Retraction

Retracted: Design of an Underground Transmission Line Condition Fault Monitoring System for Power Grids Based on Data Analysis Algorithms

Journal of Sensors

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] L. Wang, W. Jiang, and W. Dong, "Design of an Underground Transmission Line Condition Fault Monitoring System for Power Grids Based on Data Analysis Algorithms," *Journal of Sensors*, vol. 2022, Article ID 1161022, 11 pages, 2022.

Research Article

Design of an Underground Transmission Line Condition Fault Monitoring System for Power Grids Based on Data Analysis Algorithms

Liunai Wang[1](ORCID), Wen Jiang,[2] and Weiwei Dong[2]

[1]*Information Communication Branch of Hainan Power Grid Co. Ltd., Haikou, Hainan 570100, China*
[2]*Hainan Power Grid Co. Ltd., Haikou, Hainan 570100, China*

Correspondence should be addressed to Liunai Wang; wangln@hn.csg.cn

With the continuous expansion of the scale, capacity, and coverage of the modern transmission network, the role of the power system in the national economy is increasingly prominent, and the disconnection of the power system will have a huge impact on society and people's lives. Due to the long transport distance and wide coverage of transmission lines, natural conditions and human factors have caused great difficulties in line operation and maintenance. How to effectively improve the operation and maintenance of transmission lines to ensure the stability and safety of the power grid has become a common problem for the power industry and scientific researchers to discuss. In recent years, the information society has stepped into the era of big data, and big data has developed rapidly, becoming a hot area favored by academia and industry, and is widely used. Through big data analysis, potential operation rules can be discovered from a large amount of grid information, providing maintenance personnel with corresponding maintenance decision support. Using big data technology for transmission line fault analysis can effectively reduce accident processing time and avoid accident expansion. Therefore, this paper combines the underground transmission line fault of power grid with the fuzzy KNN algorithm model to apply the underground transmission line state intelligent monitoring system and conducts the study of real-time data collection and fault diagnosis analysis of the underground transmission line fault of power grid, and this paper conducts the transmission line fault analysis experiment, which fully confirms the feasibility and effectiveness of the algorithm model proposed in the paper and concludes that the data analysis model proposed in this paper. The proposed data analysis model has good innovation and practical feasibility.

1. Introduction

With the continuous development of China's economy, the material living standard and cultural level of the people are increasing, and the requirements for the safety and stability of the electric power system are getting higher and higher. The electric power system is a complex and huge artificial system with a wide geographical distribution of lines and equipment, large energy transmission, fast electric energy transmission, reliable communication and scheduling, high requirements for real-time operation, and an instant increase in major faults [1]. These characteristics make it generate a large amount of data, fast growth, and rich data when it is in operation, which fully meet the characteristics

of big data. Therefore, with the rapid growth of data in the power system, the traditional power data processing technology can no longer meet the analysis and processing of a large amount of information in the power sector [2], and the application of power data analysis technology will certainly become an inevitable requirement for the informatization and intelligent development of the power industry.

The transmission line is an important part of the power grid, which closely combines the facilities of power generation, power transformation, power supply, and distribution with the users in all aspects [3]. The working condition of this equipment has a great influence on the stable operation of the power grid and has a great impact on the reliability of power supply, power supply quality, and usage experience.

As the scale, capacity and coverage of the power system continue to expand, the power system has become an important part of the national economy and people's life, and power outages in the grid will have a huge impact on the society and people's economy [4–7]. Therefore, to ensure the reliable operation of the transmission line operation plays the same role in improving the stability and safety of the power grid, which is a key link in the power grid that cannot be ignored. However, due to the long transport distance and wide coverage of transmission lines, natural conditions and human factors cause great difficulties for line operation and maintenance [8]. How to effectively improve the operation and maintenance of transmission lines to ensure the stability and safety of the power grid has become a common problem for the power industry and scientific researchers to discuss.

In recent years, the information society has stepped into the era of big data, and big data has developed rapidly and become a hot area favored by academia and industry and is widely used. Through data analysis, potential operation rules can be found from a large number of power transmission systems, thus providing maintenance personnel of power systems with corresponding maintenance decision support [9]. Using big data technology for transmission line fault analysis can effectively reduce accident processing time and avoid accident expansion. This paper mainly introduces data mining technology, machine learning technology, data mining technology, and statistical technology.

(1) In machine learning and data mining algorithms, the learning modes are classified as supervised learning (decision tree, simple Bayesian, least squares regression, logistic regression, etc.), unsupervised learning (singular value decomposition, principal component analysis, independent component analysis, etc.), unsupervised learning (graph theory inference algorithm), and reinforcement learning (time difference learning)

(2) Statistical analysis, t-test, cluster analysis, principal component analysis, correlation analysis, regression analysis, ANOVA, and other statistical analyses of the distribution of the data using statistical indicators, graphs, etc.

This thesis focuses on classifying and categorizing the collected electrical quantities of underground transmission lines by the method of data analysis and applying them to the analysis of asymmetric faults in underground transmission lines in order to analyze and deal with grid faults. Therefore, the data analysis method is applied to the intelligent diagnosis of faults in underground transmission lines of power systems.

2. Introduction to Related Theories

2.1. Data Analysis Method.
In the computing process of big data, data analysis is an important topic that requires a lot of analysis of a large amount of data to obtain more intelligent, in-depth, and valuable data. The rapid development of big data determines the priority of data analysis algo-

rithms in the field of big data and whether the final data is valuable or not. Data analytics includes visual analytics, data mining algorithms, predictive analytics, semantic engines, data quality, and data management [10].

