IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 November 2023, accepted 7 December 2023, date of publication 15 December 2023,
date of current version 27 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343393

RESEARCH ARTICLE

Design of Line Loss Rate Calculation Method for Low-Voltage Desk Area Based on GA-LMBP Neural Network Model

ZETAO JIANG[1], GU LI[2], YONGZHI CAI[1], JIAN LI[1], AND KE ZHANG[1]
[1]Metrology Center, Guangdong Power Grid Company Ltd., Guangzhou 511545, China
[2]Guangdong Power Grid Company Ltd., Guangzhou 510080, China

Corresponding author: Zetao Jiang (leo19930731@163.com)

This work was supported by China Southern Power Grid Company Ltd., under Grant 035900KK52220006 and Grant GDKJXM20220254.

ABSTRACT The distribution network line loss computation method needs to be enhanced in light of the ongoing growth of the national power grid. The study classifies and segments the station data using a decision tree model and a multi-feature volume weighted station clustering algorithm. It then uses a back propagation neural network as a substrate, along with the Levenberg-Marquard algorithm and genetic algorithm for optimization. Collect and organize relevant data on line loss rates in low-voltage substation areas, including information on energy meters, meter boxes, and lines. Next, construct a genetic algorithm neural network model and use the backpropagation algorithm for training. Evaluate the accuracy and stability of various models by comparing the error between predicted and actual line loss rates through experiments. Finally, optimize the neural network parameters and network structure to improve the model's prediction accuracy and robustness. The experimental data showed that compared to the density-based spatial clustering algorithm for noisy applications, the contour coefficient metrics of the proposed multi-feature volume weighted station clustering algorithm improved by 0.05 and the average consumption time of the algorithm was reduced by 75%. Compared to the back-propagation neural network model optimized by the Levenberg-Marquard algorithm, the root-mean-square error of the neural network model optimized by the addition of the genetic algorithm for the calculation of the line loss rate of the four station samples was reduced by 72%, 55%, 53% and 37%, and the values of R2 were improved by 8.72%, 13.59%, 7.91% and 11.69%, respectively. The testing results demonstrated that the neural network model has good generalization capabilities and a high degree of curve fitting. Also, the relative errors of the calculation of the station area's line loss rate are mainly within the range of 0% and 10%. For the growth of energy conservation in the country, this innovative technology offers a new way to determine and manage line loss of the station area.

INDEX TERMS GA, LMBP, low voltage station, station clustering, feature quantity, line loss calculation.

I. INTRODUCTION

The low-voltage distribution system, which may serve consumers' needs for direct power delivery, is the last link in the distribution network's chain for transmitting electrical energy. Distribution lines' stability and efficiency have an impact on the quality of the power supply [1]. However, the lengthy queue and electric equipment in the circuit from the power supply station to the user's home will inevitably result in the consumption of electricity, leading to the generation of transmission losses. Resistance in power lines, equipment, and other components leads to higher power losses and energy waste, ultimately harming power users' experience, as well as the profitability of power providers, and hindering societal productivity [2]. Statistical analyses of the data show that China's distribution network has a low line loss ratio (LLR) in high-voltage power transmission of 35kV and above, which is equal to the international level [3]. Line loss in low voltage distribution of 10 kV and below, however, is higher than the international standard of 6% due

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai.

to the wide disparity between urban and rural areas, the complicated structure of the power grid, the variety of loads, and the relative concentration of time of energy consumption. The low-voltage distribution network in China, according to statistics, is responsible for close to 50% of the line loss in the whole power system. Particularly at 10kV, line loss is a serious problem that needs to be addressed [4], [5]. The administration of China's low-voltage distribution network is currently plagued by a number of issues, such as the variety of power supply characteristics in the station area, which makes line loss computation (LLC) challenging and results in manual labour taking longer to complete and line loss data with lower accuracy. Big data technology provides rich data for distribution network LLC analysis and provides a new direction for distribution network LLC. The study proposes the design of a low voltage station (LVS) LLR calculation method based on the Levenberg-Marquard back-propagation neural network (GA-LMBP) model optimized by genetic algorithm (GA), and conducts research in two aspects, namely, station clustering (SC) and machine learning-based LLR computation, with the goal of enhancing the level of station line loss management. The study is broken up into four sections: an overview of the clustering and neural network research; the design of the Low-Voltage Station Area Line Loss Rate (LVSA-LLR) calculation method based on the GA-LMBP neural network model; the analysis of the application of the LLR calculation method; and the study summary.

II. RELATED WORKS

Cluster analysis, commonly referred to as cluster analysis, is a crucial algorithm for data mining and a statistical analytic technique for classification issues. The K-means technique uses distance as a gauge of similarity between data objects. It is a division-based clustering algorithm. For interpolated separable fractional density fitting, Qin's research team developed a machine-learning K-means clustering algorithm, and the findings indicated that the approach was 10 times faster than the Fock exchange calculation [6]. It was demonstrated experimentally that Chen et al.'s adaptive orthogonal decomposition agent model, built using the K-means algorithm, classified blade shapes using a clustering algorithm, and created experimental samples using perturbations in blade control parameters in less time than 1/360th of the computational hydrodynamic time [7]. Krleza's team proposed a statistical hierarchical clustering algorithm to reduce the complexity of traditional clustering algorithms, which accomplished data prediction by tracking population evolution and data flow analysis [8].Yang's research team proposed a path-based node similarity measure that combined a new initial clustering centre and an adaptive heuristic clustering algorithm to find the most influential population. Experiments on real datasets showed that the accuracy of the method can reach 86% [9]. A zoning optimization design method was put forth by Li's team that combines K-means and GA, clusters the multi-energy model of the building zoning, and assesses resource availability by combining the clustering of resource data. The experimental results showed that the method was effective [10].

In station LLC, artificial neural network models (NNM) can usually be used for calculation and prediction, modelling neurons in the human brain in order to complete the processing, learning and prediction of information. In order to process line loss data, Hassanpour's research team used artificial NNM for predictive control, along with analysis of principal components to ensure the model's validity and autoencoder-based strategies to calculate the setpoints. The results showed that the artificial neural network increased the model's predictive control's efficiency by 59% [11]. The results showed that the model's prediction accuracy could reach 96% [12]. Hu's scientific research team for the prediction of electric power radiation proposed a ground-based cloud motion prediction method, using GAs to optimize the BP neural network's parameters in conjunction with data training to create a new ultra-short-term prediction model. Lilienkamp et al. constructed a deep-learning based ground surface motion model for energy parameter prediction using U-Net neural network, and experimental data showed that the data prediction time of this method was 1s [13]. Jeantet's research team utilized machine learning to obtain recognition behaviors and presented a fully convolutional neural network for image segmentation. Results from experiments conducted on the labeled data set indicate that the model achieved an 88% score on the predicted AUC evaluation metric [14]. Kin's team proposed four different types of integrated models for the integrated prediction model of energy consumption, based on the way of combining layers and classification features [15].

In summary, many researchers have done different studies and designs for clustering algorithms and neural network prediction models, but the applicability of these methods still needs to be improved. Therefore, the study uses multi-feature volume weighted SC algorithm and decision tree model to classify and segment the data, and proposes LVSA-LLR based on GA-LMBP NNM, which is expected to improve the efficiency of LLC.

III. DESIGN OF LVSA-LLR BASED ON GA-LMBP NNM

In this chapter, LVSA-LLR is constructed using clustering algorithm and NNM, the first section of this chapter is the design of station line loss feature identification method and the second section of this chapter is the design of station LLC method.

