

¹ Yu Ling*² WeiYu Zou³ GengBin Zhang

Research on Transmission Line Tree Fault Risk Prediction Based on GA-BP Neural Network and Digital Twin



Abstract: - With the increasing complexity and management requirements of power inspection, the use of advanced technologies such as artificial intelligence and digital twins and their application in the field of transmission line inspection has gradually become a cutting-edge direction for identifying typical problems such as tree faults and hidden dangers in transmission lines. The article presents a novel approach for the construction of a digital twin of transmission lines, integrating laser point cloud data, multispectral data, and relevant environmental data generated by tree barriers. This is achieved through the utilization of an optimized GA-BP algorithm, facilitating high-precision prediction and analysis of the growth height and risk level of tree barriers. This offers a promising digital solution for the refined inspection business of transmission lines. The experimental results show that the BP neural network optimized by GA has improved the prediction accuracy by 30% compared to conventional methods, and has higher prediction accuracy. By integrating multiple sources of data such as laser point clouds, a twin transmission line is constructed, combined with advanced artificial intelligence technology, which can not only achieve observability and measurability of transmission lines, but also accurately deduce and predict the growth of tree obstacles in the future, improving the safe operation and management level of the power grid. This is the technological development direction for empowering production business with transmission lines and even digital power grids in the future.

Keywords: Multi source data fusion, Digital twin, Transmission line inspection, Tree obstacle risk prediction, Laser point cloud.

I. INTRODUCTION

Transmission lines are mostly distributed in mountainous areas and forests. In recent years, the number of faults caused by ultra-high trees under transmission lines has been increasing, and tree barriers have become a typical hidden danger in power grid companies' transmission lines. The traditional manual inspection method, due to the complex terrain of mountains and forests, involves a large workload, high costs, and poses safety hazards to personnel. In some areas, the use of satellite remote sensing or unmanned aerial vehicle inspection methods has poor accuracy, and most of them only investigate the current hidden dangers of tree obstacles, lacking the ability to accurately deduce and predict future situations [1].

Existing scholars have conducted research on plant growth budgets. For example, Wu et al. (2024) [2] proposed using laser point cloud data to describe the relationship between eucalyptus age and height using the Richards function, but the prediction accuracy is insufficient and far from the measured values. The comparative analysis of the mixed effects model, generalized model, and BP artificial network model in reference [3] shows that when incorporating the variability of stand height diameter relationship into the model, it is clear that mixed effects nonlinear regression and backpropagation neural network modeling methods are useful tools for predicting tree height in forest resource management. However, the fusion level of various types of data is not high, which affects the final prediction accuracy. Özçelik et al. (2018) [4] proposed using laser point cloud data and multispectral data to model and predict tree height through support vector regression, reducing errors caused by small sample data and overcoming the difficulty of selecting tree growth equations. Tao et al. (2012) [5] conducted research on artificial eucalyptus trees as a case study, utilizing continuous eucalyptus growth inventory data as the dataset. This data was employed in conjunction with digital twin and artificial neural network technology to predict and analyze the diameter at breast height and height of eucalyptus trees, thereby significantly improving prediction accuracy. Leslie et al. (2018) [6] proposed using laser point cloud data and Richard's function to describe the relationship between eucalyptus age and height, but the prediction accuracy is insufficient. The above research indicates that current research lacks a solution that is observable, measurable, has strong generalization ability, can support

¹ Guangdong Power Grid Co., Ltd. Project Management Center, GuangZhou City, GuangDong Province, 510699, China

² Guangdong Power Grid Co., Ltd. Project Management Center, GuangZhou City, GuangDong Province, 510699, China

³ Guangdong Power Grid Co., Ltd. Project Management Center, GuangZhou City, GuangDong Province, 510699, China

* Corresponding Author Email: 1684660581@qq.com

multiple transmission line inspection business scenarios, and can accurately deduce and predict future tree obstacle hidden danger situations.

Therefore, the article proposes a transmission line tree obstacle risk prediction method based on GA-BP neural network and digital twin. On the one hand, by integrating multi-source data such as laser point cloud and multispectral, a transmission line digital twin platform is constructed to support the development of transmission line monitoring and inspection services such as tree obstacle analysis; Secondly, the optimized genetic algorithm error backpropagation neural network algorithm GA-BP (Genetic Algorithms Back Propagation) is used to construct a transmission line tree obstacle growth analysis model. By optimizing the neural network weights and thresholds, the optimal solution is obtained, which greatly improves the accuracy of tree obstacle risk prediction. Through research, it has been found that in the field of tree obstacle intelligent analysis, similar studies currently mainly use single technical methods such as BP, CNN, ResNet, or VGG16, with an average recognition accuracy of 64%. The experimental results show that the method described in the article has an average recognition accuracy of 98%, which is 30% higher than conventional methods.