(1) *Visual Analytics*. The essence of big data operators, which includes both analysts and general users, is the visual analysis of data. This is because the characteristics of big data can be displayed through visualization methods, thus making it more acceptable to users. Big data visualization and analysis systems are able to use three-dimensional representation techniques for more complex data, thus achieving a three-dimensional display of a large amount of data. Data visualization techniques include geometric, pixel, icon, hierarchical image, and distribution based techniques

(2) *The Algorithm of Data Fetching*. Data mining is the core of data analysis, which shows the characteristics of data hidden in the data according to the data type and format. Various data mining algorithms such as cluster analysis, segmentation processing, and isolated point analysis can discover the potential value in big data faster

(3) *Conduct Forward-Looking Research*. First, dig the main characteristics of big data and establish a scientific data analysis model; then introduce new data for forecasting and prediction. Data analysis is an important application area of data analysis, and through visualization of data and data mining, it enables analysts to make predictions about data

(4) *Semantic Analysis*. Due to the diversity of unstructured data, it is very difficult to analyze the data, and a series of data analysis tools are needed to extract and parse the information from the "documents"

(5) *Data Quality and Data Management*. Due to the large volume and variety of data, high quality data and its effective manipulation are very important, as well as best practices in data management. The process and tools for standardized processing of big data also fully guarantee the reliability and value of data analysis

2.2. Application of Data Analysis Algorithms.
With the advent of the big data era, the means of data analysis are increasingly used in various industries. At present, the most familiar one is the data analysis in the business field. Companies collect users' search terms, tag keywords, other semantic inputs, and other information to build a predictive model to make judgments about users' needs and thus provide better services to consumers [11–13]. For example, Target, a well-known U.S. retailer, uses data analytics to obtain valuable information from customer profiles so that it can anticipate when customers want their children to be born. In addition, the use of data analysis methods can predict the churn of customers. Walmart uses big data computing to make more accurate predictions and analysis of future products, while in
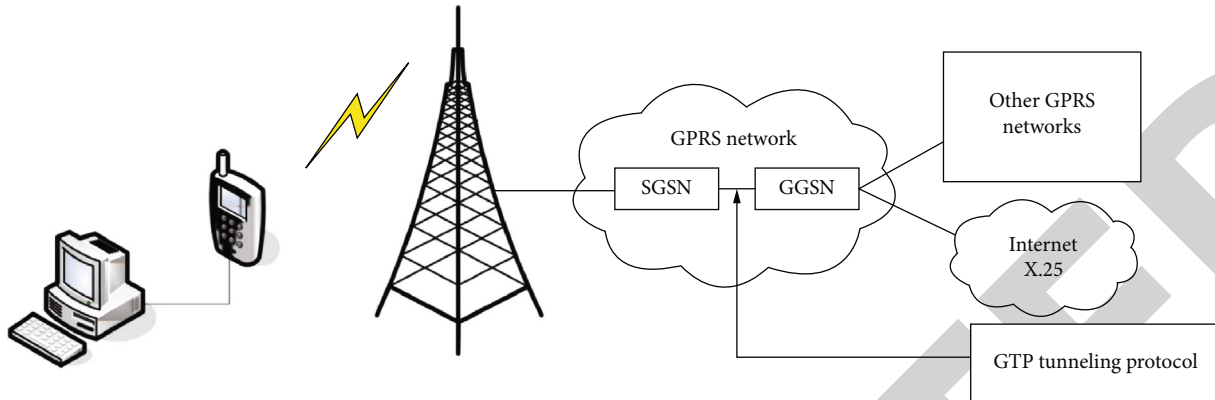
Figure 1: GPRS network structure.

the auto insurance industry, customer needs and driving skills are extracted and analyzed.

In the financial field, the application of big data computing is mainly based on financial decision making, such as high frequency trading (HFT). Currently, most stock trading uses data analytics to thoroughly analyze news on social media and the internet to make trades in the next few minutes. For example, Boshiwa's "Fingerwise" is a "Gold 100" financial management software based on Anthem. This method uses a large amount of online e-commerce transaction data to predict the future profitability and prosperity of a certain industry based on user behavior, industry growth, prices, and other characteristics and uses this as a basis to select 100 stocks [14]. This allows the index to better reflect the changes that occur with the dynamics of the market, thus allowing it to better track the market.

The application of big data algorithm technology to the medical field will help to improve the overall level of medicine and enhance the ability to develop medicine. This data processing technology has the computing power to decode DNA in a matter of minutes [15]. IBM Watson is currently working on a perceptual operating system to determine whether Watson can "understand" the natural language of doctors. At the same time, the rapid and accurate analysis of large amounts of medical research data provides answers.

The application of data analysis algorithms in the field of transportation is to collect, store, analyze, classify, and query various traffic chokepoint data in real time, providing more accurate and faster analysis to understand current traffic conditions and provide insight into potential factors affecting complex traffic [16]. In the remote monitoring and dispatching of freight vehicles, the use of data analysis can provide timely and efficient services for the location, condition, and delivery of freight vehicles, thus improving the efficiency of cargo transportation. At the same time, it can also help government regulatory authorities to detect abnormal dangerous goods transportation vehicles in time and provide rapid rescue and safety supervision.

Due to the rapid application of data analysis methods in business, finance, medical, and transportation fields, it makes the diagnosis and analysis methods of smart grid have become an important type of technology in big data application. However, in the electric power industry, data analysis

and data mining technology are currently the most commonly used method in the electric power system, which mainly focuses on the diagnosis, analysis, and evaluation of faults in the power grid [17].

2.3. Underground Transmission Lines. An underground transmission line is a type of power line that is laid on the ground to carry electricity. In urban residential areas or areas where overhead transmission lines need to be erected across rivers, straits, etc., underground transmission technology is used after comprehensive consideration of various aspects such as technology, economy, and environmental protection.

The actual underground transmission line is a power line consisting of transmission wires supported by sulfur hexafluoride gas insulation and epoxy resin isolation poles. From the 1970s, the transmission technology of low-temperature, low-resistance cables came into being. With the continuous development of superconductivity technology, people also began to research on superconductivity transmission [18]. Underground transmission lines are mainly laid in the form of direct burial, tunnel, and concealed trench. In the route selection, road facilities, road conditions, existing burial conditions, geological conditions, water level conditions, and other factors should be taken into account to minimize the total length of the line.

The transmission capacity of the cable depends mainly on the allowable current. The permissible current is in turn determined by the permissible temperature of the cable insulation. The allowable temperature is the combined effect of resistance loss, dielectric loss, and base temperature when guiding the line through the wire, so that the wire temperature cannot be higher than the heat-resistant limit of the cable insulation. In the engineering design, it should be designed according to three working conditions, namely, long time energized, short time energized [19], and short circuit. In order to increase the allowable current, oil, water, and air are usually used for forced cooling of the power cable.