A. DESIGN OF LINE LOSS CHARACTERISTICS RECOGNITION METHOD FOR TERRACE AREA

Different clustering algorithms are used in different scenarios, and the clustering model that integrates the actual working requirements of power supply is the prerequisite for the implementation of the LLC method in low-voltage distribution networks [16]. For precise clustering analysis of

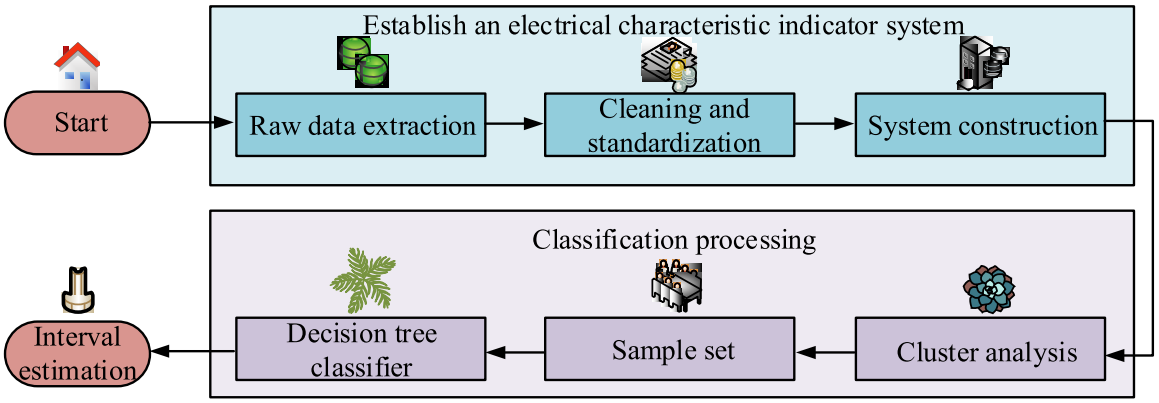

FIGURE 1. A technical model for precise identification of line loss features in low-voltage substations.

samples from low-voltage (LV) stations, this study integrates a standard clustering algorithm with a multi-feature weighted SC improvement algorithm. This combination effectively extracts information regarding line loss patterns, leading to the development of a technical model for the precise identification of line loss characteristics in LV stations, as demonstrated in Figure 1.

The basic framework of multi-feature weighted SC improvement algorithm is K-means algorithm, which integrates the needs of power grid clustering application to design a multi-feature quantity weighted clustering method suitable for station class division. The weights in the clustering delineation process need to be decided by combining the importance of each feature quantity to achieve the station class level [17].The objective function of K-means optimization is shown in Equation (1).

$$\min_{U,Z} L(U,Z) = \sum_{p=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ip}(x_{ij}-z_{pj})^2,$$
$$s.t.\ u_{ip} = \{0,1\},\ \sum_{p=1}^{k} u_{ip} = 1 \tag{1}$$

In Equation (1), the objective function to be optimized by the algorithm is $\min_{U,Z} L(U,Z)$, the cluster assignment matrix U and the cluster centre matrix Z are the independent variables, the number of clustering clusters in the dataset is k, the number of data points in the target dataset to be clustered is n, and the eigenvector dimensions are m. the sample clusters are classified as u_{ip}. The optimization problem is shown in Equation (2) when the cluster assignment matrix is fixed.

$$\min_{Z} L(U^t,Z) = \sum_{p=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ip}^t(x_{ij}-z_{pj})^2 \tag{2}$$

In Equation (2), the optimization objective function after fixing is $\min_{Z} L(U^t,Z)$ and the allocation matrix is fixed as U^t.The extreme value of the cluster centre matrix variable is solved to obtain the cluster centre position as shown in Equation (3).

$$z_{pj}^{t+1} = \frac{\sum_{i=1}^{n} u_{ip}^t x_{ij}}{\sum_{i=1}^{n} u_{ip}^t} \tag{3}$$

In Equation (3), the cluster centre z_{pj}^{t+1} at moment $t+1$ is the position of the cluster centre of mass at moment t. When the cluster centre matrix is fixed, the optimization problem contains only variable U. The result of the optimization problem for the sample data is shown in Equation (4).

$$u_{ip}^{t+1} = \begin{cases} 1, & \sum_{j=1}^{m}(x_{ij}-z_{pj})^2 \le \sum_{j=1}^{m}(x_{ij}-z_{sj}), \quad 1 \le s \le k \\ 0, & \sum_{j=1}^{m}(x_{ij}-z_{pj})^2 > \sum_{j=1}^{m}(x_{ij}-z_{sj}), \quad others \end{cases} \tag{4}$$

In Equation (4), when sample i is assigned to p clusters, u_{ip}^{t+1} takes the value of 1. The study suggests an improved k-clustering algorithm (KCA) based on weighting of line loss features in the station based on the K-means algorithm. This algorithm introduces dynamic weight coefficients matrices in the loss function as the independent variables, and its importance reference weight matrix as input parameters. The improved objective function is shown in Equation (5).

$$\min_{U,W,Z} L(U,W,Z) = \sum_{p=1}^{k}\sum_{i=1}^{n} u_{ip}\sum_{j=1}^{m}\omega_j^{\beta}(x_{ij}-z_{pj})^2 + \lambda\sum_{j=1}^{m}\left\|\omega_j-\omega_j^0\right\|^2 \tag{5}$$

In Equation (5), the improved objective function is $\min_{U,W,Z} L(U,W,Z)$, the matrix of dynamic weight coefficients is $W = [\omega_1, \omega_2, \cdots, \omega_m]$, the matrix of reference weights for the importance of each dimension of the sample data is

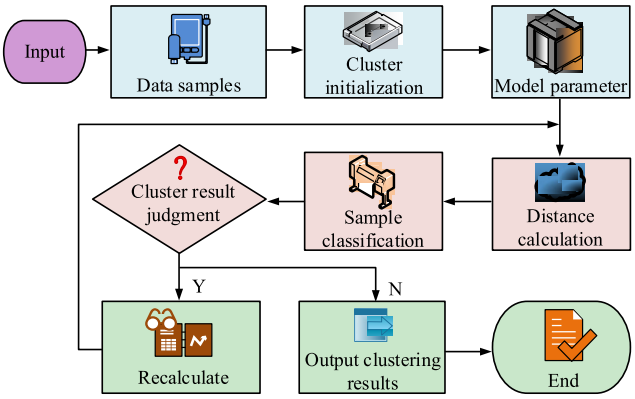

FIGURE 2. Improved K-clustering algorithm process based on weighted line loss features in substation areas.

$W^0 = [\omega_1^0, \omega_2^0, \cdots, \omega_m^0]$, and the penalty factor is B. The constraints are shown in Equation (6).

$$\begin{cases} \sum_{p=1}^{k} u_{ip} = 1, & 1 \le i \le n \\ u_{ip} \in \{0, 1\}, & 1 \le i \le n, 1 \le p \le k \\ \sum_{j=1}^{m} \omega_j = 1, & 0 \le \omega_j \le 1 \end{cases} \tag{6}$$

In Equation (6), the cluster division judgement of the sample is affected by the value and range [18]. The independent variables of the improved KCA based on the weighting of station line loss characteristics are updated using alternating direction multipliers. Additionally, as stated in Equation (7), the iterative equation for the minimal value of the goal function is derived by iteration.

$$u_{ip} = \begin{cases} 1, \sum_{j=1}^{m} \omega_j^{\beta} (x_{ij} - z_{pj})^2 \le \sum_{j=1}^{m} \omega_j^{\beta} (x_{ij} - z_{sj}), \ 1 \le s \le k \\ 0, \quad \sum_{j=1}^{m} \omega_j^{\beta} (x_{ij} - z_{pj})^2 > \sum_{j=1}^{m} \omega_j^{\beta} (x_{ij} - z_{sj}), \quad others \end{cases} \tag{7}$$

In Equation (7), the cluster class division iteration is u_{ip}. The iterative values of the cluster centre matrix and weight matrix are obtained in the same way, and the dynamic weight matrix can be solved by the augmented Lagrangian method. The flow of the improved KCA based on the weighting of station line loss features is shown in Figure 2.

Firstly, the first sample cluster centre needs to be randomly selected in the dataset, and then the remaining cluster centers are selected using the roulette wheel method to obtain the cluster initial centre vectors. After the weight matrix is randomly initialized, the reference weight matrix and penalty factor are input into the model. The objective function is iterated to get the minimum value, the matrix is judged to be updated and the clustering results are obtained. The iteration of the cluster centre matrix gets the new cluster centre matrix and the same method gets the new weight matrix.Entropy weight method and mutual information method are commonly used assignment methods, and the study combines the advantages of both of them to determine the final weights by combining the assignment. The reference weights of the electrical feature quantities are shown in Equation (8).

$$\omega_j^0 = \frac{\xi_j \gamma_j}{\sum_{j}^{M} \xi_j \gamma_j}, j = 1, 2 \cdots M \tag{8}$$

In Equation (8), the weight of each electrical feature quantity obtained by the entropy weighting method is ξ_j, and the weight of each electrical feature quantity obtained by the mutual information method is γ_j. The work employs the decision tree model to design the classifier in order to more clearly explain the laws of the clustering structure and to provide straightforward and understandable classification rules to categorize incoming data [19]. The training set of clustering results based on the improved KCA weighted by station line loss features is carried out using a decision tree model, where the Gini index is chosen for the model loss function, and in order to avoid over-fitting, leaves with a sample size of less than 5 are pruned, and in this way, a method for determining the data category is constructed. The Gini impurity coefficient corresponding to the nodes is calculated as shown in Equation (9).

$$gini = \sum_{i=1}^{k_1} p_i(1 - p_i) = 1 - \sum_{i=1}^{k_1} p_i^2 = 1 - \frac{\sum_{i=1}^{k_1} n_i^2}{(\sum_{i=1}^{k_1} n_i)^2} \tag{9}$$

In Equation (9), the Gini impurity coefficient corresponding to the node is *gini* to measure the purity of the node or leaf, the number of sample types in the data set is k_1, the percentage of the number of samples of category i in the node or leaf is p_i, and the number of samples of category i in the node or leaf is n_i.The study concludes from the analyses that the LLR data obeys the normal distribution, and unbiased estimation of the overall variance is calculated as shown in Equation (10).

