Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 8, No. 4, 2024
www.scholink.org/ojs/index.php/asir

Original Paper

The Fault Diagnosis Method for Photovoltaic Modules Based on Machine Learning

Qiheng Li[1], Yubo Cui[1], Jingyu Zhang[1], Yandong Liu[1] & Li Zhang[1]

[1] Yangtze Normal University, Fuling, Chongqing 408100, China

Received: October 3, 2024 Accepted: October 22, 2024 Online Published: October 31, 2024

doi:10.22158/asir.v8n4p92 URL: http://doi.org/10.22158/asir.v8n4p92

Abstract

Based on the urban natural gas pipeline accident statistics and semi-quantitative risk evaluation index system, this paper applies Bayesian network to establish a network model between various types of risk factors and the risk of natural gas pipeline failure. The EM algorithm was used to learn from the statistical accident data to obtain the parameters of the model. Based on the principle of evidential reasoning in reverse, the probability of occurrence of all risk indicators can be obtained when the probability of occurrence of urban natural gas pipeline accidents is 100%, the index weight is obtained by normalizing the occurrence probability. On this basis, this paper develops an efficient urban natural gas pipeline integrity risk identification and management software. The software can realize the basic data management of urban natural gas pipeline system, pipeline relative risk value calculation, pipeline risk level calculation and other functions, and the results are visualized. Finally, the practicability and effectiveness of the model and software are verified by a case of natural gas pipeline evaluation in a block.

Keywords

Machine learning, PV module, Fault Diagnosis

1. Introduction

The main problem in the development of photovoltaic power generation is that photovoltaic modules are prone to failure. The three major factors (Yang Guangqiao, Liu Bingfeng, & Xu Xiaolong, 2024) of multiple photovoltaic modules, large size of photovoltaic panels, and harsh environmental conditions in photovoltaic usage areas make it easy for photovoltaic panels to malfunction during installation and use. Failure to replace them in a timely manner can reduce power generation efficiency (Jaffery, Z. A., Dubey, A. K., Irshad, et al., 2017), increase return cycles, or even burn out the combiner box, affecting the normal operation of inverters and causing forest fires or power outages, resulting in property losses and a series of other problems.

At present, the fault detection of photovoltaic modules mainly relies on manpower. Staff need to carry their own testing instruments to the site to conduct batch fault detection on each photovoltaic module, record the fault points, and the workload is large. In addition, due to the fact that solar photovoltaic panels are mostly distributed in clusters in the form of arrays and have dense circuits, it is difficult to individually test the components. Finally, at present, the backup guarantee system in the photovoltaic industry is not perfect (Lu, X., Lin, P., Cheng, S., et al., 2019; Tsanakas, J. A., Ha, L. D., & Al Shakarchi, F., 2017), and there is a shortage of maintenance personnel. Many companies outsource the completed photovoltaic systems to other personnel for subsequent maintenance (Wang, W., Liu, A. C., Chung, H. S., et al., 2016), which not only increases costs but also reduces the efficiency of fault detection and troubleshooting, which is not conducive to the development of related enterprises.

In response to the above situation, it is not difficult to find that fault detection of photovoltaic modules in this field is a major challenge that troubles the industry. Specifically, it is reflected in the difficulty of detecting faults and fault areas, the difficulty of operation and maintenance, and the high cost of labor. It is also inconvenient to detect photovoltaic panel clusters separately (Ali, M. H., Rabhi, A., Hajjaji, A. E., et al., 2017). Therefore, this project proposes to apply machine learning to this field to reduce labor costs and improve detection efficiency.