Compared with overhead lines, the biggest advantage of the cable is that it does not occupy the line corridor; at the same time, it operates more safely because the cable is buried underground and is not disturbed by atmospheric and other natural factors. However, the cost of this method is higher, it will be affected by the ground electromagnetic field in the
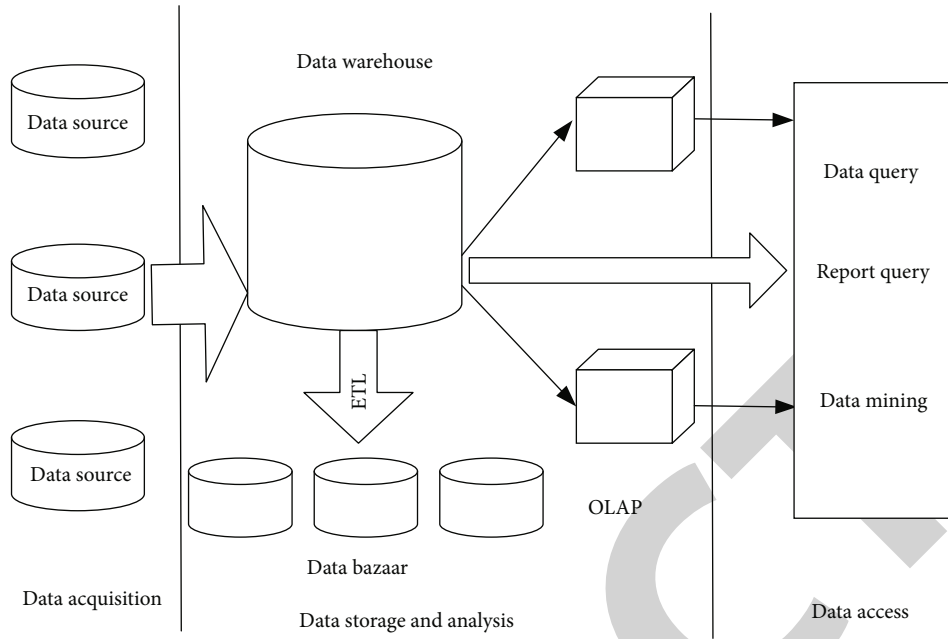
FIGURE 2: Traditional substation equipment condition monitoring platform architecture.
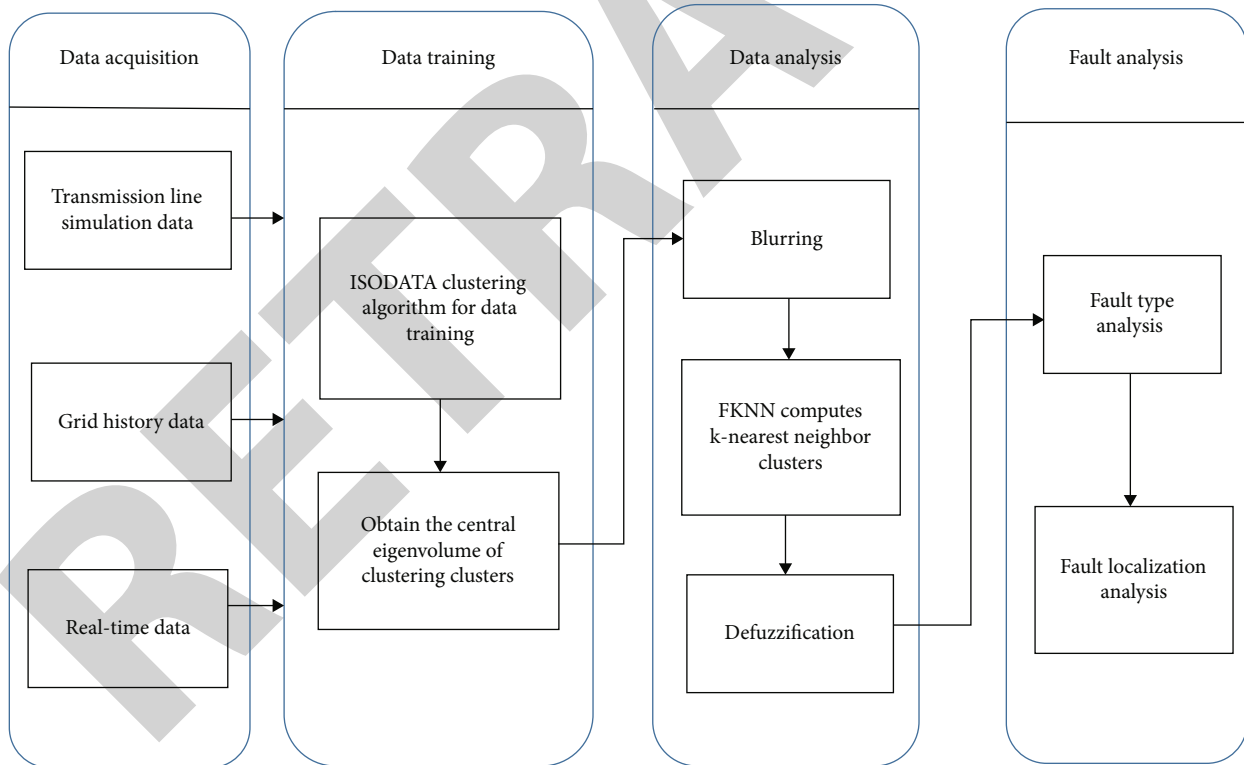

FIGURE 3: FKNN-based fault analysis model for transmission lines.

process of use, and it is easy to produce chemical corrosion and difficult to determine the fault site, so corresponding technical measures must be taken [20].

In the power system, due to the large capacitive current in the power system, the transmission capacity and transmission distance of the power system have certain limitations, and reactive power compensation is required, thus increasing the investment in the power system. The use of cables for DC transmission has higher economic benefits.