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} \tag{10}$$

In Equation (10), the sample variance is σ, the sample LLR data is x_i, the sample LLR mean is $\bar{x}$, and the number of class samples is n. The study makes a judgement on line loss data outliers based on Lajda's criterion, i.e., where the distance from the overall mean is greater than three times the sample variance, a kernel of data points is required and data culling is considered.

B. DESIGN OF STATION LLC METHOD BASED ON GA ALGORITHM OPTIMIZATION

Station theoretical LLC provides an effective basis for loss reduction programme development, line loss assessment and

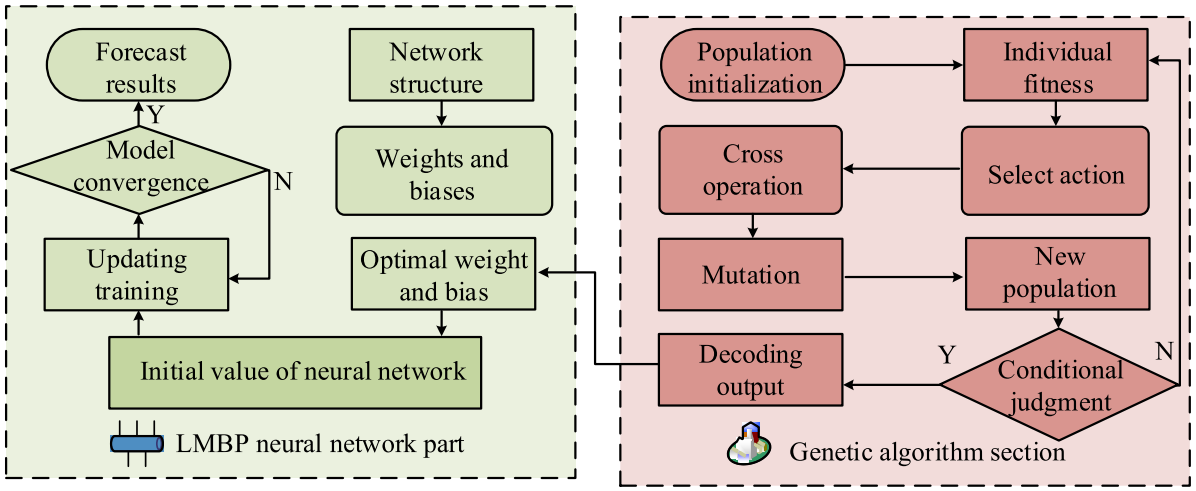

FIGURE 3. Optimization of LMBP neural network process using GA.

distribution planning, which helps to improve the economy of distribution network operation [20]. Traditional LLC methods have low accuracy, incomplete access to input data, and high manual dependency, while deep NNM will be used as a new generation of artificial intelligence technology to calculate line losses [21].By analyzing the fundamental principles and optimization algorithms of neural networks, this study utilizes a backpropagation (BP) neural network as the underlying model of LLC. The neural network iteration is optimized through the Levenberg-Marquard (LM) algorithm, and parameter optimization of the neural network is achieved by GA in conjunction with the clustering model to construct the complete LLC model. The core of the GA is to consider the iterative optimization of candidate solutions as an evolutionary process, combining natural selection and population fitness, and finally obtaining the optimal solution for a specific environment.GA has three main characteristics: it is a general search algorithm because it is independent of the problem's specific domain and data gradient information. It is not dependent on the display of the function form, which increases its applicability and capacity for self-learning; and it iterates with the population as a whole, which has a wider search coverage and boosts search efficiency [22]. Figure 3 depicts the GA optimization procedure for the LMBP neural network.

During population initialization, the LMBP neural network structure is used to determine the number of model parameters. Random generation generates the initial parameters, while real number coding is used to perform individual weights and bias coding, resulting in individual chromosomes and the formation of the initial population. After the LMBP NNM is established, the weights and bias are used as the model parameters, and the output values of the line loss are obtained from the inputs of the training samples and the amount of features. The GA fitness function is the inverse of the mean square error of the output of the NNM, as shown in Equation (11).

$$f_i = \frac{1}{\sum_{k=1}^{l} (d_k - o_k)^2 + \zeta} \tag{11}$$

In Equation (11), the adaptation value of the i th individual in the GA is f_i, the actual station line loss of the kth station is d_k, and the neural network calculates the result as o_k, and the smaller constant in the denominator, ζ, is to avoid the case where the denominator is zero. The core of the selection operation is the roulette selection while method, combined with the individual fitness to determine the probability of genetic inheritance, the resulting offspring have excellent genes, the genetic probability is calculated as shown in Equation (12).

$$C_i = \frac{f_i}{\sum_{i=1}^{S} f_i}, i = 1, 2 \ldots, S \tag{12}$$

In Equation (12), the probability of the inheritance of the ith individual is C_i, and the population size is S. two-by-two crossover recombination of the genes of the parent generation under a certain probability is the crossover operation, and the study used the real crossover method to construct a new offspring individual. The crossover operation equation is shown in Equation (13).

$$\begin{cases} \chi'_{im} = \chi_{im}(1-\alpha) + \chi_{jm} \\ \chi'_{jm} = \chi_{jm}(1-\alpha) + \chi_{im} \end{cases} \tag{13}$$

In Equation (13), the m th allele of the paternal chromosomes i and j are χ_{im} and χ_{jm}, respectively, and the genes after crossover recombination are χ'_{im} and χ'_{jm}. The random crossover probability is α, which takes the range [0, 1]. An important aspect of biological evolution is genetic variation, and population diversity and optimal search space can be increased by the variation operator. The mutation operation changes the genes in the chromosome of the new individual

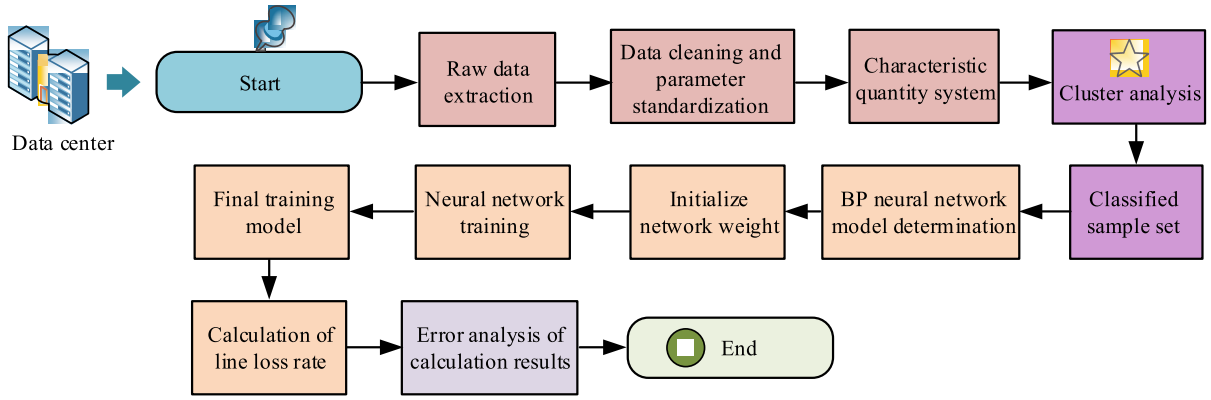

FIGURE 4. Improved K-clustering algorithm based on weighted characteristics of substation line loss and GA-LMBP neural network for substation line loss calculation process.

by a small probability on the basis of the offspring. To prevent the algorithm from falling into a local optimum and converging prematurely, the mutation probability is usually chosen to be an appropriate value in the range of 0.0001 to 0.5. The population undergoes mutation to complete one evolution, then returns to fitness computation and continues to optimize until the end conditions are met, and then the optimal weights and biases are obtained by decoding the genetic information. Deep learning provides a new direction for LV station area LLC without the limitation of traditional mathematical models. Figure 4 depicts the station LLC process based on enhanced KCA and GA-LMBP neural networks weighted by station line loss attributes.

