Ensemble deep learning for automated classification of power quality disturbances signals

Jidong Wang [a], Di Zhang [b], Yue Zhou [c],[*]

[a] *Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin, 300072, China*
[b] *State Grid Shenyang Electric Power Supply Company, Shenyang, 110811, China*
[c] *School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK*

ABSTRACT

The automatic classification of power quality disturbances (PQD) is of great significance for solving power quality problems. In this study, we propose an ensemble deep learning framework to realize intelligent classification of PQ disturbances. Specifically, based on the characteristics of the sequence of disturbance signals, the Long Short Term Memory (LSTM) network is used to classify the signals. In addition, the Bagging theory is used to integrate the training results of multiple LSTM networks to improve the generalization of the network. Our contribution lies in the combination of deep learning and ensemble learning to extract the classification representation of PQD signals. In view of the large number of unlabeled power quality disturbance samples in the power grid, the active learning strategy is adopted to select the most representative samples from the data set, which can enhance the model performance with less labeled data. Finally, experiments were conducted in different noise environments. Compared with the existing multi-label learning models, this method achieves better classification performance with good calculation speed. Furthermore, the proposed active learning strategy is able to train the classification model with fewer labeled samples, reducing the manual labeling costs.

1. Introduction

With the increasing application of solid-state switches, non-linear devices and various renewable power generation connected to the smart grid, there are increasing fluctuations in voltage, current and frequency in power systems. These interferences bring great challenges for the safe and economic operation of power systems [1]. On the other hand, with the increasing power consumption due to economic development and electrification of heat and transport, electricity consumers and electrical equipment also put forward higher requirements on the quality of power supply. Therefore, accurate and effective classification of power quality disturbance (PQD) signals is of increasing significance for power quality (PQ) evaluation and management [2].

In the past few decades, researchers have conducted detailed explorations on the detection and classification of PQD signals. Conventional disturbance classification methods consist of three steps, i.e., signal processing, feature selection and classifier design [3]. For signal analysis, many methods and improved algorithms have been proposed based on signal processing technology, including short-time Fourier transform, (STFT), S transform (ST), Wavelet transform (WT),

Variational Mode Decomposition (VMD), Empirical Mode Decomposition (EMD), and Hilbert Huang Transform (HHT). The signal processing methods are dedicated to improving the time-frequency resolution of the signals, but the above methods have different degrees of limitations. STFT is difficult to weigh the time-frequency resolution, and the effect of capturing signal transient characteristics is not good [4]. The series of WT methods are more suitable for the signal processing of PQDs, but its classification performance depends heavily on feature selection and classifier design [5]. EMD lacks the support of mathematical theory and is prone to false components [6]. Besides, all these methods are prone to be affected by the noise and the computation burden is heavy. Therefore, a signal processing algorithm with wider applicability and better anti-noise ability should be adopted.

Feature selection is a critical step to find representative features from PQD signals. The extraction of key features can significantly increase the classification speed of the classifier while ensuring accuracy. Currently, various manual features have been designed to represent PQD signals. The design of the features lacks reference, it is still not very clear why these features are chosen, what the correlation between the features is, and whether these features help distinguish different types of PQD

---

signals. To solve the above problems, some feature optimization algorithms have been proposed. The particle swarm optimization (PSO) algorithm is used to optimize the parameters to filter out the optimal feature subset in [7]. In [8], a Probabilistic Neural Network based Artificial Bee Colony (PNN-ABC) is proposed to optimize the feature selection. The non-optimality of manual feature selection and the time-consuming feature of the optimization process highlight the need of a more effective method to select meaningful and distinct morphological features and realize automatic construction and optimization of the feature set.

Besides the limitations in signal processing and feature selection as described above, the three steps of traditional PQD classification algorithms are carried out independently with a limited combination of signal processing methods and classifiers, limiting the classification accuracy. Therefore, the disturbance identification process should be optimized as a whole to simplify the analysis and improve the calculation efficiency while ensuring the accuracy.

