EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

# Smart meter-based outage detection method for power distribution systems

Xuan Wang[1], Zhiqiang Shi[1*], Bing Liu[1], Wenbiao Xiao[1], Shuai Cheng[1]

[1]State Grid Huaian Power Supply Company, Huaian, Jiangsu, 223001, China

## Abstract

This paper proposes a new data-driven method for power outage detection. By capturing the changes in data distribution of smart meters (SM), it can detect power outages in partially visible distributed systems. First, a mechanism based on breadth-first search (BFS) is proposed, which decomposes the network into a set of regions to find the location information where power outages are most likely to occur. Then, the SM data for each region, generating a generative adversarial network (GAN), is used in order to extract unsupervised manner implicit temporal behavior under normal conditions. After network training, anomaly scoring technology is used to determine whether the real-time measurement data is the data of a power outage event. Finally, in order to infer the location of a power outage in a multi-area network, a regional coordination process with interdependence be-tween cross-regions is used. At the same time, the concept of entropy is used to provide performance analysis for the algorithm in this paper. This method has been verified on the distribution feeder model with actual SM data. Experimental results show that the algorithm is effective and feasible.

Received on 11 February, 2024; accepted on 17 May 2024; published on 30 May 2024

**Keywords**: outage detection, generative adversarial network, smart meter, breadth-first search



doi:10.4108/ew.5767

★*Corresponding author. Email: Ssszhiqiango@163.com

## 1. Introduction

The electrical system faces a difficult problem with outage detection, which is crucial in distribution networks where there are frequently many outages. At present, the number of smart meters has increased significantly, and smart meters have basically achieved full coverage in the country. When such a large number of smart meters are really put into operation, it is necessary to strictly monitor the operational reliability of smart meters, and timely diagnose and repair meter failures.

The US Energy Information Administration said that, on average, each customer experienced a power outage of almost four hours in 2016. The distribution system operator (DSO) implements a cutting-edge outage management system combining contemporary software tools and protective equipment with two-way communication capabilities in order to shorten outage time, increase system dependability, and boost customer satisfaction. As a result, the DSO is able to gather network data in real-time. However, because of their high cost, devices with intelligent communication capabilities in the distribution system have not been widely adopted. Due to the complex and diverse components of smart meters, damage and various faults will inevitably occur during operation, such as appearance faults, clock unit faults, metering performance faults, etc. Due to the accuracy of measurement, reliability of communication and safety of use, smart meters bring security risks and economic losses to the power system. Therefore, it is an important task to find and deal with smart meter faults in time and extend its service life to ensure the normal operation of smart grid. The customer's fault call and manual inspection are still required for the identification of outage events in the distribution system. However, these two approaches are not only time-consuming and expensive, but also labor-intensive [1], [2], [3]. Many academics have suggested various power loss detection methods in attempt to address these issues. The evaluation mechanism for the Delphi approach is utilized in [4] to pinpoint several outage areas. In order to quickly restore power, reference [5] monitors relay state. The K-means clustering analysis is used in [6] to partition the electric energy data in order to assure accurate classification of the data without overlapping parts. The signal conditioning circuit is employed in [7] to implement power outage

monitoring. When the data uploaded by the current detection card is less than the threshold, it is considered that a power outage event occurs. This approach, meanwhile, is susceptible to mistakes. In [8], switch transposition and current mutation are used to realize power outage monitoring, which is not robust as in previous literatures. To identify and pinpoint the position of damaged lines in a fully visible network using signals from SM, a multi-label support vector machine classification approach is suggested in [9]. In [10], a hierarchical architecture is developed to provide a monitoring technique for anomalous line status. The difference between normal and erratic power use is tracked in [11] to develop a fuzzy Petri nets approach for outage detection. In [12], a real-time dynamic outage detection scheme based on alternating current power flow model and statistical change detection theory is proposed, the proposed method can capture system dynamics since it retains the time-variant and nonlinear nature of the power system. The method is computationally efficient and scales to large and realistic networks. In [13], proposes a graph-guided quickest change detection approach that leverages the grid topology and performs quickest change detection in the spectral domain of the graph underlying grid's topology. In [14], by formulating the sparse line outage detection problem in a standard compressive sensing form, a binary matching pursuit algorithm is proposed to solve the problem. In [15], [16], [17], [18], although many methods have been provided in many literatures, power failure cannot be monitored.

To address these shortcomings, this paper proposes an adversarial generative network (GAN)-based approach. It is possible to identify outage events in the observable distribution system by capturing abnormal SM measurements as a result of outage events. Compared to the previous works, the suggested approach tackles three difficult problems. 1) Unsupervised learning is employed in the proposed model. Additionally, anomaly scores are used to evaluate how outage occurrences differ from regular data; as a result, data with high anomaly scores are regarded as outage events. 2) Effective outage detection requires capturing high-dimensional spatio-temporal relationships in the measurement data. However, traditional data distribution estimators are limited by the high-dimensional data. The strategy suggested in this paper trains GANs to implicitly extract the underlying distribution of the data rather than explicitly building a sophisticated data likelihood function. Two deep neural networks are coupled in each GAN (DNNs). 3) A mechanism based on breadth-first search (BFS) is presented to decompose the large-scale distribution network into a set of intersecting regions, taking into account the system's local observability. Each region is determined by two adjacent observable nodes. Experimental results show that the proposed method has good robustness. Compared with the traditional method, the outage identification accuracy at low samples can be improved.

## 2. Real-time Data Processing

Over 6,000 consumers from the Midwestern United States' hourly energy usage (kWh) and voltage data are included in the historical Advanced Metering System (AMI) data used in this article. Electricity data is collected from January 2015 through May 2018, in which over 95% of these customers are residential and commercial loads. Removing faulty data and lost data as a result of communication problems during data pre-processing. The protective device isolates the faulted area from the load at the lower end of the fault when a radial system failure takes place, as a result, the upper observation nodes that are not faulty to deviate from the normal samples. Define the outage detection area based on the above phenomenon. Generally, two observable nodes can be used to detect outages in the lower path. To illustrate, a typical distribution feeder with two observable nodes (node $n$ and node $n+N$) is given in figure 1, where the voltage drop $\Delta V$ between these two nodes can be expressed as

$$\Delta V = |V_n| - |V_{n+N}| \approx \left| \sum_{i=n+1}^{n+N} Z_{(i-1,i),abc} \bullet I_{i-1,i} \right| \qquad (1)$$