3. Application Method Design

3.1. Real-Time Information Collection. The underground transmission line status intelligent monitoring system of
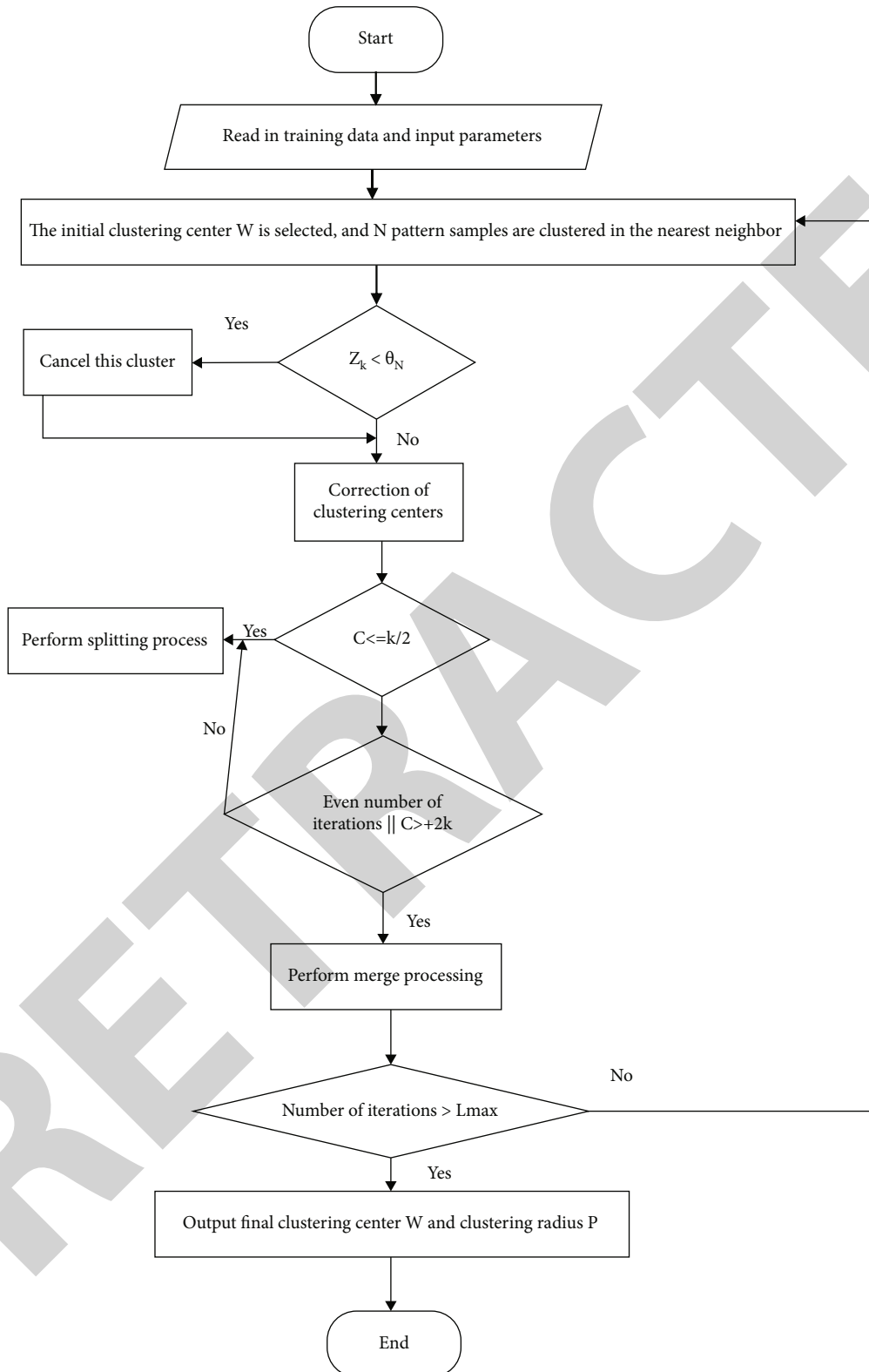
Figure 4: Flowchart of ISODATA clustering algorithm.

the power grid uses data collection terminals composed of different types of terminals. Composed of each type of terminal, the data is collected by the sensors of each terminal at regular intervals and transmitted to the monitoring host through the communication module installed on the collection terminal [21], which analyzes and processes the data and finally stores it in the database. In order to ensure the safe operation of this system, a highly reliable communication network must be

established in order to obtain high-precision and real-time monitoring information. The communication network must have the following characteristics: real-time, reliability, continuity, low information volume, and full use of transmission lines or reliance on public information networks.

Since underground power systems are generally built underground, the underground environmental conditions are unstable, and the monitoring terminal equipment is generally far away from the monitoring center of the power system. In terms of communication technology, GPRS is a better way of communication, which is used as a communication method. The monitoring terminal in the system transmits the collected data and image information through data frames or images to the main monitoring software according to the user's requirements, which is analyzed by the preprocessor and finally stored in the database. The GPRS network structure is shown in Figure 1.

3.2. Underground Transmission Line State Intelligent Monitoring Platform Architecture. The underground transmission line state intelligent monitoring platform is mainly composed of data acquisition layer, data storage and management layer, and data access layer. Its logical structure is shown in Figure 2.

The data acquisition of the underground transmission line condition intelligent monitoring system of the power network is an important part of the intelligent monitoring of the operation condition of the power system. Through ETL (extraction, transformation, cleaning, and loading), the historical data from the traditional correlation database and the condition monitoring data of power equipment are loaded into the data warehouse.

ETL will be distributed in different business systems such as production management, project management, and condition monitoring to extract, clean, transform, and load data, thus forming high-quality, high-value data.

(1) Data extraction (data)

Historical data and continuously updated monitoring data are extracted from multiple condition monitoring systems and sorted by topic. This data is the most important part of decision analysis for smart substations, and each network provincial company constantly updates monitoring data to minimize the impact on the source system, depending on the frequency of data collection.

(2) Data conversion, mainly to resolve inconsistencies in the status monitoring information of the devices, must be integrated

The data is aggregated, and the type or form of the data is converted to uniformity in the type and format of the data.

