and evolve into permanent fault eventually. The identification of CIF can help to reduce the likelihood of a permanent fault and enhance power supply reliability. This paper fully considers the randomness and uncertainty of CIF waveforms, and proposes a CIF identification method using power disturbance waveform feature learning. Firstly, the shallow features are extracted to characterize the transient components of different disturbance current waveforms, by conducting a stationary wavelet transform. Then, a dropout deep belief network (DDBN) is constructed using the extracted shallow features by pre-training and fine-tuning. Finally, the well-constructed DDBN model is used to identify CIF from other similar disturbance events. The performance of the proposed method is verified by simulation data and experimental data, for different disturbance events, such as sub-cycle cable incipient fault and multi-cycle cable incipient fault, as well as other over-current disturbance events such as capacitor switching and inrush current. Both the accuracy and generalization ability are higher than other methods. The proposed method provides new insights on possible applications on monitoring of cable status and timely warning of cable faults.

INDEX TERMS Cable incipient fault (CIF), fault identification, power disturbance, dropout deep belief network (DDBN), deep features.

I. INTRODUCTION

Cable free of faults is a precondition for uninterrupted power supply. Most cable faults can be divided into four stages, as shown in Fig. 1 [1], [2]. A cable is healthy in Stage 1. With the increase of in-service time and the accumulation of defects, water vapor will penetrate into the cable joints due to insufficient interface pressure and other reasons, forming tiny air gaps at the interface. Then, the local field strength of the cable will be distorted, causing partial discharge (PD). Due to the increase of PD activities, intermittent arc discharge will occur in the cable. Here we define cable incipient fault (CIF) as a fault that is formed by a series of intermittent arc discharge before a permanent fault [3], [4]. In Stage3, the cable insulation has not been completely deterioration, and the gas generated during the carbonization of the insulating

material has the effect of deionization, causing the fault to extinguish automatically [5]. The CIF may occur repeatedly until the insulation breaks down completely and evolves into a permanent fault, as shown the Stage 4. In order to ensure the reliability and stability of power supply, current research on cable fault includes both post-fault (Stage 4) and pre-fault (Stage 1-3) stages. For Stage 4, researchers have mainly conducted research on the location and types of permanent faults [6], [7]. However, for Stage 1-3, the main focus has been on how to alert of a potential fault to prevent permanent fault [8], [9].

PD is a high-frequency pulse signal with a frequency of multiple MHz, which is caused by excessive concentration of local electric fields inside or on the surface of cable insulation, and can be used to prevent cable fault by detecting PD signal [9]. However, it is difficult to accurately collect PD signal for it is weak and the discharge modes are diversified. In order to detect PD signal accurately, special sensors, such

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu¹.

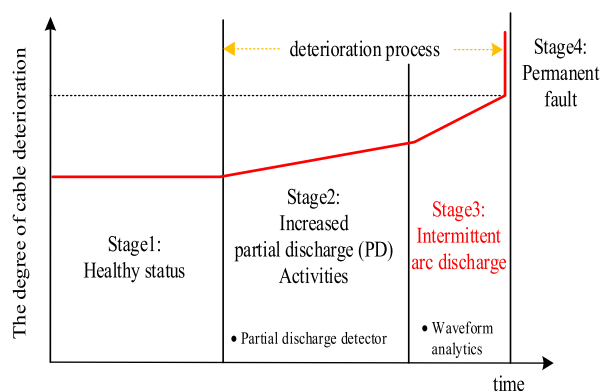


FIGURE 1. Deterioration process of cable.

as high frequency current transformer and capacitor coupler, are needed. These devices are expensive, and the detection accuracy is also easily affected by field conditions, which its widespread use, so online detection of partial discharge signal is still in research stage.

Incipient faults are often called “intermittent faults”, and if they cause short circuit arc, they are also called “arc faults”. Compared with the special equipment for measuring PD signals, the traditional measurement equipment, such as potential transformer (PT) and current transformer (CT), which can measure waveforms during the fault. PT and CT are widely installed at substations. It is economical in data acquisition, and can also support the research in the field of equipment fault diagnosis using the obtained waveform data [3]. Therefore, the research of CIF identification method using the measured waveforms is a good choice for cable fault anticipation.

The fault features of CIF are weak, which can be regarded as a weak signal. Some scholars try to enhance the weak features, and then use the enhanced features for weak signal detection and identification [10], [11]. However, the waveforms of some overcurrent disturbances are similar to that of CIF, which makes it difficult to identify CIF through the enhanced features.

For CIF identification in medium voltage distribution network with a voltage level of 10kV, researchers have done related researches. The published methods are mainly classified into two types: fault features-based and model-based. Fault features-based methods extract over-current features which are directly compared with preset criteria. Reference [3] adopted wavelet transform to decompose current waveforms, and fault duration and magnitude are used as characteristics to identify CIF. In [12], an identification method considering the change of traveling wave polarity during fault was proposed.

Model-based methods include explicitly analytical (based on physics features) and implicit analytical (based on waveform data) methods. Explicitly analytical methods identify CIF through established analytical model. Reference [13] used Kalman filter to calculate residual signals of voltage and current for identifying CIF. Furthermore, considering

the feature of fault voltage waveform distortion, an identification method was proposed by using voltage distortion rate [14].

A large number of monitoring devices are installed in the distribution network, and it is possible to collect power disturbance waveform, which can be used to monitor operation status of various equipment [15]. In this case, the machine learning based CIF identification approaches are proposed. Reference [16] proposed a method based on small sample learning to identify incipient fault, but the method cannot fully extract the features from the fault data, and the identification performance may be affected when the data set changes. Reference [17] proposed a CIF based on cumulative SUM and adaptive linear neuron (ALN), where the cumulative SUM was used to detect the transients in disturbance current waveforms, and then ALN can identify the CIF from the disturbance waveforms. Reference [18] used Kullback-Leibler divergence and S-transform to extract features, and then proposed a support vector machines (SVM)-based model to identify CIF. Reference [19] extracted the features of CIF by performing wavelet transform on the fault current waveforms, and then constructed the CIF identification model by using extreme learning machine (ELM). However, ALN, SVM and ELM are shallow architecture methods, which make it difficult to accurately learn the non-linear relationships from different disturbance waveform and may cause false identification. Compared with traditional machine learning methods, deep learning can perform adaptive learning on input data by simulating the learning process of the brain, and obtain deep features that can reflect the type of data [20], [21]. Reference [22] proposed a CIF identification method based on stacked auto-encoder and S-transform, but the diversity of field data will result in poor generalization ability of the method.

For the CIF identification problem with strong randomness, weak fault characteristics and multi-overcurrent disturbance types, the key is to accurately extract fault features. Therefore, this paper proposes a CIF identification method using power disturbance waveform feature learning, which represents the process of obtaining the deep features between power disturbance features and disturbance types using the deep learning method. Firstly, Stationary wavelet transform (SWT) is used to extract the shallow features for characterizing the local discontinuity in disturbance waveforms and improving the identification performance. Then, dropout deep belief network (DDBN) model is constructed by stacking multiple dropout constrained Boltzmann machines (DRBM), and the deep features between shallow features and disturbance types can be further learned by pre-training and fine-tuning DDBN model parameters. Finally, the CIF and other overcurrent disturbances can be identified by the deep features learned from DDBN model.

The main contributions of this paper are summarized as follows:

- The proposed method uses SWT to extract the shallow features from the disturbance waveforms, which can

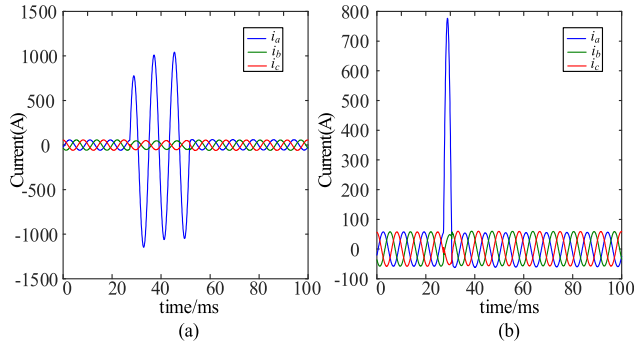


FIGURE 2. Current waveform of CIF. (a)MCIF, (b)SCIF.

characterize the local discontinuity in disturbance waveforms and improve the identification efficiency of the CIF identification model.

- This paper proposes a power disturbance waveform feature learning method using DDBN. Effective features can be learned by pre-training and fine-tuning the DDBN model even if the disturbance waveform is changed, which improves the generalization ability of the model.

- The proposed method constructs a CIF identification model using DDBN. The CIF can be identified through the deep features. It is beneficial to prevent overfitting and solve the problem that the identification accuracy decreases due to waveform uncertainty.

The remainder of this paper is structured as follows. Section II presents the method of extracting shallow features. Section III describes the network structure of DDBN. Section IV proposes a CIF identification method. Section V analyzes the validation results using simulation data recorded in PSCAD/EMTDC. Section VI introduces the test platform built in the laboratory, and the performance of the proposed method is verified using experimental data. Finally, the conclusions are drawn in Section VII.