The study proposes that the LLC method consists of two steps, classification and calculation, and the data need to be pre-processed before line loss analysis, cleaning and standardization of the electrical index data, and screening the system of line loss characteristic quantity of the station combined with the correlation degree [23].After data preprocessing, the sample type of station is classified by the improved KCA based on the weighting of line loss features of the station, followed by the training of the GA-LMBP NNM as a way to get the calculated value of the LLR of the station, and finally, the results of the LLC are subjected to an error analysis. The nonlinear component of the neural nodes is introduced through the activation function, which is the key to approximate the nonlinear function of the BP neural network. ReLU is a universal activation function, and Equation (14), which depicts the activation function, shows that the network is more sparse when there are neuron outputs of zero.

$$\mathrm{Re}LU(x) = \max(0, x) \tag{14}$$

In Equation (14), the activation function is $\mathrm{Re}LU(x)$, which indicates that the modified linear unit takes the maximum value between 0 and x. However, the ReLU activation function also causes dead neurons, two variant forms PRelu and Selu can solve such problems.The Sigmoida function is a biological S-curve that is commonly used as a threshold function for neural nodes, and the Sigmoida function expression is shown in Equation (15).

$$f(x) = \frac{1}{1 + e^{x}} \tag{15}$$

In Equation (15), the Sigmoida function is $f(x)$. The interval mapping of the input data can be done by Tanh activation function which is shown in Equation (16).

$$\tanh = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{16}$$

In Equation (16), the Tanh activation function is tanh. The Sigmoid activation function and the Tanh activation function lead to higher computational intensity during backpropagation of errors, resulting in phenomena like gradient vanishing. The Sigmoid function and its combined form, however, produce better results for the classifier. Therefore, the study uses the ReLU activation function and its variant forms as the activation function for the input and hidden layers.

IV. ANALYSIS OF LVSA-LLR APPLICATION BASED ON GA-LMBP NNM

This chapter examines the use of the LVSA-LLR based on the GA-LMBP NNM. It is divided into two sections, the first of which examines the use of the station line loss feature identification method and the second of which examines the use of the station LLC method.

A. APPLICATION ANALYSIS OF LINE LOSS CHARACTERISTICS IDENTIFICATION METHODS IN TERRACE AREAS

For the line loss feature identification method proposed by the study, the study uses the historical data of 710 stations of a power supply company for analysis, and the traditional clustering algorithm as a comparison, and the feature quantity of station LLR is selected as the gray correlation analysis results.SC before the need to determine the reference weight penalty factor and k value of each feature quantity, and the combination of weighting method to obtain the reference weight of the feature quantity is demonstrated in Table 1.

TABLE 1. Reference weight of feature quantity obtained by combination weighting method.

Characteristic quantity	Line length	Station capacity	Total Users	Load rate
Entropy weight method weight	0.239507	0.25718	0.191472	0.31184
Mutual information method weight	0.279365	0.256505	0.228888	0.235242
Final weight	0.269518	0.265737	0.176534	0.28822

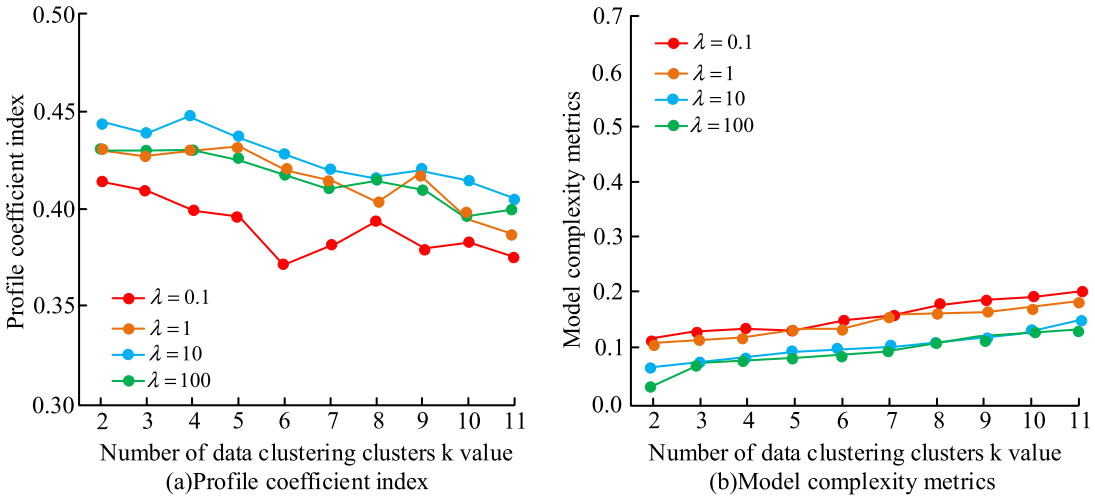

FIGURE 5. Change curve of model clustering indicators.

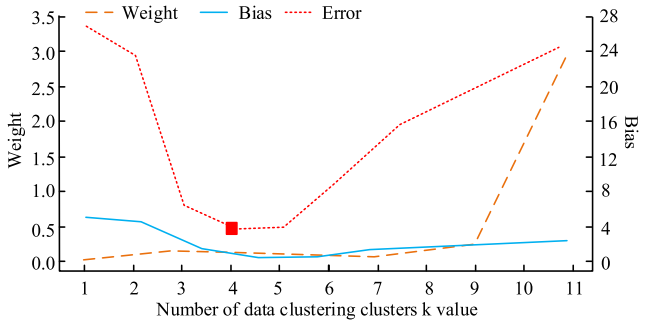

FIGURE 6. Bias weight test results.

In Table 1, the weights obtained by entropy weighting method and mutual information method are given respectively, and the final weights are obtained according to the combined assignment method. To obtain the optimal clustering model, the total profile coefficient and the model convergence time are used to determine the setting of the penalty factor and the number of clusters k value of the data set clustering clusters. The change curve of the model clustering index is given in Figure 5.

Figure 5(a) shows the change of contour coefficient with the value of k. It can be seen that the model contour coefficient shows a general trend of decreasing as the value of k increases, and the model contour coefficient index is the highest when the penalty factor λ is taken as 10. At the same time, when k = 4, the maximum value of the model contour coefficient index can reach 0.448, indicating that the clustering effect at this time is the best. Figure 5(b) indicates the change of model time complexity with the value of k. As k-value increases, the model complexity rises gradually, and at the same time, the larger the value of λ, the faster the convergence of clustering, and the stronger the stability of the algorithm. Under comprehensive analysis, the penalty factor λ is taken as 10 and k value is taken as 4 as the optimal parameters of the model. To better understand how the value of k impacts the number of clusters in clustering, this study employedbias-weight trade-off to show case model fitting. The results of the bias weight test are presented in Figure 6.

Figure 6 shows that when k = 4, the deviation variance is at its minimum, and the model fits best at this point. Tooptimize the model parameters, this study utilized bias weight balancing and time complexity analysis methods. In summary, it is essential to use appropriate methods to determine the optimal parameters of the model. The results indicate that as k-values increased, the model complexity also increased, resulting in faster clustering convergence and stronger algorithm stability. Therefore, when the value of k is high, the model's fitting leads to over-fitting, and it becomes necessary to decrease the k value to avoid this issue.Conversely, when the value of k is low, the model's fitting is inadequate, and it becomes necessary to increase the k value to enhance the model's fitting ability. In conclusion, it is found that the greatest model fitting effect is achieved when the penalty factor is set to 10 and

TABLE 2. District cluster center feature quantity.