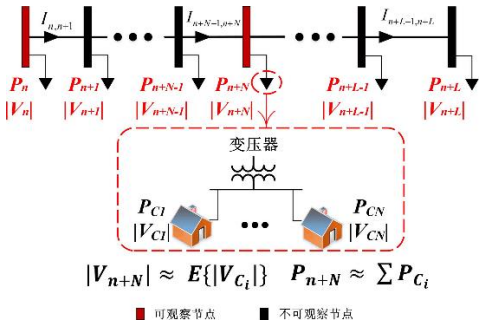


**Figure 1.** Schematic area under normal conditions

where $|V_n|$ and $|V_{n+N}|$ are expressed as the voltage at the observation point, $I_{i-1,i}$ is the branch current, $Z_{(i-1,i),abc}$ is the phase impedance matrix between bus i-1 and bus i. And the dimensionality of the variables in Equation 1 depends on the number of phases in the distribution line. For example, $|V_n|$, $|V_{n+N}|$ and $I_{i-1}$ are all three-phase feeder, is a 3×1 vector. $Z_{(i-1,i),abc}$ is a 3×3 vector. The above equation is expressed in the form of nodal power.

$$\Delta V = \sum_{i=n+1}^{n+N} \sum_{j=i}^{n+L} K_{i-1,i} \otimes l_{i-1,i} \otimes \frac{P_j}{\cos \phi_j} \qquad (2)$$

where $n+L$ is the total length of this path. $K_{i-1,i}$ is the voltage drop factor, $L_{i-1,i}$ is the line length between two nodes $i-1$ and $i$. $P_j$ and $cos\varphi_j$ are the node power consumption and power factor at node $j$, respectively. When an interruption occurs at the unobservable node $s$ below node $n$, the value of voltage drops $\triangle V_o$ after power failure can be expressed as follows.

$$\Delta V_o \approx \Delta V + \sum_{i=n+1}^{\min(s,n+N)} K_{i-1,i} \otimes l_{i-1,i} \otimes \frac{\Delta P_s}{\cos \phi_s} \qquad (3)$$

where $\triangle P_s$ denotes the negative value of the outage event. It can be seen that when an outage event occurs at the lower end of any two nodes, the value of the voltage drops at the two observable nodes changes. And the variation of them is proportional to $\Delta P_s$, which can be confirmed by the real AMI data. As shown in Figure 2, Pag1 and Pag2 are the total power consumption of the first and second observable nodes in the region. The three-dimensional gap between the joint data of two observable nodes under normal conditions and the data of one specific outage condition is shown. Considering that a fault event at any location at the lower end of the two nodes causes a deviation in their underlying joint measurement data distribution during normal operation, the fault detection region is defined as follows.

Definition 1: In a radiation network, an outage detection region, $\Psi_i$, can be defined as $\Psi_i = \{S_{o1}, S_{o2}, Z_{\Psi i}\}$, where $S_{o1}$ and $S_{o2}$ are two observable nodes, with the former at the upper end of the latter. $Z_{\Psi i}$ is the set of all branches at the lower end of $S_{o1}$. From Definition 1, it is clear that for a specific distribution system, different zone selection strategies lead to different zone divisions, which affect the performance of outage detection and localization. Therefore, this paper proposes a BFS region selection method using the tree structure of the distribution system. Specifically, the nodes at the current depth are used to select regions, and then the nodes are moved at the next depth level. The proposed method has two advantages: (1) It is able to obtain the optimal set of regions for maximizing outage location information in any part of the observable network. (2) The proposed BFS-based algorithm introduces an effective topological ordering, which greatly simplifies the fault location identification process.

Definition 2: In a radiation network, node $B$ is defined as a directly observable lower node of any node $A$ when the following two conditions are satisfied: 1) node $B$ is located at the lower end of node $A$; 2) the path connecting $A$ and $B$ contains only unobservable nodes.

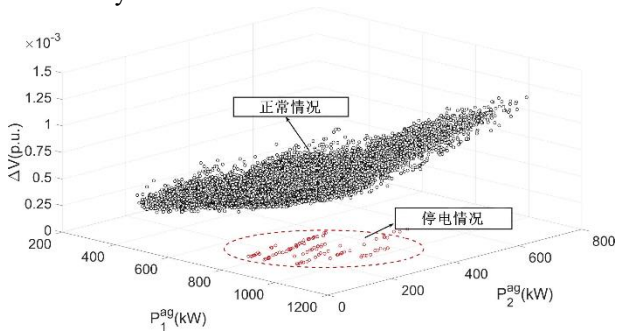


**Figure 2.** Joint distribution data during normal and outage

The proposed algorithm consists of the following steps:

Step 1: A partially visible distribution system $g$, containing $M$ branches, $B_g = \{b_1,..., b_M\}$, and a set of $O+1$ observable nodes $S_g = \{S_r, S_1, S_2,..., S_O\}$, where $S_r$ denotes the root node of the network.

Step 2: Define $g$ and initialize the set of regions and the set of neighboring nodes: $\Psi_g$ and $N(g) = \{\emptyset\}$. where $\Psi_g$ is the ordered set and new elements will be added to the right side of the set. Initialize the candidate node set $SB=\{S_r\}$. The region counter $k$ is set to 1.

Step 3: If $N(g)$ is the empty set, randomly select and remove a node $S_{o1}$ from $S_B$. if not empty, randomly select and remove a node $S_{o1}$ from $N(g)$.

Step 4: Find all the observable nodes at the lower end of $S_{o1}$ (see Definition 2) and randomly select a node from it, denoted as $S_{o2}$. If $N(g) = \{\emptyset\}$, add all the visible nodes at the lower end of $S_{o1}$ to $N(g)$; otherwise, add them to $S_B$.

Step 5: Choose a region $\Psi_k$ with new $S_{o1}$ and $S_{o2}$ and include all branches at the lower end of $S_{o1}$ into $Z_{\Psi k}$ (see Definition 1). Add $\Psi_k$ to the right side of the existing interval.

Step 6: $k = k+1$. Then return to Step 3 until $N(g)$ is empty.

Step 7: Output the ordered set of all network regions, $\Psi_g=\{\Psi_1,.... , \Psi_w\}$. where $w$ denotes the number of selected regions.

For better understanding, an example is given in Figure 3, where $B_g= \{b_1,..., b_{36}\}$ and $S_g = \{S_r, S_1, S_2,..., S_8\}$. In the first iteration ($k = 1$), both $\Psi_g$ and $N(g)$ are empty sets. In step two, the root node is chosen as the first observable node: $S_B = \{S_r\}$. In Step 3, $S_{o1}$ is randomly selected and removed from $S_B$ since $N(g)$ is empty. Thus, $S_{o1}=S_r$ and $S_B=\{\emptyset\}$. In step IV, $S_1$ and $S_2$ are the lower observable nodes of $S_r$. Since $N(g)$ is empty, these two nodes are added to $N(g)$. Then, $S_{o2}$ is randomly selected from $\{S_1, S_2\}$. in this example, $S_{o2}= S_1$. in step V, the first region is defined based on the selected $S_{o1}$ and $S_{o2}$, which are added to the set $\Psi_g$; $\Psi_g = \{\Psi_1\}$. where $\Psi_1= \{S_r, S_1, Z_{\Psi 1}\}$. The algorithm then returns to step three and re-iterates. The region $w$ can be expressed as a function of the number of observable nodes: $w = O+1-Oend$, where $O$ is the number of all visible nodes and Oend is the number of observable nodes without any observable lower nodes. We can conclude that the above indicates that the approach requires the installation of sensors at internal nodes in order to form meaningful zoning. This requirement is consistent with recent smart grid monitoring equipment. In the current distribution system, metering devices are generally installed at selected locations, such as the root node and other major utilities, which can be used to obtain zoning. On the other hand, in many distribution systems, monitoring devices are installed only at the terminal nodes. To deal with the region selection problem, an approximate method is provided in this paper. Definition: internally active nodes are a subset of the internal nodes of the network with non-zero current injection, and internally inactive nodes do not have any injection. The method uses a portion of the measurement data of the observed terminal nodes to represent the recently unobserved internally inactive nodes. The principle is that the voltage drops between an internally inactive node and the nearest terminal node is