Deep learning is an emerging and promising classification framework, which can learn features and capture patterns from multi-level abstract signals (such as PQD signals) [9]. The Long Short Term Memory (LSTM) network is a special recurrent neural network (RNN) in the field of deep learning. It has achieved great performance in dealing with text classification, sentiment analysis, and time series prediction. Considering PQD signals are in nature time series data, the LSTM network is proposed to be used for PQD classification in this paper. The main work and contributions of this paper are as follows:

(1) Based on the characteristics of PQD signals, a new method for PQD classification based on ensemble learning and deep learning is proposed, which combine the three steps of traditional methods without using additional feature selection module, thus simplifying the classification steps.

(2) Based on the Bagging algorithm, we trained multiple LSTM base classifiers and improved the traditional majority voting strategy. The best-worst weighted voting strategy is proposed to better improve the network generalization performance. Performance of the proposed method is verified in comparative experiments.

(3) To solve the problem of less measured data on labels, the active learning algorithm is adopted in sample selection to enhance the utilization efficiency of unlabeled samples. By actively selecting the samples with the most abundant information, the need of annotation for deep neural network can be reduced.

The rest of this paper is organized as follows. The proposed Bagging-LSTM is presented in Section 2, in which the weighted voting strategy is embedded in the Bagging-LSTM. We then introduce the active learning strategy in Section 3. Comparative experiments are presented in Section 4. Finally, the conclusion of this paper is given in Section 5.

2. Bagging-LSTM network framework

2.1. Regular LSTM

LSTM is a special recurrent neural network applied in the field of deep learning, which is an improvement of the RNN model. The key of the LSTM model is to introduce a memory unit for cyclic information transmission, recording all historical information up to the current moment. Therefore, compared with the short-term memory of traditional RNN, LSTM has long-term memory capabilities: a gating mechanism (involving forgetting gate, input gate and output gate) with a value between (0,1) is used to control the transmission path of the internal information in the model [10].

The structure of the cyclic unit is shown in Fig. 1, where x_t, f_t, i_t, o_t, c_t, a_t and h_t represent the input, forget gate, input gate, output gate, memory storage unit, candidate state, and output of the hidden layer at time t, respectively. c_{t-1} and h_{t-1} are the output of memory storage unit and hidden layer at time t-1. δ is the logistic sigmoid function, and tanh is the activation function.

We denote the connection weight of the forgetting gate, input gate, output gate, and memory unit as W_{xf}, W_{xi}, W_{xo}, and W_{xc} respectively. We also use W_{hf}, W_{hi}, W_{ho} and W_{hc} to represent the weights between the hidden layer and the forget gate, input gate, output gate, and memory unit. b_f, b_i and b_o represent the bias. $\odot$ represents the product of vector elements. The LSTM loop structure unit controls the flow of information by controlling the degree of opening and closing of the forgetting gate, input gate and output gate. The specific process is shown in the following Step 1 to Step 6.

Step 1: The forgetting gate f_t takes the input x_t of the current layer and the output h_{t-1} of the hidden layer at the previous moment as input, and the output result of the forgetting gate is multiplied by c_{t-1} to control how much information needs to be forgotten in the internal state c_{t-1} at the previous moment. The forget gate is expressed as

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{1}$$

Step 2: The input gate selectively saves the current input information, and the output result i_t is used as the information to be updated. The input gate is obtained as

$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right) \tag{2}$$

Step 3: The output gate o_t controls the internal state c_t at the current moment, to control how much information needs to be output to the external state h_t. The output gate is expressed as
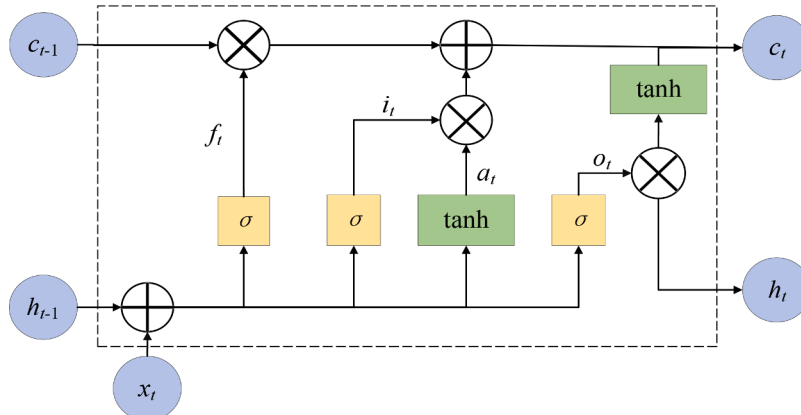

Fig. 1. Structure diagram of LSTM neural network.