In order to detect faults in photovoltaic arrays, many scholars have proposed various diagnostic methods. The image-based methods mainly include infrared thermal images and current voltage signal images. The principle based on infrared thermal imaging method is that when thermal spots appear in the photovoltaic array, the output characteristics are not abnormal or the abnormalities are not obvious, and it is difficult to detect faults only from the output characteristics. However, using infrared thermal imaging can detect early abnormal thermal spots in the photovoltaic array. This method can detect anomalies in photovoltaic arrays as early as possible and has an early fault warning function. The disadvantage is that the equipment and operating costs are high, which is not conducive to widespread use. The I-V characteristic analysis method mainly uses the comparison of the shape with the normal I-V characteristic curve to classify faults. Its main methods include parameter extraction method, difference comparison method, key point analysis method, etc. However, most of these methods have good accuracy in aging and shadow faults. In addition, most of these methods discriminate by quantitatively comparing with normal components, which not only has low accuracy but also requires high discriminator requirements, making it difficult to promote and apply in large-scale photovoltaic arrays. In recent years, the popularity of artificial intelligence algorithms has brought new ideas to solve such real-time fault diagnosis problems, including kernel based extreme learning machines (KELM), support vector machines (SVM) (Chen, Z., Wu, L., Cheng, S., et al., 2017), ANN (Yi, Z. & Etemadi, A. H., 2017), probabilistic neural networks (Mekki, H., Mellit, A., & Salhi, H., 2016), random forests (RF) (Zhu, H., Lu, L., Yao, J., et al., 2018), etc., which have been applied in the field of fault diagnosis of photovoltaic modules/arrays. Compared with traditional threshold judgment methods, these methods greatly improve the accuracy of photovoltaic fault diagnosis. With the development of deep learning, there are more deep level network structures.

By providing stronger feature extraction capabilities and non-linear function fitting abilities, this paper further applies deep learning algorithms to the field of photovoltaic fault diagnosis, making it more accurate and stable.

2. Research on Fault Modeling of Photovoltaic Arrays

2.1 Establishment of Photovoltaic Array Model

The principle of establishing a photovoltaic array model for solar photovoltaic power generation is that photons in sunlight are irradiated onto the photovoltaic cell, and the energy of the photons is absorbed by the material. When the energy of a photon is high enough, it will excite electrons in photovoltaic materials, causing them to transition from the valence band to the conduction band, thereby forming free electrons. Inside a photovoltaic cell, there is usually a built-in electric field (formed by a P-N junction) that separates free electrons from the remaining holes (positive charges). The separated electrons will move towards one side of the battery, while the holes will move towards the other side, forming an electric current. When a photovoltaic cell is connected to an external circuit, electrons flow to generate current and provide electricity. Ultimately, light energy is converted into electrical energy for supply to loads or storage. At present, most of the models mentioned in the literature are constructed based on the I-V characteristics of photovoltaic cells/modules, with the most commonly used models being the single diode model and the dual diode model (Yang Guangqiao, Liu Bingfeng, & Xu Xiaolong, 2024). The specific structure is shown in the following figure:

