Data-Driven Traction Substations' Health Condition Monitoring via Power Quality Analysis

Jingyi Xie *

College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, China

Electrified railway traction substations are an important part of the transportation system, the health of its operation condition indirectly affects the national economy. Generally, traction substations' conditions are studied from their power quality, while the nonlinearity of loads and effects from the outside environment are factors mainly affecting the accuracy of condition monitoring. In order to recognize the status of traction substations intelligently and govern them with fast measurements, this paper proposed a data-driven approach for recognizing types of power quality problems, and developed a system with intelligent governance strategies. The proposed approach contains two parts. Firstly, a double discrete Fourier transform (DDFT) algorithm was developed to extract valid feature vectors from power data. Then, a well-known data-driven method, support vector machine (SVM), was applied to build classifiers. Finally, based on classification results, a strategy library for power quality problems was built. Industrial data of a real traction substation in Wuhan, China, was tested for the experiment. Compared with traditional methods, the proposed approach is validated to be useful in improving the classification performance of power quality problems, and fast and effective for governance in traction substations.

Keywords: data-driven, condition monitoring, traction substation system, power quality analysis, DDFT

INTRODUCTION

As the major infrastructure in transportation systems, electrified railway traction substations are a powerful guarantee for national economic development and social progress (Hu et al., 2017). Due to the electric locomotive being a type of nonlinear load, it has nonlinearity, impact, and asymmetric distribution features in three phases of power electricity. Currently, there are serious challenges in the process of maintaining the power quality traction substation, guaranteeing the safe and stable operation of electrified railway (Hu et al., 2016). For example, the major power quality problems faced by traction substations include harmonics, voltage fluctuations, flicker, as well as negative-sequence current (Bitoleanu et al., 2016). These problems seriously reduced the power quality of both traction substations and superior power systems, and subsequently impact the safety and economic operation of the power system (Liu and Hu, 2017). Therefore, to ensure the stable power supply of traction substation systems and the safe operation of electrified railway, monitoring their health status and making targeted compensation measures are very important.

Power quality problems are dangerous not only to the traction power system but also to other power systems. Currently, the study of power quality problem analyses in literature are mainly based on voltage signals (Khadkikar et al., 2017), for example, voltage analysis via Fourier transformation, wavelet analysis, and so on. Considering the particularity of traction

substations, higher computation speed is required to guarantee the real-time decision-making. It is not feasible to analyze the entire signal sequence as the object. Currently, according to global research, mathematical transformations were commonly used in power quality detection, such as the short-time Fourier transform, wavelet transform, s-transform, and bilinear time-frequency transform (De Yong et al., 2015; Mahela et al., 2015; Shen et al., 2022). Then, the transformed features have been utilized for fast and effective analysis. These transformation methods achieve some positive results to a certain extent. However, taking the actual requirements of rapid compensation in traction power quality into account, these methods still have some restrictions in practical applications, e.g. their complex calculations. On the other hand, according to the results of power quality analysis, how to make timely and effective control measures is also essential to traction substation systems. In power quality governance measures, the mixed compensation methods are usually applied, namely the coordinated control of fixed units and dynamic compensation devices (Mikkili and Panda, 2015; Lam et al., 2017). However, the power quality governance of traction substations is different from that of traditional power systems. It has something to do with the real-time operation status, and is related to factors such as weather, geographical distribution and so on, so it is a typical non-stationary random process. As the intelligent scheduling requirements have been proposed in recent years, when and what types of compensation methods are the major questions. Therefore, an intelligent governance system for a traction substation is required to solve power quality problems.

Aiming to tackle the mentioned issues above, in order to solve power quality problems in electrified railway traction substations intelligently, fast, and effectively, an intelligent system for substations' power quality problems is proposed based on data-driven method in this paper. It is mainly based on Double Discrete Fourier Transform (DDFT) with a sliding window and intelligent classification methods. Contributions of the proposed method are summarized as follows. First, considering power quality problems vary with time, only the latest period of data is used to analyze the power quality problem in this paper, instead of using a long sequence of historic data to extract features in traditional transformation methods. Meanwhile, a sliding window is proposed to guarantee the analysis process in real time. Second, in order to ensure less information is lost in the sliding window, DDFT is applied in the features analysis. DDFT can not only extract the basic physical characteristics of power quality at the first level as other transformation methods do, but also analyze the historical variation characteristics at the second level. In this way, less information will be lost, and the dimension of analyzed variables will not be increased dramatically. So DDFT could facilitate the rapid detection of power quality problems in traction substations. Third, advanced machine learning models are built for multi-category classification, they can be used to establish the power quality governance

