A Failure Detection Method based on SVM Model for Solar Power Generation Equipment

Longyun Qi[1,2], Xingshen Wei[1,2,*], Weiping Zhu[3], Jian Zhou[1,2], and Yongjian Cao[1,2]
[1]NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing, Jiangsu, China
[2]Nanjing NARI Information and Communication Technology Co. Ltd., Nanjing, Jiangsu, China
[3]State Grid Jiangsu Electric Power Co., Ltd., Nanjing, Jiangsu, China
qilongyun@sgepri.sgcc.com.cn, weixshen@gmail.com, jszhuweipin@163.com, {zhoujian3, caoyongjian}@sgepri.sgcc.com.cn
*corresponding author

Abstract—Solar energy has emerged as a cornerstone in the quest for renewable energy sources, with its low carbon footprint and abundant availability propelling its adoption. The proliferation of solar power generation devices across the globe is a testament to the commitment to a sustainable energy future. These devices, however, are subject to a myriad of challenges that can impair their operation. Environmental factors such as extreme weather, dust accumulation, and shading, along with human-induced damages or technical malfunctions, can precipitate faults that degrade the performance of solar arrays and, by extension, the robustness of the power grid they support. The intermittent nature of solar power already poses a challenge to grid stability; device failures exacerbate this issue, potentially leading to fluctuations in power supply and even outages. To mitigate these risks and enhance the reliability of solar power systems, we have introduced a sophisticated hierarchical failure detection strategy grounded in Support Vector Machine (SVM) algorithms. This innovative approach organizes solar power devices into clusters, optimizing the monitoring process. At the helm of each cluster, a dedicated failure detector scrutinizes operational data transmitted by the devices, employing SVM classification to discern the functional status of each unit. We have subjected our hierarchical SVM-based failure detection method to rigorous testing within a controlled simulation environment. The empirical evaluation of our system reveals a marked improvement in detecting device anomalies, as evidenced by substantial gains in Detection Accuracy (DA) and a reduction in the False Positive Rate (FPR). These advancements signify not only a stride forward in fault diagnosis in solar power networks but also a step toward ensuring the continuous, reliable delivery of clean energy. Our method promises to bolster the resilience of solar power infrastructure, thereby supporting the broader integration of renewable energy sources into the modern electrical grid.

Keywords–failure detection; solar power generation device; SVM classifier; power supply safety

1. Introduction

To tackle the increasing depletion of petrochemical energy, solar energy [1], [2], [3] has become a vigorously developed renewable energy source globally. Solar energy is an ideal renewable green energy [4], and solar power generation is a crucial way to address the growing scarcity of petrochemical energy and reduce environmental pollution. Solar power (as shown in Figure 1) generation not only has the advantages of being environmentally friendly and renewable but also is not restricted by geographical location, can be used on-site, is easy to store, allows for flexible design in terms of scale, and is convenient to integrate with buildings [5], [6].

It has witnessed substantial advancements, with a notable global trend towards increased efficiency and cost reduction in photovoltaic (PV) technology. High-end solar panels have achieved conversion efficiencies exceeding 22%, thanks to ongoing research into multi-junction cells and perovskite materials. Concurrently, the cost of solar installations has decreased significantly, enhancing competitiveness with fossil fuels. This is attributed to economies of scale, improved technologies, and optimized supply chains. Energy storage systems, particularly batteries, are increasingly being coupled with solar panels to manage the intermittency of solar power, ensuring a more stable energy supply. Additionally, to circumvent land constraints, floating solar farms are gaining popularity, providing the dual benefits of energy generation and water conservation. In China, the world's leader in solar PV production and deployment, the pace of solar farm construction continues unabated, especially in desert regions. Chinese firms are pouring resources into R&D to further enhance solar technology, with the government scaling back subsidies to foster a more market-driven industry. Solar projects increasingly incorporate large-scale battery systems to provide consistent power output. Moreover, solar energy is central to China's rural electrification efforts, ensuring clean energy access to off-grid communities. As a global solar market player, China's influence extends beyond its borders through the export of solar technology and expertise.