II. MATERIALS AND METHODS

A. Overall research approach

The article takes the risk prediction of transmission line tree faults as the research object, and proposes a research approach using data fusion, artificial intelligence, and digital twin technology, as shown in Figure 1.

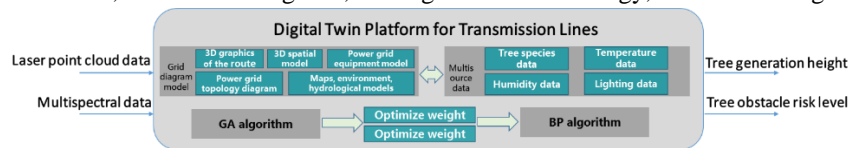


Fig.1 Overall research approach

Firstly, the high-precision 3D information contained in point cloud data is fused with multispectral data, and data processing is carried out. For point cloud data, feature variable extraction, single tree segmentation, point cloud classification, ground point separation, and point cloud denoising are the main tasks. Multispectral images mainly obtain spectral information of ground objects and expand and calculate spectral indices and texture information based on information obtained from different sources. After fusion processing of laser point cloud data and multispectral data, rich spectral and structural information is obtained [7].

Secondly, the GA algorithm is used to optimize the weights and thresholds of the BP neural network, and a tree obstacle risk prediction model is constructed to achieve the evaluation and prediction of tree obstacle height and risk level [8].

Finally, a digital twin platform for transmission lines is constructed, which integrates power system diagrams and data from multiple sources, and uses the GA-BP neural network algorithm to build observable and measurable capabilities for transmission lines and channels, supporting the power system equipment management department to achieve lean management of transmission lines[9].

B. Key technology research

1) Laser point cloud and multispectral data fusion

The overall design of laser point cloud and multispectral data fusion is divided into three stages: the data processing stage, the data fusion and modeling stage, and the result output stage, as shown in Figure 2.

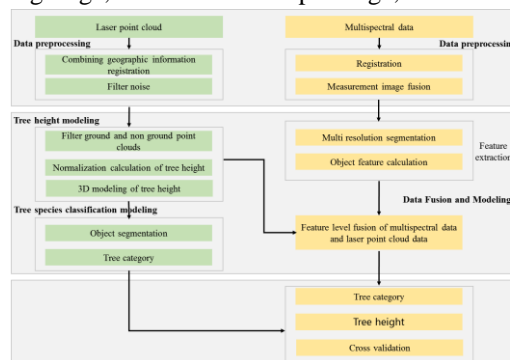


Fig.2 Roadmap of Data Fusion Technology

Data processing stage: Laser point cloud data includes discrete data with three-dimensional features within the transmission line channel, and points in the channel that are clearly not in the three-dimensional structure or have height anomalies need to be removed; multispectral data is registered based on unified coordinates and laser point cloud data to eliminate residual geometric inconsistencies and pixel misalignment; image fusion methods are used to fuse multispectral images with panchromatic images to improve the spatial resolution of multispectral images[10].

Data fusion and modelling phase: Generate ground points from laser point clouds and classify non-ground points according to object categories. Wire, insulator, drainage pipe and other parts are classified and identified by the point cloud data; after the classification is completed, the single tree segmentation function is used to perform single tree segmentation on the trees in the area, further detect and classify the scenes of roads, power lines and other ground buildings, and visualize them based on height and physical features; obtain the tree height of a single tree from its segmentation results, and stratify the forest point cloud based on the height information to extract trees at different levels in the region. Combining multispectral data to perform tree object feature calculation and clarify the image features of different tree species[11].

Output result: Compare multispectral data with point cloud data. Use software to analyze different data and compare whether control points coincide. If the degree of overlap is low, move or rotate the point cloud data to achieve fusion of multispectral data and laser point cloud data, complete cross-validation of different types of data, and output tree categories and tree heights [12].

2) GA-BP model optimization

The paper proposes to construct a tree hazard risk prediction model by optimizing the GA-BP neural network. Six factors (air temperature and humidity, light intensity, soil moisture, tree species, current tree height and line height) are selected to predict the growth height and risk level of various tree obstacles. The neural network model shown in Figure 3 is constructed.