FIGURE 1 | The framework of intelligent governance system.
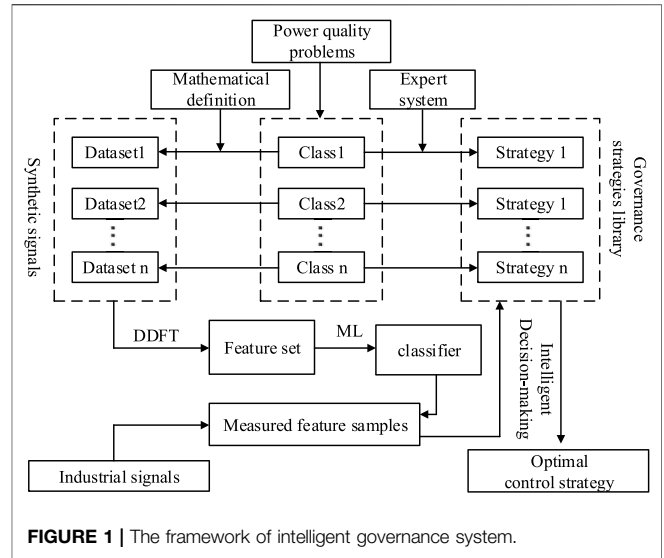
system combined with the expert system model. Then, in the power quality governance of real traction substations, the proposed system could analyze the power quality problems rapidly, then make a fast decision automatically based on the strategies library, and finally realize the effective governance of traction substation's power quality problems intelligently.

FRAMEWORK OF THE PROPOSED APPROACH

According to the description above, to maintain the safe and stable operation of electrified railway, the main purpose existing in a traction substation's power quality problems is to build a system for rapid analysis and intelligent governance. This paper proposed an intelligent system based on DDFT and machine learning (ML) algorithms, which mainly consists of two parts: fast detection of power quality problems based on a DDFT algorithm, and intelligent control and targeted governance based on ML. The framework of the main idea of the proposed approach is shown in **Figure 1**.

It can be seen from **Figure 1** that the intelligent governance system can be divided into two parts: the training part and testing part. In the training part, in order to achieve the intelligent and rapid governance system for traction substation, some typical signals of power quality problems are generated firstly based on their mathematical definition models. The frequency-domain features of these power quality samples are extracted through the DDFT algorithm. On the other hand, based on the classes of different power quality problems, the expert system is used to determine their effective control measures or governance strategies, and these strategies consist of the governance library. Finally, taking the extracted features as inputs and the strategies library as the outputs, classification models are built based on ML algorithms. The optimal classifier is applied for the intelligent governance of power quality problems. In the testing part, the application

objects are practical electric railways. The industrial data is often voltage signals of traction substation. The same as the training process, the frequency-domain features of industrial data should be extracted by DDFT at first. Then the feature vectors are input as the optimal classifier for decision-making. The results are the effective strategies based on the classifier and strategies library. Ignoring the classification process of determining what exact types of power quality problems the industrial signal belongs to, the optimal governance strategies can be given out directly according to the proposed system. In this way, it realizes intelligent governance of the traction substation's power quality problems, and keeps the safe operation of electric railways.

INTELLIGENT SYSTEM AND METHODOLOGY

Notion of Power Quality Problems

The operation of electric railway causes several power quality problems to traction substations and outlet substations of power systems, including voltage fluctuation, voltage unbalance, voltage harmonics, and so on (Shen et al., 2021). In order to quantify these problems in practice, models of some common power quality problems including voltage rise, voltage drop, voltage flicker, and transient impact are defined (Singh et al., 2014; Wong et al., 2014; Shi et al., 2016), as expressed in (1–3).

1) Signal of voltage rise and drop

$$f(t) = \begin{cases} \sin(\omega t) & t < t_1, t > t_2 \\ \alpha \sin(\omega t) & t_1 < t < t_2 \end{cases} \quad (1)$$

Where, t_1 and t_2 represent the start and end time of disturbance respectively; ω is the angle frequency of carrier voltage on power frequency; α is the normalized voltage amplitude, when $\alpha > 1$, $f(x)$ represents the voltage rise signal, when $\alpha < 1$, $f(x)$ represent the voltage drop signal.