However, utilizing solar energy to provide stable, continuous electrical energy output still presents some challenges. For instance, widely distributed solar power generation devices are prone to faults due to various natural and human-induced factors, posing a challenge to power supply safety. Fault detection methods are essential to address the issue of power supply safety caused by faults in power generation devices [7], [8]. They can timely and accurately identify faulty power generation devices for prompt maintenance and updates, ensuring the stability and continuous power output of the entire

2835-2823/24/$31.00 ©2024 IEEE
DOI 10.1109/ISSSR61934.2024.00070

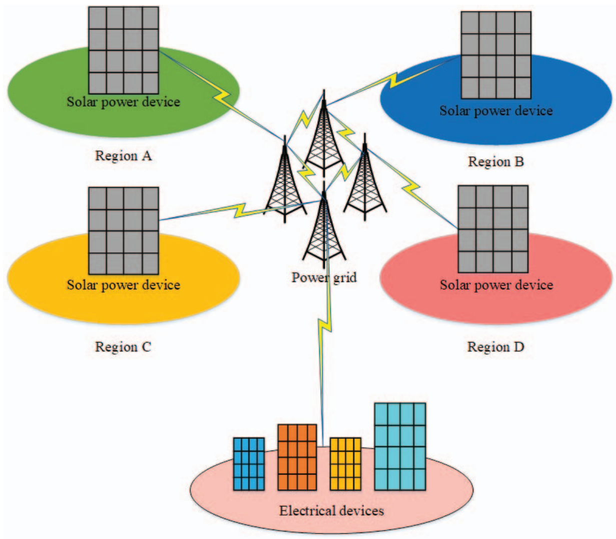

Figure 1. Solar power model

power generation network. Currently, there are several methods applied in the fault detection of power grid equipment. For example, in [9], an expert system method is employed to detect faults in circuit breakers. This approach establishes an expert knowledge base by describing the experiential rules of maintenance personnel and compares the established rules with monitoring data, completing the fault diagnosis. In [10], an artificial immune network classification algorithm is used for transformer fault diagnosis. The algorithm utilizes information from antigen and memory antibody categories in the immune network, constructs the learning of fault samples through the construction of the artificial immune network, obtains the memory antibody set of fault sample features, and performs classification through the nearest neighbor method to achieve transformer fault diagnosis. In [11], a Petri net model is established for transmission lines and bus-bars, and the Petri net model can locate faults when they occur. These fault detection methods typically require the establishment of complex fault detection models and a large amount of data for model training. This requires more computational, storage, and network performance for the entire distributed solar power generation network, and the accuracy and time of detection are also affected by multiple parameters.

In this paper, we propose a fault detection method based on SVM model. SVM classification, as a binary classification method, can handle both linear and nonlinear data. Moreover, SVM classification models [12] are characterized by their simplicity and the absence of a need for extensive data to train the model. Therefore, SVM classification is suitable for processing linear data output by distributed solar power generation devices, thereby achieving fault detection of solar power generation devices. Through simulation experiments, we demonstrate that our proposed method can quickly identify faulty solar power generation devices and ensure the accuracy of fault detection. The innovations of this paper mainly include the following aspects:

- We propose a fault detection method based on SVM for solar power generation devices, which can achieve timely and accurate fault detection to meet the requirements of power supply safety.
- By adopting a hierarchical detection architecture, we can promptly process equipment operating data locally, providing assurance for timely subsequent recovery measures after the occurrence of faults in power generation devices.
- Through the establishment of a simulation experiment platform, we verify the detection speed and accuracy of our proposed fault detection method.

The remainder of this paper is organized as follows: Section 2 introduces the system model; Section 3 describes the implementation of the fault detection method; Section 4 validates the performance of our proposed fault detection method through experiments; Section 5 concludes the paper.