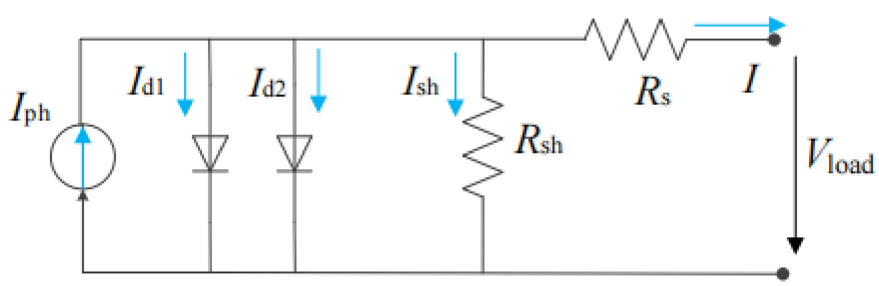


Figure 1. Dual Diode Photovoltaic Model

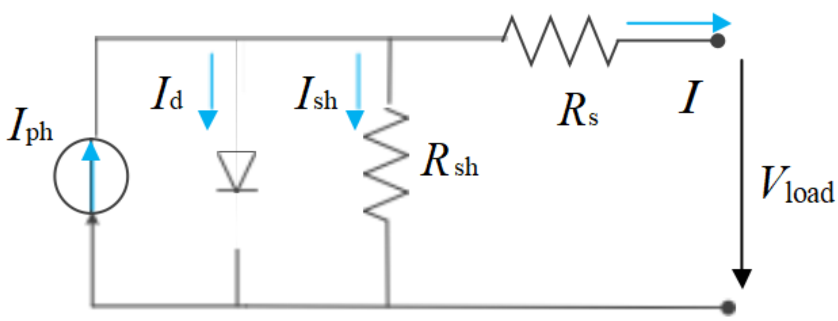


Figure 2. Single Diode Photovoltaic Model

2.2 Photovoltaic Cell Simulation Model

Based on the dual diode mathematical model of the photovoltaic cell mentioned above, this section constructed a simulation model of the photovoltaic cell according to the mathematical model (Mekki, H., Mellit, A., & Salhi, H., 2016) mentioned above. By encapsulating the battery model in the simulation software (Simulink) and setting the required environmental parameters, the output characteristics were simulated completely according to the actual environmental parameters of the photovoltaic panel, as shown in Figure 3.

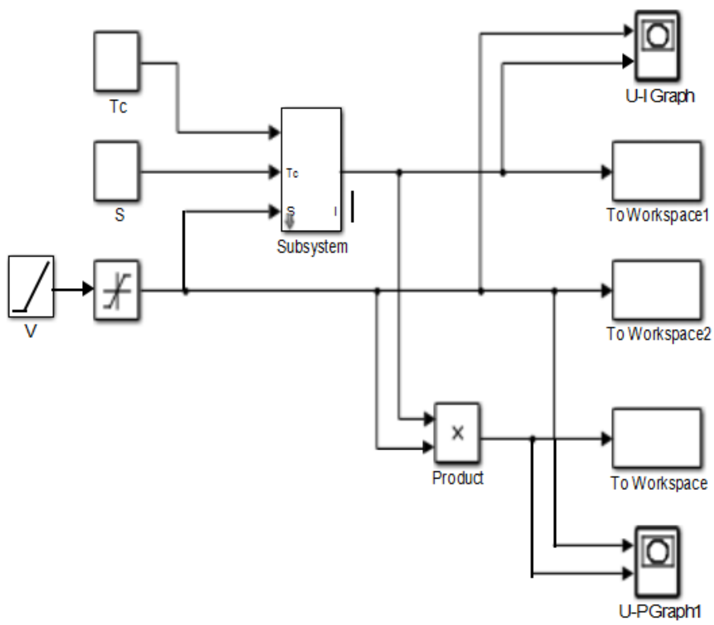


Figure 3. Photovoltaic Cell Simulation Structure Diagram

2.3 Photovoltaic Module faults and Their Mathematical Models

This paper collected voltage and current data for five types of faults and under normal conditions through simulation. Among them, normal is regarded as a type of "fault", and in subsequent deep learning based photovoltaic module fault diagnosis, it is also regarded as classifying six types of faults. In order to facilitate the display of data collected from photovoltaic modules, we have installed a voltmeter and three ammeters S1, S2, and S3 for each type of fault. Displaying different data shows that the readings of the voltmeter and the three ammeters are different under different fault conditions. By collecting this data, it provides data support for the subsequent deep learning of photovoltaic module fault diagnosis.

3. Fault Diagnosis Method for Photovoltaic Modules Based on Convolutional Neural Network

3.1Convolutional Neural Network

Convolutional neural network is a type of feed forward neural network that includes convolutional computation and has deep structure, and is one of the representative algorithms of machine learning.

Convolutional neural networks have representation learning ability and can perform translation invariant classification of input information according to their hierarchical structure, hence they are also known as "translation invariant artificial neural networks". The research on convolutional neural networks began in the 1980s and 1990s, with time delay networks and LeNet-5 (Zhu, H., Lu, L., Yao, J., et al., 2018) being the earliest convolutional neural networks to appear; After the 21st century, with the proposal of deep learning theory and the improvement of numerical computing devices, convolutional neural networks have developed rapidly and have been applied in fields such as computer vision and natural language processing. Convolutional neural networks are constructed to mimic the visual perception mechanism of living organisms, and can perform both supervised and unsupervised learning. The shared convolutional kernel parameters within the hidden layers (Kong, X., Xu, X., Yan, Z., et al., 2018) and the sparsity of inter layer connections enable convolutional neural networks to learn lattice features such as pixels and audio with less computational complexity, stable performance, and no additional feature engineering requirements for data.