2) Signal of voltage flicker

$$f(t) = \begin{cases} \sin(\omega t) & t < t_1, t > t_2 \\ (A + m\cos\Omega t)\sin(\omega t) & t_1 < t < t_2 \end{cases} \quad (2)$$

Where, m is the amplitude of amplitude-modulated (AM) wave, Ω is the angle frequency of AM wave.

3) Signal of transient voltage

$$f(t) = \begin{cases} \sin(\omega t) \\ \sin(\omega t) + \sum_{i=2}^{N} A_i \sin(i\omega t + \varphi_i) + Ae^{-\lambda t} \end{cases} \quad \begin{matrix} t < t_1, t > t_2 \\ t_1 < t < t_2 \end{matrix}$$
$$(3)$$

It can be seen that the transient voltage consists of multi-frequency signals, such as mutations with damping attenuation $Ae^{-\lambda t}$ and transient high-frequency signals $\sum_{i=2}^{N} A_i \sin(i\omega t + \varphi_i)$, A represents the amplitude and φ_i represents the phase deviation of the ith high-frequency signal. All these components lead to the power quality problems.

Double Discrete Fourier Transformation

According to the analysis of formula (1-3), the definition of power quality problems mainly considers the harmonics in the frequency domain and amplitude characteristics of the time domain. Conventional Fourier transform mainly analyzes signal characteristics from the perspective of frequency domain (Boashash, 2015). However, it has limitation in rapid detection of the traction substation's power quality problems in real-time. Wavelet transformation has the ability to express both time-domain and frequencydomain characteristics (Shen et al., 2020), but it is not easy to choose a suitable mother wavelet. After a comprehensive comparison, on the basis of discrete Fourier transform (DFT), this paper proposed the double discrete Fourier transform (DDFT) as the major tool to solve the electric power quality detection problems of traction substations.

To satisfy the real-time analysis of traction power quality, the DDFT algorithm is developed based on sliding-window iterative DFT (Zhan et al., 2016). Results of DFT usually reflect real-time information in the frequency domain, but they cannot distinguish all the power quality problems, which have an impact on the amplitude of fundamental frequency signals, e.g. voltage rise, voltage drop, and voltage flicker. Therefore, the proposed DDFT algorithm tries to extract fundamental sequences $\{X(k,t)\}$ from the historical results of DFT transformation. The sequence is then used for Fourier transform for the second time. The purpose of this process is to obtain fundamental frequency information, which can be used to distinguish between temporary voltage rise, drop, and flicker problems.

Figure 2 shows the technological process of DDFT. It can be seen that the original signal is analyzed based on Fourier transformation by taking sliding windows as units. In this way, the computational cost is reduced, and the calculation is speeded up for real-time analysis. After that, the fundamental sequence in the historical windows is extracted at the same size of sliding window, and processed with DFT algorithm once again. Therefore, the DDFT algorithm can not only extract the variation of the fundamental signal and other frequency information, but also maintain the real-time signal analysis, both of which are essential for the power quality analysis in traction substations.