2. System Model

We assume that the distributed power grid consists of N solar power generation devices. Solar power generation devices (equivalent to nodes in this paper) are randomly distributed in an $A*B$ square meter area. Each node has a unique identifier, and if a node fails and exits the network, rejoining the network will result in the reassignment of an identifier. Each node has a monitoring module to monitor the operational status of the node, including temperature, humidity, GPS information, pressure, and other information. In addition, nodes also have a transmission module that allows them to send or receive node monitoring data with neighboring nodes. By relying on multiple routes through neighboring nodes, node monitoring data can eventually be uploaded to the cloud. Since nodes have power generation capabilities, it is assumed that nodes are not subject to energy limitations.

We assume that the communication links between any two nodes are unreliable communication links, and the communication links between them belong to fair-lossy links [13]. This means that no message can be copied or modified between them, and new messages cannot be created. If a normal node p continuously sends a message m to a normal node q, node q will eventually receive message m.

We assume the existence of some local-global time (not known to every node), referred to as global stable time. Under this assumption, nodes function normally in sending, receiving, and processing data. In addition, the steps of continuous data processing by nodes require a time greater than 0.

3. Implementation of Failure Detection Method

3.1 Node Clustering

In the network assumed in this paper, nodes are randomly distributed, but the GPS information of each node is known. Based on this, we consider using the LEACH algorithm [14] for node clustering in the network. In the LEACH algorithm, each node has an opportunity to be elected as a cluster head and randomly generates a random number between 0 and 1.

If this random number is less than the threshold P, the node is selected as the cluster head. The calculation formula for threshold P is:

$$P_i(t) = \begin{cases} \frac{k}{N-k\cdot(r \bmod \frac{N}{k})}, C_i(t) = 1 \\ 0, C_i(t) = 0 \end{cases} \tag{1}$$

$$P_{CH}(i,r) = \begin{cases} 1, R(i) < P_i(t) \\ 0, otherwise \end{cases} \tag{2}$$

Where r represents the number of election rounds, k represents the number of cluster heads elected, N represents the total number of nodes in the network, $C_i(t) = 0$ represents that node i has already been elected as a cluster head, and $C_i(t) = 1$ represents that node i has not been elected as a cluster head, i.e., only nodes that have not been cluster heads in previous rounds can become cluster heads in the current round. In Equation 2, $R(i)$ is a random number between 0 and 1, $P_{CH}(i,r)$ representing the probability that node i becomes a cluster head in round r.

The improvement plan of this LEACH algorithm mainly considers node processing capability and neighbor density. By adding a cost function C_{total} for each node when selecting cluster heads, this function can be expressed as:

$$C_{total} = C_{cpu} + C_{density} + C_{distance} \tag{3}$$

Where C_{cpu} represents the node processing capability and can be calculated using the following formula:

$$C_{cpu} = \frac{IC_{total}}{CPI \times IC \times 1/f_{CPU}} \tag{4}$$

Where IC_{total} represents the total number of executed instructions, CPI represents the average number of clock cycles required to execute each instruction, IC represents the number of times the instruction is executed, and f_{CPU} represents the clock frequency. $C_{density}$ represents the node density and can be calculated using the following formula:

$$C_{density} = \{neighbor(i)|d_{ij} < d_0, i \neq j\} \tag{5}$$

Where d_{ij} represents the distance between node i and node j, and d_0 represents the threshold measuring the distance between nodes.

In the traditional LEACH algorithm, a new set of cluster heads needs to be selected in each round. This frequent selection of cluster heads can lead to a significant increase in communication overhead and may result in network congestion. Therefore, in this improvement plan, cluster heads are selected only when a cluster head fails, to minimize the frequency of cluster head selection. The process of clustering of our method is shown in Algorithm 1. For all the nodes, they can be selected as the cluster head if $C_i(t) = 1$. Through the comparison between $R(i)$ and $P_{CH}(i,r)$, the node can be selected as the cluster head eventually. For the non-cluster head nodes, they need compute the distance between themself and cluster head. Finally, these nodes can find the nearest cluster head as their cluster head and join the cluster.

