IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR ELECTRIC POWER SYSTEMS

Received May 21, 2019, accepted June 7, 2019, date of publication June 12, 2019, date of current version June 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2922367

Power Quality Disturbance Classification Based on Compressed Sensing and Deep Convolution Neural Networks

JIDONG WANG, (Member, IEEE), ZHILIN XU, AND YANBO CHE, (Member, IEEE)
[1]Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China

Corresponding author: Yanbo Che (ybche@tju.edu.cn)

This work was supported by SGCC program: Research on Extensive Application and Benefit Evaluation of Typical Power Substitution Technology Considering Power Quality Influence under Grant 52182018000H. Innovation fund program of Lanzhou Jiaotong University-Tianjin University (Research on Management Strategies of Power Quality Problems from Thermostatically Controlled Appliances).

ABSTRACT By analyzing the recovery and reconstruction process of various power quality single disturbances and composite disturbance signals, we proposed a set of acquisition methods suitable for power quality disturbance (PQD) signals. The proposed acquisition method is applied to the compression sensing (CS) technology for data compression, the demand for the acquisition device memory is reduced, and the transmission rate is increased. An end-to-end intelligent classification framework is designed, which can directly classify the collected data without any time-consuming data pre-processing operations. The model is designed with noise adaptation module, which can cope with the error of compressed sensing recovery and has also showed good classification performance in noise data. Simultaneously, the model applies a lot of easy-to-implement techniques, which makes the trained model have better generalization ability and classification effect. The proposed method is verified by both simulation and measured data. The method showed superior performance compared to the existing disturbance identification methods based on the classification results.

INDEX TERMS Power quality, classification, compressed sensing, deep convolutional neural network, batch normalization.

I. INTRODUCTION

With the large number of distributed components, electric vehicles and other non-linear components connected to the grid, various power quality disturbances (PQD) have emerged [1]. More and more power quality data are collected due to the application of smart meters and various intelligent acquisition devices. Concurrently, the generation of massive power information imposes a large burden on the data transmission and storage of the current power grid. In order to reduce the pressure on communication and the problem of storage at the collection end, the application of lighter physical quantity is the development trend of smart grid. The introduction of compressed sensing (CS) technology has an important impact on the development of smart grid. Compressed sampling technology has been widely used in medicine and other fields, and the application research in smart grid is increasing. Aiming at handing the difficulty in recovering data collected by compressed sensing technology, the large amount of data in smart grid and the long-time of recovery, this paper proposes a method for direct classification of power quality data collected by compression technology.

Compressed sensing technology [2] breaks through the limitations of the sampling theorem, and can recover the original signal completely by collecting data much smaller than the sampling theorem. Research on data recovery and reconstruction algorithms in compressed sensing technology has always been challenging. At present, the recovery and reconstruction algorithm of compressed sensing technology is computationally intensive and slow in operation, and it is difficult to apply to a large number of data reconstruction problems. In addition, the power quality disturbance data volume is large, the recovery reconstruction algorithm of the compressed sensing technology runs slowly, and is not suitable for recovery and reconstruction of a large number

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

of signals. Therefore, the deep learning algorithm is used to directly classify the compressed signals, and according to the classification results, the specific types of disturbance signals are analyzed. The reconstruction and processing of signals for a particular class of disturbance signals can save a significant amount of process for recovering signals.

The design of the classifier is a critical part for the classification of power quality disturbances. At present, the classifiers commonly used in power quality disturbance classification are: Probabilistic Neural Network (PNN) algorithm [3], [4], Decision Tree [5]–[7], Support Vector Machine (SVM) [8]–[10], Artificial Neural Network [11], [12] and so on. These traditional artificial intelligence methods have achieved certain results in practical applications, but their ability to extract features and process large amounts of data is still poor. A feature extraction process is required which affects their classification speed. In recent years, with the development of machine learning algorithms and the improvement of hardware computing power, more intelligent algorithms have been applied to the field of power quality disturbance classification. The rapid development of deep learning has made rapid classification possible. Although the training process of the deep learning algorithm is slow and time-consuming, the algorithm trained by a large amount of data has high recognition accuracy and fast classification. Therefore, training an efficient classifier in the power system can quickly analyze the power quality disturbance.

In this paper, a power quality disturbance recognition method based on compressed sensing and deep convolutional neural network is proposed. The contributions of work are listed as follows:

1) By analyzing single power quality disturbance and composite power quality disturbance, a compressed sensing method suitable for power quality disturbance data acquisition is proposed.
2) Aiming at the disturbance data extracted by compressed sensing, a deep learning framework is designed, which can reduce the influence of noise on recognition accuracy, high recognition accuracy and fast classification.
3) By combining the compressed sensing technology with the deep learning algorithm, the rapid classification of power quality disturbance can be realistic and it provides novel route for disturbance recognition.

The rest of the paper is organized in the following sequence. Section II develops a power quality compression sampling model. Section III describes the principle of convolutional neural networks (CNN). Section IV proposes an algorithm for identifying PQD compressed data. Section V simulation experiment verifies the feasibility of the algorithm. Section VI concludes the paper.

II. COMPRESSED SENSING THEORY OF PQD DATA ACQUISITION

In this section, we develop a data acquisition model based on compressed sensing.

A. COMPRESSED SENSING THEORY

Compressed sensing theory [13]–[17] shows that when the original signal x itself is sparse, or the signal is sparse on an orthogonal base Ψ, We can use the stochastic stationary observation matrix Φ for compression sampling based on spatial transformation. Obtaining a measured value y that maintains the original signal structure and is much smaller than the signal length, then the original signal x is accurately reconstructed by solving the numerical optimization problem.

The CS theory directly obtains the data compressed expression, omitting the intermediate steps of acquiring the N-dimensional signal, and its linear observation model is as shown in (1).

$$y = \Phi x = \Phi \psi x = \Theta S \tag{1}$$

where: x is the $N \times 1$ dimension original signal;Φ is the $M \times N$ dimension ($M \ll N$) observation matrix; there are only K ($K \ll N$) non-zero elements in S, which is an N-dimensional K-sparse vector; $\Theta = \Phi\psi$ called the sensing matrix. CS recovers the sparse vector S from the observation vector y, thereby accurately reconstructing the original signal $\hat{X} = \psi \hat{S}$.

For the above problem, the sparse solution can be obtained by l_0-norm optimization as shown in (2):

$$\hat{S} = \arg\min ||S||_0 \quad (s.t. Y = \Phi\Psi x) \tag{2}$$

where: $||S||_0$ is the zero norm of S, but as a nonlinear programming problem, it is difficult to solve it, as a result of the sparse domain Ψ and Φ non-coherent observation matrix, can be used to solve l_1- norm to get the same result as shown in (3):

$$\hat{S} = \arg\min ||S||_1 \quad (s.t. Y = \Phi\Psi x) \tag{3}$$

However, to properly restore the sparse vector $\hat{S}$, must satisfy: 1) The dimension M of the observation vector y has $M = O\left[K \log (N)\right]$, where O is the operator of computational complexity; 2) The observation matrix Φ needs to satisfy the restricted equidistance characteristic criterion (RIP), that is, there is a constrained equidistance constant $\delta_K \in (0, 1)$, therefore, for any K sparse signal, equation (4) holds.

$$(1 - \delta_K) ||X||_2^2 \leq ||\Phi X||_2^2 \leq (1 + \delta_K) ||X||_2^2 \tag{4}$$

This can be transformed into a convex optimization problem so that the convex relaxation algorithm can be used to solve this problem. In this paper, the reconstruction of PQD signals is reconstructed using the Sparsity adaptive matching pursuit algorithm [18].