3.2 The Structure of Convolutional Neural Networks

Convolutional neural networks typically include the following important components: Convolutional layers: using convolutional kernels to slide over the input image, extracting local features, and generating feature maps. Activation layer: Typically, ReLU and other activation functions are used to introduce nonlinearity, enabling the model to learn more complex features. Pooling layer: By down sampling (such as max pooling or average pooling), the dimensionality of the feature map is reduced, reducing computational complexity while preserving important information (Mehmood, A., Sher, H. A., Murtaza, A. F., et al., 2021). Fully connected layer: In the final stage of the network, the extracted features are integrated and typically used for classification tasks. The structure is shown in the following figure.

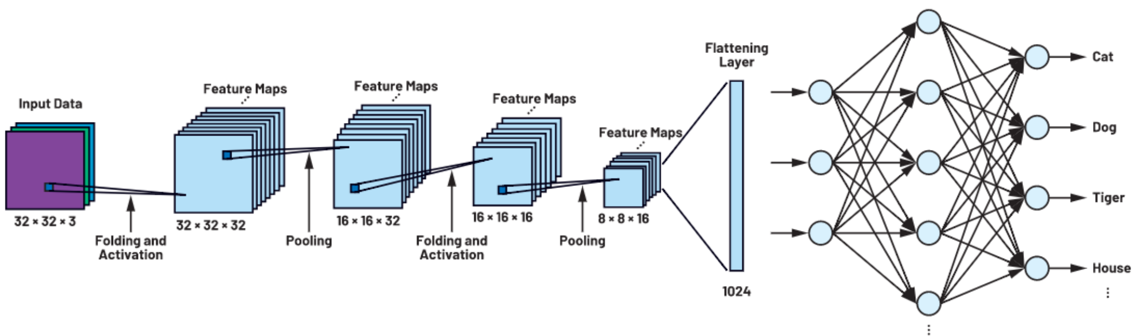


Figure 4. Convolutional Neural Network Structure Diagram

3.3 Convolutional Layer

The convolutional kernel used in convolutional layers, also known as neurons, is mainly used to perform scanning convolution operations on the local input values and obtain their features through scanning. The important feature of this layer is the sharing of weights, and the same core performs traversal scans with a determined scan length. Through this step, the parameters of the convolutional layer network can be reduced, thereby avoiding over fitting problems caused by too many parameters and occupying too much memory due to too many parameters. The parameter calculation formula is:

$$y^{l(i,j)} = \sum_{j'}^{c-1} w_i^{l(j')} \Theta x^{l(j+j')} + b_i^{l(j')}$$

In the formula, $w_i^l(j')$ represents the *jth* weight in the i-th convolution kernel of layer l; *xl* (*j* +*j*') represents the jth position of this layer after convolution; Bil (j ') represents the bias term. Θ represents convolution operation. Convolution operation refers to a mathematical operator that generates a third function by the interaction of two functions f and g, representing the area of the trapezoid formed by the product of the function f and g through flipping and translation. The function of convolution is to act on the area by moving a certain step, so that the entire area is scanned and a certain area is obtained again. In this article, we choose to traverse all data using translation to obtain their weight values and bias parameters.

3.4 Fully Connected Layer

The fully connected layer is the classifier, which flattens the output of the previous layer as the input of that layer, and then fully connects the input and output. Its schematic diagram is shown in Figure 5, and its main purpose is to classify the features extracted by the filter by tiling. To make the model more robust, this article uses data regularization, which is to standardize the features uniformly. When selecting parameters, the activation functions of both the hidden layer and the output layer (Pillai, D. S. & Rajasekar, N., 2018) are determined as Soft plus, mainly because it can handle more fault features.