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{3}$$

Step 4: The output gate o_t is multiplied by the state of the memory unit processed by the tanh layer. The output h_t of the hidden layer is calculated as

$$h_t = o_t \tanh \odot (c_t) \tag{4}$$

Step 5: The memory unit c_t records the historical information up to the current moment, which can be calculated by

$$c_t = f_t \odot c_{t-1} + i_t \odot a_t \tag{5}$$

Step 6: The candidate state is calculated by

$$a_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{6}$$

2.2. Ensemble learning

L. K. Hansen first proposed the concept of ensemble learning, which improved the generalization performance of learners by constructing multiple neural networks [11]. The main learning modes of the ensemble include Bagging, Boosting and Stacking. Bagging focuses on reducing the overall variance, while Boosting and Stacking focus on reducing model bias. To use an artificial neural network or a deep learning method in the classification of PQD signals, it is usually more likely to obtain an overfitting model rather than an underfitting model. Therefore, reducing the variance is the main consideration to improve the performance of the artificial neural networks model, and thus a Bagging-like ensemble model is used in this study [12].

The Bagging algorithm is shown in Fig. 2. At first, T sample sets are obtained by random sampling T times, and T independent weak learners (also known as base classifiers) are developed respectively. Then, through the voting strategy of T weak learners, the final strong learners can be obtained. To ensure an ensemble model makes sense, the diversity of the base classifiers is a key factor. Instead of structural variations, different training datasets are fed into LSTM to meet the diversity requirement. Bootstrapping method is commonly used in the random sampling algorithm. For the original training set of M samples, by random sampling of M times with replacement method, a sample set containing M samples is obtained. In this way, the training datasets of each base classifier is different from the original training set and other sample sets, which guarantees the diversity of each base classifier in

ensemble learning.

2.3. An improved bagging algorithm

In traditional Bagging algorithms, the majority voting method is usually used to make decisions, which makes all the base classifier have the same decision-making power and ignores the performance differences of the base classifiers [13]. This is the main limitation of the majority voting methods. Generally, the selected classifiers have different competence. Therefore, the weighted voting method is used to aggregate the selected classifiers decision. In this method, the output of each classifier is weighted by coefficients that affect the combination process. Assuming w_i is the weight of the i th classifier, the weighted majority voting is represented as follows:

$$WMV(x) = \max_{j \in \Omega} \sum_{i=1}^{L} w_i C_{ij} \, and \sum_{i=1}^{L} w_i = 1 \tag{7}$$

where $C=[C_1, C_2, \ldots, C_L]$ is a collection of L classifiers; x is the input; C_{ij} is the output of the j-th category of the i th classifier; w_i is the weight of the i th classifier.

In this study, the best-worst weighted voting method is adopted as a measure to quantify the weights [14]. The basic idea of this method is to identify the worst and the best members on the ensemble using their estimated errors on the validation set. The relative accuracy of the i th individual learner, a_i, is expressed as

$$a_i = 1 - \frac{e_i - e_b}{e_w - e_b} \tag{8}$$

where e_w and e_b are the maximum and minimum error rates of all individual learners (error rate = 1 - accuracy rate); e_i is the error rate of the i th individual learner.

Then the weighted of every classifier is expressed as

$$w_i = \frac{a_i}{\sum\limits_{i=1}^{L} a_i} \tag{9}$$

2.4. Bagging-LSTM framework for classification of PQD signals

The network structure of the LSTM and the improved Bagging algorithm are introduced in the previous sections, based on which we propose a PQD signal classification framework. The framework of the proposed Bagging-LSTM is presented in Fig. 3.