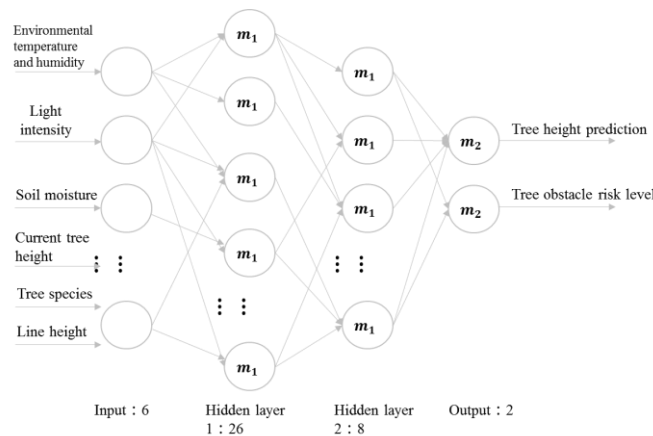


Fig.3 Schematic diagram of tree obstacle risk prediction model

The principle of the model is explained as follows:

a) The tree obstacle neural network includes an input layer, a hidden layer, and an output layer, and the BP model calculates two processes, including forward propagation of signals and backward propagation of errors[13].

b) The forward propagation process of the signal is as follows:

Assuming x_j is the input of the j-th node in the input layer. w_{ij} is the weight between the i-th node in the hidden layer and the j-th node in the input layer. θ_i is the threshold of the i-th node in the hidden layer. $\phi(x)$ is the excitation function of the hidden layer. w_{ki} is the weight between the k-th node in the output layer and the i-th node in the hidden layer, $i=1, \dots, q$. a_k is the threshold of the kth node in the output layer, $k=1, \dots, L$. $\psi(x)$ is the excitation function of the output layer. O_k is the output of the kth node in the output layer.

The net_i is the input of the i-th node in the hidden layer, and its calculation formula is as follows:

$$net_i = \sum_{j=1}^M w_{ij}x_j + \theta_i \tag{1}$$

The output of the i-th node in the hidden layer is y_i , and its calculation formula is as follows:

$$y_i = \phi(net_i) = \phi\left(\sum_{j=1}^M w_{ij}x_j + \theta_i\right) \tag{2}$$

The input of the k-th node in the output layer is net_k , and its calculation formula is as follows:

$$net_k = \sum_{i=1}^q w_{ki}y_i + a_k = \sum_{i=1}^q w_{ki}\phi\left(\sum_{j=1}^M w_{ij}x_j + \theta_i\right) + a_k \tag{3}$$

The output of the kth node in the output layer is o_k , and its calculation formula is as follows:

$$o_k = \psi(net_k) = \psi\left(\sum_{i=1}^q w_{ki}y_i + a_k\right) = \psi\left(\sum_{i=1}^q w_{ki}\phi\left(\sum_{j=1}^M w_{ij}x_j + \theta_i\right) + a_k\right) \tag{4}$$

c) Backpropagation of errors:

First, the output error of each layer of neurons is calculated layer by layer, starting from the output layer. Then, the weights and thresholds of each layer are adjusted using the error gradient descent method to ensure that the final output of the modified network is close to the expected value. The specific calculation method can be found in reference [14] and will not be repeated in this article.

The article proposes to use the GA algorithm to optimize the weights and thresholds of the BP neural network to obtain the optimal solution. By using different types of data to train the BP model, this tree barrier growth model can evaluate and predict the height and risk level of tree barriers [15]. Comparing the height of tree obstacles with the height of the line, the closer the distance, the higher the risk level. The specific pseudocode for implementing the GA-BP algorithm is shown below:

Tab.1 GA-BP algorithm pseudocode

Algorithm 1: GA-BP algorithm pseudocode

```
function GA-BP()
//1# Initialize BP neural network parameters, such as weights and thresholds
BP.initial();//BP neural network initialization
//2# Initialize genetic algorithm parameters, such as population size and iteration
times
Population.initial((population_size, chromosome_length):
//3# Generate initial population
Population.Random();//Randomly generate a population
//4# Evaluate the fitness of each individual in the population
Population.fitness();//Evaluate fitness
//5# Select operation
Population.selection(population, fitness_values, num_parents);
//6# Cross operation
Population.crossover(parents, population_size, chromosome);
//7# Mutation operation
Population.mutation(offspring, mutation_rate):
//8# Update BP neural network parameters
Population.update((population_size, chromosome_length):
end function
```

3) Construction of a digital twin platform for transmission lines

Through the construction of a digital twin platform, the monitoring business of transmission lines has been realized to be observable, measurable and predictable. As shown in Figure 4, the overall architecture is divided into a platform layer, a model layer and a physical layer.