Multi-Classes Classification Based on Machine Learning Models

For a given voltage signal in the traction substation $\{x_n\}$, a series of transformed features are obtained through the DDFT algorithm. To determine the class that the power quality signal belongs to is a must for the decision-making of selecting appropriate control strategies and power compensation devices. This paper generated some typical power quality signals, and extracted their characteristics by DDFT. By regarding the mechanism of identifying power quality problems as the process of choosing a classifier, and collecting feature samples to be used to train the classifier, the classifier would complete the recognition of power quality
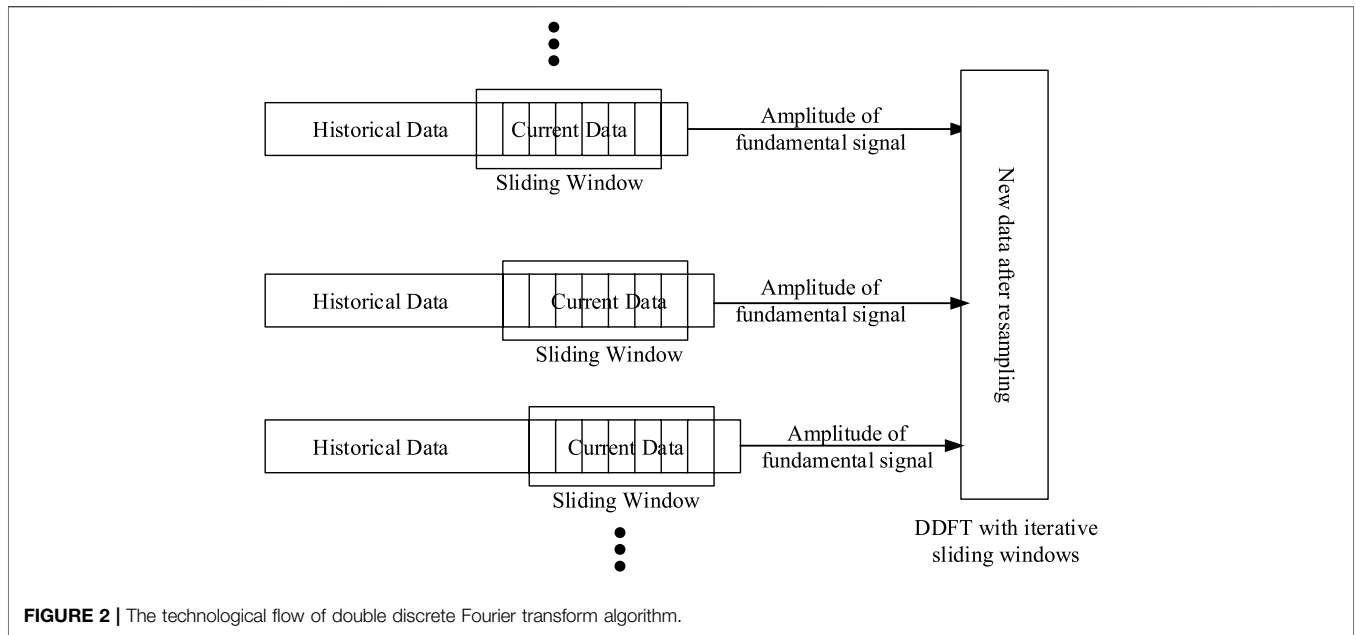
FIGURE 2 | The technological flow of double discrete Fourier transform algorithm.

TABLE 1 | Feature values of different power quality signals.

Feature label	Normal voltage	Voltage rise (20%)	Voltage drop 20%	Voltage flicker	Transient voltage
X0	19.9848	22.4087	17.5729	19.6173	19.9848
X1	0.0194	1.2042	1.2085	0.4559	0.0194
X2	0.0265	0.3623	0.3821	1.2708	0.0265
X3	0.0115	0.2228	0.2498	0.5574	0.0115
X4	0.0195	0.2696	0.2423	0.2507	0.0195

problems. According to the above analysis, this paper applied ML-based algorithms to build the classifier.

EXPERIMENTS AND DISCUSSION

Training Part

1) synthetic power quality signals.

To train an optimal mechanism for power quality governance in a traction substation, the first task is to obtain the training dataset. In this paper, the training data is generated through the definitions of typical power quality problems. According to the definitions in (1-3), 10 sets of power quality signals with different parameters are constructed based on each mathematical definition.

2) Features extraction.

For different types of power quality problems, it is subjective and not applicable to determine the classes that each power quality signal belongs to through by observation directly. It absolutely does not work in the analysis of actual operation of electric railways. In order to discriminate different types of power quality problems quantitatively, the DDFT algorithm is proposed to extract features of voltage signal for fast analysis, instead of the

original signal sequences. Due to the amplitude of high-frequency harmonics always being low, to reduce the dimension and computation cost, only the DC component and 5–50 Hz harmonic components are taken as the main feature variables. **Table 1** gives out the feature values of four typical power quality signals.

In **Table 1**, feature values of four typical power quality signals are presented, features of the corresponding original signals are also extracted and displayed. Here, X0 represents the DC component, and X1-X5 represent the 5–50 Hz harmonic components, respectively. It is seen that voltage rise and drop signals mainly reflect difference with other signals on the DC component and 5 Hz harmonics component. The voltage flicker signals mainly have some influence on the 10 Hz harmonics component. For more thorough feature analysis, a nonlinear model is required.