The algorithm uses the LSTM network as the base classifier, and integrates the results of multiple LSTM networks with differences through the Bagging algorithm to form a strong classifier. To a certain extent, the
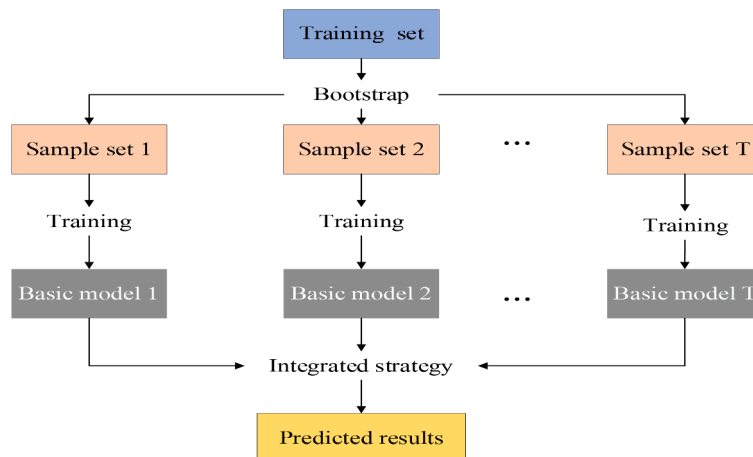

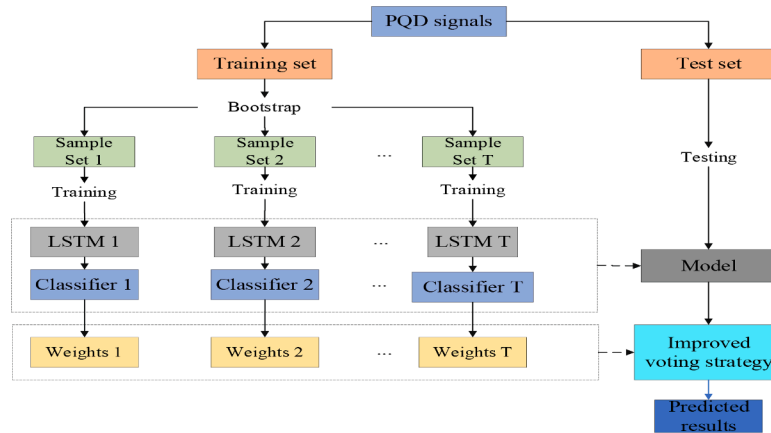**Fig. 2.** Illustration of Bagging algorithm.

Fig. 3. The classification model based on Bagging-LSTM.

Bagging-LSTM model can be seen as a kind of extension of LSTM. In order to verify the superiority of the ensemble deep learning, the model is developed on the basis of the original LSTM network instead of other complex models. The accuracy of the Bagging-LSTM model can be significantly improved and the generalization of the network can be effectively improved. The specific implementation steps are as follows:

1) According to the 'ten-fold cross validation principle', 10% of the training data of the PQ disturbance signal is split as a verification set.
2) Bootstrapping: T training samples are obtained by random sampling with replacement from the training set.
3) For the t-th training sample, the LSTM model is used to train the t-th classifier, which is the t-th base classifier.
4) Train the base classifier and collect the accuracy data from using the validation set corresponding to the T-th training sample, so as to calculate the weight of the T-th weak classifier.
5) Link all base classifiers with the corresponding weight distributions, and use the best-worst weighted voting strategy to predict the category of the test set and calculate the accuracy.

3. Active learning

Most of the sample data sets used to train the classifier are generated by MATLAB software in the existing PQD classification research, and a number of training samples are randomly generated, so that we can easily label the sampled signals obtained. However, the detectors deployed in the power grid in practice are collecting signals on the transmission lines all the time, and labeling the collected samples of power quality events requires manual operation by experts. This work is costly and time-consuming, and it is easy to misjudge under the condition of visual fatigue. Therefore, how to train the classification model with as few labeled samples as possible, so that the training time and size of the model can be reduced, is a problem that we need to further explore.