II. SHALLOW FEATURES EXTRACTION

An intermittent arc forms between the core and sheath of a cable during CIF period, resulting in an increase in the fault phase current. Compared with voltage waveform, the current waveform contains much more information related to cable operating state. Thus, the current waveform is utilized for constructing CIF identification model. Due to short duration of the fault, CIF will be automatically cleared before the relay protection device operates. CIF can be divided into sub-cycle cable incipient fault (SCIF) and multi-cycle cable incipient fault (MCIF). The duration of SCIF and MCIF are about 1/4 cycle and 1-4 cycles, respectively [3].

Assuming CIF occurs in phase A, the fault phase current waveforms of MCIF and SCIF are shown in Fig. 2. The current waveform of CIF contains healthy and fault periods. When the current waveform is used directly as the input of DDBN, the learning efficiency is low. Thus, it is necessary to accurately extract fault waveform features.

In this paper, we define the extracted features as shallow features, which indicates that the mapping relationship between the input data and the disturbance type is not fully mined. At the same time, it is also for comparison with the features automatically extracted by the subsequent deep learning model. Through feature extraction and modeling of the current waveform, MCIF and SCIF can be identified from a variety of over-current disturbances accurately.

Reference [23] analyzes the field data of CIF: “In the fault period, the fundamental frequency, the 2nd and 3rd harmonic are the most important components.” Therefore, the fundamental frequency, 2nd and 3rd harmonic component are extracted, and the features are named f_1 , f_2 , and f_3 , respectively. However, it is difficult to distinguish CIF from other disturbances accurately using physical features only. There are many methods prove that statistical features can be extracted to characterize local discontinuity in non-stationary waveforms [24], [25]. Discrete Wavelet Transform (DWT) is widely used to extract time and frequency domain features of power disturbance waveforms. However, DWT will down-sample the data when it decomposes the waveform layer by layer. Down-sample means that in the process of decomposing the waveform, the coefficients obtained need to be sampled at intervals, which causes the length of the obtained coefficients to be inconsistent with the length of the original data. Although high-frequency information of the original data is retained, when the feature is weak, it is easy to cause loss of the disturbance feature information.

Compared with DWT, SWT has good time-frequency characteristics and does not perform down-sample in the process of decomposing the waveform. Thus, SWT is proposed to analyze the disturbance waveforms and extract the statistical features [26]. The statistical features can be extracted from the low-frequency coefficient and high-frequency coefficient obtained by using SWT. The extracted features include [27], [29]:

Mean (μ), variance (Std), root mean square (RMS), crest factor (CF), skewness (SK), kurtosis (KT), and shape factor (SF), information entropy (IE), energy entropy (EE) and the maximum of singular value (λ_m).

$$\mu_i = \frac{1}{J} \sum_{j=1}^J \{c_{ij}, d_{ij}\} \tag{1}$$

$$Std_i = \sqrt{\frac{1}{J-1} \sum_{j=1}^J \{(c_{ij} - \mu_i)^2, (d_{ij} - \mu_i)^2\}} \tag{2}$$

$$RMS_i = \sqrt{\frac{1}{J} \sum_{j=1}^J \{(c_{ij})^2, (d_{ij})^2\}} \tag{3}$$

$$CF_i = \left\{ \frac{\text{Max}(c_{ij})}{RMS_i}, \frac{\text{Max}(d_{ij})}{RMS_i} \right\} \tag{4}$$

$$SK_i = \left\{ \frac{E(c_{ij} - \mu_i)^3}{Std_i^3}, \frac{E(d_{ij} - \mu_i)^3}{Std_i^3} \right\} \tag{5}$$

$$KT_i = \left\{ \frac{E(c_{ij} - \mu_i)^4}{Std_i^4}, \frac{E(d_{ij} - \mu_i)^4}{Std_i^4} \right\} \quad (6)$$

$$SF_i = \left\{ \frac{RMS_i}{\mu_i} \right\} \quad (7)$$

$$IE_i = - \sum_{j=1}^J \left\{ c_{ij}^2 \log_2(c_{ij})^2, d_{ij}^2 \log_2(d_{ij})^2 \right\} \quad (8)$$

$$En_i = \sum_{j=1}^J \left\{ (c_{ij})^2, (d_{ij})^2 \right\}, EE_i = -En_i \log_2(En_i/En) \quad (9)$$

$$T = U \Lambda V, \lambda m_i = \max \{ \Lambda \} \quad (10)$$

where i and j represent the number of decomposition layers and the degree of dispersion, respectively. c_{ij} is the low-frequency coefficient and d_{ij} is the high-frequency coefficient. T represents c_{ij} and d_{ij} , Λ represents a matrix consisting of singular values λ , U and V represent orthogonal matrices obtained by singular value decomposition, respectively.

It has been proved that ‘daub4’ mother wavelet is more sensitive to non-stationary signals [30]. Thus the ‘daub4’ mother wavelet is selected to decompose the input disturbance waveforms layer by layer. Here, the number of layers is not a parameter of disturbance waveforms, but a parameter of the SWT algorithm.

The principle of determining the number of decomposition layers is to keep the basic frequency of the power grid in the sub-band. The basic frequency of the power grid in this paper is 50 Hz, and the sampling frequency is 10 kHz. When the number of decomposition layers is 8, the fundamental frequency can be located in the last high frequency sub-band. Therefore, the total number of layers selected for decomposition in this paper is 8.

Finally, the extracted features including physical and statistical features are as follows:

$$\begin{aligned} F_{c1} &= [f_1, f_2, f_3] \\ F_{t1} &= [\mu_{d1}, \dots, \mu_{d8}, \mu_{c8}] \\ F_{t2} &= [Std_{d1}, \dots, Std_{d8}, Std_{c8}] \\ F_{t3} &= [RMS_{d1}, \dots, RMS_{d8}, RMS_{c8}] \\ F_{t4} &= [CF_{d1}, \dots, CF_{d8}, CF_{c8}] \\ F_{t5} &= [SK_{d1}, \dots, SK_{d8}, SK_{c8}] \\ F_{t6} &= [KT_{d1}, \dots, KT_{d8}, KT_{c8}] \\ F_{t7} &= [SF_{d1}, \dots, SF_{d8}, SF_{c8}] \\ F_{t8} &= [IE_{d1}, \dots, IE_{d8}, IE_{c8}] \\ F_{t9} &= [EE_{d1}, \dots, EE_{d8}, EE_{c8}] \\ F_{t10} &= [\lambda m_{d1}, \dots, \lambda m_{d8}, \lambda m_{c8}] \end{aligned} \quad (11)$$

where, c_8 is the 8-th layer low-frequency coefficient, and d_i ($i = 1, 2, \dots, 8$) is the i -layer high-frequency coefficient.

The input data of DDBN should be normalized between 0 and 1, which is conducive to updating the network parameters. In this paper, the feature vectors are normalized using the

min-max normalization method. Taking the normalization of F_{t1} as an example [31], [32]:

$$F_{t1}' = \frac{F_{t1} - F_{t1_min}}{F_{t1_max} - F_{t1_min}} \quad (12)$$

where F_{t1}' is the normalized data, F_{t1_min} is minimum and F_{t1_max} is maximum value in F_{t1} .

Then, the normalized data F ($F \in R^{1 \times 93}$) can be obtained, which is used as the input of DDBN.

$$F = [F'_{c1}, F'_{t1}, \dots, F'_{t10}] \quad (13)$$

III. DROPOUT DEEP BELIEF NETWORK

In this section, the extracted features are used as the input data of DDBN, where the value of each feature represents a neuron, so the input layer of the DDBN model contains 93 neurons. DDBN performs adaptive learning on the extracted features for obtaining the deep features.

Restricted boltzmann machine (RBM) is the basic model of standard deep belief network (SDBN). RBM has two layers of networks: a random visible layer unit $\mathbf{v} = \{v_i\}_{i=1}^n$ and a random hidden layer unit $\mathbf{h} = \{h_j\}_{j=1}^m$. DRBM is a variant of RBM with a vector of binary variables $\mathbf{r} = (r_j)_{j=1}^m$. If $r_j = 1$, the hidden layer unit is retained, otherwise the neuron is discarded, reducing the interaction of hidden layer units [33].

DRBM is an energy-based model. If the elements in \mathbf{v} and \mathbf{h} are known, the energy function of the RBM can be expressed as follows:

$$\begin{aligned} E(\mathbf{v}, \mathbf{h} | \mathbf{r}, \theta) &= - \sum_{i=1}^n b_i v_i - \sum_{j=1}^m c_j (r_j h_j) \\ &\quad - \sum_{i=1}^n \sum_{j=1}^m v_i W_{ij} (r_j h_j) \end{aligned} \quad (14)$$

where, $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{c}\}$ is the network parameter of the RBM, \mathbf{W} is the connection weight matrix between the visible layer and the hidden layer. \mathbf{b} and \mathbf{c} represent the bias matrix of the visible layer and the hidden layer. W_{ij} is the element of \mathbf{W} , which represents the connection weight between the i th visible layer neuron and the j th hidden layer neuron, b_i and c_j are the element of \mathbf{b} and \mathbf{c} , which represents the bias of i th visible layer neuron and the j th hidden layer neuron, respectively.