Category	Line length/m	Station capacity/KVA	Total users	Load rate(%)	Line loss rate(%)	Number of samples
1	375.9395	400	84.4151	18.7409	1.9680	159
2	1665.9109	200	181.7831	40.7784	4.3741	105
3	383.3587	210	53.2866	28.5177	2.9755	285
4	581.5929	320	147.0313	59.5038	4.3252	161

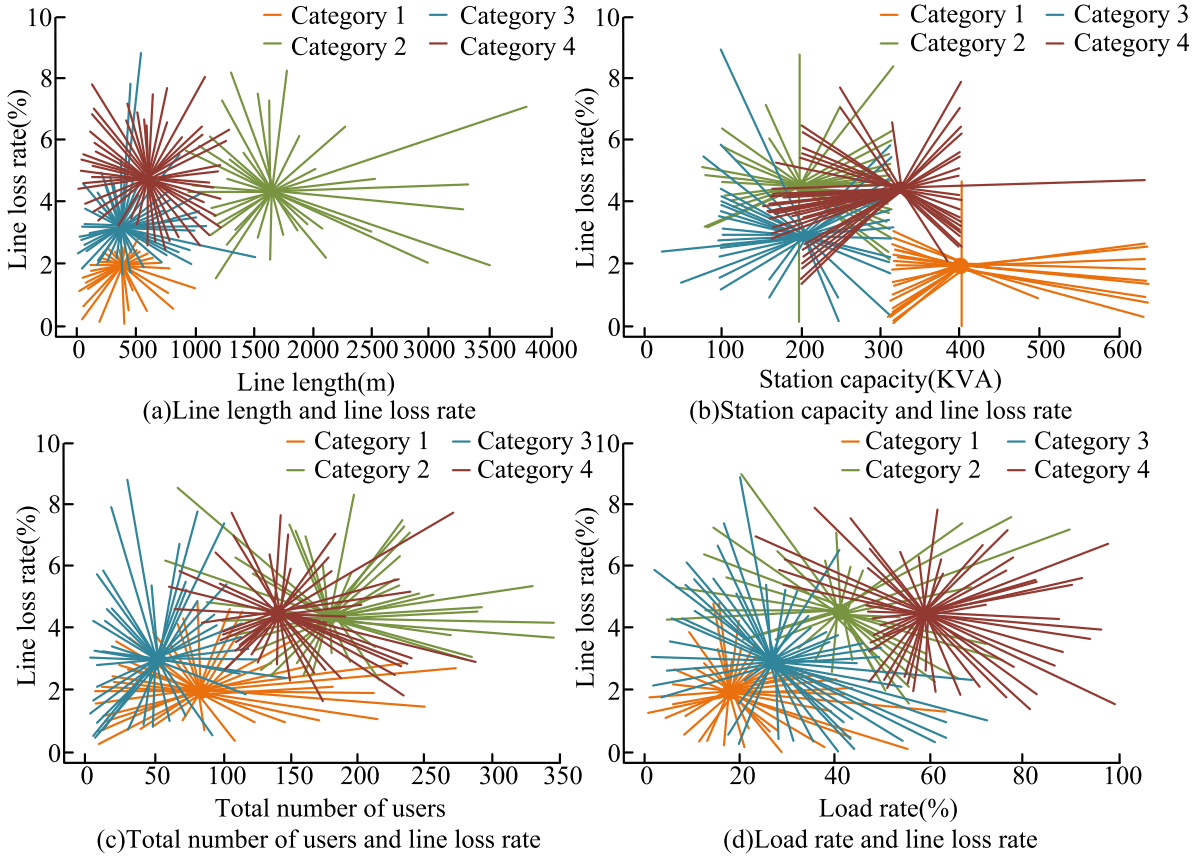

FIGURE 7. Visualization relationship between various feature quantities and line loss rate.

the k-value is set to 4. When the value of k is sizable, the model experiences over-fitting, while small values of k result in insufficient fitting of the model.Compared to the density-based spatial clustering of noisy applications (DBSCAN) algorithm, the contour coefficient metrics of the multi-feature volume weighted SC algorithm proposed in the study are improved by 0.05, and the average consumption time of the algorithm is reduced by 75%. The SC central feature quantity is shown in Table 2 when the penalty factor λ is taken as 10 and k value is taken as 4.

In Table 2, it is seen that the SC central feature quantity line length is in the range of 300-1700m, station capacity is in the range of 200-400, total number of subscribers is in the range of 80-200, load factor is in the range of 18-60%, LLR is in the range of 1.9%-4.5%. The revised KCA model's visualization of the correlation between each feature quantity and LLR based on the weighting of station line loss features is depicted in Figure7.

Figure7 shows the visualization of the clustering results for line length, station capacity, total number of users and load factor, respectively, versus LLR. The length of the line, its load factor, and the total number of users for category 1 stations are 375.9395 meters, 18.7409%, and 84.4151, respectively. These are small characteristic values, and the station has a capacity of 400 KVA. Additionally, it has a simpler electricity consumption structure with the lowest average LLR of 1.9680%. The same analysis shows that category 2 has the highest average LLR of 4.3741% because of the smaller capacity of the station and the larger line length and total number of users.The average LLR of category 3 and 4 stations is 2.9755% and 4.3252% respectively, which is greater than the LLR of category 1 stations and less

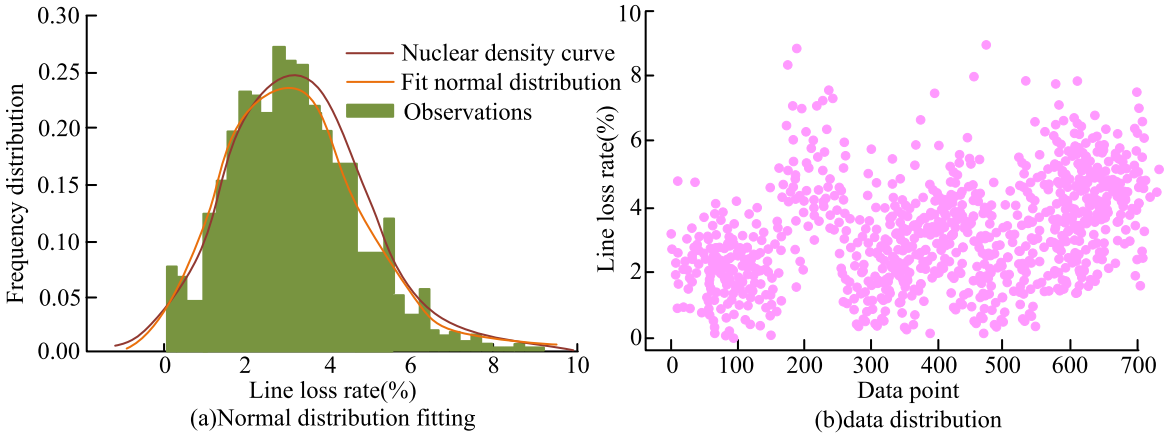

FIGURE 8. Normal distribution of sample data.

than the LLR of category 2 stations.Numerical descriptions of the LLR data were examined for statistical conformity to the normal distribution, with results indicating its adherence to the distribution curve. Sample data was also found to be normally distributed, as illustrated in Figure 8.

Figure 8 shows that the two distributions likely belong to the same overall based on the results of the KS test (P=0.2604) obtained through simulation analysis. Afterward, the classification model trains the clustering results, and the 15 sub-classifications are obtained by subdividing the four main station area sample categories. Some sample data are not ideal for normal distribution fitting, so LLR is used to reasonably estimate the interval. The results of the 15 sub-classification intervals are presented in Figure 9.

In Figure9, each of the 15 subcategories is labeled with A-O. The maximum and minimum values of the LLR data within the sample categories provide reliable estimation intervals, and the significant category it falls under typically adheres to a normal distribution, with 95% of data symmetrically distributed around the mean. If the data falls outside of the normal intervals, it is deemed anomalous, and the corresponding intervals are the anomalous intervals. Locations exceeding the reasonable estimation interval but not reaching the abnormal interval are regarded as hazardous intervals, where outliers may occur and further verification of the data is required. In summary, the combination of improved KCA and decision tree classification based on the weighting of line loss characteristics of the station area, completes the classification of low-voltage stations and the refinement of the judgement of the line loss characteristics, which provides the basis for the identification of line loss characteristics.

B. ANALYSIS OF THE APPLICATION OF LLC METHODS IN TERRACE AREAS

To verify the application effect of the station LLC method, the study still uses 710 low-voltage station data of a power supply company to conduct experiments, and evaluates the model's calculation accuracy and algorithm operation speed in two aspects, respectively.To confirm the effectiveness of the LMBP algorithm proposed in the research, the experiment conducted a comparison with three algorithms: standard BP, adaptive BP, and elastic BP. Table 3 compares the time requirements of various optimization strategies.

In Table 3, when the training error is 0.01, all four algorithms can converge, and the average time consumed by the LMBP algorithm is 3.242s, which is reduced by 17.494s compared to the standard BP.When the training error is 0.001, the standard BP algorithm is not convergent, and the average time consumed by the LMBP algorithm is 6.326s, which is reduced by 9.592s compared to the adaptive BP.When the training error is 0.0001, the adaptive BP algorithm is not convergent and the average time taken by the LMBP algorithm is 10.422s, which is reduced by 7.799s compared to the elastic BP.The comprehensive analysis shows that the computational accuracy and convergence time of the LMBP algorithm are better than the heuristic algorithm. The reasonable selection of training batch and neurons helps the model performance, and the error comparison of GA-LMBP model under different training batch and number of neurons is shown in Figure 10.