The active learning technology can interact with experts to select the most valuable uncalibrated data for experts to label, so that the model can obtain better performance with less labeled data. The active learning algorithm is a cyclic iterative process, and its model is mainly composed of four parts, as shown in Fig. 4, including target classification model, labeled sample set, unlabeled sample pool, and human experts [15]

We use the uncertainty sampling strategy to select the unlabeled sample, which is widely used in active learning. In this strategy, the classifier gives each unlabeled sample a confidence score as the evaluation score to determine its uncertainty. The confidence score is expressed as:

$$x_{LC} = \text{argmax}_{i=1,2,\ldots,n} 1 - P(\widehat{y}|x_i) \qquad (10)$$
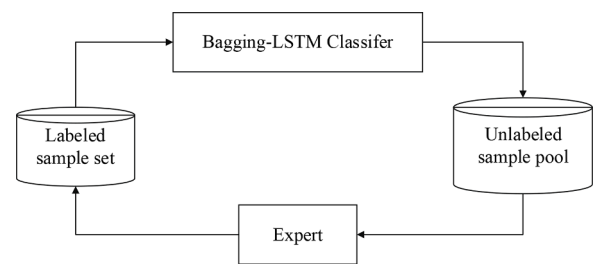

Fig. 4. Illustration of active learning algorithm.

where $\widehat{y}$ is the category with the highest predicted probability of the model for sample x_i. Then sort the confidence scores of the unlabeled samples, and select the sample with the lowest trust level, which is most likely to be misclassified by the model for expert annotation. Then add these labeled samples to the labeled sample set and use the updated labeled sample set to train the model again. In this way, the labeled sample set is updated, and the model is continuously trained until the model performance reaches a certain standard, and then the iteration is stopped. The active deep learning model is fine-tuned incrementally to speed up the convergence of the model.

4. Experimental result

4.1. PQD data preparation

Fifteen PQD signals including pure sine waves are used to evaluate the classification performance of the proposed algorithm. The PQD signal is composed of 10 single types, including pure sine waveform (C1), sag (C2), swell (C3), interruption (C4), harmonic (C5), pulse transient (C6), oscillation transient (C7), flicker (C8), notches (C9) and spikes (C10). Five complex types of PQD are harmonic and sag (C11), harmonic and swell (C12), harmonic with interruption (C13), sag with flicker (C14) and swell with flicker (C15). The parameter changes conform to the parameter equation of the IEEE-1159 standard [16]. The basic mathematical models of single disturbance are referred to [17].

The sampling frequency is set to 6.4 kHz, and the data length of a single sample is 1280 points. The label data is represented by a one-hot encoding, such as [1,0,0,0,0,0,0,0,0,0,0,0,0,0,0] (indicating that the sample belongs to the first type). 12,000 samples have been generated. However, there is always noise in the data collected in the power grid. In order to enhance the anti-noise performance of the proposed method, Gaussian white noise is randomly added to the synthetic PQDs data of different levels. The signal-to-noise ratio (SNR) varies from 20 dB to 50 dB Then perform Ten-fold cross-validation, and split 10% of the training

data into a validation set.

4.2. Optimizing the bagging-LSTM neural network

We build the Bagging-LSTM neural network and set the model parameters as follows: the number of base classifiers is 3, the epoch is 100, the batch-size is 64, and the number of LSTM hidden neurons is 64. The cross-entropy loss and Adam optimizer are used to update the model parameters. The learning-rate is 0.001. The length of the PQD input is determined by the timestep. The network is trained by using the data without noise interference. The consuming time represents the time it takes for the network to update the parameter weight once. Table 1 shows the performance comparison of the network under different timesteps.

To explore the impact of the timesteps, the accuracy, loss and training time with different timesteps are verified. We can see that, the training time increases as the timestep value increases. However, the training accuracy and loss first increase and then decrease, which indicates that choosing appropriate timestep is very important to the performance of the network. A too short timestep may not contain enough disturbance information, resulting in low accuracy, while a too long timestep can cover complete information but with low efficiency. Balancing the accuracy of network classification and training time, we choose timesteps as 10.

In order to verify the superiority of the proposed improved voting strategy, we compared the best-worst weighted voting strategy used in our study (Vote 1) and the traditional majority voting strategy (Vote 2). The base classifiers trained by different samples are named as LSTM 1, 2, 3 respectively. The comparison results of the two voting strategies are shown in Fig 5.

The results are the average of 10 training sessions. It can be seen from Fig. 5 that the accuracy of using the voting strategy proposed in this paper is higher than that of the traditional majority voting strategy. In other words, the Bagging algorithm participates in decision-making through multiple classifiers, and obtains better generalization performance than a single classifier.