The change of neuron state in DRBM can be represented by probability, the joint probability distribution of can be obtained:

$$P(\mathbf{v}, \mathbf{h} | \theta) = \frac{1}{Z(\theta)} \exp(-E(\mathbf{v}, \mathbf{h} | \theta)) \quad (15)$$

$$Z(\theta) = \sum_{\mathbf{v}, \mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h} | \theta)) \quad (16)$$

where, $Z(\theta)$ represents the normalization function.

Since the neurons in each layer of DRBM are independent of each other, when the state of the neurons in the visible layer is given, the activation probability of the neurons in the

hidden layer can be calculated as:

$$P(h_j = 1 | \mathbf{v}, \theta) = r_j f \left(\sum_{i=1}^n W_{ij} v_i + c_j \right) \quad (17)$$

where, f represents the sigmoid activation function, $f(x) = 1 / (1 + \exp(-x))$, $h_j = 1$ represents the j th neuron of the hidden layer is activated.

Similarly, when the state of the neurons in the hidden layer is given, the activation probability of the neurons in the visible layer can be calculated as:

$$P(v_i = 1 | \mathbf{h}, \theta) = f \left(\sum_{j=1}^m W_{ij} (r_j h_j) + b_i \right) \quad (18)$$

When a set of N training samples is given, by adjusting the parameters of the DRBM, the DRBM can fit the training sample set to the greatest extent, which is mainly achieved by maximizing the log-likelihood function:

$$\hat{\theta} = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \sum_{s=1}^N \log P(\mathbf{v}^s | \mathbf{h}, \theta) \quad (19)$$

In order to train RBM and obtain RBM parameters, Hinton proposed a method using contrastive divergence, which can be completed with one Gibbs sampling [34], [35]. The calculation steps of W_{ij} in RBM parameters can be obtained as follows, and the calculation formulas of other parameters b_i and c_j similar.

1) Set the initial values of RBM parameters θ . The element W_{ij} in \mathbf{W} can be initialized as a random number in the normal distribution $N(0, 0.01)$, $\Delta W_{ij} = 0$. Set the learning rate and number of iterations.

2) Initialize the training sample as the neuron state in the visible layer \mathbf{v}^0 , and calculate the state of the hidden layer \mathbf{h}^0 by eq. (17).

3) Reconstruct the visible layer by eq. (18).

4) After 1 Gibbs sampling, \mathbf{v}^1 is obtained to update the W_{ij} , and the maximum likelihood value is calculated by the gradient ascent algorithm:

$$\Delta W_{ij} = \frac{\partial \log P(\mathbf{v})}{\partial W_{ij}} = P(h_i = 1 | \mathbf{v}^0) v_j^0 - P(h_i = 1 | \mathbf{v}^1) v_j^1 \quad (20)$$

5) The expression for updating the W_{ij} , is as follows:

$$W_{ij} = \alpha W_{ij} + \varepsilon \Delta W_{ij} \quad (21)$$

where, ε represents the learning rate. α represents the momentum, which can be used to prevent the algorithm from converging to a local optimum.

6) The learning rate is usually constant in the process of updating parameters, resulting in the same update rate in the early and late stages of the iteration, which is not conducive to the convergence of the algorithm. Therefore, this paper adopts an adaptive learning rate based on natural exponential decay,

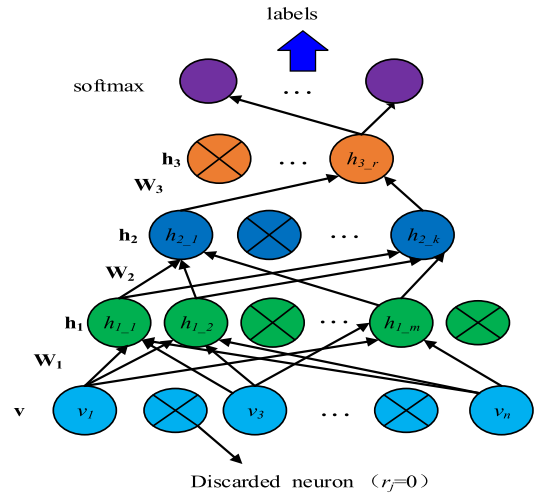


FIGURE 3. Structure of DDBN.

which helps to find the optimal solution. The expression for updating the W_{ij} becomes:

$$W_{ij} = \alpha W_{ij} + \varepsilon^{(l-1)} e^{\chi} \Delta W_{ij} \quad (22)$$

where, l represents the iteration number. χ represents the decay rate. Through the above steps, the elements of the weight W_{ij} can be calculated.

Single DRBM has limited ability to extract features of the input data, multiple DRBMs are stacked to form dropout deep belief network (DDBN), which can prevent over-fitting, reduce the complex interaction in neurons and improve generalization ability. The structure diagram of DDBN model consists some DRBMs and a softmax classifier, as shown in Fig. 3.

The input layer of DDBN is the visible layer of the first DRBM, and then the activation probability of the lower DRBM is used as the input to the next DRBM, by analogy, the softmax classifier is the last layer of DDBN. Softmax classifier is a logistic regression model that can map the neurons in the output layer of the DDBN model to (0,1), and finally the mapping results are used to identify different disturbance types.

The parameter of dropout_Fraction represents the proportion of neurons that are randomly discarded in layer-by-layer training to total neurons. After adopting the parameter of dropout_Fraction, DDBN can randomly delete hidden neurons in the neural network and obtain a network model containing different activated neurons. When training each network model, the update of network parameters can be made independent of the interaction between hidden neurons, and the network can learn more robust features. Therefore, effective features can still be learned through the model even if the CIF waveform data is changed, which improves the generalization ability of the model.

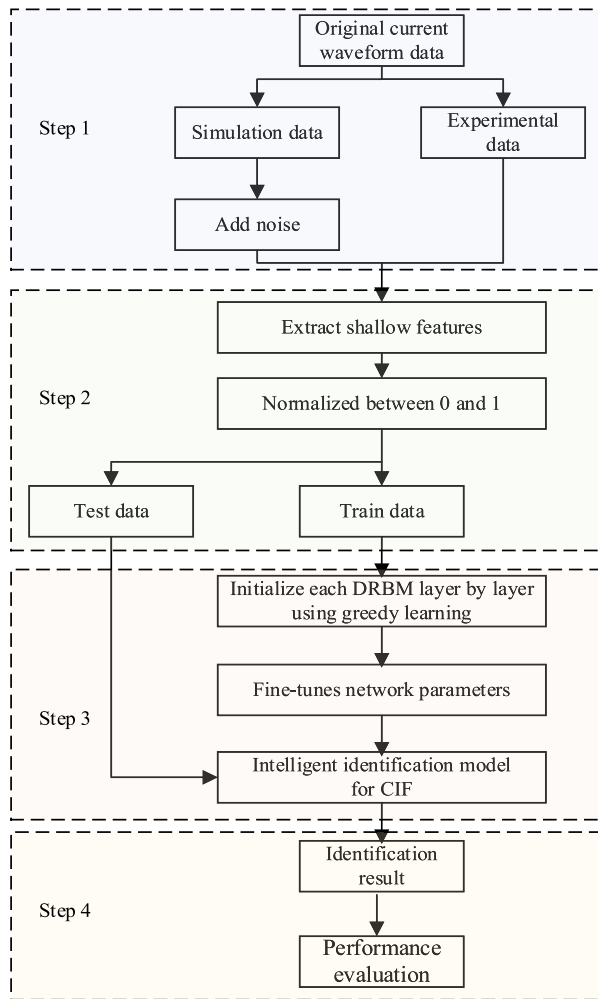


FIGURE 4. Flowchart of CIF identification method.

IV. THE PROPOSED CIF IDENTIFICATION METHOD

The flowchart of the proposed CIF identification method is shown in Fig. 4. The proposed method consists of four steps. The first step is to generate data sets. The second step is to extract shallow features from the data sets. The third step is to construct the intelligent identification model for CIF. The final step is to evaluate the performance of the proposed method. Detailed steps are as follows:

A. GENERATION OF DATA SETS

The simulation model is built in PSCAD/EMTDC based on analyzing the features of various over-current disturbances. Then, the current waveform of CIF (SCIF and MCIF), other over-current disturbances (constant impedance fault (Imp), capacitor switching (Cap) and inrush current (Inr)) and healthy condition (Hc) can be recorded. Since field data contains noise, white Gaussian noise of a certain signal-to-noise ratio (SNR) is added to the simulation data to improve feasibility.

In order to further verify the generalization ability of the proposed method to different data, a CIF test platform is built in the laboratory to measure experimental data.

B. EXTRACTION OF SHALLOW FEATURES

The fundamental frequency, 2nd and 3rd harmonic component are extracted by FFT. And SWT is used to decompose the data for extracting statistical features. As presented in Section III, the ‘daub4’ mother wavelet is used to decompose the data layer by layer, and statistical features are extracted from the obtained coefficients of each layer. The whole set of features are normalized between 0 and 1 by (12) to obtain \mathbf{F} , which is used as the input of DDBN.