(3) Data cleaning

The so-called "data cleaning" is to eliminate some unclean data and then eliminate some unsuitable data, so as to reduce the waste of memory and reduce the scanning overhead.

```
1: Input: An unlabeled test set(X,Y)
2: For i = 1, ⋯, n do
3:     getDistance(test,clusterCenter);
4: End for
5: For j = 1, ⋯, k do
6:     getTopMinDistance;
7:     saveClusterCenter;
8:     getMaxMembership(test,clusterCenter);
9:     testLabel = ClusterCenterLabel;
10: End for
11: Output: Accuracy
```

ALGORITHM 1: Fuzzy KNN algorithm pseudocode.

(4) Data loading

The main goal of data loading is to load the data collection of deleted data into the needed data warehouse according to the table type of custom data model with powerful capabilities such as data recovery, error reporting, and data backup.

3.3. FKNN-Based Fault Analysis Model for Transmission Lines. Using the basic principles of fuzzy mathematics, a mathematical model for transmission line fault analysis is established. In the training phase, the ISODATA clustering algorithm is used to train the sample data with the existing failure type markers. In the data testing, the fuzzy theory is used to fuzzify the training data and test data and compare them with the sample data [22]. The KNN algorithm is used to perform clustering operation on k adjacent test data and calculate their belonging degree, so as to determine the fault type of the test data and analyze its fault location. The transmission line fault analysis model used in the paper is shown in Figure 3.

As shown in Figure 3 above, the FKNN-based transmission fault analysis model in this paper is divided into four major components: data acquisition, data training, data analysis, and analysis. Among them, the data acquisition module of the power system mainly includes the historical and real-time data of the power system and the simulation of the three-phase voltage and three-phase current simulated by the intelligent monitoring system. The data vectors obtained from the three-phase voltage and three-phase current signals are mapped into clusters containing fault types. Since the number of fault cases in actual operation is very small, some of the required fault and nonfault data are obtained by the simulation software and combined with the large amount of data in the power system database for online fault analysis using a fuzzy KNN classifier.

3.4. ISODATA Algorithm-Based Data Training. In the data training phase, the ISODATA clustering algorithm is used to perform supervised learning on the training data to obtain typical cluster centers. ISODATA is a common method for cluster analysis, also known as dynamic clustering or iterative self-organizing data analysis. The algorithm is an unsupervised classifier that adds markers to the training data to control the fusion and splitting of clusters to obtain more
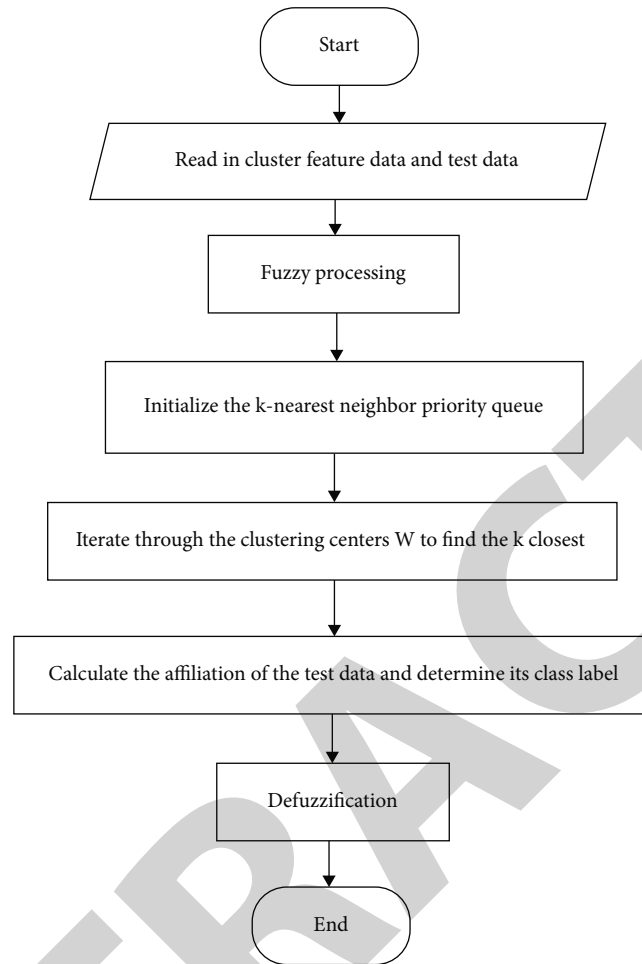
FIGURE 5: Flowchart of fuzzy KNN data analysis.

accurate core features for clustering. The flowchart of the ISODATA clustering algorithm during the data training phase is shown in Figure 4.

As shown in Figure 4, the ISODATA algorithm first selects the initial clustering center W_j, $j = \{1, 2 \cdots c\}$ divides the N pattern samples X_i, $i = \{1, 2, \cdots N\}$ into each initial set of clusters according to the nearest neighbor principle, $D_i = \min \{\|x_i - w_j\|, i = 1, 2, \cdots, N, j = 1, 2, \cdots, C\}$. According to the minimum sample number threshold θ_N in the clustering domain, remove the clusters Z_j with sample number less than θ_N, and correct each clustering center $W_j = 1/m \sum_{x_i e_{z_j}} x_i$, $\{i = 1, 2, \cdots, m, j = 1, 2, \cdots, C\}$. The split operation determines whether to split the clusters based on a predetermined value k of the number of cluster centers and the classification marker c. The clusters that require the split operation are split, and the cluster centers W are updated. The merge operation merges two clusters belonging to the same class c based on the minimum distance threshold qc between all cluster centers and the classification token c and updates the cluster centers W. At each iteration, the number of iterations $L_{\max}$ for each parameter is determined, and if at the previous iteration, the algorithm is complete and the final cluster center W and the radius p of each cluster are output;

otherwise, return to step 1 and continue with the next iteration until the result converge.

3.5. Data Analysis Based on Fuzzy KNN Algorithm. Based on the analysis of the data characteristics of the transmission lines, the data were fuzzy processed using fuzzy mathematics, and the affiliation degree was applied to represent the fault level of the lines. Finally, the KNN algorithm was used to cluster the experimental data in K neighborhoods. Finally, the classification affiliation degree of the test data Xi was derived based on the affiliation relationship of the test data Xi. On this basis, the detected fault types were classified, and the fault sites of the lines were analyzed according to the fault distance parameters in the training data.