4.3. Comparing the bagging-LSTM with existing methods

The proposed Bagging-LSTM method is compared with four other deep learning methods, which are briefly introduced as follows:

(1) Deep Convolutional Neural Network (DCNN): The deep convolutional neural network with six standard convolutional layers proposed in [5] is used as a reference method. The convolution kernel is set to 3, the stride is 1, and the number of convolutional layer filters is set to 32, 32, 64, 64, 128 and 128. so that the DCNN has a huge number of parameters.
(2) Convolutional Auto-encoder (CAE) proposed in [9] is used as another reference method. The down-sampling factor is set to 2, 8 and 2, and the parameter setting of the decoder is symmetrical to the encoder.
(3) Gated recurrent unit (GRU): GRU is also an RNN. It cancels the separate memory cell, and shortens the training time compared to LSTM. The GRU used is almost the same as the Bagging-LSTM except that the LSTM layers are replaced with GRU layers.

Table 1
The performance with different timesteps.

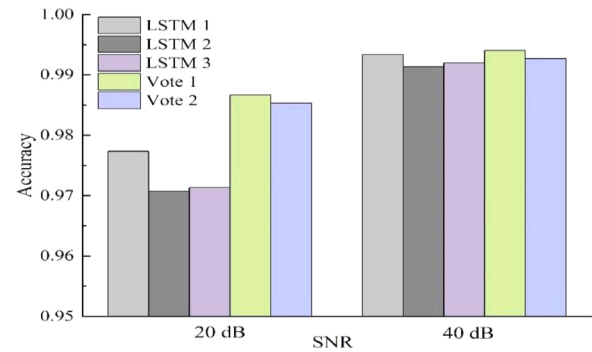| Timestep	Accuracy/%	Loss	Consuming time(us)
5	98.533	0.07742	108
10	99.260	0.03046	168
20	99.467	0.02326	270
40	99.133	0.04238	509

Fig. 5. Comparison of different voting strategies.

(4) CNN-LSTM: [18] designed and proposed a combination of a CNN with two standard convolutional layers and LSTM to classify PQD signals. The size of the convolution kernel is set to 3. The number of convolutional filters are 64 and 128, The number of memory nodes in the LSTM is 50.

For fair comparison, the neural networks in the reference methods are trained and compared on the same computer and data set. The neural networks are trained through 100 iterations. As is clear from Table 2, DCNN can achieve higher classification accuracy, but the training time is longer. Compared with the Bagging-LSTM network, the GRU network has a slightly shorter training time, because the separate storage unit is eliminated, and the accuracy rate is slightly lower than that of the Bagging-LSTM network. Table 2 demonstrates that Bagging-LSTM proposed in this paper is an optimal choice for PQDs classification, which has higher classification precision and less training time cost..

In order to further compare the Bagging-LSTM network with other traditional methods, Table 3 lists the performance of these methods for complex PQDs. The traditional methods require design handcrafted features before classification. The number of features in these methods is relatively random. For example, [17] selected 19 features while [1] selected only 9 features. Compared with these traditional methods, the Bagging-LSTM can automatically extract valid features, which simplifies the classification process. This makes the PQD classification more integrated and automated.

4.4. Active learning verification

In order to verify the effectiveness of the active learning strategy, we compare the uncertainty-based active learning sampling method (US) with the random selection (RS) method. Both of the two sampling methods evaluate the performance indicators through ten-fold cross validation, and the classifier is Bagging-LSTM network. The algorithm queries 0.5k samples every time, and each curve in the Fig. 6 shows the average after ten random training. The vertical axis is the performance evaluation index commonly used in active learning: F1-score and loss. Fig. 6 shows that in the early stage of training, the performance of active learning sampling is close to random sampling. As the number of active queries increases, the training result of the active learning becomes better, showing its advantages.

Table 2
The performance of different methods.

Method	Accuracy under different noise levels(%)				Training time(min)
	20 dB	30 dB	40 dB	50 dB	
DCNN	97.67	98.40	99.06	99.43	176.7
CAE	97.60	98.86	99.00	99.06	5.2
GRU	98.06	98.86	99.00	99.17	1.8
CNN-LSTM	97.73	97.53	98.35	98.93	156.6
Bagging-LSTM	98.67	99.20	99.40	99.66	3.5

Table 3
Comparison of the proposed method with other methods.