Determine the ratio of training and testing sets is 3:1. Three-quarters of the data sets for each event type are used to train DDBN network parameters to obtain an intelligent model for identifying CIF from other disturbance events. The remaining samples are used to test the performance of the constructed model.

C. CONSTRUCTION OF INTELLIGENT IDENTIFICATION MODEL FOR CIF

DDBN adopts two steps to train and optimize model parameters according to the characteristics of the input data, including forward pre-training and back fine-tuning. In the process of pre-training, these discarded neurons are excluded from the network temporarily so its parameters are not updated when the neurons are out of the network.

In this paper, a suitable cross-validation method is proposed to select the DDBN model parameters with the optimal loss function evaluation results.

D. PERFORMANCE EVALUATION

After the construction of the DDBN model is completed, it represents that the model can be used to identify CIF. The model is used to identify CIF from the data set containing multiple disturbance types, and the performance is evaluated by corresponding evaluation indicators, such as Accuracy, Precision, Recall, etc. The follow-up evaluation indicators in this paper are the average values obtained by 10 tests using the proposed cross-validation method.

V. SIMULATION DATA VERIFICATION

The simulation model is built in PSCAD/EMTDC, as shown in Fig.5. Monitoring points are installed at the outlet of each feeder to record the current data of different feeders. And then MATLAB is used to extract the features of the obtained data and construct the DDBN model. The proposed method is executed on a personal computer configured with a Windows 7 64-bit operating system and an Intel (R) Core (TM) i7 CPU with 4 GB of RAM.

The parameters of each disturbance type are shown in Table 1. Different disturbance data can be recorded by connecting different disturbance modules at different feeders and positions, and the connection of disturbance modules is simulated by closing the circuit breaker [19]. In order to record a large number of disturbance waveforms, the multi-run module in PSCAD/EMTDC model is used to traverse the parameter range. and each waveform obtained corresponds

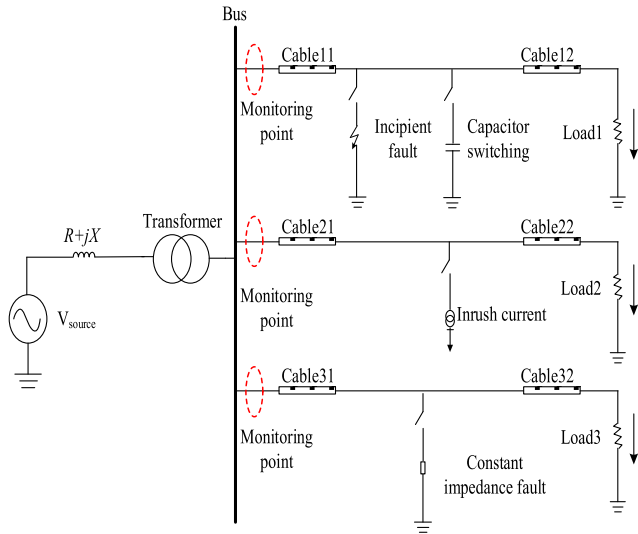


FIGURE 5. Simulation model in PSCAD/EMTDC.

TABLE 1. Parameters of different disturbance types in pscad/emtdc.

Method	Description	Value
Sbase	system capacity	100MVA
Vsource	System parameters	voltage level 110kV
R+jX	system impedance	0.35 + j0.68 pu
Fault location	Fault parameters	distance between the fault point and the monitoring point 500~20000m (Increment every 500m)
τ	arc time constant	0.2~0.4ms (Increment every 0.2ms)
u0	Arc parameters	arc characteristic voltage 300~1500V (Increment every 200V)
r0	arc characteristic resistance	0.001~0.013 Ω (Increment every 0.003 Ω)
C	Capacitor parameters	capacitor value 1~7000mF (Increment every 100mF)
S	Transformer parameters	transformer capacity 1~7000kV·A (Increment every 100kV·A)
Z	Impedance parameters	impedance value 1~7000 Ω (Increment every 100 Ω)

to a set of parameter values. Therefore, the disturbance waveforms are generated whose parameter values have the corresponding ranges, rather than obtaining different waveforms by selecting a group of parameters in Table 8. In this paper, 16800 samples are recorded through the simulation model, that is, 2800 samples for each event. Then the 93 shallow features are extracted from each sample.

TABLE 2. Identification result of different data sets.

Group number	Training data	Testing data	P_{Acc} (%)	P_{Pre} (%)	P_{Rec} (%)	F_1 (%)
1	R_1-R_{21}	$R_{22}-R_{28}$	99.5	100	99	99.75
2	R_2-R_{22}	$R_{15}, R_{23}-R_{28}$	98.9	98.3	98	98.15
3	R_6-R_{26}	$R_1-R_{55}, R_{27}-R_{28}$	99.2	99.1	99	99.05
4	R_7-R_{27}	R_1-R_{66}, R_{28}	98.7	98.4	97	97.7
5	R_8-R_{28}	R_1-R_7	98.3	97.9	98	97.95

A. IDENTIFICATION PERFORMANCE EVALUATION INDICATORS

The target of the proposed method is to identify CIF based on the CIF identification model from the data sets containing CIF, different types of disturbances (capacitor switching, constant impedance fault, inrush current) and health conditions. Therefore, this is a multi-classification problem. The detailed results can be shown by confusion matrix, whose horizontal axis represents identification type and vertical axis represents actual type. Some evaluation indicators can be obtained from confusion matrix, including P_{Acc} (Accuracy), P_{Pre} (Precision), P_{Rec} (Recall) and F_1 (F_1 -Score) [36].

$$\begin{aligned}
 P_{Acc} &= \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100\% \\
 P_{Pre} &= \frac{T_p}{T_p + F_p} \times 100\% \\
 P_{Rec} &= \frac{T_p}{T_p + F_n} \times 100\% \\
 F_1 &= \frac{2T_p}{2T_p + F_p + F_n} \times 100\% \quad (23)
 \end{aligned}$$

Since the goal of this paper is to identify CIF from a variety of disturbances, the CIF samples are taken as positive (T) and the non-CIF samples are negative (F), then T_p means actual type is T and identification type is T. F_p means actual type is F and identification type is T. F_n means that actual type is T and the identification type is F. T_n means actual type is F and identification type is F.

B. THE IDENTIFICATION RESULTS OF DIFFERENT DATA SETS

The model based on the data-driven method depends on training of input data. In order to test the generalization ability of the proposed model, different data sets need to be used to test the identification performance.

Suppose that \mathbf{R} represents a sample matrix of a disturbance type, whose row represents a sample, and column represents the features extracted from Section II. The number of rows and columns of \mathbf{R} are 2800 and 93 respectively. 100 rows in \mathbf{R} are randomly selected to form the new matrix \mathbf{R}_i ($i = 1, 2, \dots, 28$).

Table 2 lists the test results of five different data sets, and the values of these indicators are average values for

TABLE 3. Identification accuracy using different levels of SNR.

Waveform Type	P_{Acc} for different levels of SNR (%)		
	40dB	30dB	20dB
SCIF	99	95	91
MCIF	98	96	93
Imp	100	96	94
Cap	99	92	90
Inr	100	97	93
Hc	100	98	97
Average	99.3	95.67	93

identification of SCIF and MCIF. The results show that the proposed method can learn the hidden features from the data after the input data set is changed, and improves the identification performance and generalization ability of the model.

C. THE COMPARISON RESULTS WITH DIFFERENT LEVELS OF SNR

In order to test the noise immunity of the proposed method, white Gaussian noise with different levels of Signal-to-noise ratio (SNR) is added to the simulation data. Then the data after feature extraction is used as the input of the DDBN model. The identification accuracy for different types is given in Table 3.

It can be concluded from Table 3 that the proposed method obtains an average identification accuracy of 99.3% when SNR is 40dB. The average identification accuracy decreases when SNR is 30dB, but it still exceeds 95%. In a 20dB noise environment, the identification accuracy of SCIF, Cap and Inr decreases significantly. The main reason is that the first two disturbance types have a short duration. It is difficult for DDBN to accurately obtain useful information under strong noise environment. The waveform of Inr contains a large number of harmonic components, resulting in overlap of noise and disturbance features, which reduces the identification accuracy. But the average identification accuracy can still achieve 93% when the level of SNR is 20dB, which fully reflects good noise immunity of the proposed method.

Due to the weak disturbance features in MCIF and Cap, if the power of the added white Gaussian noise is too large, the disturbance features in the original waveform may be buried in the noise. As a result, it may not be able to simulate the field data well. In this paper, we select noisy data with 40dB as the basic data for subsequent analysis to construct a model for intelligently identifying CIF after analyzing the noise content in the monitoring waveform and the identification performance.

D. THE COMPARISON RESULTS WITH OTHER METHODS

This part first compares the proposed method with other pattern identification methods to evaluate the performance

TABLE 4. Comparison result with different pattern identification methods.