Algorithm 1 Cluster Head Election Algorithm

```
Input: N, k, r
Output: P_ch(i,r) = 1 or P_ch(i,r) = 0

for all nodes do
    Compute P_ch(i,r)
    if C_i(t) = 0 then Abandoning cluster head elections
    else if C_i(t) = 1 then
        Select a value R(i) randomly
        Compare R(i) and P_ch(i,r)
        if R(i) < P_ch(i,r) then
            Node i becomes the cluster head
        end if
    end if
end for

for each non-cluster head node do
    for all l > 0 do
        Select CH_l
        Compute the distance between oneself and CH_l
        Join the nearest cluster
    end for
end for
```

3.2 SVM Model

Vapnik [15] proposed a region separation algorithm, called SVM. The main purpose of SVM is to find the optimal hyperplane that divides the data into two classes. In this technique, its principle is to define a decision function $f : X \to \{-1, 1\}$, while having a normal set of data $\{(x_i, y_i); x_i \in X \text{and} y_i \in (-1, 1)\}$ and $x \in X$. For each new point $x \in X$, we can use the decision function to estimate whether it belongs to the right class $((-1)\text{or}(+1))$. To minimize the structural risk, this decision can be made by relying on empirical risk.

It is supposed that there are the following empirical data $(x_1, y_1) \ldots (x_i, y_i) \ldots (x_m, y_m) \in R \times \{\pm 1\}$. In the case of linear classification, a hyperplane, which is computed by the SVM algorithm, separates at best the samples of two classes. In that case, the function f is linear in x_i with the following general form: $f(x_i =< \omega, x_i >) + b$ [30]. As shown in Figure 2, these data can be separated by an infinity of hyperplanes. However, the middle hyperplane of them is optimal. The limiting condition of this hyperplane is as follows:

$$y_i(\omega x_i + b) \geq 1 \text{for} i = 1 \cdots m \tag{6}$$

The optimization problem is described in the following way: $\min_{x \in X} f(x)$ under constraint $g_i(x) \leq 0$. The Lagrangian method is used to resolve this problem. So, the dual problem becomes:

$$\begin{cases} \sum_i^L \alpha_i - \frac{1}{2}\sum_{i,j}^L \alpha_i \alpha_j y_i y_j < x_i, x_j > \\ \sum_i^L \alpha_i y_i = 0 \end{cases} \tag{7}$$

where α_i are lagrangian parameters ($\alpha_i \neq 0$ for

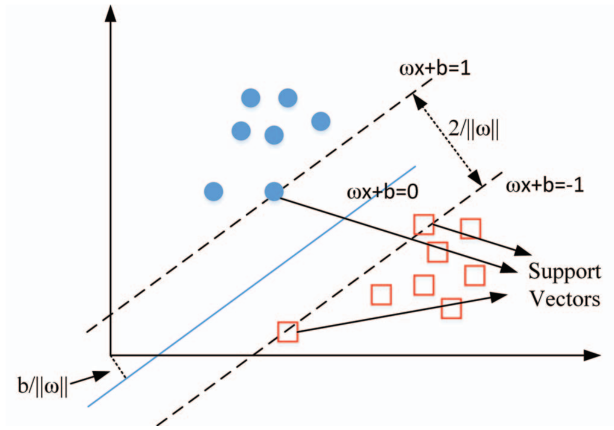

Figure 2. Linear separation

$y_i(<\omega, x_i> + b) = 1$ and $\alpha_i = 0$ for $y_i(<\omega, x_i> + b) > 1$).
So, the weight hyperplane vector $\omega = \sum_{i=1}^{L} \alpha_i y_i x_i$ and the offset $b = 1 - <\omega, x_i>$ can be obtained, where x_i is the support vector of the known class (here its class is 1), and we obtain the hyperplane function.