usually negligible. Using this approximation, the proposed method achieves a reasonable partitioning of regions when only metering terminal buses are available. When the set of regions is obtained, each branch in the system will belong to at least one region, while no two regions have exactly the same set of branches. For example, the branches of region $\Psi_6$ in Figure 3 are also covered by regions $\Psi_1 ... \Psi_5$.

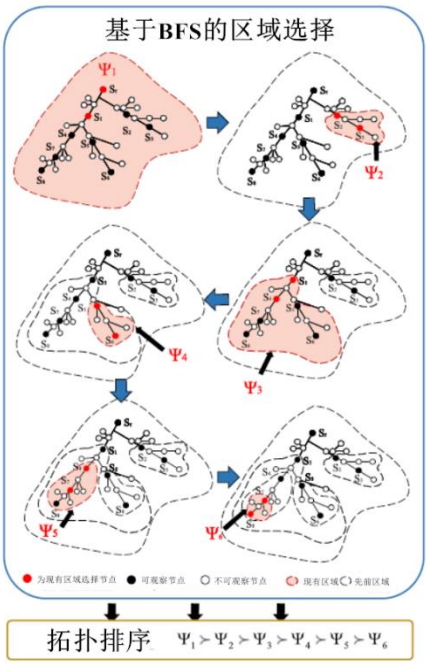


**Figure 3.** BFS-based region selection and ranking method

## 3. GAN-based Area Monitoring

In this paper, we use a parameter-free unsupervised learning method, GAN, to quantify the deviation of the data distribution of power outage events under normal conditions. The advantage of GAN is that it can implicitly represent complex data distributions without constructing a high-dimensional likelihood function, thus solving the problem posed by dimensionality. In addition, GAN does not have a priori parameter structure on the data distribution, which ensures the performance of outage detection. Also, our approach is not prone to blackout data shortage problem because the model training uses only the data under normal conditions. When the training is completed, the GAN-based anomaly scores are assigned to real-time measurements to detect power outage events in the region.

For each region, the GAN is trained to learn the joint distribution of the measured variables ($X$= $\{\triangle V^t, P^t_n, P^t_{n+N}\}^T_{t=1}$) for a time window of $T$ (Figure 1). Where $P^t_n$ and $P^t_{n+N}$ are the power consumption of two observable nodes in the region. $\triangle V^t$ is the voltage difference between the two nodes at time $t$. The purpose of defining time windows on the observed variables is to exploit the temporal relationships between consecutive data samples in the distribution network for more effective anomaly detection. Based on the results of the grid search method, $T$ is chosen to be 3 hours. Although the training process of GAN is an offline process, the computational complexity of the grid search method does not affect the real-time performance of the algorithm. The training set consists of historical SM data for each region and is denoted as $X_{\Psi I}$ in the interval $\Psi_i$. Considering that seasonal variations in customer behavior may mislead to detect the boundary overlap between normal and fault, the dataset is decomposed into separate seasons to train different GAN models. Each dataset is randomly divided into three separate subsets for training (70% of the total data), validation (15% of the total data) and testing (15% of the total data).The GAN relies on two interconnected DNNs that are trained simultaneously by an adversarial process: the generator $G$ and the classifier $D$. This is shown in Figure 4 (part a). The interaction between these two DNNs can be modeled by game theory.

$$\min_{\theta_G}\max_{\theta_D} V(D,G) = \mathrm{E}_{x\psi_i \sim p\psi_i(x\psi_i)}[\log(D(x\psi_i))] + E_{z\sim pz(z)}[\log(1-D(G(z)))] \quad (4)$$

where $\theta_G$ and $\theta_D$ are the learning parameters of $G$ and $D$, and $pX_{\Psi I}$ is the underlying probability density function of the relevant historical data acquired by the two observable nodes. At each iteration, $D$ is trained to maximize the probability of assigning the correct label. On the other hand, $G$ is trained to generate artificial samples that maximize the probability of mislabeling the classifier $D$. The input to $G$ is defined as $z$, representing a uniformly distributed $p_z(z)$ noise signal. In this case, $d$ = 4 shows the best performance on the validation set. After several iterations of training, $G$ and $D$ will reach a unique global optimum. The training process is performed offline, and the steps are shown in Algorithm 1. In this paper, a random search algorithm is used to calibrate the hyper parameter set of GAN. Thus, $G$ consists of three parts:4 neurons in the input layer, 8 neurons in the two hidden layers, and 9 neurons in the output layer. The $D$ also has three parts:9 neurons in the input layer, 8 neurons in the two hidden layers, and 1 neuron in the output layer. In addition, $\{\alpha, m, n_D\}$ are chosen as 0.01, 100, and 1, respectively. Minibatch stochastic gradient descent is used only in order to be able to update $\theta_G$ and $\theta_D$.

Arithmetic 1 GAN training of the interval $\Psi_i$

Requirement: Normal data for region $\Psi_i$

Requirement: Learning rate $\alpha$, $m$ of batch $m$, Initial learning parameters of $G$ and $D$: $\theta_G$ and $\theta_D$

for $t$=0, …, $n_D$ do

Generate sample batches

$P_z \rightarrow \{(z_j)\}^m_{j=1}$

Obtain sample batches from historical data

$p_{X\Psi i} \rightarrow \{x_{\Psi i}(j)\}^m_{j=1}$

Update classifier parameters using gradient descent

$$\delta_{\mathrm{D}} = \frac{1}{m}\sum_{j=1}^{m}[-\log D(x_{\psi_i}(j)) - \log(1 - D(G(z_i)))]$$

$$\theta_{\mathrm{D}} := \theta_{\mathrm{D}} - \alpha * \nabla\theta_{\mathrm{D}}\delta_{D}$$

end for

Update the generator parameters using the gradient descent method $\delta_{\mathrm{G}} = \frac{1}{m}\sum_{j=1}^{m}[-\log D(G(z_j))]$