B. DETERMINATION OF PQD DATA ACQUISITION COMPRESSION BASE AND OBSERVATION MATRIX

The collection of power quality disturbances needs to follow two principles: 1) On the one hand, in order to reduce the amount of data collected, assume ($M \ll N$);2) On the other hand, in order to ensure data quality, it is desirable that the compressed PQD data is sufficient to represent the actual power quality disturbance, that is, the recovery matrix $\hat{X}$ is as close as possible to X. X is the PQD data that satisfies the Nyquist sampling theorem

The satisfaction of the two principles requires a detailed analysis of the PQD. Since the PQD signal itself is not a sparse signal, it is necessary to determine a sparse basis such that the PQD signal is sparse on the sparse basis. Analysis of 9 common single perturbation findings, PQD signals are mostly low-frequency signals, and PQD is different in the measured signals. In order to simplify the simplicity of the device, reference [13] uses DCT as the sparse basis.

DCT is a simple and practical sparse base. Its advantage is its strong applicability. It can be applied to the acquisition of all PQD signals, and its structure is simple and easy to implement in hardware. The transformation matrix representation of DCT is as shown in equation (5):

$$DCT\ (N \times N) = \left[\psi_{i,j} = C \cos \left(i\,(1+2j)\,\frac{\Pi}{2N} \right) \right] \tag{5}$$

where $i \in \{0, \ldots, N-1\}$ and $j \in \{0, \ldots, N-1\}$ represent the row number and column number of the matrix, and C is defined by equation (6):

$$C = \begin{cases} \sqrt{1/N} & for\ i = 0 \\ \sqrt{2/N} & for\ i \neq 0 \end{cases} \tag{6}$$

Since the DCT matrix is orthogonal, its inverse matrix can be obtained simply by performing a matrix transposition operation as shown in (7).

$$IDCT = DCT^{-1} \tag{7}$$

The DCT sparse the signal is by concentrating most of the information in some low frequency components. The remaining high frequency components tend to have lower values and they can be discarded without significant loss.

Figure 1 shows the compression reconstruction of sag signal, the sparse representation on the IDCT basis and the error of compression reconstruction. It can be seen from the compression reconstruction results that the PQD can be completely reconstructed, and the reconstruction error is within the allowable range.

To directly classify PQD using compressed data, first determine the magnitude of the M value in the observation matrix. In order to determine the magnitude of the M value, this paper analyzes the relationship between the compression reconstruction of a variety of single PQDs and the compression and reconstruction M values of multiple composite perturbations and the reconstituted power. There is an error in the reconstruction. Setting the average reconstruction error per point is less than 0.0007, which means that the signal is completely reconstructed successfully. Show in Figure 2 is the relationship between the magnitude of the M values of the eight disturbance signals and the reconstructed power.

It can be found from Figure. 2 that when the disturbance type is a single sag and other disturbances, the signal of

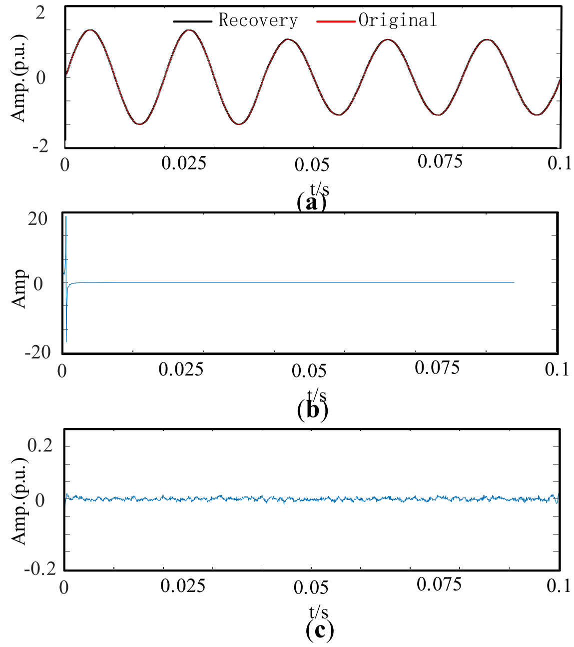

FIGURE 1. (a) Sag compression reconstruction. (b) IDCT sparse representation. (c) compression reconstruction error.

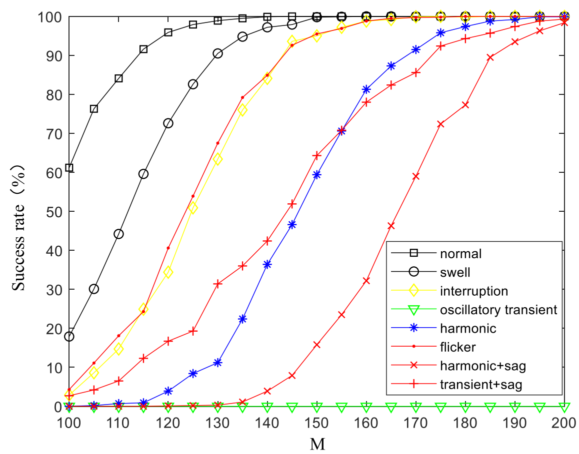

FIGURE 2. Eight kinds of disturbance recombination power and M value relationship.

M = 170 can be completely reconstructed successfully. When the disturbance is transient oscillation and compound disturbance, M = 200 cannot guarantee 100% reconstruction success. In order to comprehensively consider the power quality disturbance type, the reconstruction can be successfully performed, and the composite disturbance and transient oscillation disturbance are further analyzed.

Further analysis shows that when the reconstruction error of each point is 0.0007, the composite perturbation can be completely reconstructed successfully when M = 250, while the transient oscillation perturbation recombination power is still zero. In order to ensure that the M value is in a small range, the reconstruction error of the transient oscillation at each point is 0.002, which is the signal reconstruction success. The reconstruction result is shown in Figure 3. I As described by Figure 3, when the M value is 290, the transient

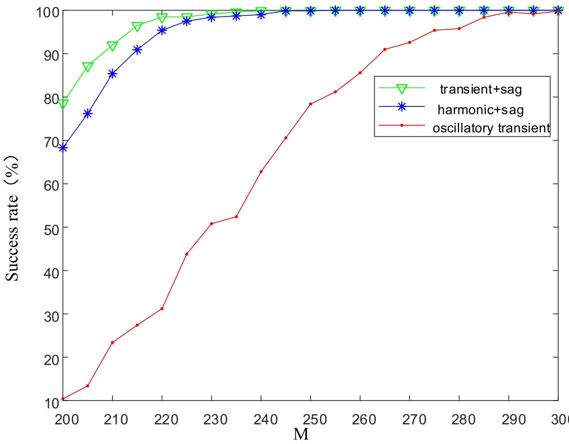

FIGURE 3. Three kinds of disturbance recombination power and M value relationship.

oscillation signal can achieve complete reconstruction success. Analysis of PQD found that the characteristics of power quality disturbances are complex, even if the same type of interference is very different. Therefore, the M value needs to reserve a certain margin to ensure that the signal can be completely reconstructed successfully. In this paper, using M = 300 as the dimension of the compressed data ensures that the reconstruction error is as small as possible, within the thousandth. The compression ratio is 1:4, and the dimension of the collected data is reduced by 4 times.