$$f(x) = <\omega, x> + b = \sum_{i \in SV} \alpha_i y_i < x_i, x > + b \tag{8}$$

For the nonlinear classification, it is not useful to the separator hyperplane of linear classification. Thus, it is useful in this case by employing nonlinear SVM. The principle of nonlinear SVM is to find a space with the biggest dimension where the projection of examples is linearly separable (as shown in Figure 3), which is a Hilbert space H based on a scalar product that can be replaced by a kernel function of the starting space (space of observations).
We suppose:

$$\emptyset : R^P \to H; x \mapsto \emptyset(x_i) \tag{9}$$

here a kernel function $K(x_i, x_j)$ is used to replace the scalar product $< \phi(x_i), \phi(x_j) >$, so the problem of optimization is translated into

$$\begin{cases} \max_{\alpha_i} \sum_{\alpha_i}^{L} \alpha_i - \frac{1}{2} \sum_{i,j}^{L} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \sum_{i}^{L} \alpha_i y_i = 0 \\ c \geq \alpha_i \geq 0 \end{cases} \tag{10}$$

where C is the tolerance constant. We can obtain the decision function:

$$f(x) = <\omega, x> + b = \sum_{i \in SV} \alpha_i y_i K(< x_i, x >) + b \tag{11}$$

We use the following Gaussian kernel function for classification.

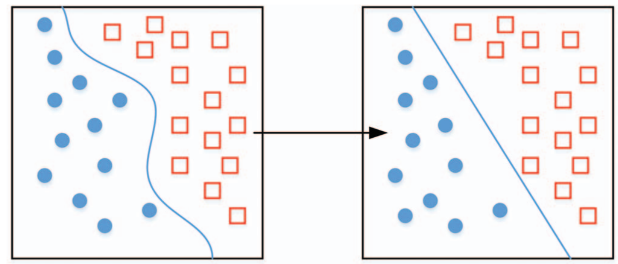

Figure 3. Non-linear separation

$$K(x|x') = exp(-\frac{\|x - x'\|}{2\sigma^2}) \tag{12}$$

3.3 Node Failure Detection

To monitor the status of nodes in the assumed network described in this paper, we propose a hierarchical failure detection method, as illustrated in Figure 4. In this detection method, the system is divided into three layers: the bottom layer consists of regular nodes, followed by the cluster head node layer, and finally, the top layer is the cloud service layer. In each cluster, regular nodes send operational status information (such as voltage, current, temperature, humidity, etc.) to the cluster head node. The SVM classifier located at the cluster head categorizes this information to determine the operational status of the node. Simultaneously, the operational status information from the cluster head node is uploaded to the cloud service center via the base station. The SVM classifier at the cloud service center categorizes this information to determine the operational status of the cluster head node.
According to the above analysis, we propose a node failure detection algorithm, as shown in Algorithm 2. In this algorithm, regular nodes periodically send operational status messages to the cluster head node, with a time interval set as Δt. For the cluster head node, it has two main functions. First, it receives operational status messages from regular nodes, extracts the information, and classifies the nodes using its deployed SVM classifier. If it detects a node failure, it sends the information of the failed node to the cloud service center. Second, it periodically sends operational status messages to the cloud service center, with a time interval also set as δ. For the cloud service center, upon receiving operational status messages from the cluster head node, it extracts the information and classifies the cluster head node using its deployed SVM classifier. If it detects a failure in the cluster head node, it initiates the next round of cluster head node selection and re-clustering of nodes.

4. Experiments

In this section, to verify the effectiveness of the proposed failure detection method, we used MATLAB to build a simulation platform and compared it with other failure detection methods. In the experiment, we assumed a total of 200 nodes randomly