$$\theta_{\mathrm{G}} := \theta_{\mathrm{G}} - \alpha * \nabla\theta_{G}\delta_{G}$$

end while

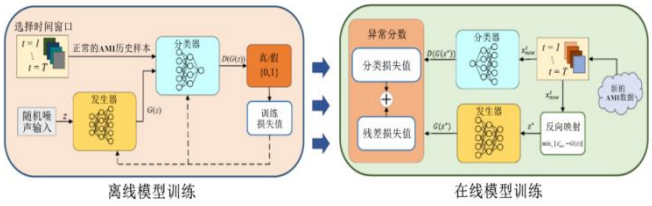


**Figure 4.** GAN-based learning and testing structure

As shown in Figure 4(b), in order to detect potential outages in each region, the data of online SM is evaluated based on the anomaly scores of GAN. The anomaly consists of the residual $\delta_R(\bullet)$, and the classification error $\delta_D(\bullet)$. When a new data $x^t_{new}$ is obtained, the residuals are expressed as

$$\delta_R(x^t_{new}) = \min_z \left| x^t_{new} - G(z) \right| \tag{5}$$

After training, the generator $G$ learns the mapping of the data distribution from the latent space $z$ to the regional measurements. Therefore, if xtnew comes from the normal case, its residual value is zero. To obtain the optimal solution $z^*$ during the test, the solver fmincon is used in this paper. if the residual value deviates from the normal residual value indicates an outage event in the region. The classification error $\delta_D(x^t_{new})$ is defined according to the trained classifier $D$.

$$\delta_D(x^t_{new}) = -\log D(x^t_{new}) - \log(1 - D(G(z^*))) \tag{6}$$

The GAN-based anomaly score in region $\Psi_i$ is defined as a weighted sum of two error metrics.

$$\zeta\psi_i(x^t_{new}) = (1-\lambda)\bullet\delta_R(x^t_{new}) + \lambda\bullet\delta_D(x^t_{new}) \tag{7}$$

Where $\lambda$ is a user-defined weighting factor whose value is 0.1 in this paper. to determine the critical threshold of the anomaly score, all training samples of region $\Psi_i$ are trained to obtain the GAN-based anomaly score $\zeta_{\psi i}$. the sample mean $\mu_{\psi i}$ and sample variance $\sigma_{\psi i}$ of the anomaly score of the training data samples are calculated to determine the range of the normal operation anomaly score. When power failure occurs, the real-time measurement data samples are expected to have anomaly scores higher than this range. Due to the use of a rolling window, power outages can be detected in a single interval. The length of the time interval depends on the resolution of the smart meter data. The details of the anomaly identification process will be elaborated in the next section.

Using the trained GAN and the GAN-based anomaly scoring method, outages within each region can be detected by comparing the anomaly scores of the new samples with the critical threshold. Since the granularity of location information depends on the number of candidate branches, it is necessary to reduce this number as much as possible. To achieve this goal, a GAN-based region coordination method is proposed. The method consists of the following steps by integrating the anomaly scores of multiple regions.

Step 1: Each region is assigned a GAN, $\Psi_i \in \Psi_g$, while learning measurement data using historical data from two observable nodes of Algorithm 1 in each region.

Step 2: After training in each region $\Psi_i$, obtain the anomaly scores of the training samples in the region and determine the sample mean and sample variance of the anomaly scores($\mu_{\psi i}$ 、 $\sigma_{\psi i}$).

Step 3: Observe the anomaly scores of all data in the set $\Psi_g$.

Step 4: Select the first region on the right side of the set $\Psi_g$ and denote it as $\Psi_a$. where $a=argmax_\xi\ \xi$, satisfied $\zeta_{\psi\xi} > \mu_{\psi\xi} + h\bullet\sigma_{\psi\xi}$. where $h$ is a user-defined threshold factor.

Step 5: Output the set of candidate branches with outage potential locations: $B_c = Z_{\psi a} \backslash \{Z_{\psi a+1} \cup Z_{\psi a+2} \cup \ldots \cup Z_{\psi w}\}$, where $A \backslash B$ is denoted as the elements of the set $A$ that are not in the set $B$.

The minimum candidate branches affected by the outage can be obtained based on the results of regional coordination, thus maximizing the outage information that will help maintenance personnel to find the fault location quickly. Given the unbalanced nature of the distribution network, the proposed method is applicable to each phase, and therefore, it is necessary to obtain the set of regions for all three phases. In the power grid, faults such as on-load tap-changer and capacitor switching occur frequently in the distribution system, and these faults have a significant impact on the actual data distribution. Therefore, the proposed outage detection method needs to be customized to consider the impact of these events (as shown in Figure 5).

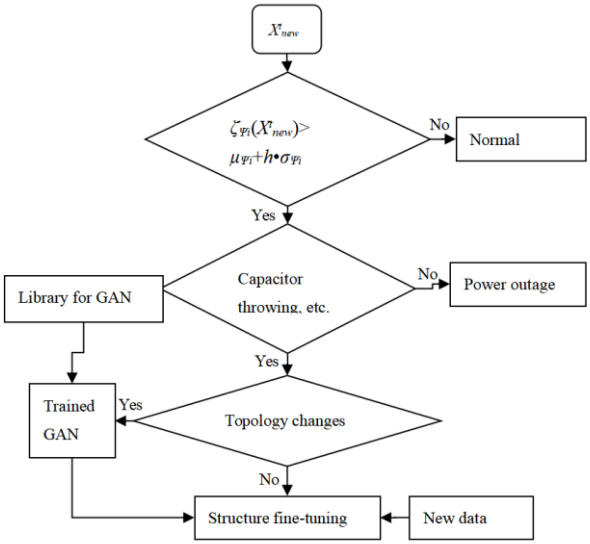


**Figure 5.** The outage identification framework of this paper

## 4. Improved outage recognition architecture

The framework proposed in this paper has the following three properties.

### 4.1. Effective topological ordering of regions

The framework introduces an effective topological ordering among regions, which can be used to simplify the outage locations in large-scale networks and can be expressed as $\Psi_i > \Psi_j$. This implies that $\Psi_i$ has a higher topological ordering than $\Psi_j$. Note that $\Psi_g=\{\Psi_1,...,\Psi_w\}$ follows a valid topological ordering, i.e., it is $\Psi_1>...>\Psi_w$. When a power outage occurs, the anomaly fraction of the subset of regions will increase above the normal range. To infer the branch where the outage event occurs, all normal regional branches below the threshold are removed, which helps to select the smallest set of branches directly from thousands of candidate branches.