3) Establishing classifiers and strategies library.

To build a suitable classification model for analyzing power quality problems, the DDFT algorithm is applied to extract feature vectors x_i at first. Taking feature vectors x_i = [X (0,t), X (1,t), $\cdots$ X (N-1,t)] as inputs, the class labels of power quality problems as outputs, different ML classification models are established, such as SVM, ELM, NN (Ouyang, 2021; Ouyang
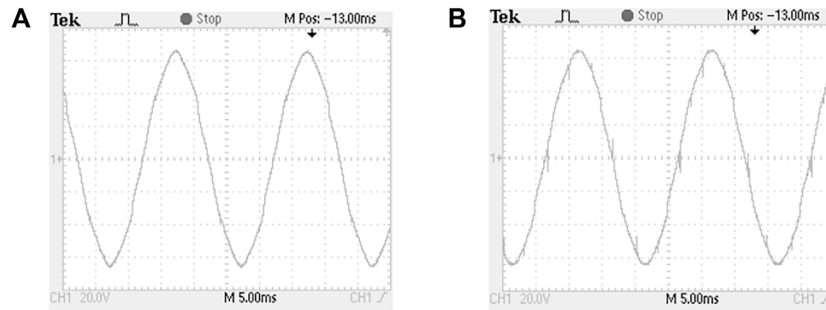
FIGURE 3 | Two signals with typical power quality problems in an actual traction substation.

et al., 2021; Tang et al., 2021). Finally, the optimal ML model is selected. The essence of the proposed method is to construct classifiers based on ML algorithms, and to the constructed classifier to distinguish *power quality* problems automatically.

Testing and Validation Part

According to the above description, some typical power quality signals are generated as training data, and used to build classification models. Meanwhile, the governance strategies are given out according to expert systems and form the strategies library. Then, the intelligent governance system for traction substations' power quality problems is completed by combining classifiers and strategies libraries. In order to validate performance of the proposed system, including how effective it is to use the DDFT algorithm to extract features of power quality problems, the performance of using ML to classify different power quality problems, and the effectiveness and feasibility of the intelligent governance system in traction substation are discussed.

For analyzing the effectiveness of the proposed approach quantitatively, this paper introduces the confusion matrix (Xiong et al., 2017) in performance evaluation. Here, four commonly used indicators are defined in (4), such as Recall (R), Precision (P), Accuracy (Acc), and Error Rate (ER) (Ohsaki et al., 2017).

$$\begin{cases} Acc = \dfrac{TP + TN}{TP + FP + TN + FN} \\ R = TP/(TP + FN) \\ P = TP/(TP + FP) \\ ER = 1 - Acc \end{cases} \quad (4)$$

In this paper, taking an actual traction substation in Wuhan China as an example, there are in total 1,000 voltage signals collected as the testing dataset. The frequency of these signals is 50Hz, and each signal has a period of 50 ms. First, necessary data preprocessing is needed, and the dataset is normalized. Then the power quality problems of the collected signals are tested according to the given intelligent governance system. According to the records of operators in traction substations, the distribution of these signals is displayed as 43 voltage

variation, 16 voltage flicker, 19 voltage transient singles and 922 other signals. In other signals, most of the signals are normal signals, and a few of the signals are other undefined signals in this paper. Two typical signals collected from actual traction substations are shown in **Figure 3**.

In **Figure 3**, the variation of voltage signals for two typical power quality problems are shown. **Figure 3A** shows the signal of voltage flicker, and **Figure 3B** shows the impact of transient voltage signal. It is known that it is difficult for operators to determine what power quality problems a collected signal belongs to, and the period of a set of signals is too fast to capture. This is also the reason why an intelligent system is required to be developed for governing the power quality problems in traction substations.

To evaluate the performance of the proposed approach in the classification of power quality problems, 120 sets of actual signals are used for validation. Since normal signals take up more than 90% of the testing datasets, for classifying different types of power quality problems effectively, only 42 sets of normal signals are selected, and combined with other typical signals as the validation dataset. According to the description in **Figure 1**, the validation data is processed by DDFT to extracted feature vectors, then feature vectors are input into the obtained classifier. Based on the classification results, the suitable governance strategies could be determined according to the strategies library to control the harmful effects of power quality problems. In other words, the performance of classifying power quality problems of voltage signals of traction substation is directly related to the performance of the proposed intelligent governance system. Therefore, statistical analysis is completed to discuss the performance of the proposed approach. According to the definitions in (**Eq. 4**), values of four indicators are shown in **Table 2**.