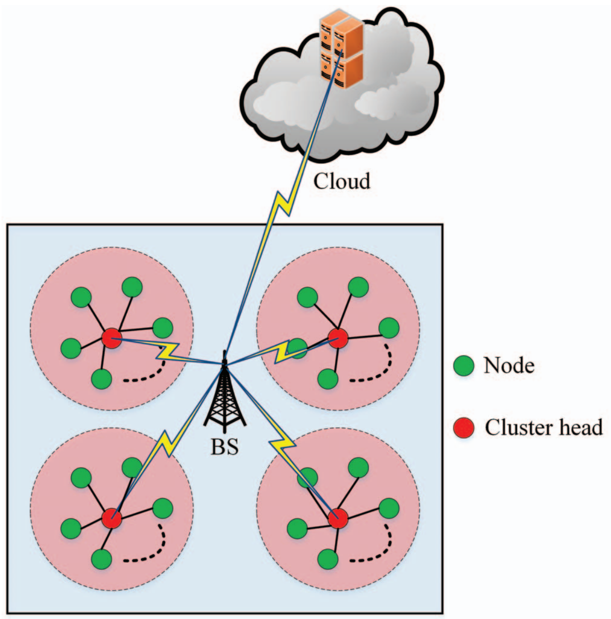

Figure 4. Hierarchical failure detection method

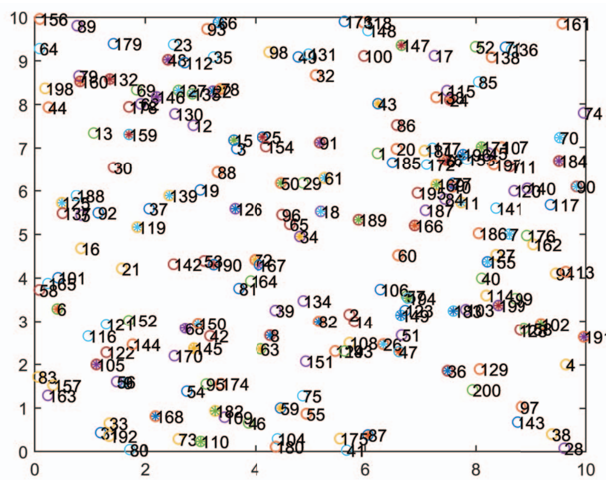

Figure 5. Network with 200 nodes

distributed in a $10*10KM$ area, with the base station located at the center of the region, as shown in Figure 5.

For evaluating the performance of the failure detection methods, we used accuracy metrics. In the comparative experiment, we chose a Bayesian-based failure detection method as a benchmark, evaluating our proposed failure detection method and the Bayesian-based method using two accuracy metrics. The first metric is Detection Accuracy (DA), calculated as follows:

$$DA = \frac{Number of faulty node detected}{Total number of faulty node} \tag{13}$$

Algorithm 2 Node Failure Detection Algorithm

Input: information from node or cluster head, Δt
Output: node $\rightarrow$ normal; node $\rightarrow$ failure; cluster head $\rightarrow$ normal; cluster head $\rightarrow$ failure

for node at time t **do**
  Send information to cluster head
end for

for cluster head **do**
Task 1:
  if receiving the information from node **then**
    Extract the information
    Classify node $\rightarrow$ normal or node $\rightarrow$ failure
    Send alarm information to Cloud if node $\rightarrow$ failure
  end if
Task 2:
  for all nodes at time t **do**
    Send information to Cloud
  end for
end for

for Cloud **do**
  if receiving the information from cluster head **then**
    Extract the information
    Classify cluster head $\rightarrow$ normal or cluster head $\rightarrow$ failure
    Reselect the cluster head if cluster head $\rightarrow$ failure
  end if
end for

The second metric is False Positive Rate (FPR), calculated as follows:

$$FPR = \frac{Number of non-faulty node detected as faulty}{Total fault free nodes} \tag{14}$$

To simulate node failures and evaluate the performance of our proposed failure detection method and the Bayesian-based method, we conducted comparative experiments using fault injection. In the comparative experiment, we set the node failure rate to range from 0.05 to 0.4.

Since the proposed method uses an improved LEACH method to cluster nodes, the size of each cluster has a noticeable impact on the accuracy of failure detection, as shown in Figures 6 and 7. From Figures 6 and 7, it can be observed that as the number of nodes in each cluster increases, both methods show an increasing trend in DA and a decreasing trend in FPR. The reason for this phenomenon might be that with the increase in data uploaded by nodes, the SVM model and Bayes model can be better trained, enhancing the accuracy of the models and consequently improving the accuracy of failure detection. However, it can also be seen from the figures that our proposed method has better accuracy, possibly because the SVM classifier can handle linear data more effectively.