### 4.2. Maximum outage location information extraction.

In locally observable systems, the algorithm is able to obtain the optimal set of regions because it maximizes the information of outage locations. First, the outage location information is evaluated in $\Psi_g$ using entropy. The set $\gamma^g(b_j)$ is defined as $\gamma^g(b_j)=\{\forall\Psi_i : b_j \in Z_{\Psi i},\ \Psi_i \in \Psi_g\}$. Thus $\gamma^g(b_j)$ is the set of all regions of $\Psi_g$ containing $b_j$. For each $\Psi_g$, the set of undetectable branching sets is defined as $U(\Psi_g) = \{u_1, ..., u_V\}$ where $u_k=\{b_{kn}, \ldots, b_{kn} : \forall b_{ki}, b_{kj}, \gamma^g(b_{ki})= \gamma^g(b_{kj})\ \}$. Thus, $u_k$ defines a set of branches that are covered by identical regions and cannot be distinguished from each other at the outage locations. Considering the set of outage information $U(\Psi_g)$ can be measured using the concept of entropy, expressed by the following equation.

$$H(U(\Psi^g)) = -\sum_{i=1}^{V} \frac{|u_i|}{M} \log \frac{|u_i|}{M} \tag{8}$$

where $|u_i|$ is the cardinality of the set $u_i$. A higher entropy value implies more distinguishable branches and thus more information about the location of the outage. The theoretical upper limit of entropy is $\log(M)$, which indicates that any individual branch can be distinguished using two regions that intersect exactly in that branch. The theoretical lower bound value of entropy is zero, which means that all branches are covered by the same set of regions.

### 4.3. Framework Robustness

Bad data from AMI generates high anomaly scores, so it is essential to mask them in the outage detection algorithm. In this paper, a bad data detection mechanism is integrated using the redundancy of existing regions in $\Psi_g$. The basic idea is that since the bad measurement data is not actually generated by the outage event. If the cross-regions do not share data from the same measurement devices, the possibility of bias in the anomaly scores assigned to several cross-regions at the same time is small. To enhance the robustness to bad data, a set of redundant regions is chosen for $\Psi_a$. This set consists of regions with a lower topological ordering than $\Psi_a$. Denoted as: $\Psi_R = \{\Psi_{r1}, \ldots, \Psi_{rn}\}$,where $\Psi_a \subset \Psi_{ri}$, $\forall\Psi_{ri} \in \Psi_R$. $\Psi_a$ is defined as bad data if $\zeta_{\Psi ri} \leq \mu_{\Psi ri}+h\bullet\sigma_{\Psi ri}$. The number of remaining regions$|\Psi_R|$ depends on the reliability of the algorithm. If each region receives abnormal data with probability $\eta$, then the probability of bad data being classified as an outage event increases with $\eta|\Psi_R|$.

## 5. Experimental results

The proposed fault detection method was tested on a real distribution feeder using hourly SM data (for a total of three years). To provide convincing results, the most complex real distribution feeder was selected from the dataset. The topology of this network is shown in Figure 6, and the feeder consists of 164 nodes and about 800 customers. It is assumed that there are 6 observable nodes in this feeder (node 8, node 22, node 31, node 83, node 109 and node 158). Where 5 regions $\{\Psi_1,...,\Psi_5\}$ are defined based on these nodes, which include the lower branches of node 8, node 22, node 31, node 83 and node 109 ($\Psi_1>\Psi_2...>\Psi_5$) respectively.

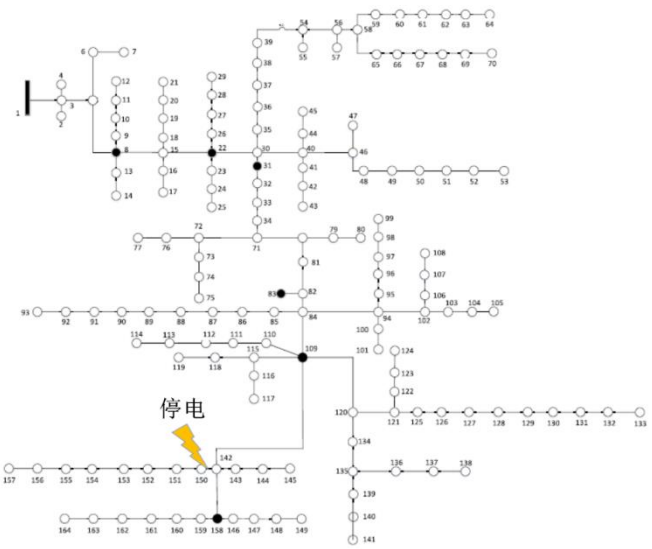


**Figure 6.** Feeder topology of 164-nodes

## 5.1. Performance of the GAN model

In order to verify the performance of GAN, the loss values of $G$ and $D$ are calculated in this paper, which can be used to verify whether the model has converged. The loss values are calculated according to the objective function of GAN. In the training process, $G$ is trained so that $\log(D(G(z)))$, while $D$ is trained to maximize the probability of correct labels. The results of the GAN model after training are shown in Figure 7. The research in this paper was conducted on a standard PC with an Intel Xeon CPU running at 3.70 GHz and 32.0 GB RAM. the average computation time for training each GAN on the available SM dataset is about 840 seconds. Multiple GANs can be trained independently and in parallel, reducing system reconfiguration and adaptive time after capacitance switching. Since the training process is offline, this parallel training approach can be easily scaled to large distribution systems.