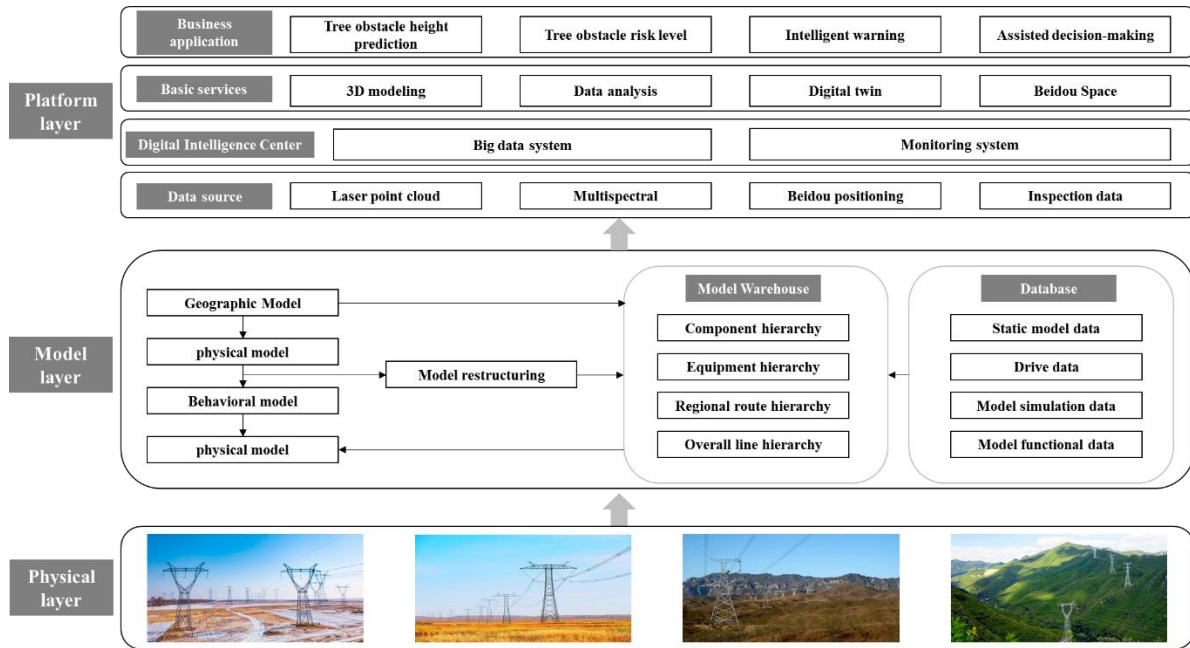


Fig.4 Architecture diagram of a digital twin platform for transmission lines

The platform layer consists of four main parts: business applications, digital twin-scene platform, digital intelligence center, and data sources. By connecting with data sources to extract data, it supports various business applications.

The model layer integrates the twin models into units by adding some spatial relationships and related constraints, so that the digital twin transmission line models are assembled and merged together, and multi-level transmission lines are synthesized according to the relevant attribute requirements of the model. A geometric model is obtained through path planning, measurement layout, geometric feature information, etc., and a physical model is constructed based on the material properties and related physical characteristics of the equipment. The behavioral dimensions of the model, as well as the coupling and path relationships of each component, are constructed [16].

The physical layer refers to the physical transmission lines that cover various voltage levels. The physical layer and model layer achieve virtual real mapping by integrating multi-source data such as environmental data, power grid data, graph modeling data, and IoT monitoring data. At present, research on dynamic modeling technology for massive data based on time-series databases is still in its early stages, but scenario-based modeling technology has matured and been applied [7].

III. EXPERIMENTAL SIMULATION AND DISCUSSIONS

A. Data preparation

Develop a tree obstacle growth risk prediction model, store the data in MySQL database, and conduct analysis on point cloud data, multispectral data, and sensor data modeling. Seven types of data, including laser point cloud, multispectral data, air temperature and humidity, light intensity, soil moisture, current tree height, and tree category, need to be collected. Taking the current mainstream LiDAR as an example, the laser point cloud data structure is shown in Table 2.