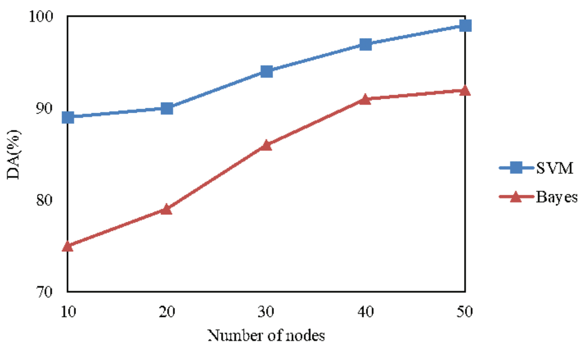

Figure 6. DA vs. number of nodes

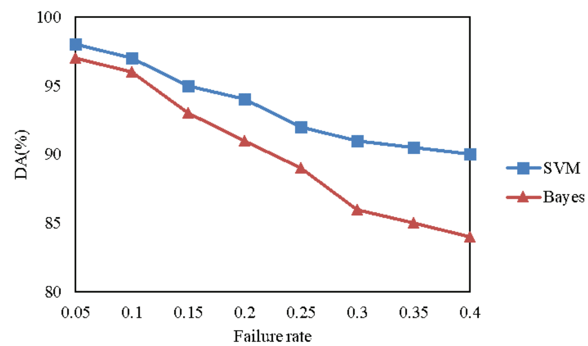

Figure 8. DA vs. failure rate

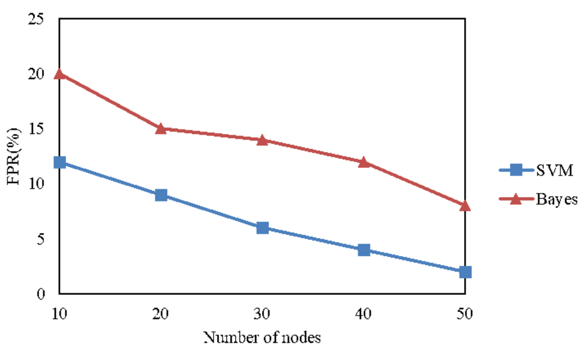

Figure 7. FPR vs. number of nodes

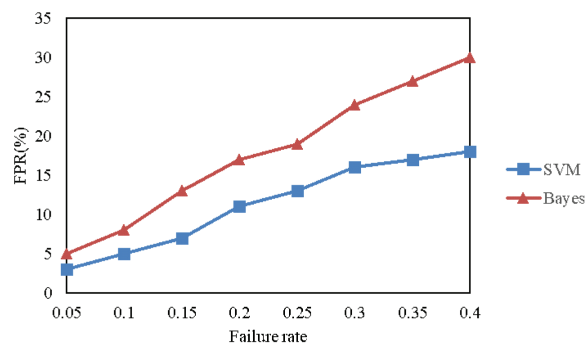

Figure 9. FPR vs. failure rate

In addition to being influenced by the number of nodes in each cluster, the accuracy of the failure detection method is also affected by the real failure probability of nodes. By injecting faults, we achieved different failure rates for nodes to validate the performance of our proposed failure detection method and other methods. As shown in Figures 8 and 9, as the node failure probability increases, both methods show a decreasing trend in DA and an increasing trend in FPR. This may be due to the increase in node failure rate, which affects the classification of both failure detection methods, leading to the incorrect identification of some failed nodes. However, our proposed method demonstrates better DA and FPR as the node failure probability increases. This may be because the SVM model can accurately identify nodes that have failed with fewer data.

5. Conclusion

The failure detection method is one of the basic components for maintaining the high availability of a system. In this paper, we proposed an important hierarchical failure detection method based on SVM classifiers. This method clusters nodes in the system and then deploys failure detectors at the cluster head nodes and the cloud, thus identifying failed nodes. Additionally, this paper verified the performance of the proposed failure detection method and other failure detection methods through building a simulation environment. Based on the experimental results, our proposed failure detection method outperforms other failure detection methods in terms of DA and FPR, making it more suitable for failure detection in distributed solar power generation devices.