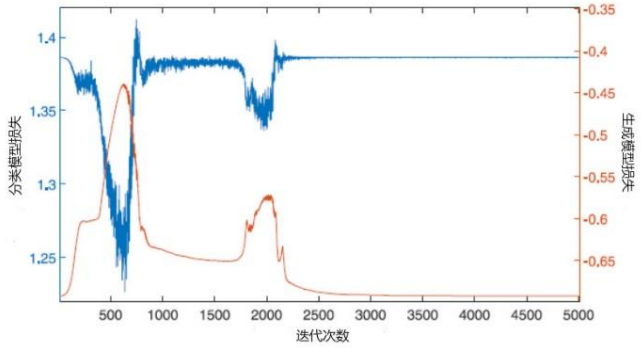


**Figure 7.** Training results of GAN model

As in Figure 6 the blackout is located between node 142 and node 164. The outages were simulated to evaluate the performance of the method under three different outage scenarios. The first case was a small event where about 20 customers were affected (40 kW per hour). The second case was a medium-sized outage where about 50 customers were affected (100 kW per hour). The third case is a large event with about 80 customers (150 kW per hour). For each case, the GAN is trained using historical SM data from five regions. The three outage cases are simulated in OpenDSS, where the voltage drop is calculated based on the simulation results. Also, to represent the standard measurement deviation, error samples were generated from a normal distribution with zero mean and 1% variance, and the error samples were added to the voltage values. Figure 8 shows the histogram of anomaly scores for a region under normal and outage conditions. The mean and variance for the normal case are 1.263 and 7.7e-5, while the mean and variance for the outage are 1.33 and 2.7e-4. As can be seen from Figure 8, it is easy to distinguish between normal and blackout, due to the large difference in anomaly scores. Also, Figure 9 gives the training set and test set anomaly scores for the normal case of the system. When the outage occurs in the region, the real-time anomaly scores reach quite high values. Figure 10 shows the distribution of the change in anomaly scores for individual regions when outages of different magnitudes occur outside the region. The plot depicts the histogram of $\Delta\zeta = \zeta_n - \zeta_{out}$, where $\zeta_n$ is the anomaly score at normal times and $\zeta_{out}$ is the anomaly score when the outage occurs outside the zone. For out-of-area outages, there is no change in the anomaly score for the area, which indicates that outages inside and outside the area can be correctly distinguished based on the anomaly score. To evaluate the outage detection performance of the method in a multi-zone network, accuracy (Accu), precision (Prec), recall, and F1 are used for scoring. The performance values for the three outage scenarios and different regions are given in Table 1. The performance of the proposed outage detection method improves with increasing event size due to the large bias in the conventional data distribution. For medium and large outage cases, all metrics are greater than 0.9. The average performance of the GAN with different size training datasets is also tested (as shown in Figure 11). As can be seen from the figure, the performance of the GAN can achieve acceptable detection accuracy with smaller training sets (about 700 data samples, i.e., about 3 days of data). To continue to demonstrate the superior performance of the method in this paper, another test was conducted with more smart meter data, resulting in a finer region. It is assumed that the feeder has 33 nodes (nodes 8,9,12,18,21,22,26,29,31,35,39,41,43,48,53,73,75,83,85,90, 93,95,99,106,108,109,110,114,125,129,134,141,158, respectively). Nineteen regions were obtained using the region selection method: $\{\Psi_1,..., \Psi_{19}\}$. The values of the statistical indicators are shown in Table 2. from this table, it can be seen that most of the statistical indicators are above 0.9, which proves the good detection performance. When no

power outage occurs in these regions, the accuracy of these regions is maintained ideally.

**Table 1.** Outage detection performance analysis

| District | Case | Accu | Recall | Prec | F1 |
|---|---|---|---|---|---|
| $\Psi_1$ | 1 | 0.752 | 0.645 | 0.82 | 0.72 |
| | 2 | 0.913 | 0.967 | 0.872 | 0.917 |
| | 3 | 0.928 | 0.997 | 0.876 | 0.932 |
| $\Psi_2$ | 1 | 0.835 | 0.784 | 0.874 | 0.826 |
| | 2 | 0.943 | 1.000 | 0.898 | 0.946 |
| | 3 | 0.943 | 1.000 | 0.898 | 0.946 |
| $\Psi_3$ | 1 | 0.673 | 0.506 | 0.768 | 0.607 |
| | 2 | 0.912 | 0.984 | 0.860 | 0.917 |
| | 3 | 0.914 | 0.988 | 0.860 | 0.919 |
| $\Psi_4$ | 1 | 0.922 | 0.884 | 0.964 | 0.922 |
| | 2 | 0.953 | 0.939 | 0.966 | 0.952 |
| | 3 | 0.981 | 0.995 | 0.968 | 0.981 |
| $\Psi_5$ | 1 | 0.834 | 0.738 | 0.913 | 0.816 |
| | 2 | 0.960 | 0.991 | 0.934 | 0.961 |
| | 3 | 0.965 | 1.000 | 0.934 | 0.966 |

**Table 2.** Outage Detection Performance Analysis (19 regional cases)

| District | Case | Accu | Recall | Prec | F1 |
|---|---|---|---|---|---|
| $\Psi_1$ | 1 | 0.752 | 0.645 | 0.82 | 0.72 |
| | 2 | 0.913 | 0.967 | 0.872 | 0.917 |
| | 3 | 0.928 | 0.997 | 0.876 | 0.932 |
| $\Psi_2$ | 1 | 0.949 | 0.955 | 0.944 | 0.949 |
| | 2 | 0.950 | 0.956 | 0.944 | 0.951 |
| | 3 | 0.951 | 0.958 | 0.944 | 0.951 |
| $\Psi_3$ | 1 | 0.922 | 0.929 | 0.916 | 0.923 |
| | 2 | 0.922 | 0.930 | 0.916 | 0.923 |
| | 3 | 0.917 | 0.920 | 0.915 | 0.917 |
| $\Psi_4$ | 1 | 0.835 | 0.784 | 0.874 | 0.826 |
| | 2 | 0.943 | 1.000 | 0.898 | 0.946 |
| | 3 | 0.943 | 1.000 | 0.898 | 0.946 |
| $\Psi_5$ | 1 | 0.933 | 0.932 | 0.934 | 0.933 |
| | 2 | 0.931 | 0.928 | 0.934 | 0.931 |
| | 3 | 0.936 | 0.938 | 0.935 | 0.936 |
| $\Psi_6$ | 1 | 0.973 | 0.972 | 0.973 | 0.973 |
| | 2 | 0.975 | 0.977 | 0.974 | 0.975 |
| | 3 | 0.976 | 0.978 | 0.947 | 0.976 |
| $\Psi_7$ | 1 | 0.945 | 0.940 | 0.950 | 0.945 |
| | 2 | 0.975 | 0.940 | 0.950 | 0.945 |
| | 3 | 0.946 | 0.942 | 0.950 | 0.946 |
| $\Psi_8$ | 1 | 0.902 | 0.908 | 0.898 | 0.903 |
| | 2 | 0.905 | 0.914 | 0.898 | 0.906 |
| | 3 | 0.906 | 0.916 | 0.900 | 0.907 |
| $\Psi_9$ | 1 | 0.673 | 0.506 | 0.768 | 0.607 |
| | 2 | 0.912 | 0.984 | 0.860 | 0.917 |
| | 3 | 0.914 | 0.988 | 0.860 | 0.919 |
| $\Psi_{10}$ | 1 | 0.929 | 0.929 | 0.930 | 0.929 |
| | 2 | 0.930 | 0.931 | 0.930 | 0.930 |
| | 3 | 0.929 | 0.930 | 0.930 | 0.929 |
| $\Psi_{11}$ | 1 | 0.922 | 0.884 | 0.964 | 0.922 |
| | 2 | 0.953 | 0.939 | 0.966 | 0.952 |
| | 3 | 0.981 | 0.995 | 0.968 | 0.981 |
| $\Psi_{12}$ | 1 | 0.940 | 0.940 | 0.940 | 0.940 |
| | 2 | 0.940 | 0.940 | 0.940 | 0.940 |
| | 3 | 0.940 | 0.941 | 0.940 | 0.940 |
| $\Psi_{13}$ | 1 | 0.960 | 0.960 | 0.960 | 0.960 |
| | 2 | 0.961 | 0.962 | 0.960 | 0.961 |
| | 3 | 0.958 | 0.956 | 0.959 | 0.957 |
| $\Psi_{14}$ | 1 | 0.962 | 0.962 | 0.963 | 0.962 |
| | 2 | 0.962 | 0.961 | 0.962 | 0.962 |
| | 3 | 0.963 | 0.964 | 0.963 | 0.963 |
| $\Psi_{15}$ | 1 | 0.945 | 0.946 | 0.944 | 0.945 |
| | 2 | 0.945 | 0.947 | 0.944 | 0.945 |
| | 3 | 0.946 | 0.948 | 0.944 | 0.946 |
| $\Psi_{16}$ | 1 | 0.834 | 0.738 | 0.913 | 0.816 |
| | 2 | 0.960 | 0.991 | 0.934 | 0.961 |
| | 3 | 0.965 | 1.000 | 0.934 | 0.966 |
| $\Psi_{17}$ | 1 | 0.929 | 0.930 | 0.928 | 0.929 |
| | 2 | 0.928 | 0.928 | 0.927 | 0.928 |
| | 3 | 0.934 | 0.940 | 0.928 | 0.934 |
| $\Psi_{18}$ | 1 | 0.976 | 0.972 | 0.979 | 0.975 |
| | 2 | 0.977 | 0.974 | 0.979 | 0.976 |
| | 3 | 0.978 | 0.977 | 0.980 | 0.978 |
| $\Psi_{19}$ | 1 | 0.911 | 0.908 | 0.914 | 0.911 |
| | 2 | 0.916 | 0.918 | 0.915 | 0.916 |
| | 3 | 0.919 | 0.924 | 0.915 | 0.920 |
| 均值 | 1 | 0.905 | 0.881 | 0.922 | 0.899 |
| | 2 | 0.940 | 0.952 | 0.932 | 0.941 |
| | 3 | 0.944 | 0.957 | 0.931 | 0.945 |