The average absolute percentage error Mape metric is used in Figure 10 to evaluate the model effect. Figure 10(a) illustrates a comparison of the error in the GA-LMBP model across various training batches. It becomes apparent that the error trend decreases with an increase in training batches. Nevertheless, once the training batch exceeds 60, there is no noticeable change in the Mape indicator, and increasing the batch size results in prolonged convergence time. Consequently, it is preferable to establish the training batch at 60. Figure 10(b) displays the error comparison of the model with different numbers of neurons and the number of training set to 200; for example, N128-128 denotes that the number of neurons is 128 and that the number of neurons in the hidden layer is 128.In this figure, a single hidden layer's sparse structure limits its ability to fit data; therefore, the study uses a two-layer N128-32 neuron model that trains under various activation functions. There are two hidden layers,

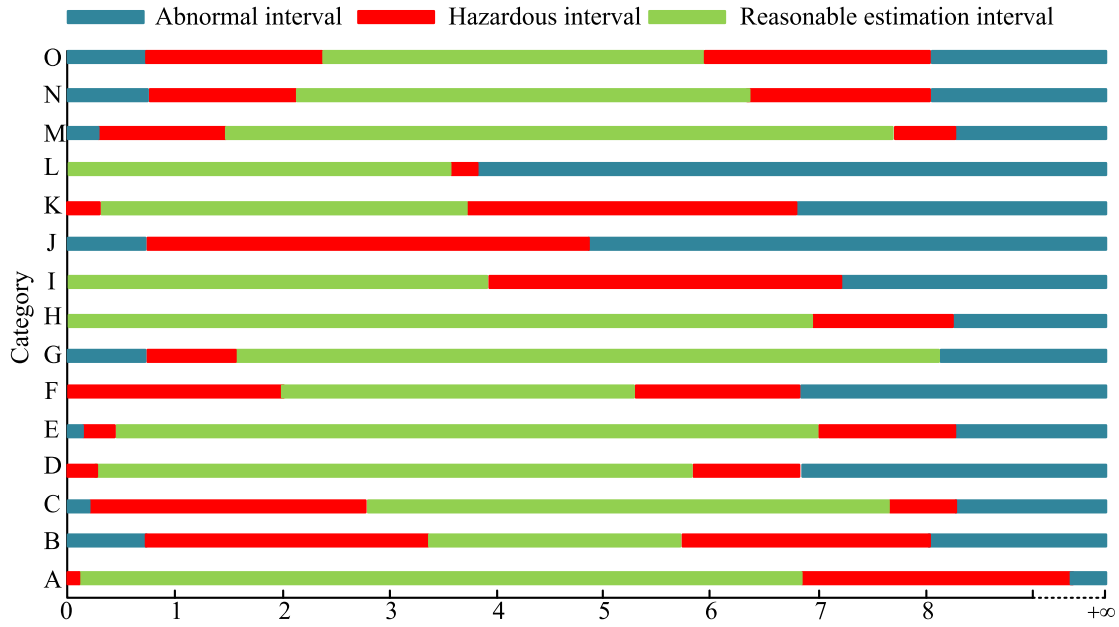

FIGURE 9. 15 sub classification interval results.

TABLE 3. Comparison of time consumption of different optimization algorithms.

Algorithm	Error 0.01 convergence time	Error 0.001 convergence time	Error 0.0001 convergence time	Optimal algorithm
Standard BP	20.736s	/	/	LM
Adaptive BP	10.506s	15.819s	/	
Elastic BP	6.785s	9.929s	18.221s	
LMBP	3.242s	6.326s	10.422s	

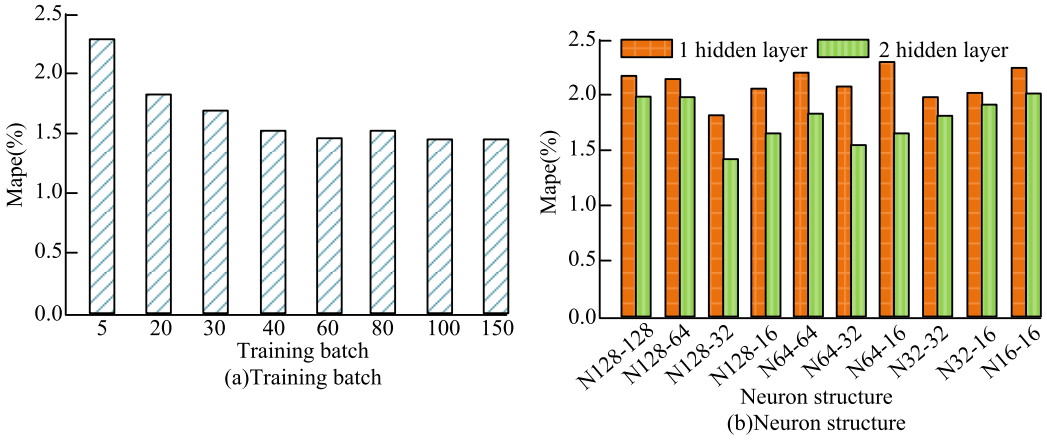

FIGURE 10. Comparison of errors in the GA-LMBP model under different training batches and number of neurons.

128 input neurons, and 32 neurons in the hidden layer. The Mape index for this configuration is 1.429%, and at this point, the error is the smallest. Table 4 displays the time-consuming comparison.

Table 4 shows that when ReLU is used as the activation function for all layers, the Mape metric is reduced to 1.326%, and when Sigmoid is used as the activation function for the output layer, the gradient vanishes. The study uses ReLU as the activation function for the entire NNM because the results of ReLU activation function are superior to those of PRelu and Selu activation functions. The standard BP, LMBP, and GA-LMBP networks were trained and tested in the study, and the computational errors of various network algorithms are reported in Table 5. The study separated the data from

TABLE 4. Comparison of training time for N128-32 neuron model under different activation functions.

Input layer and hidden layer	Output layer	Mape(%)	Training time consumption(s)
ReLU	Sigmoid	6.212	7.952
ReLU	Tanh	2.523	7.891
ReLU	ReLU	1.326	7.922
PRelu	PRelu	2.102	9.263
Selu	Selu	1.971	8.188

TABLE 5. Calculation errors of different network algorithms.

Category	Algorithm	Relative error E<5%	Relative error 5%≤E<10%	Relative error 10%≤E<20%	Relative error E≥20%
	Standard BP	9	4	9	10
1	LMBP	15	6	5	6
	GA-LMBP	22	7	2	1
	Standard BP	4	6	9	3
2	LMBP	9	5	6	2
	GA-LMBP	14	5	3	0
	Standard BP	13	17	11	16
3	LMBP	27	13	10	7
	GA-LMBP	39	12	4	2
	Standard BP	8	6	7	11
4	LMBP	16	7	5	4
	GA-LMBP	21	8	2	1

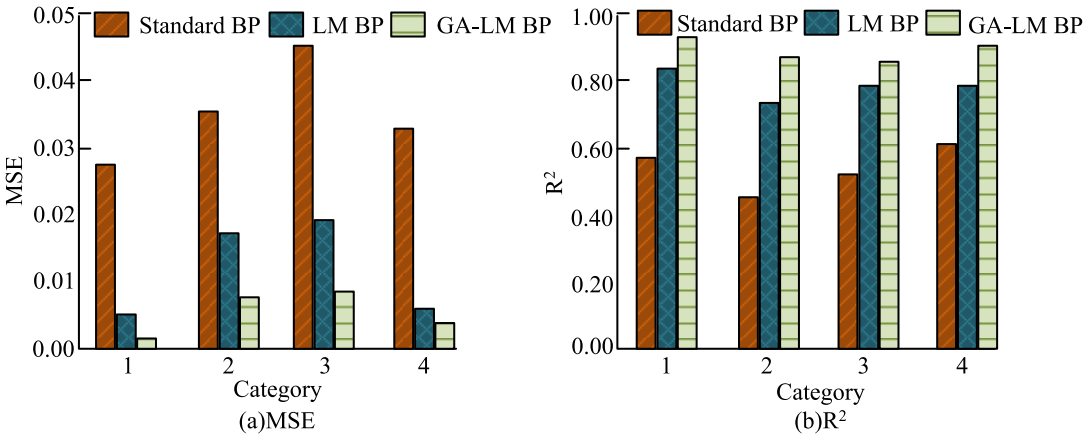

FIGURE 11. Comparison of generalization ability and fitting degree of different models.

the 710 station regions into training and testing sets in a ratio of 4:1.

In Table 5, the percentage of GA-LMBP neural network LLC errors less than 5% in the test data of the four categories of station samples are 68.75%, 63.64%, 68.42%, and 65.63%, respectively, which are higher than the standard BP and LMBP NNM. Compared to the standard BP neural network, the GA-LMBP neural network combination optimization proposed in this study decreases the average error of the LLR calculation for the station samples of the four categories by 9.62%, 6.72%, 8.59%, and 10.01%, respectively. As a result, the computational accuracy is higher and closer to the realistic situation.Figure 11 compares the generalization capacity and degree of fitting of several models.