Method	Num. of PQD	Num. of features	Accuracy(%)	
			20 dB	30 dB
ST and PNN [1]	9	4	–	98.63
ST and RF [2]	15	4	95.9	99.70
ST and NSGA-II [3]	15	26	96.43	97.27
DWT and PNN [8]	16	9	93.60	95.20
DWT and H-ELM [17]	15	19	95.2	97.67
Bagging-LSTM	15	Auto	98.67	99.20

Table 4
Comparison of the proposed method with other method.

Num. of samples added	F1-score(%)	
	Active learning	Random sampling
1000	96.733	94.400
2000	97.933	96.067
3000	98.400	97.333
4000	98.467	96.933
5000	99.200	98.333

Fig. 6. Active learning verification.

Fig. 7. The transient oscillation signal after data enhancement.

Table 5
Performance of Bagging-LSTM and reference methods.

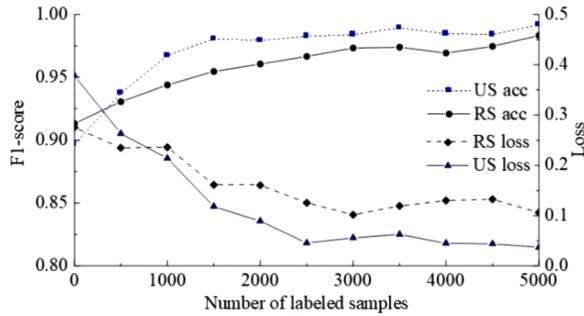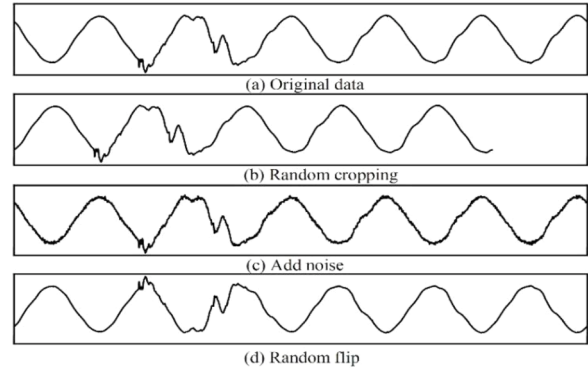| Methods	Accuracy (%)	Test time (s)
DCNN	95.20	8 ms
CAE	94.41	141 us
GRU	93.75	66 us
CNN-LSTM	94.52	3 ms
Bagging-LSTM	95.83	62 us

The F1-score after several queries is presented in Table 4. It is seen that the performance of the active learning algorithm based on uncertainty sampling becomes better than that of random sampling. When more than 3000 labeled samples are added, the F1-score of active learning reaches 0.9840, while random sampling needs to add nearly 5000 labeled samples to achieve similar performance. In this sense, the active learning algorithm can save about 40% of the manual labeling workload.

4.5. Testing using practical data

A set of practical signals are used to test the performance of the Bagging-LSTM network. The measurement data set is taken from the IEEE PES database for PQD classification [19, 20]. The sampling rate of the signals is 256 points per cycle. The length of signals is 1536. Since the neural network model requires a large amount of data for training and optimization, the quality of data directly affects the performance of the model. The current data volume is not enough to train a good network, so we use data augmentation methods to increase the amount of training data. Data augmentation is an effective way to expand the size of data samples by flipping, random cropping, adding Gaussian noise, etc. Fig. 7 shows the result of data enhancement of the transient oscillation signal.

With the enhanced data set to train the proposed Bagging-LSTM model, the Table 5 shows the classification result comparison. The average accuracy of Bagging-LSTM is 95.83%, which is lower than the simulation results presented in the previous sub-section. The main reason is that the data is still not enough, in spite of the data enhancement used. The interference of real data signals is more complicated, and the distribution between categories is uneven, further decreasing the accuracy. Although the classification result is not ideal, it can be seen that the Bagging-LSTM performs better than the other methods.