Method	Type	P_{Acc} (%)	P_{Pre} (%)	P_{Rec} (%)	F_1 (%)
DDBN	SCIF	99	100	99	99.5
	MCIF	100	100	100	100
	P_{Ave}	99.5	100	99.5	99.75
BPNN	SCIF	92	94.4	92	88.04
	MCIF	71	83.53	71	76.76
	P_{Ave}	81.5	88.97	81.5	82.4
SDBN	SCIF	81	100	81	89.5
	MCIF	100	100	100	100
	P_{Ave}	90.5	100	90.5	94.75
SAE	SCIF	84	100	84	91.3
	MCIF	100	87	100	93
	P_{Ave}	92	93.5	92	92.15

of identifying CIF, including: Back-propagation Neural Network (BPNN) [37], Standard Deep Belief Network (SDBN) [38] and Stacked Autoencoder (SAE) [39].

The corresponding confusion matrix can be obtained through the results of different methods, as shown in Fig. 6. The detailed results of SCIF and MCIF are given in Table 4, and P_{Ave} represents the average value evaluation indicators.

It can be drawn from Table 4 : the average identification accuracy of DDBN is 99.5%, which is higher than that of BPNN, SDBN and SAE. The average identification accuracy of BPNN is 81.5%. For disturbance types with large similarity in input data, such as MCIF and Imp, BPNN misidentifies 27% of MCIF samples as Imp type, and 9% of the Imp samples are misidentified as MCIF type. The main reason is that the training process of BPNN belongs to “shallow learning” methods, which excessively relies on the features of input data and cannot fully mine the hidden features of input data. The identification performance of SDBN and SAE are improved compared with BPNN. However, due to the randomness of CIF, these methods are prone to overfitting and lead to false identification. For example, the current waveforms of SCIF are similar to that of capacitor switching, and the current waveform of MCIF are similar to that of constant impedance fault. SDBN misidentifies 19% of SCIF samples as Cap type, and SAE misidentifies 15% of SCIF samples as MCIF type.

The proposed method only misidentifies 1% of SCIF samples as Cap type. It can be concluded that the proposed method has higher identification performance for identifying CIF than other pattern identification methods. This great improvement from the proposed method prevents over-fitting of the model and reduces the complex interaction in neurons, which can obtain different scale information and learn more useful and identifiable deep features hidden in data.

In addition, the proposed method is also compared with the wavelet transform method [3] and the voltage arc feature

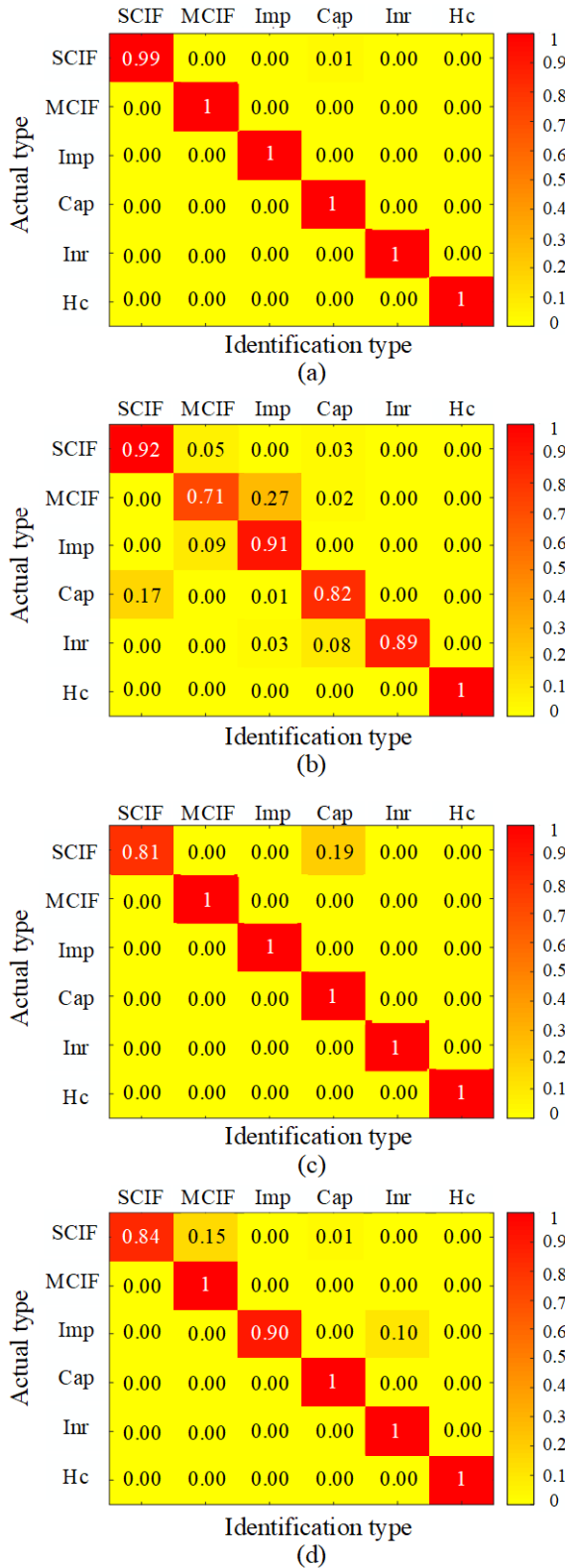


FIGURE 6. Confusion matrix of different methods. (a) DDBN, (b) BPNN, (c) SDBN, and (d) SAE.

method [14]. 1800 sets of current waveforms are used for analysis, including 900 sets of CIF waveforms (MCIF and

TABLE 5. Comparison result with cif identification methods.

Method	P_{Acc} (%)	P_{Pre} (%)	P_{Rec} (%)	F_1 (%)
DDBN	99.6	100	98	98.9
Ref. [3]	81.44	85.46	75.78	80.33
Ref. [14]	88.45	86.81	90.67	88.7

TABLE 6. Comparison result with different sampling frequency and other waveform decomposition methods.

Waveform decomposition method	P_{Acc} (%)	Computing time (s)
SWT (10kHz)	99.5	483.69
SWT (5kHz)	93.1	364.87
SWT (15kHz)	97.2	607.15
DWT	87.6	427.59
Waveform data (without feature extraction)	70.1	5021.36

SCIF) and 900 sets of non-CIF waveforms. The difference between these waveforms is the different simulation parameters. The test results are listed in Table 5.

It can be seen from Table 5 that the accuracy of the proposed method is 18.16% and 11.15% higher than that of the methods proposed in [3] and [14], respectively. At the same time, the value of P_{Pre} , P_{Rec} and F_1 are much higher than other two methods. The main reason is that the proposed method uses DDBN model to learn useful information from different power disturbance waveform, which can overcome the influence of CIF waveform uncertainty on features-based method. Those deep features learned by DDBN are more identifiable than traditional features, and can be better correlated with the disturbance type, which greatly improves the identification performance of CIF.

E. THE ADVANTAGES OF THE PROPOSED METHOD

1) THE ADVANTAGE OF EXTRACTING SHALLOW FEATURES
This paper proposes a method of extracting shallow features (physical and statistical features) from the CIF current waveform. In order to prove the advantages of feature extraction, the input data of DDBN is converted into features extracted by Discrete wavelet transform (DWT) and original data without feature extraction.

In addition, different time series lengths will have a greater impact on the time domain characteristics. In order to explain the effect of time domain length, waveforms containing different sampling frequency are used as original data. Here, we consider the case where the sampling frequency are 5kHz, 10kHz and 15kHz, respectively. Table 6 gives the test results which are obtained according to the steps in Fig 4.

The result shows that the proposed method using SWT to extract shallow features obtains higher identification accuracy (99.5%) than that of DWT and original data. Compared with SWT, the computing time of the method using DWT

to extract features is about 60s shorter, but the identification accuracy is 11.9% lower. The main reason is that when the DWT is used to decompose the waveform, the length of the coefficients of each layer is reduced, and the weak disturbance features such as sub-cycle incipient fault and capacitor switching are lost. On the contrary, SWT can greatly retain the disturbance information in the original waveform during the decomposition process, which improves the identification accuracy.

The identification accuracy of the proposed method is 6.4% and 2.3% higher than that of the sampling frequency of 5kHz and 10kHz, respectively. After the sampling frequency is reduced, the computing time will be reduced due to the reduction of the waveform length, losing the disturbance information existing in the waveform at the same time. For example, the duration and amplitude of SCIF are very small, the weak features may be lost when the sampling frequency is low, reducing the identification accuracy.

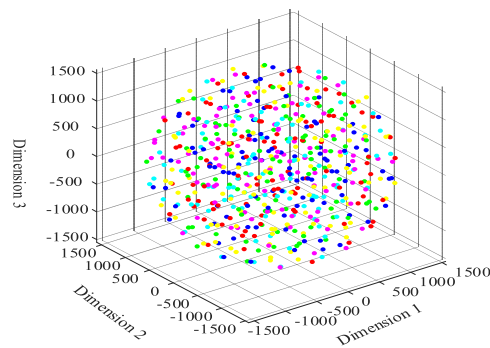
When the original data is used as the input of DDBN, the identification performance is worse than that of the other two decomposition methods. The main reason is that the amount of original data is large, and there is a certain amount of confusion in different disturbances, which causes the problems of long computing time and low accuracy. Thus, it is difficult to use original data to identify CIF.