Figure11(a) shows the comparison of the generalization ability of different models, and the GA-LMBP NNM has the smallest RMS error for the same number of iterations. Compared to the LMBP NNM, the GA-LMBP NNM reduces the root mean square error of the LLR computation for the four station samples by 72%, 55%, 53%, and 37%, and the model has a higher computational accuracy. Figure11(b) shows the comparison of the fitting degree of different models, and a larger value of R2 indicates a better fitting effect. Compared with the LMBP NNM, the GA-LMBP NNM improves the value of R2 by 8.72%, 13.59%, 7.91% and 11.69% in the calculation results of the four station samples, respectively. The comprehensive analysis shows that the GA-LMBP NNM has the best performance under the generalization ability and degree of fit indicators. The relative error of LLR calculation for 4 types of station data when the number of model iterations is 200 is shown in Figure 12.

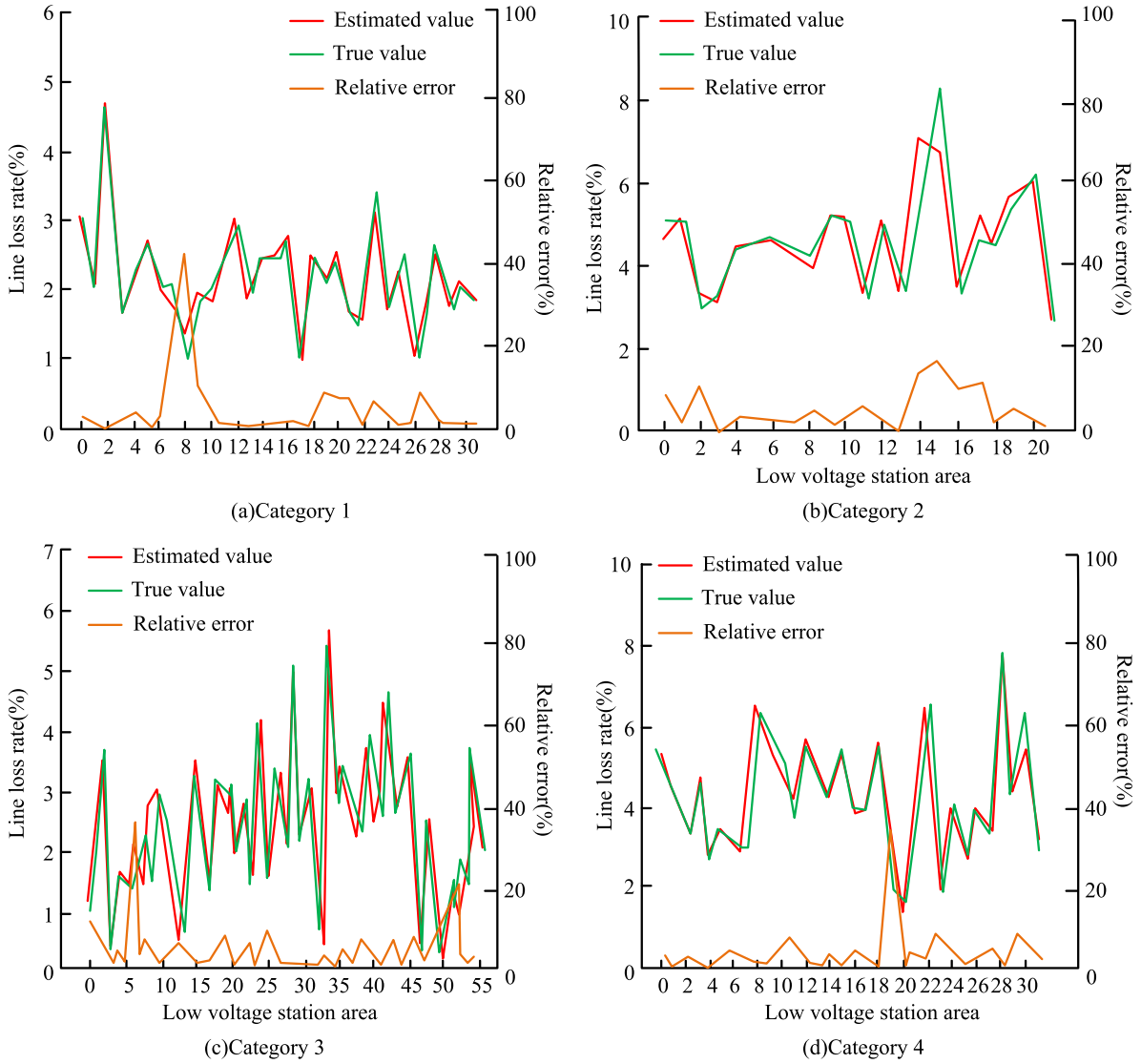

FIGURE 12. The relative error of the model in calculating the line loss rate of data in four types of stations.

TABLE 6. Verification of abnormal sample data.

Station data	Line length/m	Station capacity/KVA	Total users	Load rate(%)	Actual line loss rate(%)
1-1	110	400	75	18.84	0.97
1-2	95	400	62	12.09	1.23
3-1	376	210	56	28.84	1.87
3-2	317	210	48	24.86	2.51
3-3	200	210	8	32.93	1.26
3-4	173	210	5	27.69	1.43
4-1	425	320	69	42.63	1.88
4-1	406	320	64	38.96	2.73

In Figure12, the GA-LMBP NNM is generally close to the real value of the LLR calculation curve for the station area. And most of the relative errors of the LLR calculation for the four types of stations are in the range of 0%-10%, with a good overall fitting effect. At the same time, there will be a situation where the relative error of some sample data is large. At this time, the anomalous data need to be verified. The anomalous data contains data from a total of 8 station areas, including data from 2 station areas in class 1 station areas, 4 station areas in class 3 station areas, and 2 station areas in class 4 station areas. The verification of the anomaly sample data is demonstrated in Table 6.

Based on the analysis of power theory, it is established that the capacity of the station area has a negative correlation with LLR. On the other hand, Table 6 shows that the line length, total subscribers, and load factor have a positive correlation with LLR. Using this as a basis for checking the data where the GA-LMBP NNM calculates the outliers, the true LLR for the eight stations is obtained as 0.97%, 1.23%, 1.87%, 2.51%, 1.26%, 1.43%, 1.88% and 2.73%, respectively.

V. CONCLUSION

Data mining in big data technology offers new technical assistance for LLC and distribution network management. The study develops the station feature clustering approach with K-means clustering algorithm as the baseline, combines with the decision tree classification method to subdivide the data types, and presents LVSA-LLR based on GA-LMBP NNM to increase LLC efficiency. The experiment exhibited that the optimal parameter of the model was assumed to be with a penalty factor λ of 10 and k value of 4.When compared to the DBSCAN algorithm, the multi-feature quantity weighted SC algorithm proposed in the study had a contour coefficient index that was 0.05 better and an average algorithm consumption time that was 75% shorter. The improved KCA, which weighted line loss features of station areas, and the decision tree classification method were combined to classify low-voltage station areas and accurately assess line loss features. This method served as the foundation for identifying line loss features.The study's GA-LMBP neural network combination optimization reduced the mean error of LLR calculation for the four categories of station samples by 9.62%, 6.72%, 8.59%, and 10.01% compared to the standard BP neural network, which was much more accurate and closer to the real situation. The GA-LMBP NNM demonstrated superior generalization ability and fitting degree when compared to the LMBP NNM. It reduced the root-mean-square errors of LLR calculation for all four station samples by 72%, 55%, 53%, and 37% while also increasing the R2 values by 8.72%, 13.59%, 7.91%, and 11.69% respectively, in contrast to the LMBP NNM. The LLR calculation curve for stations in the GA-LMBP NNM was generally close to the real value, and the relative LLR calculation errors for the four types of stations were primarily in the 0%–10% range, with a better overall fitting effect. This methodological support provided for the lean management of stations. The study's findings demonstrated that the combined BP neural network optimization model's LLR calculation accuracy and efficacy are high, which has both theoretical and practical implications.Although this study has achieved the ideal experimental results, the model calculation effect can be further enhanced through the mining of the depth of the influencing factors. This study's shortcoming is that the mining of the influencing factors of line loss in the station area is not sufficiently deep. The effectiveness of the model can be further enhanced by future research that combines laboratory and field work to expand the indications and data of line loss in the station area.