For the same signal, we collect less data. When delivering the same information, we pass less data than the existing sampling method, and the data transfer rate will be better than the existing method.

III. CONVOLUTIONAL NEURAL NETWORK

In this section, we introduce the basic structure of a convolutional neural network.

A. TRADITIONAL CONVOLUTIONAL NEURAL NETWORK FRAMEWORK

At the heart of CNN is the convolutional layer, which contains many different convolution kernels to extract various features. At the same time, the convolution layer combined with the collection layer can reduce the number of parameters and speed up the calculation. The feature quantities extracted by the last convolution kernel are passed to the fully connected layer to combine the previously extracted features to achieve the final prediction. In this way, we can obtain implicit information about the data to achieve faster, more stable predictions.

The structure of the traditional convolutional neural network is shown in Figure 4. It consists of a convolutional layer, a pooled layer, a fully connected layer and a Softmax layer, each of which has different functions.

B. CONVOLUTION OPERATION

Convolutional layer [19]: The convolutional layer convolves the input local area with the convolution kernel and then

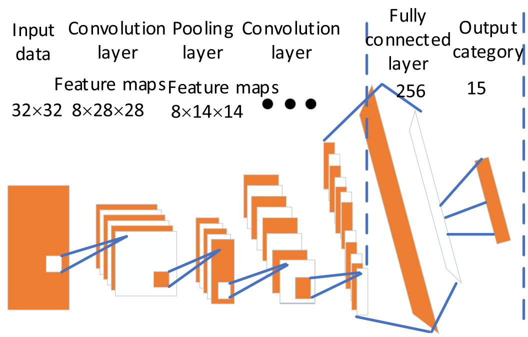

FIGURE 4. Traditional deep learning frame composition.

generates an output feature through the activation unit. Because of the same kernel extraction feature, it has the property of weight sharing. The calculation process is as shown in equation (8):

$$y^{l(i,j)} = k_i^l * x^{l(x^j)} = \sum_{j'=0}^{w-1} k_i^{l\left(j'\right)} x^{l\left(j+j'\right)} \tag{8}$$

where $k_i^{l\left(j'\right)}$ represents the j' weight of the I convolution kernel of layer 1, $x^{l(x^j)}$ denotes the local area of convolution of j in layer 1, and $^{\omega}$ is the width of the convolution kernel.

C. POOLING OPERATION

Pooling layer [19]: A pooling layer usually added after the convolution layer, is mainly used for down sampling. The pooling layer is used to scale and map data after the convolution layer feature extraction, to reduce the dimension of the data and extract important information. It can reduce the impact of data fluctuations. The common collection layer is the largest pool layer, and the function is as shown in equation (9):

$$P^{l(i,t)} = \max_{(j-1)W+1 \le t} \left\{a^{l(i,t)}\right\} \tag{9}$$

where $a^{l(i,t)}$ represents the activation function value of the t-neuron in one layer and $P^{l(i,t)}$ represents the width of the merged region.

D. ACTIVATION FUNCTION

Two types of activation functions [20] are used in this work: (1) Rectifying linear activation units (Relu) and (2) softmax. Rectified linear activation unit: After each convolutional layer, an activation function is employed. The activation function is an operation that maps the output to a set of inputs. The Relu function has attributes that add non-linearity and sparsity to the network structure. Therefore, it provides robustness for small changes such as noise in the input. Its expression is as shown in Equation (10).

$$f(x) = \begin{cases} x & if\ x > 0 \\ 0.01x & others \end{cases} \tag{10}$$

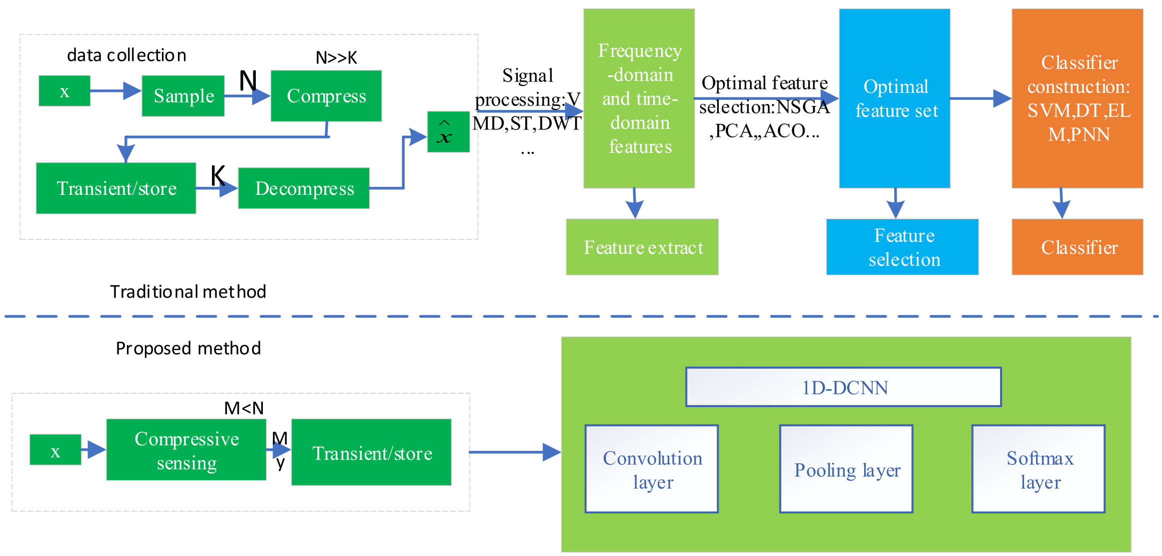

FIGURE 5. **Comparison of this method with the framework of traditional methods.**

Softmax: This function calculates the probability distribution of k output classes. Therefore, the Softmax function is used to predict the category to which the input PQD belongs.

$$p_j = \frac{e^{x_j}}{\sum_1^k e^{x_k}} \quad for\ j = 1, \cdots k \tag{11}$$

where x is the net input. The output value of p is between 0 and 1, and the sum is equal to 1.