2) THE ADVANTAGE OF DEEP FEATURES LEARNED BY DDBN

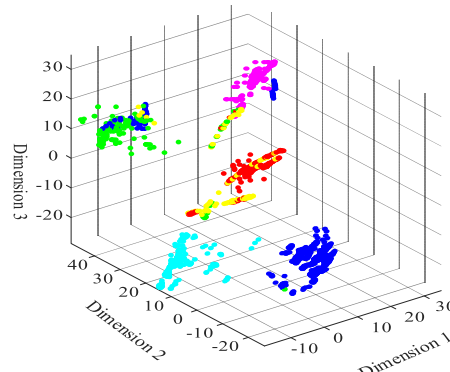
It is difficult to represent the features extracted by deep learning into a form that is easy to interpret the meaning, the clustering performance of different disturbance features can be analyzed in a visual form. The deep features are high-dimensional data. It is necessary to project the high-dimensional data into three-dimensional space through a dimensionality reduction algorithm [40]. In this paper, t-sne is used to visualize different data types, as shown in Fig. 7, and the colors in Fig.7 represent the different types.

It can be concluded from Fig. 7 that due to the similarity of some waveforms, the visualization result shows overlap in the original data, and the clustering performance of same types is low. After extracting the shallow features, the clustering performance is improved, but there is still partial overlap, such as the types of Cap, SCIF and MCIF are clustered together. And there are misidentifications between MCIF and SCIF, Imp and Inr in the visualization results obtained by SAE, and the clustering performance between each type is not high. By learning the deep features in shallow features using DDBN, different types of data can be accurately separated, and the clustering performance is improved. The main reason is that the 3-layer features learned by DDBN can represent the input data in a more identifiable way than the other features.

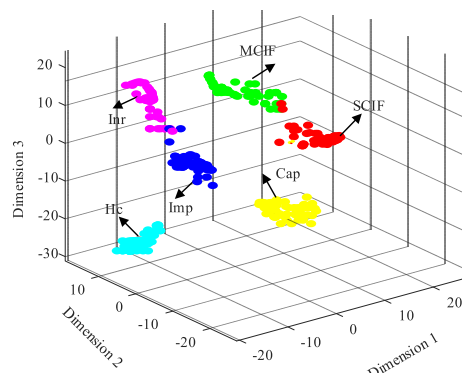
In order to further analyze the clustering performance in Fig. 7, the between-class covariance S_b , within-class covariance S_w are calculated. The formulas and of S_b and S_w can be found in [41] and [42]. Then the evaluation indicators can



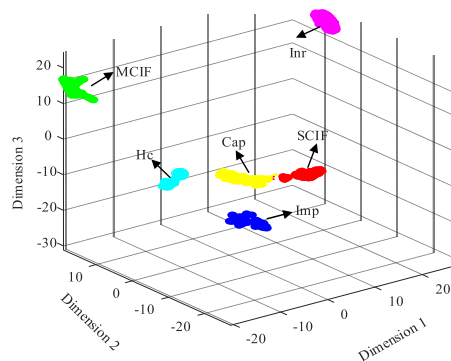
(a) Original data



(b) Shallow features



(c) The features learned by SAE



(d) The 3-layer features learned by DDBN

FIGURE 7. Visualization result using t-sne.

be calculated, and the result is listed in Table 7.

$$Q_1 = \ln \frac{|S_b|}{|S_w|}$$

TABLE 7. Comparison of clustering performance of different data

Data type	Q_1	Q_2	Q_3
Original data	0.89	1.19	2.07
Shallow features	3.31	2.16	2.62
The features learned by SAE	6.72	4.93	5.96
The 3-layer features learned by DDBN	8.76	6.39	7.39

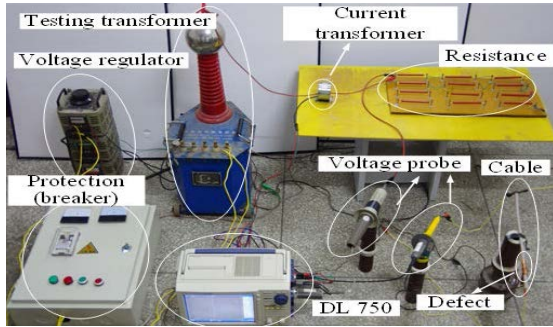


FIGURE 8. Wiring diagram of experimental data recording system in laboratory.

$$\begin{aligned}
 Q_2 &= \frac{tr(S_b)}{tr(S_w)} \\
 Q_3 &= \frac{|S_w + S_b|}{|S_w|}
 \end{aligned} \tag{24}$$

where $|S_b|$ and $tr(S_b)$ refers to the determinant and trace of the S_b matrix, respectively.

The result shows that the features learned by DDBN has high values of Q_1 - Q_3 than original data and shallow features. According to the scatter matrix criterion, the larger Q_1 - Q_3 , the better clustering performance. It is proved that deep features are more identifiable and can accurately map to types of input data.

VI. EXPERIMENTAL DATA VERIFICATION

In this section, experimental data recorded from laboratory is used to test the performance of the proposed method [14]. The wiring diagram of experimental data recording system in laboratory is shown in Fig. 8, including: (1) Power source, AC 220 V, 50Hz. (2) Voltage regulator: TSG2J-6. (3) Testing transformer: YDQ -10/100. (4) Cable: YJV22-0.6/1 Kv. (5) Waveform Recorder: Yokogawa DL 750; (6) Voltage Probe: Tektronix P6015A. (7) Current Transformer: Yuanxing TA25CL-1A /1mA, the rated load is 100 ohms. (8) Current limiting resistor: metal glass uranium resistor. (9) Protection device: Breaker.

The power source is used to supply power to the whole experimental system. The voltage regulator is used to adjust the voltage during the experiment. The test transformer realizes the voltage transformation. The voltage probe and the current transformer can measure the voltage and current respectively. The waveform recorder is used to record the

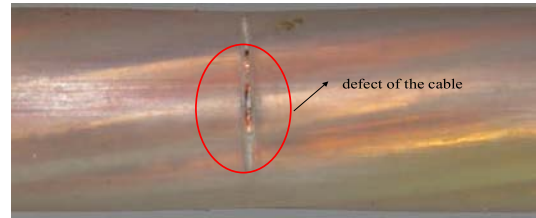


FIGURE 9. Schematic diagram of cable sample defects.

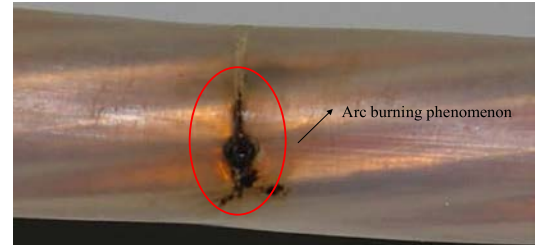
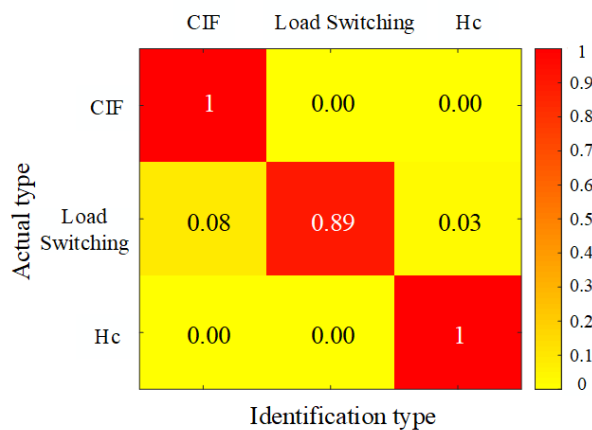


FIGURE 10. Schematic diagram of cable sample defects.

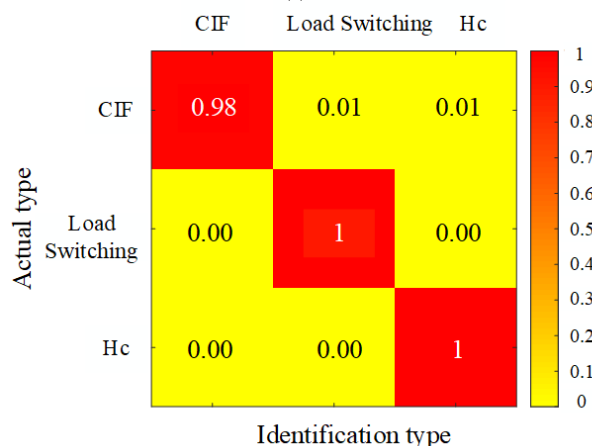
voltage and current waveforms generated during the experiment. Current limiting resistors and protection devices are used to protect the safe operation of the experiment. The tested cable is the returned XLPE cable after long-term operation in the field, which is a three-core structure. The experiment is a single-phase experiment, so one phase of the three-core cable is extracted for the experiment. In order to ensure the breakdown of the cable to form an arc fault, a knife is used to make cuts in the cable insulation. Then the metal armor outside the insulation is restored to form the defective cable, as shown in Fig.9.