Acknowledgment

This work is supported by the science and technology project of State Grid Corporation of China "Deepening research on key technologies for dynamic defense of network security in new distribution systems" (Grand No.5400-202340217A-1-1-ZN).

References

[1] A. Echevarria, Y. Rivera-Matos, N. Irshad, C. Gregory, M. J. Castro-Sitiriche, R. R. King, and C. A. Miller, "Unleashing sociotechnical imaginaries to advance just and sustainable energy transitions: The case of solar energy in puerto rico," *IEEE Transactions on Technology and Society*, vol. 4, no. 3, pp. 255–268, 2023.

[2] N. M. Haegel and S. R. Kurtz, "Global progress toward renewable electricity: Tracking the role of solar (version 3)," *IEEE Journal of Photovoltaics*, vol. 13, no. 6, pp. 768–776, 2023.

[3] S. Sadhukhan, S. Acharya, T. Panda, N. C. Mandal, S. Bose, G. Das, S. Maity, P. Chaudhuri, S. Chakraborty, P. Chakrabarti, and H. Saha, "Improvement in the performance of p-type tunnel oxide passivated contact solar cell with selective tunneling junctions underneath the contacts on front side," *IEEE Journal of Photovoltaics*, vol. 13, no. 2, pp. 236–241, 2023.

[4] D.-G. Owusu-Manu, L. Mankata, D. Edwards, H. El-Gohary, and J. Hussain, "A cognizance of green bond features preferential to renewable energy project financing in ghana," *IEEE Transactions on Engineering Management*, vol. 70, no. 3, pp. 979–990, 2023.

[5] D. Das, B. Singh, and S. Mishra, "Grid interactive solar pv and battery operated air conditioning system: Energy management and power quality improvement," *IEEE Transactions on Consumer Electronics*, vol. 69, no. 2, pp. 109–117, 2023.

[6] K. Chauhan, J. Moirangthem, S. Ly, L. M. Koh, and H. D. Nguyen, "Stochastic day-ahead scheduling of distributed energy resources: A meta modeling approach," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, vol. 5, no. 1, pp. 179–192, 2024.

[7] L. Hou, S. Zhang, Y. Wei, B. Zhao, and Q. Jiang, "A dynamic series voltage compensator for the mitigation of lcc-hvdc commutation failure," *IEEE Transactions on Power Delivery*, vol. 36, no. 6, pp. 3977–3987, 2021.

[8] S. Robak, J. Machowski, M. Skwarski, and M. Januszewski, "Improvement of power system transient stability in the event of multi-phase faults and circuit breaker failures," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 2422–2430, 2020.

[9] Y. M. Park, G.-W. Kim, and J.-M. Sohn, "A logic based expert system (lbes) for fault diagnosis of power system," *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 363–369, 1997.

[10] X. Hao and S. Cai-xin, "Artificial immune network classification algorithm for fault diagnosis of power transformer," *IEEE Transactions on Power Delivery*, vol. 22, no. 2, pp. 930–935, 2007.

[11] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—part i: Fault diagnosis with model-based and signal-based approaches," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3757–3767, 2015.

[12] S. Bourouis, A. Zaguia, N. Bouguila, and R. Alroobaea, "Deriving probabilistic svm kernels from flexible statistical mixture models and its application to retinal images classification," *IEEE Access*, vol. 7, pp. 1107–1117, 2019.

[13] J. Zhang and W. Chen, "Implementing uniform reliable broadcast with binary consensus in systems with fair-lossy links," *Information Processing Letters*, vol. 110, no. 1, pp. 13–19, 2009.

[14] R. Gantassi, Z. Masood, and Y. Choi, "Enhancing qos and residual energy by using of grid-size clustering, k-means, and tsp algorithms with mdc in leach protocol," *IEEE Access*, vol. 10, pp. 58 199–58 211, 2022.

[15] V. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.