E. FULLY CONNECTED LAYER

Fully connected layer (FC) [20]: Each node of the fully connected layer is connected to all nodes of the upper layer to integrate the features extracted from the front. Due to its fully connected nature, the parameters of the fully connected layer are also the most. The l-th dense layer has its learnable parameters D. The output value can be expressed as shown in Equation (12).

$$X_o^l = f\left(X_i^{l-1} \times D_{io}^l + B^l\right) \tag{12}$$

IV. INTELLIGENT IDENTIFICATION METHOD FOR PQD

A. PQD RECOGNITION FEAMEWORK BASED ON CS AND DCNN

As shown in Figure 5, this paper introduces a recognition framework for power quality disturbance based on CS and Deep convolutional neural network (DCNN). Compared with the existing methods, the application of this method has the following advantages:

(1) Compressed sensing data: Using compressed sensing technology to collect power quality disturbance data, it can reduce the amount of data collected and reduce storage memory and transmission energy consumption.

(2) Automatic feature extraction [21]–[23]: The biggest feature of deep learning is the function of automatically extracting features, which can automatically find, combine and extract features from data for classification of power quality disturbances. Overcoming the traditional signal analysis and feature extraction process, and reducing the process of feature selection.

(3) A closed-loop feedback [19]: During the supervisory training process, the weight of each layer is automatically updated by feedback on classification performance. This is a fully closed loop feedback without any manual operation.

B. DEEP LEARNING ALGORITHM

In order to directly classify the compressed data with power quality disturbance, a deep learning network structure is designed according to the characteristics of the compressed data. The designed network structure is illustrated in Figure. 6.

Because PQD is a one-dimensional signal, the original two-dimensional convolution operation is no longer used, and one-dimensional convolution is more suitable for PQD classification, which can extract power quality data features more accurately.

1-D Convolutional (conv) Layer [24]: The convolutional layer is a key step in the extraction of deep learning self-learning features. The important features in the power quality disturbance compression data can be extracted through the convolutional layer. The number of filters in the first layer is Fl, and Xi is the input 1-D matrix (n × 1). The filter kernel is denoted as K ($k \times 1$). The convolutional layer output of the Fl filter can be expressed as shown in Equation (13):

$$X_{o,fl}^l = f\left(\sum_{i\in m} X_I^{L-1} \times K_{io,fl}^l + B^L\right) \tag{13}$$

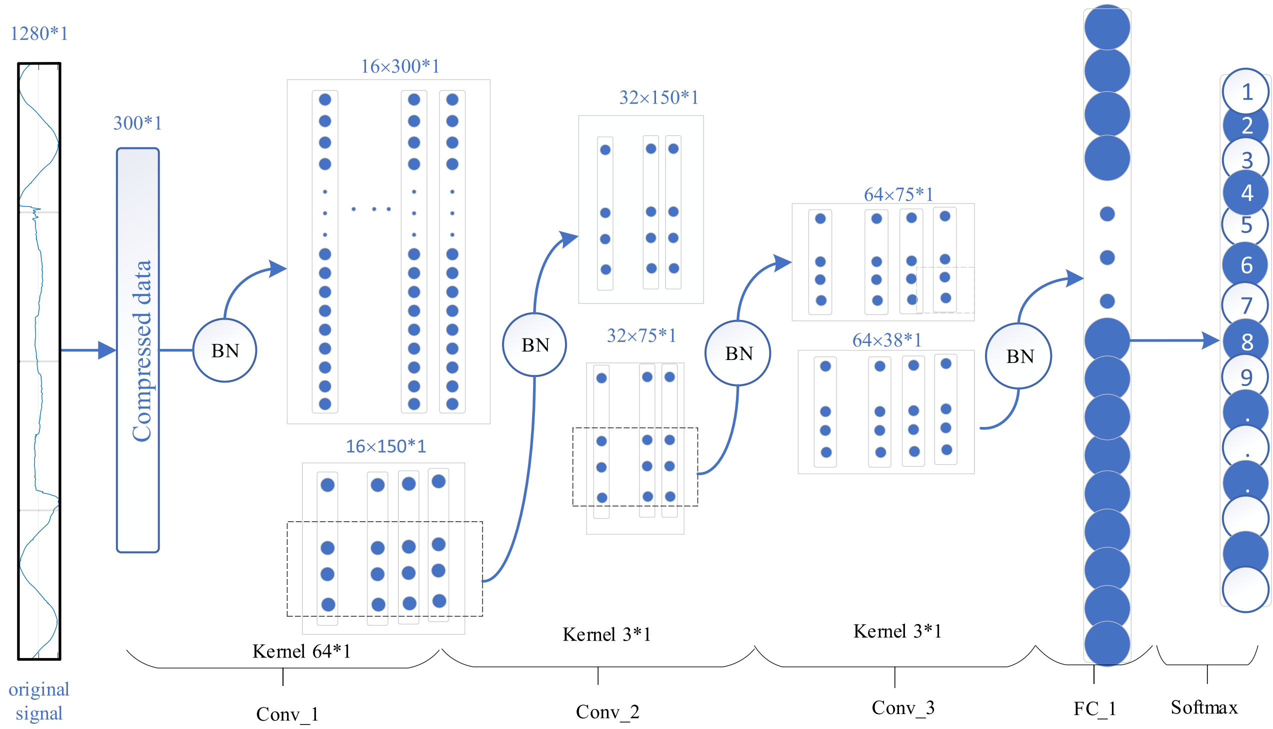

FIGURE 6. **Proposed 1D-DCNN structure.**

where $m = nk+1$ and $f(x)$ are activation functions. It implements the activation of neurons from input to output.

Power quality compression data acquisition and transmission process will produce noise interference. In order to suppress the influence of noise interference on classification accuracy, the network first layer adopts wide kernel design, and the convolution kernel is 64*1 to filter out noise. Because the wide kernel does not extract local features, the convolution in subsequent networks uses a 3*1 kernel design.

Since the acquisition of the existing power quality disturbance is completed in several cycles, the time is within 0.2 s, and the data dimension collected is small, so the level of the network structure cannot be too deep, and the network designed in the text is a 3-layer convolution layer, full connection layer and softmax layer. The shallow network structure is simpler, and its model is smaller and the parameter quantity is lesser. Therefore, the training and classification speed is fast, and the performance requirements of the device are low.

C. EASY TO IMPLEMENT TECHNIQUES FOE WEB TRAINING APPLICATIONS

Using deep learning classification training, a small trick can be applied to improve the performance of the network and prevent over-fitting of the network.

Batch normalization [19]: The processing not only improves the training speed, but also prevents over-fitting of the network. The BN layer is typically added after the convolutional layer or fully connected layer and before the activation unit. The conversion of the BN layer is described as shown in Equation (14-15):

$$\hat{y}^{l(i,j)} = \frac{y^{l(i,j)} - \mu_{\wp}}{\sqrt{\left(\sigma_{\wp}^2 + \varepsilon\right)}} \tag{14}$$

$$z^{l(i,j)} = \Upsilon^{l(i)}\, \hat{y}\, l\,(i,j) + \beta^{l(i)} \tag{15}$$

where $z^{l(i,j)}$ is the output of a neuron response, $\mu_{\wp} = E\left[y^{l(i,j)}\right] \sigma_{\wp}^2 = Var\left[y^{l(i,j)}\right]$, $\in$ is a small constant that increases the numerical stability, $\gamma^{l(i)}$ and $\beta^{l(i)}$ are the ratios and shifting parameters to be learned.