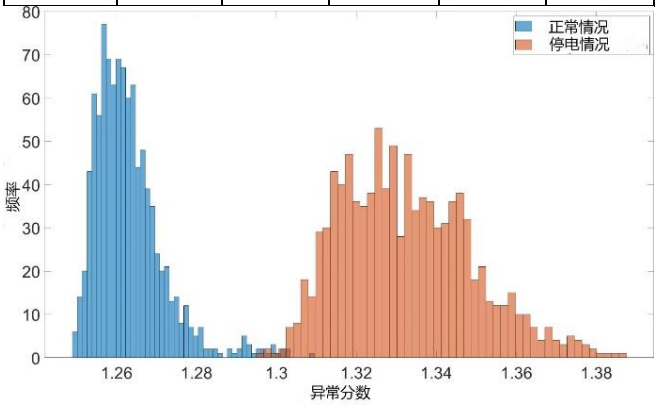


**Figure 8.** Histogram of abnormal scores under normal and outage conditions

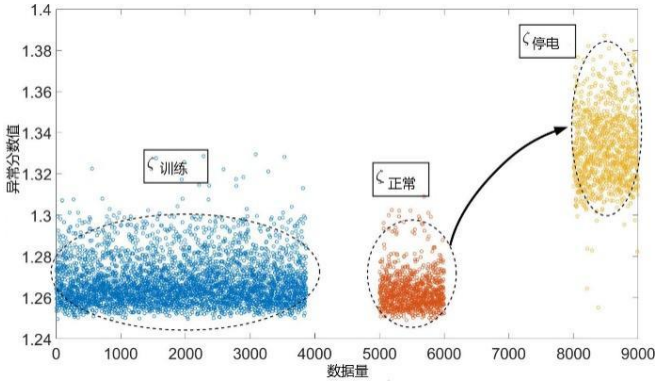


**Figure 9.** Anomaly scores of the training set

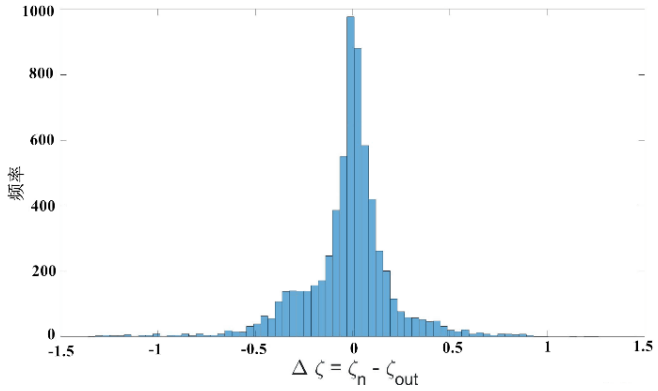


**Figure 10.** Histogram of △ζ

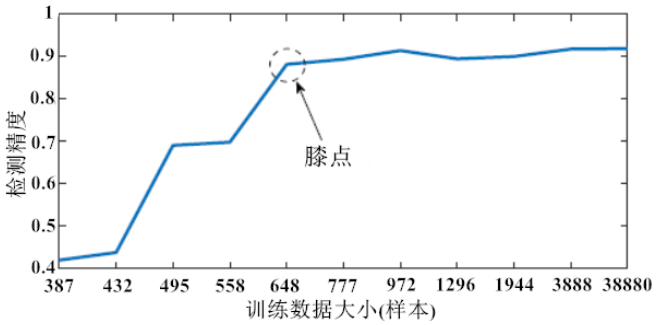


**Figure 11.** Sensitivity of fault detection accuracy to training set size

To further verify the performance of the fine-tuning algorithm, an experimental comparison was performed (shown in Figure 12). Let a certain capacitor switch occur at 12:00 noon. Due to the change in the underlying data distribution, the performance of the method drops from about 97% to about 76%. This is caused by the fact that the newly obtained training data set is very small and the critical threshold for recalculating the anomaly score. Then, as the size of the training data increases, the average accuracy of the method improves significantly. After about one day, the method in this paper reaches a similar level of accuracy as before the capacitance switch, which means that the proposed method has adapted to the change in system conditions. Compared with the results in Figure 11, the data acquisition time can be reduced from 3 days to 1 day using the fine-tuning strategy in this paper.

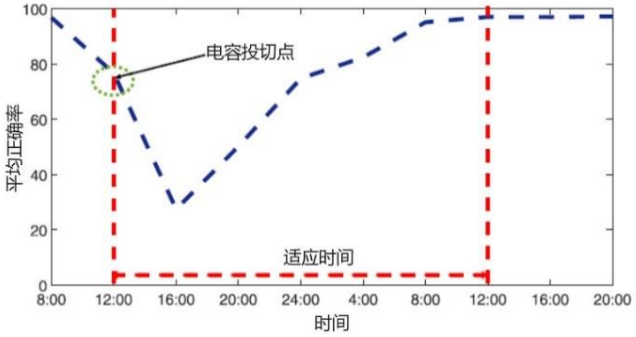


**Figure 12.** Performance of capacitor tuning strategy

To demonstrate the superiority of the proposed strategy, it is compared with a support vector machine. The results show that the proposed method can achieve good and better outage detection accuracy with a smaller number of smart meters. As shown in Figure 13, for three different outage cases, the support vector machine requires a higher level of observability (almost 10 times higher than the method in this paper) to achieve a similar detection accuracy. It is shown that the proposed scheme is able to detect outages accurately.