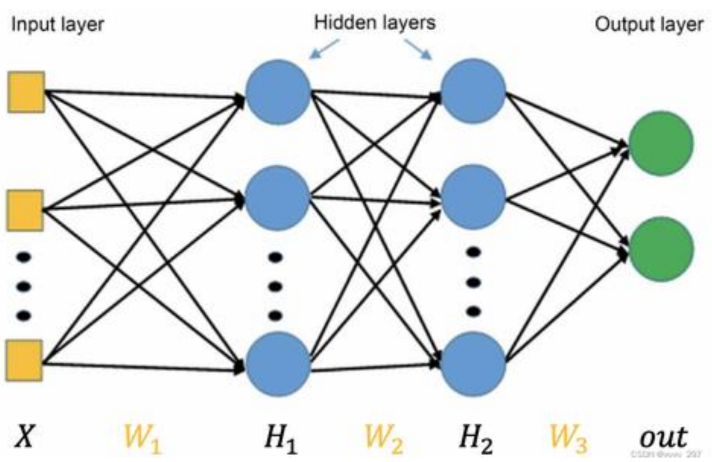


Figure 5. Fully Connected Layer Structure

3.5 Loss Function

The loss function is a function used to evaluate model error (Weerasekara, M., Vilathgamuwa, M., & Mishra, Y., 2018), and the better the loss function, the better the performance of the model is usually. Different models generally use different loss functions. This article uses the cross entropy function, which measures the observation error. The larger the error, the less close it is to the true value, and the smaller the error, the closer it is to the true value.

3.6 Photovoltaic Fault Diagnosis Method Based on Convolutional Neural Network

This section proposes a photovoltaic module fault diagnosis model based on convolutional neural networks. The specific model structure diagram includes convolutional layers, pooling layers, GRU layers, and fully connected layers. Firstly, the original one-dimensional data is used as input, and after passing through convolutional layers, pooling layers, and GRU layers, the parameters γ and β are learned through the network model. Then, the model is tested using target test data, and finally the results are output through fully connected layers and Soft plus (Saleh, M. U., Deline, C., Benoit, E., et al., 2020) layers. As the model is used for multi classification problems, the classification layer of the model uses the Soft plus activation function suitable for multi classification, and the loss function of the model uses the categorical cross entropy loss function. Cross entropy is used to evaluate the difference between the current trained probability distribution and the true distribution. It characterizes the distance between the actual output probability and the expected output probability, that is, the smaller the value of cross entropy, the closer the two probability distributions are.

4. Example Analysis

4.1 Experimental Sample Construction

We used authoritative data from NREL as the dataset for this experiment. In addition, in the previous chapter, we obtained the timing data of photovoltaic module operation through simulation, which includes fault data of photovoltaic modules. This dataset contains data of different irradiance and temperature, and different faults were simulated. Based on the external environment in which the component operates, first determine a temperature of 25℃. To fit a more realistic on-site environment, we set irradiance levels of 400, 800, and 1200W/m2 respectively. We established three datasets, namely Dataset A, Dataset B, and Dataset C, with dataset labels 0-5 representing six types of faults.

4.2 Model Training

In order to select the optimal model framework and specific parameters for each layer of the model, this section conducts a portion of model testing experiments. In order to demonstrate adaptability, a cross experiment was conducted: during the experiment, two sets from datasets A, B, and C were selected in sequence and merged as the training set, while the remaining set was used as the testing set. Each testing experiment was trained for 100 iterations, and the optimal parameters of the proposed model were found by comparing the final output results. The detailed comparison of experimental parameters is shown in Table 1.