Advanced optimizer [20]: During training, the weight of each layer is updated by a function called "optimizer", such as stochastic gradient descent (SGD), Adam and AdaGrad. The different optimizer models require different optimizers. Determine the optimizer used by the network by analyzing the impact of different optimizers on the network.

Small and small batch training: Since the characteristics of PQD are complex, even the same type of interference is very different, and the network to be trained has good generalization performance. It was found that in practice, when an approximation of the gradient is calculated using a smaller batch and then the parameters are updated, the generalization ability of the model is better.

V. SIMULATION ANALYSIS

A. PQD DATA PREPARATION AND NETWORK PARAMETERS

In order to obtain compressed data, the parametric equations of fifteen PQD signals including pure sine waves have been

TABLE 1. Mathematical models of power quality disturbances.

Label	PQ disturbances	Mathematical equations	Parameters
C1	Normal	$y(t)=A\left[1\pm\alpha\left(u(t-t_1)-u(t-t_2)\right)\right]\sin(\omega t)$	$\alpha\le 0.1; T\le t_2-t_1\le 9T$
C2	Sag	$y(t)=A\left[1-\alpha\left(u(t-t_1)-u(t-t_2)\right)\right]\sin(\omega t)$	$0.1\le\alpha\le 0.9; T\le t_2-t_1\le 9T$
C3	Swell	$y(t)=A\left[1+\alpha\left(u(t-t_1)-u(t-t_2)\right)\right]\sin(\omega t)$	$0.1\le\alpha\le 0.8; T\le t_2-t_1\le 9T$
C4	Interruption	$y(t)=A\left[1-\alpha\left(u(t-t_1)-u(t-t_2)\right)\right]\sin(\omega t)$	$0.9\le\alpha\le 1; T\le t_2-t_1\le 9T$
C5	Harmonics	$y(t)=A[\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)$ $+\alpha_5\sin(5\omega t)+\alpha_7\sin(7\omega t)]$	$0.05\le\alpha_3,\alpha_5,\alpha_7\le 0.15;\sum\alpha_i^{\,2}=1$
C6	Spike	$y(t)=\sin(\omega t)+sign(\sin(\omega t))$ $\times\left\{\sum_{n=0}^{9}K\times\left[u\left(t-(t_1-0.02n)\right)-u\left(t-(t_2-0.02n)\right)\right]\right\}$	$0\le t_1,t_2\le 0.5T$ $0.01T\le t_2-t_1\le 0.05T$ $0.1\le K\le 0.4$
C7	Flicker	$y(t)=A\left[1+\alpha_f\sin(\beta\omega t)\right]\sin(\omega t)$	$0.1\le\alpha_f\le 0.2; 5\le\beta\le 20Hz$
C8	Oscillatory transient	$y(t)=A\left[\sin(\omega t)+\alpha^{-c(t-t_1)/\tau}\sin\omega_n(t-t_1)\left(u(t_2)-u(t_1)\right)\right]$	$0.1\le\alpha\le 0.8; 0.5T\le t_2-t_1\le 3T$ $8ms\le\tau\le 40ms; 300\le\omega_n\le 900Hz$
C9	Impulsive transient	$y(t)=A\left[1-\alpha\left\{u(t-t_1)-u(t-t_2)\right\}\right]\sin(\omega t)$	$0\le\alpha\le 0.414; T/20\le t_2-t_1\le T/10$
C10	Periodic notch	$y(t)=\sin(\omega t)-sign(\sin(\omega t))$ $\times\left\{\sum_{n=0}^{9}K\times\left[u\left(t-(t_1-0.02n)\right)-u\left(t-(t_2-0.02n)\right)\right]\right\}$	$0\le t_1,t_2\le 0.5T$ $0.01T\le t_2-t_1\le 0.05T$ $0.1\le K\le 0.4$
C11	Sag with harmonics	$y(t)=A\left[1-\alpha\left(u(t-t_1)-u(t-t_2)\right)\right]$ $\times\left[\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)+\alpha_5\sin(5\omega t)\right]$	$0.1\le\alpha\le 0.9; T\le t_2-t_1\le 9T$ $0.05\le\alpha_3,\alpha_5,\alpha_7\le 0.15;\sum\alpha_i^{\,2}=1$
C12	Swell with harmonics	$y(t)=A\left[1+\alpha\left(u(t-t_1)-u(t-t_2)\right)\right]$ $\times\left[\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)+\alpha_5\sin(5\omega t)\right]$	$0.1\le\alpha\le 0.8; T\le t_2-t_1\le 9T$ $0.05\le\alpha_3,\alpha_5,\alpha_7\le 0.15;\sum\alpha_i^{\,2}=1$
C13	Interruption with harmonics	$y(t)=A\left[1-\alpha\left(u(t-t_1)-u(t-t_2)\right)\right]$ $\times\left[\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)+\alpha_5\sin(5\omega t)\right]$	$0.9\le\alpha\le 1; T\le t_2-t_1\le 9T$ $0.05\le\alpha_3,\alpha_5,\alpha_7\le 0.15;\sum\alpha_i^{\,2}=1$
C14	Flicker with sag	$y(t)=A\left[1+\alpha_f\sin(\beta\omega t)\right]\sin(\omega t)$ $\times\left(1-\alpha\left(u(t-t_1)-u(t-t_2)\right)\right)$	$0.1\le\alpha_f\le 0.2; 0.1\le\alpha\le 0.9;$ $T\le t_2-t_1\le 9T; 5\le\beta\le 20Hz$
C15	Flicker with swell	$y(t)=A\left[1+\alpha_f\sin(\beta\omega t)\right]\sin(\omega t)$ $\times\left(1+\alpha\left(u(t-t_1)-u(t-t_2)\right)\right)$	$0.1\le\alpha_f\le 0.2; 0.1\le\alpha\le 0.9;$ $T\le t_2-t_1\le 9T; 5\le\beta\le 20Hz$

used to evaluate the classification performance of the proposed algorithm. PQD consists of 10 single types, namely pure sine wave, sag, swell, interrupt, harmonics, pulse transients, Oscillatory transient, flicker, notches and spikes. Five different types of PQD include harmonic and sag, harmonic and swell, harmonic with interruption, sag with flicker and swell with flicker. The parameter changes conform to the parameter equation of the IEEE-1159 standard [25]. The perturbation data is then converted to a sparse signal according to a compressed sampling process, with sparse signals as sparse data. Generate 180000 sets of data for training the network.

All the data is generated according to Table 1, and the original data is compressed and converted into the data required for verification. The noise situation of the data and the required data samples are provided in Table 2.

TABLE 2. Datasets used in the paper.

	Number of simples	Noise grade
Train set	12000×15=180000	From 20dB to 50dB
Validation set	180000×0.1=18000	From 20dB to 50dB
Test set	4000×4=16000	20dB,30dB,40dB and 50dB

The data set uses compressed data. For more intuitive representation of the data, two kinds of disturbance compression data are given, as shown in Figure. 7.