In **Table2**, the performance of four indicators are presented. For comparison, three ML models using feature vectors extracted from DFT are also analyzed. From the results of **Table 2**, it is seen that all these methods have values of almost 85% on Recall, Precision, and Accuracy indicators, and the classifier based on ELM in power quality problems classification performs the best. Therefore, ELM could be selected for the construction of final intelligent system. On the other hand, comparing the feature vectors from DFT and DDFT, it is seen that the performance has

TABLE 2 | Performance of classification of power quality problems.

	Recall (R)	Precision (P)	Accuracy (Acc)	Error Rate (ER)
DFT + SVM	0.8975	0.8469	0.9249	0.0751
DDFT + SVM	0.9237	0.8705	0.9509	0.0491
DFT + ELM	0.9107	0.8722	0.9433	0.0567
DDFT + ELM	0.9294	0.8903	0.9526	0.0474
DFT + NN	0.9039	0.8688	0.9335	0.0665
DDFT + NN	0.9277	0.8831	0.9494	0.0506

been improved. Therefore, it is validated that the proposed approach has better performance on classification of power quality problems. While, for the governance of traction substations in the real world, the proposed system combined the advanced approach with the control strategies library, therefore it not only achieves a fast and accurate identification of power quality problems, but also makes a fast decision to govern traction substations' power quality by taking the corresponding measures from the strategies library. In this way, the proposed system can realize the intelligent governance of power quality in traction substations and guarantee the safety operation of electrified railways.

CONCLUSION

In this paper, an intelligent system based on data-driven and DDFT is proposed to govern the power quality problems of traction substations. First, definitions of several typical power quality problems are given out, and synthetic voltage signals based on these definitions are generated. By using the proposed DDFT algorithm to extract feature vectors of these signals, it shows that DDFT has advantages at extract features distinguishing different power quality problems. Using the extracted feature vectors as inputs, three ML-based classifiers

are built to discriminate three types of power quality problems, and finally ELM is selected. Combined with the control strategies library from an expert system, the intelligent governance of power quality problems in traction substation is completed. The industrial data of actual traction substations were tested with the proposed approach and traditional approach, the numerical results validated the proposed approach improved the performance of classifying power quality problems. It is validated that the proposed system can realize the intelligent and fast governance of power quality problems in traction substations, and can guarantee the safe operation of electricity railways.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

Bitoleanu, A., Popescu, M., and Suru, V. (2016). "Optimal Controllers Design in Indirect Current Control System of Active DC-Traction Substation," in *Power Electronics and Motion Control Conference (PEMC), 2016 IEEE International* (IEEE), 912–917. doi:10.1109/epepemc.2016.7752115

Boashash, B. (2015). *Time-frequency Signal Analysis and Processing: A Comprehensive Reference.* Academic Press.

De Yong, D., Bhowmik, S., and Magnago, F. (2015). An Effective Power Quality Classifier Using Wavelet Transform and Support Vector Machines. *Expert Syst. Appl.* 42 (15), 6075–6081. doi:10.1016/j.eswa.2015.04.002

Hu, H., Gao, S., Shao, Y., Wang, K., He, Z., and Chen, L. (2017). Harmonic Resonance Evaluation for Hub Traction Substation Consisting of Multiple High-Speed Railways. *IEEE Trans. Power Deliv.* 32 (2), 910–920. doi:10.1109/tpwrd.2016.2578941

Hu, K., Liu, Z., Huang, K., Dai, C., and Gao, S. (2016). Improved Differential Evolution Algorithm of Model-based Diagnosis in Traction Substation Fault Diagnosis of High-speed Railway. *IET Electr. Syst. Transportation* 6 (3), 163–169. doi:10.1049/iet-est.2015.0003

Khadkikar, V., Xu, D., and Cecati, C. (2017). Emerging Power Quality Problems and State-Of-The-Art Solutions. *IEEE Trans. Ind. Electron.* 64 (1), 761–763. doi:10.1109/tie.2016.2619663