The experimental process is as follows: (1) set up the experimental system and place the voltage regulator at the minimum output position. (2) gradually adjust the voltage regulator to increase the output voltage until the cable breakdown forms an arc fault. (3) Maintain this state until a permanent fault is formed and maintained for a period of time. (4) Finally, adjust the regulator to zero and end the experiment.

The waveforms of three types are recorded using DL750 waveform recorder: including healthy condition, arc fault and load switching. Firstly, the experiment is carried out on the defective cable, and the waveforms of health condition are recorded without breakdown of the cable. Then, the voltage is gradually increased by adjusting the voltage regulator until the cable breakdown, and the circuit changes from health condition to arc fault. When the arc fault lasts for a period of time, the insulation layer of the cable is gradually carbonized due to the burning of the arc. The material morphology in the intermittent arc discharge defect location is cross linked polyethylene, and the arc burning phenomenon is shown in Fig.10. Finally, a permanent fault is formed. Therefore, by selecting the waveforms after cable breakdown and before permanent fault, it can be ensured that all arc fault samples are in the state of cable incipient fault and intermittent arc discharge.



(a) SAE



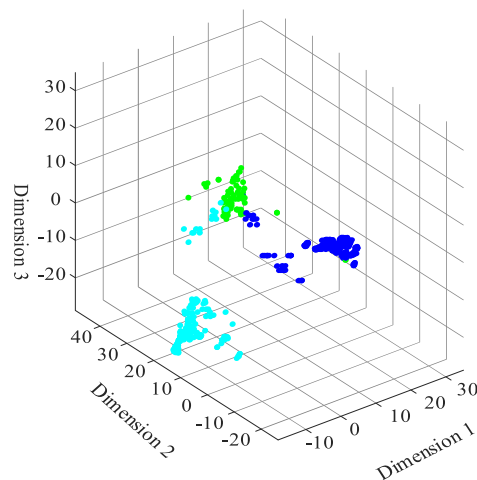
(b) DDBN

FIGURE 11. Confusion matrix result of the proposed method and SAE using the experimental data.

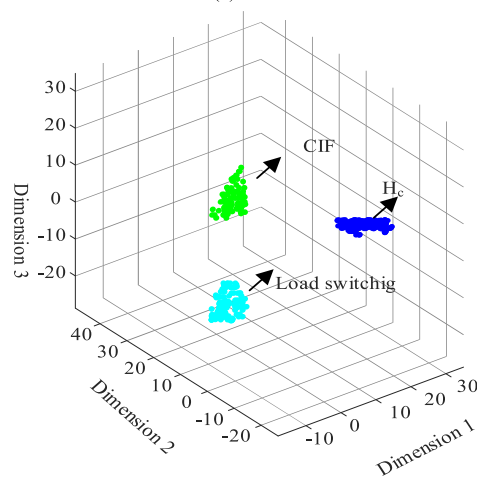
In contrast, the waveforms of load switching are recorded in a healthy cable state, which is simulated by changing the access resistance in the circuit, and there is no need to set cable defects. Finally, 1680 samples can be measured, including 560 samples of each type.

The confusion matrix results of SAE and the proposed method is shown in Fig.11. The confusion matrix results show that SAE misidentifies 8% of load switching samples as CIF type. The proposed method only misidentifies 2% of CIF samples as other types. the evaluation indicators can be obtained from confusion matrix. The detailed result is given in Table 8. It can be concluded from Table 8 that P_{Acc} of the proposed method is 99.29%, and the value of P_{Pre} , P_{Rec} , F_1 are 100%, 97.86%, 98.57%, respectively. These performance indicators are higher than other methods.

Furthermore, the t-SNE visualization result of SAE and the proposed method is shown in Fig.12. The visualization result shows that the load switching samples and CIF samples are overlap. The proposed method can separate samples of each type, and improves the clustering performance. It proves



(a) SAE



(b) DDBN

FIGURE 12. t-SNE visualization result of the proposed method and SAE using the experimental data.

TABLE 8. Comparison result with different pattern identification methods using experimental data.

Method	P_{Acc} (%)	P_{Pre} (%)	P_{Rec} (%)	F_1 (%)
DDBN	99.29	100	97.86	98.57
BPNN	70.18	75.89	71.61	73.7
SDBN	87.1	71.9	100	83.7
SAE	94.8	89.3	100	94.3

the practicability of the proposed method for experimental data.

VII. CONCLUSION

In order to prevent cable permanent fault, a method is proposed for identifying CIF using power disturbance waveform feature learning, which provides a favorable tool for the research of cable condition monitoring. The main conclusions are:

(1) The proposed method extracts the shallow features from the disturbance waveform, which improves the efficiency and accuracy of deep learning model. The results show that the computing time is about one-tenth of the time using original data.

(2) The proposed method introduces the parameter of dropout_Fraction and adaptive learning rate into network training, which ensures the update of network parameters does not depend on the joint action between hidden neurons and means more robust features can be learned. Therefore, the proposed method has strong generalization ability and can cope with the randomness and uncertainty of CIF waveforms. The identification results show that the proposed method has high identification accuracy than other methods. And the original data, the data after feature extraction and the features learned by DDBN are visualized by t-sne, the results show that the deep features learned by DDBN are more identifiable, which further proves the advantage of using deep features to identify CIF.

In this paper, the model parameters of DDBN are determined by cross-validation methods. In the future research, for the parameter of the deep learning model, optimization algorithms can be further used in subsequent research to improve the accuracy and efficiency of parameter settings. After the CIF is identified, it is also meaningful to further estimate the remaining service life of the cable. We expect to be able to build a cable life monitoring system to promote the development of achieving accurate monitoring of cable status and timely warning of cable faults, through the above research.