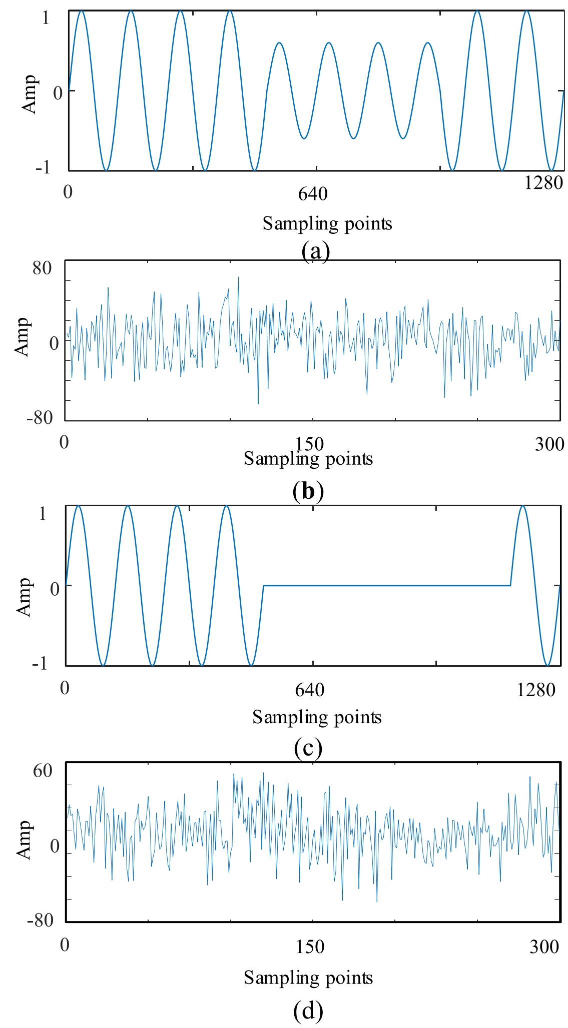

FIGURE 7. (a) Sag. (b) Sag compressed data. (c) Interruption. (d) Interruption compressed data.

It can be seen from Figure. 7 that the waveform of the compressed data has a large gap, and it can be seen from the amplitude that the sag signal is higher than the interrupt signal.

The network is trained according to the collected compressed sampling data. The trained network structure is shown in Table 3. The network parameters of each layer and the convolution kernel used are clearly stated.

B. SIMULATION DATA FOR RESULT VERIFICATION

Train and classify the network by using the data in Section 4.1. The performance of the 1D-DCNN is assessed by log loss and classification accuracy in the training set and test set. We further analyze the performance of the proposed method. In Figure 8, the loss and performance curves during model training are shown. The pink and black descent curves correspond to the loss function values of the training and validation sets during training. We further analyze the performance of the proposed method.

The model was trained through 30 iterations. It can be seen from Figure. 8 that the test accuracy and the training accuracy are almost equal after the 25th iteration, and the highest classification accuracy is 99.5% or more. Moreover, the loss is less than 0.02. Figure 8 demonstrated that the network model has good performance and there is no over-fitting phenomenon.

In order to get a good performance network, we need to determine the impact of the learning rate on the network. A variety of learning rates were set up and verified by classification results. It was found that the model training with the learning rate of 0.0001 and 0.00001 had the best training effect, though the training of the model was slow. When the learning rate was 0.001, the comprehensive performance of the training was better. Considering the efficiency and classification accuracy of the instrument, this paper uses a learning rate of 0.001. The simulation verification process is shown in Figure 9.

In Figure 10, the convergence of three different optimizers is illustrated based on the verification losses over the same period. AdaGrad converges quickly, but it quickly stops

TABLE 3. Deep learning network structure design.

NO	Layer type	Kernel size/ strides	Kernel channel size	Output size	Padding
1	Conv1	64*1/8*1	16	38*16	Yes
2	Pooling1	2*1/2*1	16	19*16	No
3	Conv2	3*1*1	32	19*32	Yes
4	Pooling2	2*1/2*1	32	9*32	N0
5	Conv3	3*1*1	64	9*64	Yes
6	Pooling3	2*1/2*1	64	4*64	No
7	FC	100	1	100*1	*
8	Softmax	15	1	15	*

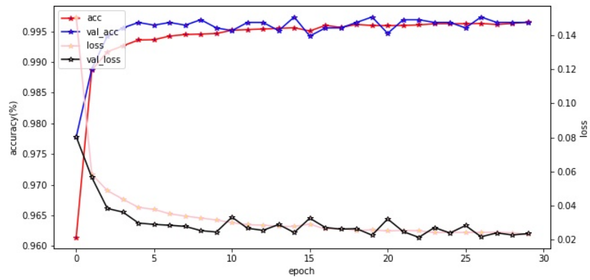

FIGURE 8. Classification results and loss.

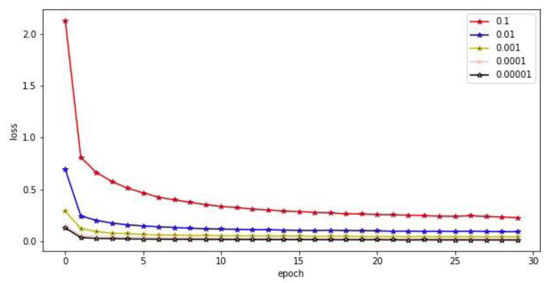

FIGURE 9. The effect of different learning rates on classification results.

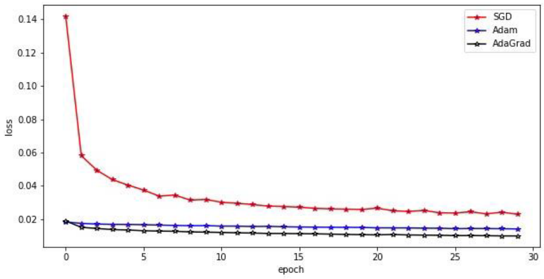

FIGURE 10. The effect of different optimizers on classification effect.

improving because of the rapid decline in learning rates. The effect of SGD is also poor. Adam and AdaGrad seem to behave almost the same, but here we choose Adam, whose stable behavior is independent of its parameters.

C. COMPARISON OF CS + DCNN WITH EXISTING METHODS

Although the algorithm in this paper takes a relatively long time in training, it has a very high speed in classification prediction. In order to verify the effectiveness of the proposed algorithm, we also verify the advantages of the proposed algorithm compared to the traditional method, and compare the gaps of the three algorithms in training time. The existing algorithms used in this paper are extreme learning machines (ELM) and hierarchical extreme learning machines (H-ELM). Since the traditional algorithm has feature extraction and feature selection process, this paper only compares the performance of the classifier. The classification time of the traditional algorithm does not calculate the feature extraction and selection process. Table 2 shows a comparison of the existing methods and the method classification time in the text.

The PNN, ELM and H-ELM in the table are extracted by EWT feature extraction method, and the extracted feature data is referenced [4]. Moreover, the calculated time is the classification time of 4500 data sets. As is shown in Table 4 the classification time of CS-DCNN is obviously faster than that of the high-performance classifier of the extreme learning machine, and the classification of the compressed data is directly used, and the feature selection process is also reduced. Therefore, in terms of classification speed, the method in this paper is far superior to the traditional classification method.