Lam, C.-S., Wang, L., Ho, S.-I., and Wong, M.-C. (2017). Adaptive Thyristor-Controlled LC-Hybrid Active Power Filter for Reactive Power and Current Harmonics Compensation with Switching Loss Reduction. *IEEE Trans. Power Electron.* 32 (10), 7577–7590. doi:10.1109/tpel.2016.2640304

Liu, Z., and Hu, K. (2017). A Model-Based Diagnosis System for Traction Power Supply System. *IEEE Trans. Ind. Inform.* 13 (6), 2834–2843. doi:10.1109/tii.2017.2735480

Mahela, O. P., Shaik, A. G., and Gupta, N. (2015). A Critical Review of Detection and Classification of Power Quality Events. *Renew. Sustain. Energ. Rev.* 41, 495–505. doi:10.1016/j.rser.2014.08.070

Mikkili, S., and Panda, A. K. (2015). *Power Quality Issues: Current Harmonics.* Taylor & Francis.

Ohsaki, M., Wang, P., Matsuda, K., Katagiri, S., Watanabe, H., and Ralescu, A. (2017). Confusion-matrix-based Kernel Logistic Regression for Imbalanced Data Classification. *IEEE Trans. Knowledge Data Eng.* 29 (9), 1806–1819. doi:10.1109/tkde.2017.2682249

Ouyang, T. (2021). Feature Learning for Stacked ELM via Low-Rank Matrix Factorization. *Neurocomputing* 448, 82–93. doi:10.1016/j.neucom.2021.03.110

Ouyang, T., Marco, V. S., Isobe, Y., Asoh, H., Oiwa, Y., and Seo, Y. (2021). Improved Surprise Adequacy Tools for Corner Case Data Description and Detection. *Appl. Sci.* 11 (15), 6826. doi:10.3390/app11156826

Shen, X., Ouyang, T., Khajorntraidet, C., Li, Y., Li, S., and Zhuang, J. (2022). Mixture Density Networks-Based Knock Simulator. *Ieee/asme Trans. Mechatron.*, 27, 159–168. doi:10.1109/TMECH.2021.3059775

Shen, X., Ouyang, T., Yang, N., and Zhuang, J. (2021). Sample-Based Neural Approximation Approach for Probabilistic Constrained Programs. *IEEE Trans. Neural Netw. Learn. Syst.*, 1–8. doi:10.1109/TNNLS.2021.3102323

Shen, X., Ouyang, T., Zhang, Y., and Zhang, X. (2020). Computing Probabilistic Bounds on State Trajectories for Uncertain Systems. *IEEE Trans. Emerg. Top. Comput. Intell.*, 1–6. doi:10.1109/TETCI.2020.3019040

Shi, S., Xiao, J., Liu, S., Tu, L., Cao, Q., Liang, Y., Zhang, G., and Mao, C. (2016). A Scheme of Dynamic Voltage Restorer to Improve Power Quality in Large Scale Industrial Enterprises. *Jpee* 04 (02), 19–26. doi:10.4236/jpee.2016.42003

Singh, B., Chandra, A., and Al-Haddad, K. (2014). *Power Quality: Problems and Mitigation Techniques*. John Wiley & Sons.

Tang, Z., Zhao, G., and Ouyang, T. (2021). Two-phase Deep Learning Model for Short-Term Wind Direction Forecasting. *Renew. Energ.* 173, 1005–1016. doi:10.1016/j.renene.2021.04.041

Wong, J., Lim, Y. S., Tang, J. H., and Morris, E. (2014). Grid-connected Photovoltaic System in Malaysia: A Review on Voltage Issues. *Renew. Sustain. Energ. Rev.* 29, 535–545. doi:10.1016/j.rser.2013.08.087

Xiong, Y., Zha, X., Qin, L., Ouyang, T., and Xia, T. (2017). Research on Wind Power Ramp Events Prediction Based on Strongly Convective Weather Classification. *IET Renew. Power Generation* 11 (8), 1278–1285. doi:10.1049/iet-rpg.2016.0516

Zhan, J., Cheng, X., Xu, Q., Wu, L., and Chen, G. (2016). "Design of Comprehensive Motor Protector Based on Improved Sliding-Window Iterative DFT Algorithm," in *Power Electronics and Motion Control Conference (IPEMC-ECCE Asia)* (IEEE 8th InternationalIEEE), 3658–3662. doi:10.1109/ipemc.2016.7512881