The pseudocode of the fuzzy KNN algorithm is shown in Algorithm 1.

3.6. Principle of Fuzzy KNN Algorithm. After developed on the mathematical basis of fuzzy set theory, fuzzy theory has been widely used, which has greatly improved the fault diagnosis and diagnosis efficiency of the actual system. In response to the problems encountered in transmission line failure analysis by the KNN method described above, the fuzzy set transformation theory is used to predict and
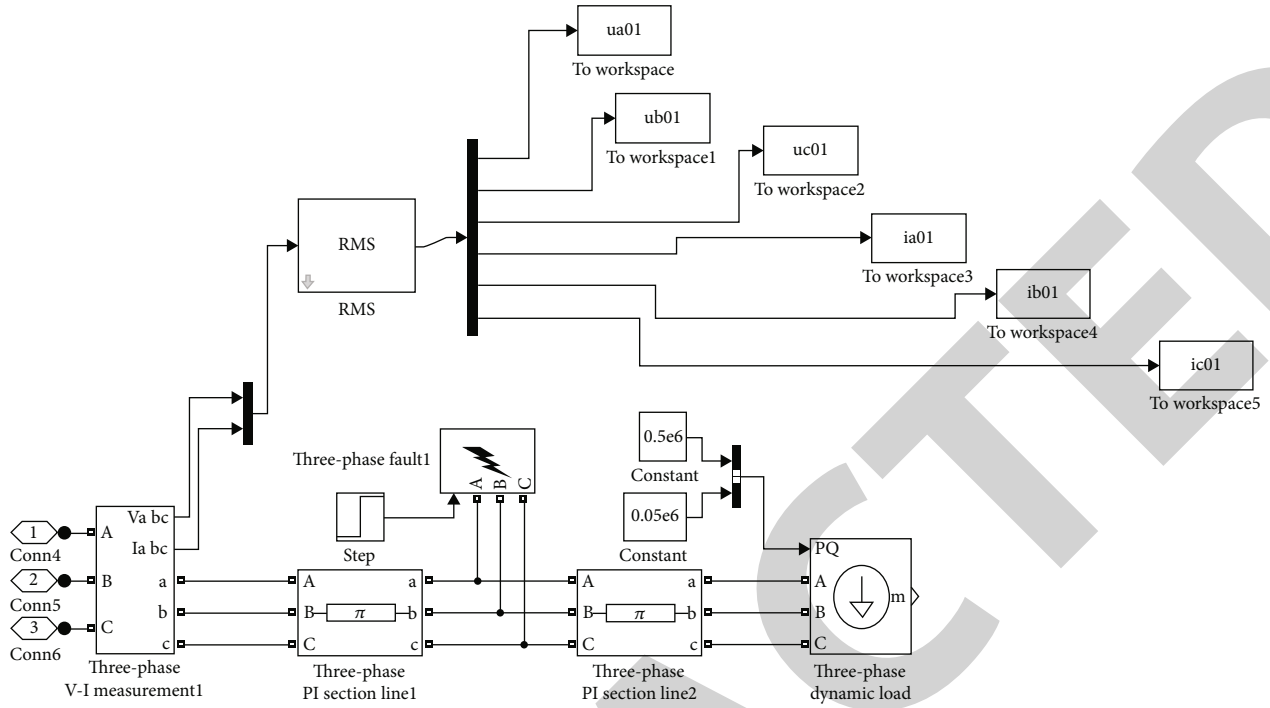
FIGURE 6: A transmission line fault simulation diagram.

TABLE 1: 10 types of transmission line asymmetric fault types.

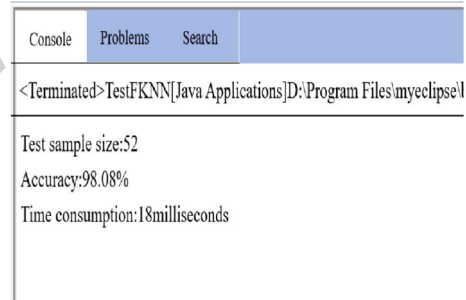| Single-phase short-circuit ground fault	Two-phase short-circuit fault	Two-phase short-circuit ground fault	Three-phase short-circuit fault
AG	AB	ABG	ABC
BG	AC	ACG	
CG	BC	BCG	

FIGURE 8: Fuzzy KNN algorithm fault category analysis results.

analyze the fault types of transmission lines by using a combined calculation method of the affiliation function, thus effectively overcoming the problem of unclear classification of classes [23].

The basic principle of the fuzzy KNN algorithm used in this chapter is that the k samples closest to the classified point are selected based on the Euclidean distance between the detection sample X and the cluster center W obtained during the training process; second, a large number of weights are added to the nearest neighbors of the model samples; finally, the subordination function of the model sample x_i is used to classify x_i. It concludes that the model sample x_i belongs to the subordinate function $u_c(x_i)$ of the classification c. The detailed representation is shown in the following expression.

FIGURE 7: ISODATA clustering algorithm training results.

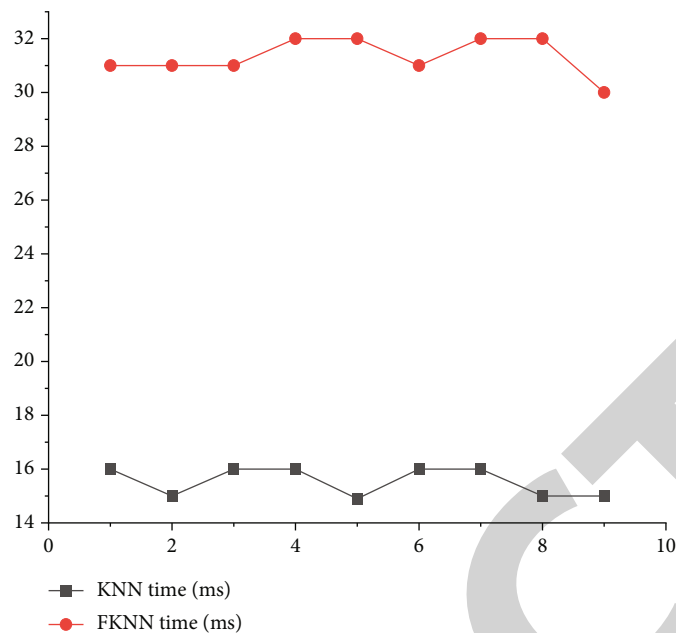$$u_c(x_i) = \frac{\sum_{k=1}^{K} u_c(w_k)/\|x_i - w_k\|^2}{\sum_{k=1}^{K} p/\|x_i - w_k\|^2}. \tag{1}$$

Figure 9: Comparison of the time results of fuzzy KNN and KNN algorithms.