Compressed sensing technology for data collection is faster than traditional acquisition methods. The classification efficiency of the classifier is verified in the paper, and the classification speed is faster than the traditional machine learning algorithm. The method in this paper is aimed at directly classification of the collected raw data, and does not require pre-processing. Therefore, online processing is theoretically feasible.

In order to verify the classification accuracy of the proposed method and the comparison with several existing methods, the same data set is classified and verified.

TABLE 4. Algorithm running time comparison.

	ELM	H-ELM	CS+DCNN
20dB	1.106	0.864	0.067
30dB	1.134	0.863	0.066
40dB	1.109	0.738	0.066
50dB	1.107	0.953	0.064

TABLE 5. Comparison of classification accuracy of six methods.

	50dB	40dB	30dB	20dB
PCA-SVM	98.50	97.41	96.30	95.10
ELM	98.61	97.20	94.91	91.30
PNN	94.67	94.30	93.60	91.20
H-ELM	98.85	97.60	96.20	93.21
PSO-H-ELM	99.00	98.10	97.60	95.22
CS-DCNN	99.99	99.94	99.81	99.68

Table 5 is a comparative analysis of CS-DCNN and existing algorithms. The comparison between the two is used to verify the feasibility of the proposed algorithm. Compared with the existing perturbation classification method, the accuracy of disturbance recognition is significantly improved. The classification results clearly showed that our method has almost 100% classification accuracy, and the signal-to-noise ratio has little effect on the classification performance. The 20dB SNR data also has high classification performance, which is more suitable for practical use. The power quality disturbance is categorized. Synchronously, it also verified the role of the first-layer network wide kernel convolution mentioned in the paper, and its anti-noise performance is good.

D. MEASURED DATA VERIFICATION AND DATA ENHANCEMENT OPERATIONS

In order to further verify the feasibility of the proposed method in the actual signal, in this part, a set of actual signals is used to test the analysis of the measured signals by the CS and DCNN classifiers. The data set is provided by the IEEE PES database [26], [27] for PQD classification. The data set has been tested for power quality classification in reference [28] to meet the needs of the experiment. The sampling rate of the supplied signal is 256 points per cycle. Each signal has a length of 1536. The measured data is compressed and reconstructed, and the reconstructed result is shown in Figure 9. As demonstrated by figure 11, the compressed reconstructed signal can fully represent the original waveform features, and the original waveform and the reconstructed signal have certain errors, but do not affect the analysis of the power quality disturbance.

Disturbance in the power system is unbalanced. For example, the sag disturbance type accounts for more than 80% of all disturbance types, and the data is unbalanced. Some disturbance signals are flicker disturbances, and the amount

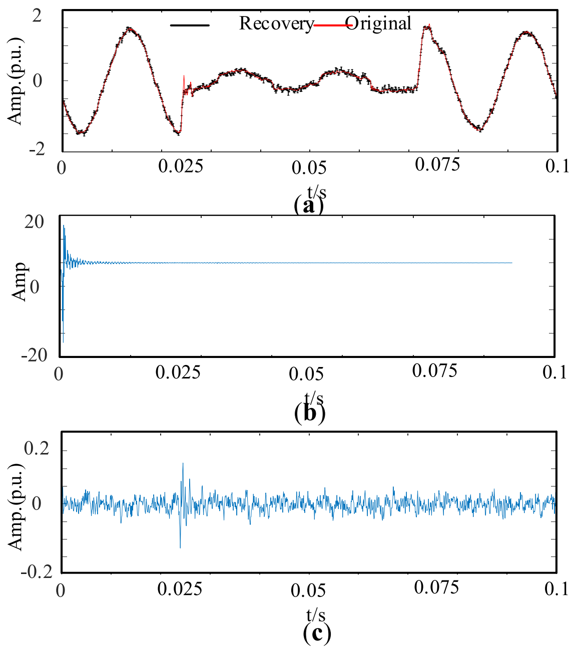

FIGURE 11. (a) Sag compression reconstruction. (b) IDCT sparse representation. (c) compression reconstruction error.

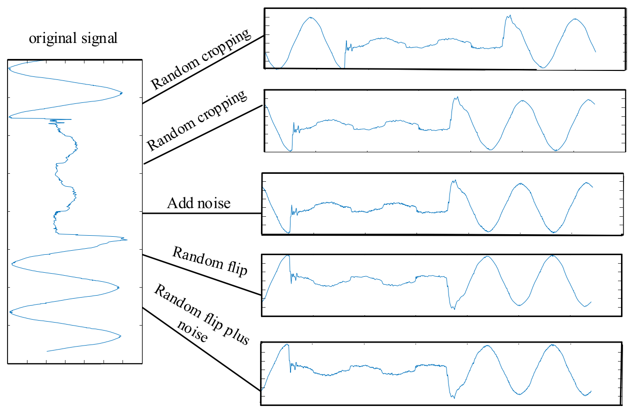

FIGURE 12. Power quality measured signal data enhancement operation.

of disturbance data is small. Directly using data for training will result in weak generalization ability and large classification errors. The amount of data in the power system is large, but the labeling of the data requires a lot of manpower and material resources. The training of deep learning networks requires a large amount of data, and the existing data volume is not enough to train a better network. In order to solve these two problems in the measured data, this paper proposes a data enhancement method for preprocessing data. In computer vision, data enhancement is commonly used to increase the number of training samples to enhance the generalization performance of the classifier. In this paper, for the data imbalance problem, the data enhancement method is adopted, the data volume is equalized, and the data enhancement operation is performed on the interference such as flicker with small data volume. The data enhancement operation mainly adopts random cropping, moderately increases random noise, inverts signals, etc., and performs random extraction and verification on all operation signals to ensure that the data after data enhancement belongs to interference type data. The data enhancement operation is shown in Figure 12.

The average accuracy of the measured data is 91%, which is much lower than the simulation result. The main reason is that the data set used has less data. If the data is over-enhanced, over-fitting will occur. Therefore, even with data enhancement operations, it is not enough to train a well-performing network. The measured data is more complicated, the same type of disturbance is more different, and the amount of data is smaller, therefore, it is difficult to train a network with better generalization performance. Although the classification result is poor, it is also within the allowable range of power quality disturbance recognition. Recognition accuracy is 1.5% higher than reference 28.

VI. CONCLUSION

A classification model based on compressed sensing and deep convolutional neural networks is proposed for complex power quality disturbance identification.

1) Analyze the single disturbance and compound disturbance of power quality, determine the acquisition method suitable for power quality disturbance signal, reduce the amount of data collected and improve the transmission efficiency in data acquisition.

2) Using the deep learning algorithm to directly classify the compressed data, the proposed end-to-end intelligent algorithm is suitable for compressing data, which can solve the compression error and eliminate the influence of high noise.

3) Simulation and measured data verify the feasibility, classification accuracy and robustness of the compression and deep learning classification power quality disturbances. The algorithm performs well in noisy environments and works directly on the original noise signal without any prior restoration methods.