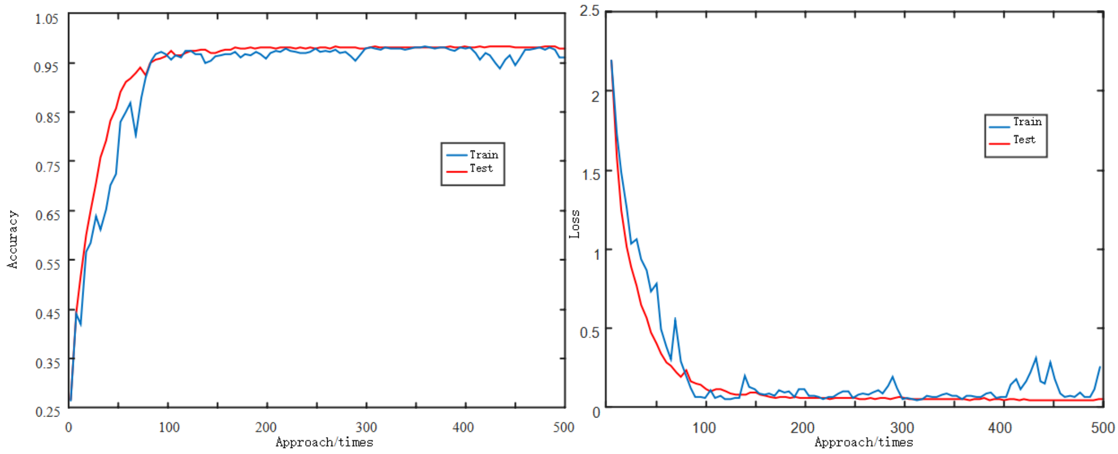
Table 1. Experimental Results under Different Parameters of the Model

Num	**Convolutional Kernel Size**	**Convolutional Kernel Num**	**GRU Neuron**	**Batch_size**	**Test Accuracy**	**Loss**
1	4	16	16	64	0.9267	0.3392
2	16	32	16	64	0.9267	0.0713
3	16	32	32	64	0.9340	0.0237
4	16	32	32	128	0.9460	0.0275
5	32	32	32	128	0.9530	0.0023
6	16	32	32	256	0.9480	0.0273
7	32	32	32	256	0.9540	0.0472
8	64	32	32	256	0.9620	0.0180

According to the comparison, it can be concluded that parameter 8 has the best effect. Therefore, parameter 8 was selected as the subsequent experimental parameter, with corresponding experimental accuracy and loss of 0.9620 and 0.0180, respectively.

4.3 Model Training

Train using the dataset mentioned earlier, with 12000 samples as training data. Divide a portion of the dataset into a test set, with 1200 samples as testing data. The loss and accuracy of the test results are shown in Figure 6.



a) Cross Validation Accuracy Results b)Cross validation loss value results

Figure 6. Visualization Display of Training Results

5. Discussion

In order to ensure the normal operation of photovoltaic arrays, improve the power generation and operation efficiency of photovoltaic power plants, and increase the economic benefits of photovoltaic power generation, this paper applies deep learning to photovoltaic module fault diagnosis. Using the strong data classification of deep learning, a deep learning model suitable for distributed photovoltaic module and large-scale photovoltaic module fault diagnosis is designed. We have mainly developed a fault diagnosis method for photovoltaic arrays based on machine learning algorithms. This method can diagnose multiple common faults of photovoltaic arrays and achieve efficient fault diagnosis of photovoltaic arrays in different application scenarios, with high model classification accuracy.

This article mainly studies the fault diagnosis methods of components in different types of photovoltaic arrays, and has also achieved certain results. At present, there have been many studies on fault diagnosis of photovoltaic power generation systems, but further research is needed on how to diagnose more quickly and accurately. Regarding the work done in this article, further research can be conducted in the following areas: for components of different sizes of arrays, the same diagnostic method can be used to ensure their safety, reliability, and speed, making them applicable to various arrays; During actual operation, the accuracy of the sampling equipment and electromagnetic interference may have a certain impact on the fault diagnosis results. Propose anti-interference diagnostic methods for different sampling environments; In view of the current lack of fault location methods for photovoltaic modules, future research can focus on locating methods for photovoltaic modules.