In the equation, the affiliation function $u_c(w_k)$ of cluster center w_k is its radius ρ_k. The classification marker of the cluster center is the class marker of all samples in cluster z_k. The affiliation function of cluster center w_k is shown in the following expression.

$$u_c(w_k) = \begin{cases} \rho_k, \\ 0. \end{cases} \tag{2}$$

The denominator part of equation (1) of the affiliation function for the same test pattern sample is summed over the $k\rho/\|x_i - w_k\|$ in that test pattern sample, that is, the nearest neighbor clustering radius ρ and the product of the inverse of the distance between the center of the cluster and the test sample. The maximum affiliation $u_c(w_k)$ of the pattern sample x_i is obtained, then, the category label of the test pattern sample is then c. The affiliation function of the test sample x_i is as shown in the following expression.

$$u_c(x_i) = \max_c u_c(x_i). \tag{3}$$

The flowchart of the fuzzy KNN algorithm is shown in Figure 5.

Flowchart 6 shows that the major difference of the fuzzy KNN compared to KNN is that the method can fuzzify the cluster center features after training. Following the cluster center affiliation relationship described in Section 2, i.e., equation (2), the degree of affiliation of cluster center W to a class c is found. Next, the distance between the affiliation degree of this cluster center and the test sample point x_i is stored as a priority queue of size k. After traversing the training cluster center W and obtaining the k neighboring cluster centers of the test data x_i, the affiliation degree of the test

data x_i belonging to the classification marker c is calculated based on the affiliation function $u_c(x_i)$ of the test data, that is, equation (1), and the test data x_i of the classification marker is the type marker of the cluster with the most affiliation degree. The classification effect of KNN can be improved by the selection of clustering centers for the training data. At the same time, the fuzzy KNN method can attenuate the effect of nonuniformity of samples on the classification effect and improve the classification accuracy.

4. Application Experimental Analysis

4.1. Experimental Data.
In practice, faults in underground power systems rarely occur. The large amount of experimental data used in this paper is mainly from the PSB model library of MATLAB/Simulink, and the powerful secondary development capability and sufficient toolkit of Simulink are used to simulate the faults of underground transmission lines.

PSB (Power System Blocker) is a group of power system modules in MATLAB software, mainly developed jointly by Hydrow Quebec and TECSIM International in Canada [24]. The software provides a similar approach to circuit modeling for implementing a circuit model into a system described by state equations and simulated in Simulink. A simulation of an underground transmission line fault used in the experiments is shown in Figure 6.

This paper focuses on the diagnosis and analysis of 10 different types of transmission line asymmetric faults. There are four main types: single-phase short circuit, two-phase short circuit, two-phase grounding, and three-phase short circuit.

These 10 types of faults are shown in Table 1.

4.2. Fuzzy KNN Algorithm Analysis Model Experiments.
The data samples used in the model experiments of the fuzzy KNN algorithm are distance parameters from 10% to 90%
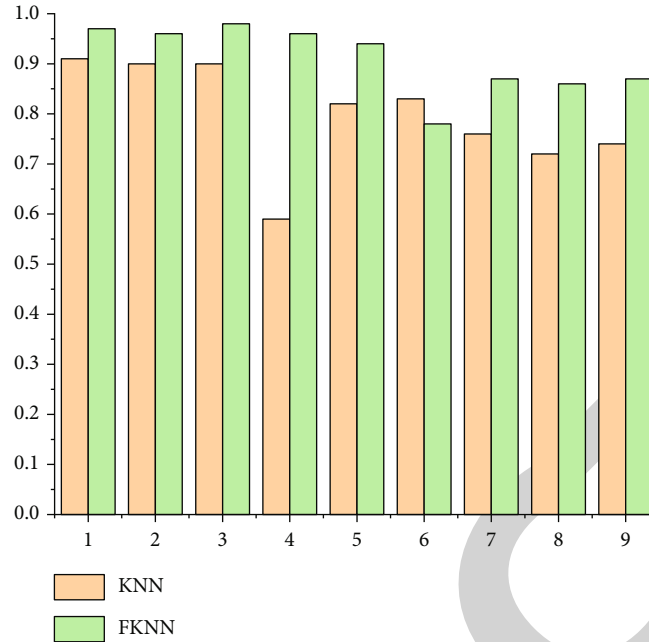
Figure 10: Comparison of the accuracy results of fuzzy KNN and KNN algorithms.

for each 10% increase; various transition resistance values of $1\,\Omega$, $3\,\Omega$, $9\,\Omega$, and $15\,\Omega$; transient abrupt parameters of three-phase voltage and three-phase current for a total of 252 entries when 10 types of faults occur suddenly. What is used in this test is the general cross-confirmation. In the entire transmission line, 80% of the fault samples were randomly selected, for a total of 200. The remaining 52 fault samples were used as test data. The distribution of ISODATA training results for one of the transmission line fault data is shown in Figure 7.

The test data were randomly selected from the sampled data, 52 unmarked data were selected for the failure analysis of fuzzy KNN algorithm. Then, different values of k are set, and the test trials are performed; finally, the k value with higher correctness is used as a comparison. The results of the fault classification diagnosis using the fuzzy KNN method with $k = 4$ are given in Figure 8.

Simulations were performed for 10 lines, which included different parts of different lines, and $1\,\Omega$, $3\,\Omega$, $9\,\Omega$, $15\,\Omega$, $10 \times 9 \times 4 \times 9 = 3240$ data. The test tuples used in the experiment are 1800 unlabeled samples randomly selected from the total sampling for fault diagnosis and analysis. In this paper, KNN and fuzzy KNN algorithms are used for fault analysis, and the result pairs of this method in terms of time efficiency are shown in Figure 9 and in terms of accuracy are shown in Figure 10.