REFERENCES

- [1] S. Kulkarni, S. Santoso, and T. A. Short, "Incipient fault location algorithm for underground cables," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1165–1174, May 2014, doi: [10.1109/TSG.2014.2303483](https://doi.org/10.1109/TSG.2014.2303483).
- [2] C. Zhou, H. Yi, and X. Dong, "Review of recent research towards power cable life cycle management," *High Volt.*, vol. 2, no. 3, pp. 179–187, Sep. 2017, doi: [10.1049/hve.2017.0037](https://doi.org/10.1049/hve.2017.0037).
- [3] T. S. Sidhu and Z. Xu, "Detection of incipient faults in distribution underground cables," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1363–1371, Jul. 2010, doi: [10.1109/TPWRD.2010.2041373](https://doi.org/10.1109/TPWRD.2010.2041373).
- [4] B. Kaszteny, I. Voloh, and C. G. Jones, "Detection of incipient faults in underground medium voltage cables," in *Proc. Power Syst. Conf.*, Mar. 2009, pp. 1–12, doi: [10.1109/PSAMP.2009.5262402](https://doi.org/10.1109/PSAMP.2009.5262402).
- [5] F. Tian and X. Peng, "The research progress of corona-resistant and tracking-resistant insulating materials," *Trans. China Electrotech. Soc.*, vol. 32, no. 16, pp. 3–13, Aug. 2017, doi: [10.19595/j.cnki.1000-6753.tces.161759](https://doi.org/10.19595/j.cnki.1000-6753.tces.161759).
- [6] K. Zhu and P. W. T. Pong, "Fault classification of power distribution cables by detecting decaying DC components with magnetic sensing," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2016–2027, May 2020, doi: [10.1109/TIM.2019.2922514](https://doi.org/10.1109/TIM.2019.2922514).
- [7] M.-F. Guo, J.-H. Gao, X. Shao, and D.-Y. Chen, "Location of single-line-to-ground fault using 1-D convolutional neural network and waveform concatenation in resonant grounding distribution systems," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021, doi: [10.1109/TIM.2020.3014006](https://doi.org/10.1109/TIM.2020.3014006).
- [8] L. Li, J. Yong, and W. Xu, "On-line cable condition monitoring using natural power disturbances," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1242–1250, Aug. 2019, doi: [10.1109/TPWRD.2018.2879964](https://doi.org/10.1109/TPWRD.2018.2879964).
- [9] H. A. Illias, M. E. Othman, M. A. Tunio, A. H. A. Bakar, H. Mokhlis, G. Chen, P. L. Lewin, and A. M. Ariffin, "Measurement and simulation of partial discharge activity within a void cavity in a polymeric power cable model," in *Proc. IEEE Int. Conf. Solid Dielectr. (ICSD)*, Jun. 2013, pp. 105–108, doi: [10.1109/ICSD.2013.6619829](https://doi.org/10.1109/ICSD.2013.6619829).
- [10] Y. Meng, S. Chen, Z. Wu, C. Wang, and X. Li, "Research on feature enhancement of DC arc fault detection in photovoltaic systems based on stochastic resonance," *Proc. CSEE*, vol. 42, no. 6, pp. 2396–2407, Mar. 2022, doi: [10.13334/j.0258-8013.pcsee.210471](https://doi.org/10.13334/j.0258-8013.pcsee.210471).
- [11] H. Cui, Y. Guan, H. Chen, and W. Deng, "A novel advancing signal processing method based on coupled multi-stable stochastic resonance for fault detection," *Appl. Sci.*, vol. 11, no. 12, p. 5385, Jun. 2021, doi: [10.3390/app11125385](https://doi.org/10.3390/app11125385).
- [12] S. Shi, Z. Hu, and M. Zhao, "Travelling waves-based identification of sub-health condition of feeders in power distribution system," *IET Gener. Transmiss. Distrib.*, vol. 12, no. 5, pp. 1067–1073, 2018, doi: [10.1049/iet-gtd.2016.1638](https://doi.org/10.1049/iet-gtd.2016.1638).
- [13] T. Ghanbari, "Kalman filter based incipient fault detection method for underground cables," *IET Gener. Transmiss. Distrib.*, vol. 9, no. 14, pp. 1988–1997, Nov. 2015, doi: [10.1049/iet-gtd.2015.0040](https://doi.org/10.1049/iet-gtd.2015.0040).
- [14] W. Zhang, X. Xiao, K. Zhou, W. Xu, and Y. Jing, "Multicycle incipient fault detection and location for medium voltage underground cable," *IEEE Trans. Power Del.*, vol. 32, no. 3, pp. 1450–1459, Jun. 2017, doi: [10.1109/TPWRD.2016.2615886](https://doi.org/10.1109/TPWRD.2016.2615886).
- [15] D. Salles and W. Xu, "Information extraction from PQ disturbances—An emerging direction of power quality research," in *Proc. IEEE 15th Int. Conf. Harmon. Quality Power*, Jun. 2012, pp. 649–655, doi: [10.1109/ICHQP.2012.6381165](https://doi.org/10.1109/ICHQP.2012.6381165).
- [16] S. Xiong, Y. Liu, J. Fang, J. Dai, L. Luo, and X. Jiang, "Incipient fault identification in power distribution systems via human-level concept learning," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5239–5248, Nov. 2020, doi: [10.1109/TSG.2020.2994637](https://doi.org/10.1109/TSG.2020.2994637).
- [17] M. Jannati, B. Vahidi, and S. H. Hosseini, "Incipient faults monitoring in underground medium voltage cables of distribution systems based on a two-step strategy," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1647–1655, Aug. 2019, doi: [10.1109/TPWRD.2019.2917268](https://doi.org/10.1109/TPWRD.2019.2917268).
- [18] G. W. Chang, Y.-H. Hong, and G.-Y. Li, "A hybrid intelligent approach for classification of incipient faults in transmission network," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1785–1794, Aug. 2019, doi: [10.1109/TPWRD.2019.2924840](https://doi.org/10.1109/TPWRD.2019.2924840).
- [19] X. Chang, Y. Liang, J. Lou, W. Zhang, B. Han, J. Zhong, and K. Li, "Detection of latent fault in medium voltage distribution cables based on guiding learning model," in *Proc. 4th Asia Energy Electr. Eng. Symp. (AEEES)*, Mar. 2022, pp. 667–673, doi: [10.1109/AEEES54426.2022.9759678](https://doi.org/10.1109/AEEES54426.2022.9759678).
- [20] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Aug. 2006, doi: [10.1162/neco.2006.18.7.1527](https://doi.org/10.1162/neco.2006.18.7.1527).
- [21] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, "Early fault detection approach with deep architectures," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1679–1689, Jul. 2018, doi: [10.1109/TIM.2018.2800978](https://doi.org/10.1109/TIM.2018.2800978).
- [22] Y. Wang, H. Lu, X. Xiao, X. Yang, and W. Zhang, "Cable incipient fault identification using restricted Boltzmann machine and stacked autoencoder," *IET Gener. Transmiss. Distrib.*, vol. 14, no. 7, pp. 1242–1250, Feb. 2020, doi: [10.1049/iet-gtd.2019.0743](https://doi.org/10.1049/iet-gtd.2019.0743).
- [23] R. Moghe, M. J. Mousavi, J. Stoupis, and J. McGowan, "Field investigation and analysis of incipient faults leading to a catastrophic failure in an underground distribution feeder," in *Proc. IEEE/PES Power Syst. Conf. Expo.*, Mar. 2009, pp. 1–6, doi: [10.1109/PSC.2009.4840203](https://doi.org/10.1109/PSC.2009.4840203).
- [24] W. Qiu, Q. Tang, J. Liu, and W. Yao, "An automatic identification framework for complex power quality disturbances based on multifusion convolutional neural network," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3233–3241, May 2020, doi: [10.1109/TII.2019.2920689](https://doi.org/10.1109/TII.2019.2920689).
- [25] F. A. S. Borges, R. A. S. Fernandes, I. N. Silva, and C. B. S. Silva, "Feature extraction and power quality disturbances classification using smart meters signals," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 824–833, Apr. 2016, doi: [10.1109/TII.2015.2486379](https://doi.org/10.1109/TII.2015.2486379).
- [26] F. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical application," *Science*, vol. 103, pp. 281–300, Dec. 1995, doi: [10.1007/978-1-4612-2544-7_17](https://doi.org/10.1007/978-1-4612-2544-7_17).
- [27] 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, doi: [10.1016/j.measurement.2016.10.013](https://doi.org/10.1016/j.measurement.2016.10.013).

- [28] W. Wei, L. Zengli, C. Lin, and S. Weiwei, "Harmonic detection of power system based on SVD and EMD," in *Proc. 9th Int. Conf. Comput. Intell. Secur.*, Dec. 2013, pp. 185–189, doi: [10.1109/CIS.2013.46](https://doi.org/10.1109/CIS.2013.46).
- [29] F. WANG, X. Quan, and L. Ren, "Review of power quality disturbance detection and identification methods," *Proc. CSEE*, vol. 41, no. 12, pp. 4104–4121, 2021, doi: [10.13334/j.0258-8013.pcsee.201261](https://doi.org/10.13334/j.0258-8013.pcsee.201261).
- [30] N. S. D. Brito, B. A. Souza, and F. A. C. Pires, "Daubechies wavelets in quality of electrical power," in *Proc. 8th Int. Conf. Harmon. Quality Power*, Oct. 1998, pp. 511–515, doi: [10.1109/ICHQP.1998.759961](https://doi.org/10.1109/ICHQP.1998.759961).
- [31] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 105524, doi: [10.1016/j.asoc.2019.105524](https://doi.org/10.1016/j.asoc.2019.105524).
- [32] S. Galeshchuk and S. Mukherjee, "Deep networks for predicting direction of change in foreign exchange rates," *Intell. Syst. Accounting, Finance Manage.*, vol. 24, no. 4, pp. 100–110, Oct. 2017, doi: [10.1002/isaf.1404](https://doi.org/10.1002/isaf.1404).
- [33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014, doi: [10.5555/2627435.2670313](https://doi.org/10.5555/2627435.2670313).
- [34] T. Gedeon, "Additive neural networks and periodic patterns," *Neural Netw.*, vol. 12, no. 4, pp. 617–626, Jun. 1999, doi: [10.1016/S0893-6080\(99\)00023-4](https://doi.org/10.1016/S0893-6080(99)00023-4).
- [35] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1800–11771, Aug. 2002, doi: [10.1162/089976602760128018](https://doi.org/10.1162/089976602760128018).
- [36] M. Piao, H. S. Shon, J. Y. Lee, and K. H. Ryu, "Subspace projection method based clustering analysis in load profiling," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2628–2635, Nov. 2014, doi: [10.1109/TPWRS.2014.2309697](https://doi.org/10.1109/TPWRS.2014.2309697).
- [37] R. Geng, X. Wang, N. Ye, and J. Liu, "A fault prediction algorithm based on rough sets and back propagation neural network for vehicular networks," *IEEE Access*, vol. 6, pp. 74984–74992, 2018, doi: [10.1109/ACCESS.2018.2881890](https://doi.org/10.1109/ACCESS.2018.2881890).
- [38] J. Qiao, G. Wang, X. Li, and W. Li, "A self-organizing deep belief network for nonlinear system modelling," *Appl. Soft Comput.*, vol. 65, pp. 170–183, Apr. 2018, doi: [10.1016/j.asoc.2018.01.019](https://doi.org/10.1016/j.asoc.2018.01.019).
- [39] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, Mar. 2018, doi: [10.1016/j.ymsp.2017.09.026](https://doi.org/10.1016/j.ymsp.2017.09.026).
- [40] G. Dimitriadis, J. P. Neto, and A. R. Kampff, "T-SNE visualization of large-scale neural recordings," *Neural Comput.*, vol. 30, no. 7, pp. 1750–1774, Jul. 2018, doi: [10.1162/neco_a_01097](https://doi.org/10.1162/neco_a_01097).
- [41] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *J. Sound Vibrat.*, vol. 335, pp. 367–383, Jan. 2015, doi: [10.1016/j.jsv.2014.09.026](https://doi.org/10.1016/j.jsv.2014.09.026).
- [42] H. Shao, H. Jiang, Z. Ke, D. Wei, and X. Li, "A novel tracking deep wavelet auto-encoder method for intelligent fault diagnosis of electric locomotive bearings," *Mech. Syst. Signal Process.*, vol. 110, pp. 193–209, Sep. 2018, doi: [10.1016/j.ymsp.2018.03.011](https://doi.org/10.1016/j.ymsp.2018.03.011).

• • •