REFERENCES

[1] G. Meerimatha and B. L. Rao, "Novel reconfiguration approach to reduce line losses of the photovoltaic array under various shading conditions," *Energy*, vol. 196, Apr. 2020, Art. no. 117120, doi: 10.1016/j.energy.2020.117120.

[2] N. B. Ngang, E. N. Aneke, and E. D. N. N. Bassey, "Reduction of power system losses in transmission lines using optimization technique," *Int. J. Eng. Sci.*, vol. 10, no. 4, pp. 58–66, Apr. 2021.

[3] B. Gjorgiev and G. Sansavini, "Identifying and assessing power system vulnerabilities to transmission asset outages via cascading failure analysis," *Rel. Eng. Syst. Saf.*, vol. 217, Jan. 2022, Art. no. 108085, doi: 10.1016/j.ress.2021.108085.

[4] L. Beňa, V. Gáll, M. Kanálik, M. Kolcun, A. Margitová, A. Mészáros, and J. Urbanský, "Calculation of the overhead transmission line conductor temperature in real operating conditions," *Electr. Eng.*, vol. 103, no. 2, pp. 769–780, Apr. 2021, doi: 10.1007/s00202-020-01107-2.

[5] S. Özyön, "Optimal short-term operation of pumped-storage power plants with differential evolution algorithm," *Energy*, vol. 194, Mar. 2020, Art. no. 116866, doi: 10.1016/j.energy.2019.116866.

[6] X. Qin, J. Li, W. Hu, and J. Yang, "Machine learning K-means clustering algorithm for interpolative separable density fitting to accelerate hybrid functional calculations with numerical atomic orbitals," *J. Phys. Chem. A*, vol. 124, no. 48, pp. 10066–10074, Dec. 2020, doi: 10.1021/acs.jpca.0c06019.

[7] X. Chen, R. Zhang, L. Jiang, and W. Yang, "Adaptive POD surrogate model method for centrifugal pump impeller flow field reconstruction based on clustering algorithm," *Mod. Phys. Lett. B*, vol. 35, no. 7, Mar. 2021, Art. no. 2150126, doi: 10.1142/S0217984921501268.

[8] D. Krleža, B. Vrdoljak, and M. Brčić, "Statistical hierarchical clustering algorithm for outlier detection in evolving data streams," *Mach. Learn.*, vol. 110, no. 1, pp. 139–184, Sep. 2020, doi: 10.1007/s10994-020-05905-4.

[9] P.-L. Yang, G.-Q. Xu, Q. Yu, and J.-W. Guo, "An adaptive heuristic clustering algorithm for influence maximization in complex networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 30, no. 9, Sep. 2020, Art. no. 093106, doi: 10.1063/1.5140646.

[10] Y. Li, C. Liu, L. Zhang, and B. Sun, "A partition optimization design method for a regional integrated energy system based on a clustering algorithm," *Energy*, vol. 219, Mar. 2021, Art. no. 119562, doi: 10.1016/j.energy.2020.119562.

[11] H. Hassanpour, B. Corbett, and P. Mhaskar, "Artificial neural network-based model predictive control using correlated data," *Ind. Eng. Chem. Res.*, vol. 61, no. 8, pp. 3075–3090, Feb. 2022, doi: 10.1021/acs.iecr.1c04339.

[12] K. Hu, L. Wang, W. Li, S. Cao, and Y. Shen, "Forecasting of solar radiation in photovoltaic power station based on ground-based cloud images and BP neural network," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 2, pp. 333–350, Oct. 2021, doi: 10.1049/gtd2.12309.

[13] H. Lilienkamp, S. von Specht, G. Weatherill, G. Caire, and F. Cotton, "Ground-motion modeling as an image processing task: Introducing a neural network based, fully data-driven, and nonergodic approach," *Bull. Seismolog. Soc. Amer.*, vol. 112, no. 3, pp. 1565–1582, Mar. 2022, doi: 10.1785/0120220008.

[14] L. Jeantet, V. Vigon, S. Geiger, and D. Chevallier, "Fully convolutional neural network: A solution to infer animal behaviours from multi-sensor data," *Ecolog. Model.*, vol. 450, Jun. 2021, Art. no. 109555, doi: 10.1016/j.ecolmodel.2021.109555.

[15] S.-Y. Kim, S.-J. Hong, E. Kim, C.-H. Lee, and G. Kim, "Application of ensemble neural-network method to integrated sugar content prediction model for citrus fruit using Vis/NIR spectroscopy," *J. Food Eng.*, vol. 338, Feb. 2023, Art. no. 111254, doi: 10.1016/j.jfoodeng.2022.111254.

[16] W. Shan, "Digital streaming media distribution and transmission process optimisation based on adaptive recurrent neural network," *Connection Sci.*, vol. 34, no. 1, pp. 1169–1180, Dec. 2022, doi: 10.1080/09540091.2022.2052264.

[17] H.-L. Minh, T. Sang-To, M. Abdel Wahab, and T. Cuong-Le, "A new metaheuristic optimization based on K-means clustering algorithm and its application to structural damage identification," *Knowl.-Based Syst.*, vol. 251, Sep. 2022, Art. no. 109189, doi: 10.1016/j.knosys.2022.109189.

[18] H. Zhang, H. Cheng, and S. Zhang, "Stochastic optimal transmission switching considering the correlated wind power," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 13, pp. 2664–2672, Jul. 2019, doi: 10.1049/iet-gtd.2018.5043.

[19] M. Deakin, T. Morstyn, D. Apostolopoulou, and M. D. McCulloch, "Voltage control loss factors for quantifying DG reactive power control impacts on losses and curtailment," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 10, pp. 2049–2062, Feb. 2022, doi: 10.1049/gtd2.12413.
[20] M. Khosravi, H. Monsef, and M. H. Aliabadi, "Network loss management and allocating the transmission losses to loads and generation units according to their transactions," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 8, pp. 1540–1551, Mar. 2020, doi: 10.1049/iet-gtd.2019.0503.
[21] Q. Liu, S. Wang, Q. Zhao, and K. Wang, "Interval power flow calculation algorithm for multi-terminal DC distribution networks considering distributed generation output uncertainties," *IET Gener., Transmiss. Distrib.*, vol. 15, no. 5, pp. 986–996, Dec. 2020, doi: 10.1049/gtd2.12074.
[22] Y. Guo, Z. Mustafaoglu, and D. Koundal, "Spam detection using bidirectional transformers and machine learning classifier algorithms," *J. Comput. Cognit. Eng.*, vol. 2, no. 1, pp. 5–9, Apr. 2022, doi: 10.47852/bonviewJCCE2202192.
[23] P. A. Ejegwa and J. M. Agbetayo, "Similarity-distance decision-making technique and its applications via intuitionistic fuzzy pairs," *J. Comput. Cogn. Eng.*, vol. 23, no. 1, pp. 68–74, Jan. 2023, doi: 10.47852/bonviewJCCE512522514.

ZETAO JIANG was born in Jieyang, Guangdong, China, in July 1993. He received the bachelor's degree in electrical engineering and automation and the master's degree in power systems and automation from the South China University of Technology, in 2016 and 2019, respectively.

Since July 2019, he has been the Data Application Specialist with the Energy Data Department, Measurement Center, Guangdong Power Grid Company Ltd. As the Project Leader, he participated in the research on the key technology project of "Research and Application of Line Loss Analysis, Diagnosis, and Carbon Reduction Technology for Adapting to New Power Systems" with China Southern Power Grid Corporation.

GU LI was born in Wuhan, Hubei, China, in June 1978. He received the bachelor's degree in power systems and automation from the Wuhan University of Water Resources and Electric Power, in 2000, and the master's degree in electrical engineering from the Wuhan University of Technology, in 2011. He is currently a Senior Engineer in power engineering technology, mainly engaged in power system measurement and line loss research.

YONGZHI CAI was born in Zhanjiang, Guangdong, China, in September 1984. He received the bachelor's and master's degrees in power systems and automation from the Guangdong University of Technology, in 2007 and 2011, respectively, and the Ph.D. degree in power systems and automation from the South China University of Technology, in 2016. His research interest includes electricity metering.

JIAN LI was born in Suizhou, Hubei, China, in March 1986. He received the bachelor's degree in electrical engineering and automation and the master's degree in power systems and automation from the South China University of Technology, in 2008 and 2011, respectively. His research interests include energy metering and electricity analysis.

KE ZHANG was born in Qiannan, Guizhou, China, in May 1982. She received the bachelor's degree in electrical engineering and automation from the South China University of Technology, in 2004, and the master's degree in power systems and automation from Tianjin University, in 2007. Her research interests include electricity metering, automation inspection and testing, and power line loss.