From the comparison of the above experimental results, as shown in Figures 9 and 10, the fuzzy KNN algorithm has the advantage of higher accuracy and more time saving in transmission line asymmetric fault diagnosis and analysis compared to the KNN algorithm. Moreover, the fault distance parameter is indicative of the location of the fault.

5. Conclusion

In the network era, with the large-scale smart grid construction in China, the massive use of various condition monitoring devices and sensors has made the condition monitoring data increase at a geometric rate, from the initial TB level to PB level. This includes not only primary system equipment but also secondary system equipment. The increase of monitoring data is not only the increase of quantity but also the variety of monitoring data, from the previous data of records and online monitoring to today's unstructured data such as status information, images, and video, gradually forming the condition monitoring big data of power equipment, which brings great challenges to the storage and analysis of data.

In this paper, a big data-based approach is used to combine fault data collection and fault diagnosis and analysis in power systems, and simulation experiments are conducted using online data collection technology and KNN algorithms. The goal is to improve the intelligent monitoring function of the power system in order to improve the real-time fault diagnosis and analysis of the power system and make its operation stable. The paper employs a large number of clustering methods, which are analyzed in depth and the algorithms of which are described in detail. In addition, the paper uses density-based logistic regression algorithm to improve the accuracy of the previously proposed fuzzy KNN transmission line fault analysis algorithm for the accuracy problems that occur in the analysis of hybrid transmission line fault data. The experimental analysis related to fault analysis in the intelligent monitoring system is carried out by simulating transmission line fault data through the professional software of power system, and the expected objectives of the experiment are achieved. The effectiveness and

accuracy of the algorithm model proposed in the paper, in the application of fault analysis for underground transmission lines, are also fully demonstrated.

Data Availability

The dataset used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] L. Minglei, "Analysis of the current situation of transmission line online monitoring technology [J]," *Science and Technology Outlook*, vol. 5, p. 89, 2016.

[2] J. Station, "Transmission line condition monitoring and diagnosis technology [M]," *China Electric Power Publishing House*, vol. 4, 2014.

[3] L. Xiaoming, "Online monitoring system for transmission lines based on internet of things technology [J]," *Electronic Technology and Software Engineering*, vol. 13, p. 47, 2014.

[4] L. Yao, *Research and Implementation of Grid Health Monitoring Architecture and Its Fault Analysis Method [D]*, Wuhan University of Technology, 2014.

[5] H. Lihua and H. Qinghua, "Research on online monitoring and fault diagnosis technology of power transmission lines [J]," *Power Safety Technology*, vol. 7, pp. 24–27, 2016.

[6] R. Zhi, "Analysis of the application of online monitoring system for high-voltage transmission lines [J]," *Science and technology innovation and application*, vol. 15, pp. 168-169, 2013.

[7] Z. Yongqian, "Research and application of online monitoring system for high voltage transmission line operation status [D]," *North China University of Electric Power*, vol. 20, pp. 25–27, 2016.

[8] G. Zhilong, *Research and Design of Online Condition Monitoring System for Overhead Transmission Lines [D]*, Tianjin Vocational and Technical Normal University, 2015.

[9] B. Dong, *Design and Implementation of Online Ice Monitoring System for Transmission Line Micro-Meteorological Zone [D]*, University of Electronic Science and Technology, 2013.

[10] M. Weijie, *Research on Safe Operation and Maintenance System of Overhead Transmission Lines [D]*, Shandong University, 2015.

[11] E. Faleiro, G. Asensio, J. Moreno, P. Simón, G. Denche, and D. Garcaí, "Modelling and simulation of the grounding system of a class of power transmission line towers involving inhomogeneous conductive media," *Electric Power Systems Research*, vol. 136, pp. 154–162, 2016.

[12] S. Yuyao, "Application of online monitoring and condition management in high-voltage transmission lines [D]. North China University of Electric," *Power*, vol. 33, pp. 55–57, 2016.

[13] L. Fangjun, *Research and Application of MVC-Based Three-Tier Architecture Pattern [D]*, Guangdong University of Technology, 2011.

[14] L. Rong, "Research on web applications based on MVC pattern [J]," *Software Guide*, vol. 1, pp. 19–21, 2010.

[15] Z. Xiaojun, "Application of MVC architecture pattern in Java development [J]," *Heilongjiang Science and Technology Information*, vol. 29, pp. 177-178, 2016.

[16] G. Yu and Sudan, "Research and application of MVC design pattern in Web development [J]," *Journal of Natural Sciences of Harbin Normal University*, vol. 4, pp. 62–64, 2015.

[17] Z. Chunhua, *Research and Application of Struts Framework Based on MVC Pattern [D]*, Jilin University, 2010.

[18] W. Qi, "The world's first e-commerce big data index linked to Boshi gold 100 index fund is about to debut [J]," *Stock Market Dynamic Analysis*, vol. 16, pp. 55–55, 2015.

[19] L. Fenglin, "Exploring the application of big data and cloud computing in the medical industry [J]," *Information and Computer: Theory Edition*, vol. 8, pp. 24-25, 2016.

[20] C. Yujia, "The application of big data in the field of intelligent transportation [J]," *China Public Safety: Academic Edition*, vol. 14, pp. 56–59, 2016.

[21] L. Zaihua and L. Mingkun, "A review of intelligent diagnosis of power system faults [J]," *Electrical Technology*, vol. 8, pp. 21–24, 2010.

[22] J. C. Tan, P. A. Crossly, and J. Goody, "Fault Section Identification on a Transmission Network Using Action Factors and Expert System Technology," in *Proceedings of 13th Power Systems Computation Conference (PSCC'99)*, pp. 820–826, Trondheim (Norway), 1999.

[23] X. J. Zhang, S. B. Liu, C. P. Yan et al., "Power system fault diagnosis based on forward and backward reasoning [J]," *Power System Automation*, vol. 22, no. 5, pp. 30–32, 1998.

[24] J. Cai-Liang, "Application of artificial intelligence technology in power system fault diagnosis [J]," *Guangdong Electric Power*, vol. 24, no. 9, pp. 87–92, 2011.