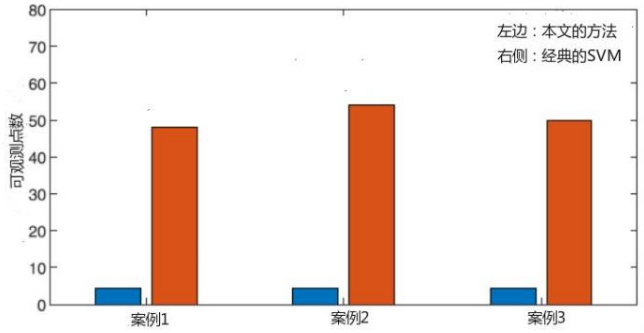


**Figure 13.** Comparison between the proposed method and SVM

## 5. conclusion

In this paper, a new data-driven approach is proposed to detect and locate outages in partially observable power grids using SM data. The GAN-based detection method proposed in this paper is able to implicitly represent the data distribution under normal conditions and identify potential online outages. The proposed multi-area outage detection mechanism is based on an unsupervised learning method, which can solve several difficulties in detection: 1) Due to the limited SM data, the system has poor observability. 2) Data imbalance caused by the scarcity of power outage data. 3) High dimensionality of data caused by spatio-temporal relationship. At the same time, the proposed region selection and ranking mechanism based on BFS ensures that the maximum amount of outage location information can be obtained in any observable system. The proposed method is

validated on real public feeders using SM data, and its performance is more effective than the existing support vector machine algorithms.

## References


[1] J. Zhao, S. Zhang, and S. Liao, "Analysis and thinking of California power outage accident in mid-August 2020," *Electr. Power Eng. Technol.,* vol. 39, no. 06, pp. 52-57, Nov. 2020, doi: 10.12158/j.2096-3203.2020.06.008.

[2] J. He, J. Tu, W. Sun, L. Ji, Y. Zhang, X. Zhang, and Q. Zhou, "Preliminary Analysis and Enlightenment of "8·14" and "8·15" power outages in California, USA," *Power grid Technol.*, vol. 44, no. 12, pp. 4471-4478, Oct. 2020, doi: 10.13335/j.1000-3673.pst.2020.1586.

[3] Q. Yu, and Q. Liu, "Application of Generalized Pareto Distribution in Power Outage Accident Analysis of China Southern Power Grid," *Pract. Understanding Math.*, vol. 50, no. 19, pp. 175-185, Oct. 2020.

[4] M. Xu, R. Niu, R. Xie, W. Cao, and Q. Wu, "Online Monitoring and Early warning of Multiple outages based on Multi-source Information Fusion," *Shandong Sci.*, vol. 33, no. 04, pp. 117-123, Aug. 2020, doi: 10.3976/j.issn.1002-4026.2020.04.015.

[5] H. Li, S. Chen, K. Chen, and Y. Li, "Research on real-time outage monitoring and protection technology based on intelligent equipment," *Comput. Digital Eng.*, vol. 48, no. 02, pp. 498-502, Feb. 2020, doi: 10.3969/j.issn.1672-9722.2020.02.045.

[6] Y. Sheng, L. Zhong, L. Zhang, Z. Zhou, and S. Gong, "Reliability monitoring and research of distribution network based on detailed outage data," *Zhejiang Electr. Power,* vol. 36, no. 12, pp. 70-74, Dec. 2017, doi: 10.19585/j.zjdl.201712014.

[7] B. Li, J. Wang, Y. Zhang, and L. Huang, "Research on outage fault monitoring method of substation monitoring background machine," *Power big data*, vol. 20, no. 11, pp. 15-17+53, Nov. 2017, doi: 10.19317/j.cnki.1008-083x.2017.11.006.

[8] X. Zhang, and T. Sang, "Research and Application of 10kV distribution Network Line Monitoring Method Based on Big Data Mining Technology," *Mech. Electr. Inf.*, no. 35, pp. 12-13, Dec. 2019, doi: 10.19514/j.cnki.cn32-1628/tm.2019.35.005.

[9] Z. S. Hosseini, M. Mahoor, and A. Khodaei, "AMI-enabled distribution network line outage identification via multi-label SVM," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5470–5472, Sep. 2018, doi: 10.1109/tsg.2018.2849845.

[10] R. Moghaddass, and J. Wang, "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5820–5830, Nov. 2018, doi: 10.1109/tsg.2017.2697440.

[11] Xiaozhou Yang, Nan Chen, Chao Zhai. A Control Chart Approach to Power System Line Outage Detection Under Transient Dynamics, IEEE Transactions on Power Systems, vol. 36, no. 1, pp. 127-135, Jan. 2021.

[12] Anmol Dwivedi; Ali Tajer. Scalable Quickest Line Outage Detection and Localization Via Graph Spectral Analysis," IEEE Transactions on Power Systems, vol. 37, no. 1, pp. 590-602, Jan. 2022.

[13] Wenqu Li, Zhiwei Liu, Wei Yao, et al. Multiple Line Outage Detection for Power Systems Based on Binary Matching Pursuit," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 70, no. 8, pp. 2999-3003.

[14] Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, NY, USA

[15] S. J. Chen, T. S. Zhan, C. H. Huang, J. L. Chen, and C. H. Lin, "Nontechnical loss and outage detection using frac-tional-order self-synchronization error-based fuzzy petri nets in micro-distribution systems," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 411–420, Jan. 2015, doi: 10.1109/tsg.2014.2345780.

[16] R. A. Sevlian, Y. Zhao, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection using load and line flow meas-urements in power distribution systems," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 2053–2069, Mar. 2018, doi: 10.1109/tpwrs.2017.2727979.

[17] H. Sun, Z. Wang, J. Wang, Z. Huang, N. Carrington, and J. Liao, "Data-driven power outage detection by social sensors," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2516–2524, Sep. 2016, doi: 10.1109/tsg.2016.2546181.

[18] W. Xu, F. Guo, L. Jia, R. Zhang, J. Zhang, "Application of big data statistical analysis in power outage monitoring of distribution network," *China Power Enterprise Management Innovation Practice (2019)*, vol. 202, no. 2, pp. 748-749, Oct. 2020, doi: 10.26914/c.cnkihy.2020.020685.

[19] X. Hu, X. Lu, Y. Zhao, Y. Zhao, and G. Liu, "A visual monitoring device based on outage maintenance of large power grid," *Electr. Technol.*, vol. 20, no. 07, pp. 70-73, Jul. 2019, doi: 10.3969/j.issn.1673-3800.2019.07.020.