Of course, there are still many shortcomings in the algorithm. In the case of compressed sampling, when the signal disturbance is more complicated, the recovery and reconstruction errors of some signals are larger, and further research on the compression acquisition device is needed.

ACKNOWLEDGMENT

This work is supported by SGCC program: Research on Extensive Application and Benefit Evaluation of Typical Power Substitution Technology Considering Power Quality Influence (52182018000H); Innovation fund program of Lanzhou Jiaotong University-Tianjin University (Research on Management Strategies of Power Quality Problems from Thermostatically Controlled Appliances).

REFERENCES

[1] X. Shi, H. Yang, Z. Xu, X. Zhang, and M. R. Farahani, "An independent component analysis classification for complex power quality disturbances with sparse auto encoder features," *IEEE Access*, vol. 7, pp. 20961–20966, 2019.

[2] F. Yang, J. Tan, J. Song, and Z. Han, "Block-wise compressive sensing based multiple line outage detection for smart grid," *IEEE Access*, vol. 6, pp. 50984–50993, 2018.
[3] H. Wang, P. Wang, and T. Liu, "Power quality disturbance classification using the S-transform and probabilistic neural network," *Energies*, vol. 10, no. 1, pp. 107–126, Jan. 2017.
[4] 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.
[5] R. Kumar, B. Singh, D. T. Shahani, A. Chandra, and K. Al-Haddad, "Recognition of power-quality disturbances using S-transform-based ANN classifier and rule-based decision tree," *IEEE Trans. Ind. Appl.*, vol. 51, no. 2, pp. 1249–1258, Mar./Apr. 2015.
[6] O. P. Mahela and A. G. Shaik, "Recognition of power quality disturbances using S-transform based ruled decision tree and fuzzy C-means clustering classifiers," *Appl. Soft Comput.*, vol. 59, pp. 243–257, Oct. 2017.
[7] P. D. Achlerkar, S. R. Samantaray, and M. S. Manikandan, "Variational mode decomposition and decision tree based detection and classification of power quality disturbances in grid-connected distributed generation system," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3122–3132, Jul. 2018.
[8] C. Aneesh, S. Kumar, P. M. Hisham, and K. P. Soman, "Performance comparison of variational mode decomposition over empirical wavelet transform for the classification of power quality disturbances using support vector machine," *Procedia Comput. Sci.*, vol. 46, pp. 372–380, Dec. 2015.
[9] D. De Yong, S. Bhowmik, and F. Magnago, "An effective power quality classifier using wavelet transform and support vector machines," *Expert Syst. Appl.*, vol. 42, pp. 6075–6081, Sep. 2015.
[10] Z. Liu, Y. Cui, and W. Li, "A classification method for complex power quality disturbances using EEMD and rank wavelet SVM," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1678–1685, Jul. 2015.
[11] P. Kanirajan and V. S. Kumar, "Power quality disturbance detection and classification using wavelet and RBFNN," *Appl. Soft. Comput.*, vol. 35, no. 4, pp. 470–481, Oct. 2015.
[12] S. Khokhar, A. A. B. Mohd Zin, A. S. B. Mokhtar, and M. Pesaran, "A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances," *Renew. Sustain. Energy Rev.*, vol. 51, pp. 1650–1663, Nov. 2015.
[13] A. Ravelomanantsoa, H. Rabah, and A. Rouane, "Compressed sensing: A simple deterministic measurement matrix and a fast recovery algorithm," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 12, pp. 3405–3413, Dec. 2015.
[14] M. Majidi, A. Arabali, and M. Etezadi-Amoli, "Fault location in distribution networks by compressive sensing," *IEEE Trans. Power Del.*, vol. 30, no. 4, pp. 1761–1769, Aug. 2015.
[15] Z. He, L. Zhou, Y. Yang, Y. Chen, C. Liu, and X. Ling, "Compressive sensing-based channel estimation for FBMC-OQAM system under doubly selective channels," *IEEE Access*, vol. 7, pp. 51150–51158, 2019. doi: 10.1109/ACCESS.2019.2898896.
[16] M. S. Manikandan, S. R. Samantaray, and I. Kamwa, "Detection and classification of power quality disturbances using sparse signal decomposition on hybrid dictionaries," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 27–38, Jan. 2015.
[17] H. Wu, M. Suo, J. Wang, P. Mohapatra, and J. Cao, "A holistic approach to reconstruct data in ocean sensor network using compression sensing," *IEEE Access*, vol. 6, pp. 280–286, 2018.
[18] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. 42nd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2008, pp. 581–587.
[19] S. Wang and H. Chen, "A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network," *Appl. Energy*, vol. 235, pp. 1126–1140, Feb. 2019.
[20] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425–440, Feb. 2017.
[21] H. Liu, F. Hussain, Y. Shen, S. Arif, A. Nazir, and M. Abubakar, "Complex power quality disturbances classification via curvelet transform and deep learning," *Electr. Power Syst. Res.*, vol. 163, pp. 1–9, Oct. 2018.
[22] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[23] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Comput. Biol. Med.*, vol. 100, pp. 270–278, Sep. 2018.
[24] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
[25] *IEEE Recommended Practice for Monitoring Electric Power Quality*, IEEE Standard 1159-2009, Jun. 2009.
[26] IEEE Power Engineering Society. *IEEE PES Working Group P1433 Power Quality Definitions*. Accessed: Feb. 2, 2001. [Online]. Available: http://grouper.ieee.org/groups/1159/2/testwave.html
[27] IEEE Power Engineering Society. *IEEE PES Working Group Power Quality Definitions*. Accessed: Oct. 12, 2012. [Online]. Available: http://grouper.ieee.org/groups/1159/3/docs.html
[28] Y. Luo, K. Li, Y. Li, D. Cai, C. Zhao, and Q. Meng, "Three-layer Bayesian network for classification of complex power quality disturbances," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3997–4006, Sep. 2018.

JIDONG WANG (M'16) received the B.S. and M.S. degrees from Shandong University of Technology and Shandong University, Shandong, China, in 1999 and 2002, respectively, and the Ph.D. degree in electrical engineering from Tianjin University, Tianjin, China, in 2005. He held postdoctoral position, from 2005 to 2007, and became an Associate Professor with Tianjin University, from 2007 to 2017. He is currently a Professor with the School of Electrical and Information Engineering, Tianjin University. His research interests include power quality, distributed generation, microgrid, and smart power consumption.

ZHILIN XU was born in Shandong, China, in 1993. He received the B.S. degree in electrical engineering and automation from Hebei University of Science and Technology, Hebei, China, in 2017. He is currently pursuing the M.S. degree in electrical engineering with Tianjin University, Tianjin, China. His main research interests include power quality and artificial intelligence.

YANBO CHE (M'16) received the B.S. degree from Department of Electrical Engineering, Zhejiang University, Hangzhou, China, in 1993. He received the Ph.D. degree from School of Electrical Engineering and Automation, Tianjin University, Tianjin, China, in 2002. He is currently an Associate Professor of Tianjin University, Tianjin, China. His research interests include power quality of power system, microgrid, and energy management.