References

Ali, M. H., Rabhi, A., Hajjaji, A. E., et al. (2017). Real Time Fault Detection in Photovoltaic Systems. *Energy Procedia*, *111*, 914-923. https://doi.org/10.1016/j.egypro.2017.03.254

Chen, Z., Wu, L., Cheng, S., et al. (2017). Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics. *Applied Energy*, 2017. https://doi.org/10.1016/j.apenergy.2017.05.034

Jaffery, Z. A., Dubey, A. K., Irshad, et al. (2017). Scheme for predictive fault diagnosis in photo-voltaic modules using thermal imaging. *Infrared Physics & Technology*, *2017*, S1350449517300488. https://doi.org/10.1016/j.infrared.2017.04.015

Kong, X., Xu, X., Yan, Z., et al. (2018). Deep learning hybrid method for islanding detection in distributed generation. *Applied Energy*, *210*, 776-785. https://doi.org/10.1016/j.apenergy.2017.08.014

Lu, X., Lin, P., Cheng, S., et al., (2019). Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph. *Energy Conversion and Management*, *196*(950-65). https://doi.org/10.1016/j.enconman.2019.06.062

Mehmood, A., Sher, H. A., Murtaza, A. F., et al. (2021). A Diode-Based Fault Detection, Classification, and Localization Method for Photovoltaic Array. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-12. https://doi.org/10.1109/TIM.2021.3077675

Mekki, H., Mellit, A., & Salhi, H. (2016). Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules. *Simulation Modelling Practice & Theory*, *67*, 1-13. https://doi.org/10.1016/j.simpat.2016.05.005

Pillai, D. S., & Rajasekar, N. (2018). An MPPT-based sensorless line–line and line–ground fault detection technique for PV systems. *IEEE Transactions on Power Electronics*, *34*(9), 8646-8659. https://doi.org/10.1109/TPEL.2018.2884292

Saleh, M. U., Deline, C., Benoit, E., et al. (2020). An overview of spread spectrum time domain reflectometry responses to photovoltaic faults. *IEEE Journal of Photovoltaics*, *10*(3), 844-851. https://doi.org/10.1109/JPHOTOV.2020.2972356

Tina, G. M., Cosentino, F., & Ventura, C. (2016). Monitoring and Diagnostics of Photovoltaic Power Plants. *Renewable Energy in the Service of Mankind Vol II. City*, *2016*, 505-516. https://doi.org/10.1007/978-3-319-18215-5_45

Tsanakas, J. A., Ha, L. D., & Al Shakarchi, F. (2017). Advanced inspection of photovoltaic installations by aerial triangulation and terrestrial georeferencing of thermal/visual imagery. *Renewable energy*, *102*(PT.A), 224-233. https://doi.org/10.1016/j.renene.2016.10.046

Wang, W., Liu, A. C., Chung, H. S., et al., (2016). Fault Diagnosis of Photovoltaic Panels Using Dynamic Current–Voltage Characteristics. *IEEE Transactions on Power Electronics*, *31*(2), 1588-1599. https://doi.org/10.1109/TPEL.2015.2424079

Weerasekara, M., Vilathgamuwa, M., & Mishra, Y. (2018). Detection of high impedance faults in PV systems using mathematical morphology In 2018 IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES). *IEEE*, *2018*, 357-361. https://doi.org/10.1109/IESES.2018.8349902

Yang Guangqiao, Liu Bingfeng, & Xu Xiaolong. (2024). *Exploration of Intelligent Management and Application of Oilfield Equipment. Petrochemical Technology*, *31*(06), 303-305.

Yi, Z., & Etemadi, A. H. (2017). Line-to-Line Fault Detection for Photovoltaic Arrays Based on Multiresolution Signal Decomposition and Two-Stage Support Vector Machine. *IEEE Transactions on Industrial Electronics*, *64*(11), 8546-8556. https://doi.org/10.1109/TIE.2017.2703681

Zhu, H., Lu, L., Yao, J., et al. (2018). Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model. *Solar Energy*, *176*, 395-405. https://doi.org/10.1016/j.solener.2018.10.054