5. Conclusion

In this paper, a Bagging-LSTM network is proposed to automatically classify complex PQD signals by combining deep learning and ensemble learning. An improved Bagging strategy is used to integrate the training results of multiple LSTM networks, so as to improve the classification accuracy and generalization. Furthermore, an active learning algorithm

is used for sample selection. By actively selecting the samples with the most abundant information, the requirement for annotation and the corresponding cost can be reduced. In addition, the data augmentation technology is further adopted to alleviate the problem of insufficient labeling by creating a modified copy of the original sample. Compared with other deep learning methods, the proposed method has the advantage of less training time and higher precision. Future work will focus on developing smarter query strategies and effective model update strategies to achieve a better trade-off between classification accuracy and labeling cost.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H.H. W.ang, P. Wang, T. Liu, Power Quality Disturbance Classification Using the S-transform and probabilistic neural network, Energies 10 (1) (2017) 1–19.
[2] N. Huang, G. Lu, G. Cai, Feature selection of power quality disturbance signals with an entropy-importance-based random forest, Entropy 18 (44) (2019).
[3] U. Singh, S.N. S.ingh, Optimal feature selection via nsga-ii for power quality disturbances classification, IEEE Trans. Ind. Inf. 14 (7) (2018) 2994–3002.
[4] Z. Zheng, Y. Yang, X. Niu, H. Dai, Y. Zhou, Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids, IEEE Trans. Ind. Inf. 14 (4) (2018) 1606–1615.
[5] S. Wang, H. Chen, A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network, Appl. Energy 235 (2019) 1126–1140.

[6] X. Yang, L. Guo, X. Xiao, Classification of multiple power quality disturbances based on TQWT and random forest feature selection algorithm, Power Syst. Technol. 44 (8) (2020) 3014–3020.

[7] C. Zhao, et al., Novel method based on variational mode decomposition and a random discriminative projection extreme learning machine for multiple power quality disturbance recognition, IEEE Trans. Ind. Inf. 15 (5) (2019) 2915–2926.

[8] S. Khokhar, A.A.M. Zin, A.P. M.emon, A.S. M.okhtar, A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network, Measurement 95 (2017) 246–259.

[9] K. Chen, J. Hu, J. He, Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder, IEEE Trans. Smart Grid 9 (3) (2018) 1748–1758.

[10] Jarkko Tulensalo, Janne Seppänen, Alexander Ilin, An LSTM model for power grid loss prediction, Electr. Power Syst. Res. 189 (2020) 0378–7796.

[11] A.S. Khwaja, A. Anpalagan, M. Naeem, B. Venkatesh, Joint bagged-boosted artificial neural networks: using ensemble machine learning to improve short-term electricity load forecasting, Electr. Power Syst. Res. 179 (2022) 0378–7796.

[12] I. Okwuchi, L. Nassar, F. Karray, K. Ponnambalam, Deep learning ensemble based model for time series forecasting across multiple applications, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 2020, pp. 3077–3083.

[13] Dipu Sarkar, Ensemble machine learning models for the detection of energy theft, Electr. Power Syst. Res. 192 (2021) 0378–7796.

[14] C. Soray, A. Nabiha, D. Nilanjian, A new hybrid classifier selection model based on mRMR method and diversity measures, Int. J. Machine Learn. Cybernetics 10 (5) (2019) 1189–1204.

[15] J. Shim, S. Kang, S. Cho, Active learning of convolutional neural network for cost-effective wafer map pattern classification, IEEE Trans. Semicond. Manuf. 33 (2) (2020) 258–266.

[16] IEEE recommended practice for monitoring electric power quality, IEEE Standard (2009) 1159–2009. Revision of IEEE Standard 1159-1995.

[17] J. Wang, Z. Xu, Y. Che, Power quality disturbance classification based on DWT and Multilayer extreme learning machine, Appl. Sci.-Basel 9 (11) (2019) 1–16.

[18] N. Mohan, K.P. S.oman, R. Vinayakumar, Deep power: deep learning architectures for power quality disturbances classification, in: 2017 International Conference on Technological Advancements in Power and Energy (TAP Energy, 2017, pp. 1–6.

[19] IEEE power engineering society, IEEE PES Working Group P1433 Power Quality Definitions (Feb.2,2001). Accessed[Online].Available, http://grouper.ieee.org/groups/1159/2/testwave.html.

[20] IEEE power engineering society, IEEE PES Working Group Power Quality Definitions (Oct.12,2012). Accessed[Online].Available, http://grouper.ieee.org/groups/1159/3/docs.html.