Tab.2 Laser point cloud data table

Data Item	Data value
Vertical direction	-15 ° to+15 °
Number of laser beams	16 lines
Data length per frame	1248 bytes
	Packet identification

	12 sets of data packets
Data composition of each frame	4-byte timestamp
	The last two bytes
	Radar model parameters

The multispectral data source is the actual data collected by a multispectral camera in a mountainous area of Hubei Province. The image size is 1280 * 960, with 6 bands and 2000 images per band, for a total of 10000 images. The collected multispectral data is shown in Figure 5. After collecting data on tree species, temperature, humidity, light intensity, etc., refer to Tables 3-7 for classification and grading.

Tab.3 Tree Species Classification Table

Tree species	Data value
Evergreen trees	001
Deciduous trees	002
Evergreen shrubs	003
Deciduous shrub	004

Tab.4 Temperature Classification Table

Temperature	Level
Below 0 °C	Severe cold
0°C-25°C	Low temperature
25°C-35°C	Suitable
Above 35 °C	High temperature

Tab.5 Humidity Classification Table

Humidity	Level
0-15%	Low humidity
15%-25%	Moderate humidity
25%-100%	High humidity

Tab.6 Light Intensity Grading Table

Light intensity	Level
0-100	Extremely weak light
100-500	Weak light
500-1000	General light
1000-2000	Medium light
2000-10000	Strong light

Tab.7 Risk Level Table

The height of tree barriers and the distance between transmission lines	Risk level and risk value
Less than 1 meter	A (Significant risk) 90-100
Less than 2 meter	B (High risk) 80-90
Less than 3 meter	C (Moderate risk) 70-80
Less than 4 meter	D (General risk) 60-70
Less than 5 meter	E (Low risk) Below 50

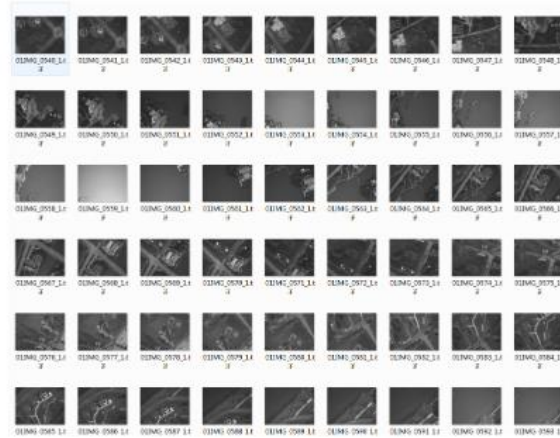


Fig.5 Multispectral raw data

B. Verification process

Using GA to optimize the weights and thresholds of BP can effectively prevent BP from falling into local optima, thereby improving the detection rate of BP. As the number of iterations of the GA algorithm increases, the fitness gradually converges globally to the optimal value, and the sliding process also gradually stabilizes, as shown in Figure 6.

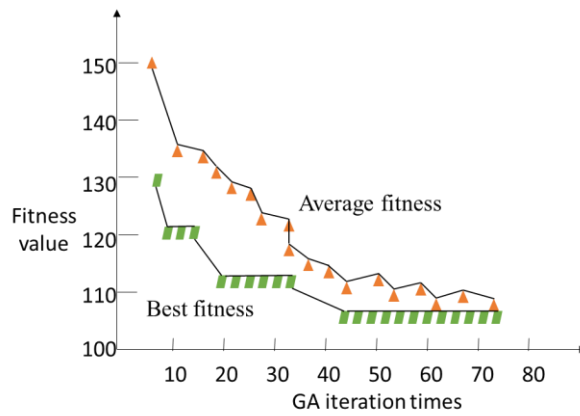


Fig.6 GA convergence result graph

To evaluate the performance of neural networks, the article uses the mean square error (MSE) as a fitness function [17]. The basic principle is to first calculate the difference between the predicted value and the actual value of the network, and then square and average it. The specific formula is as follows:

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \tag{5}$$

Among them, m is the number of samples, y_i is the true value of the sample, and \hat{y}_i is its predicted value. By solving for the minimum fitness value, the optimal weights and thresholds of the network can be obtained, thereby improving the prediction accuracy [18].

According to the comparison between the predicted results of 2564 validation data and the measured values, the regression coefficient tends to be close to 1 and the constant term tends to be close to 0, which proves that the regression effect is significant. The predicted values of tree height for different types of tree barriers tend to agree with the measured values, and the trend of change tends to be similar. After GA optimization of the parameters, the calculation accuracy of the BP neural network has been significantly improved, as shown in Figure 7.

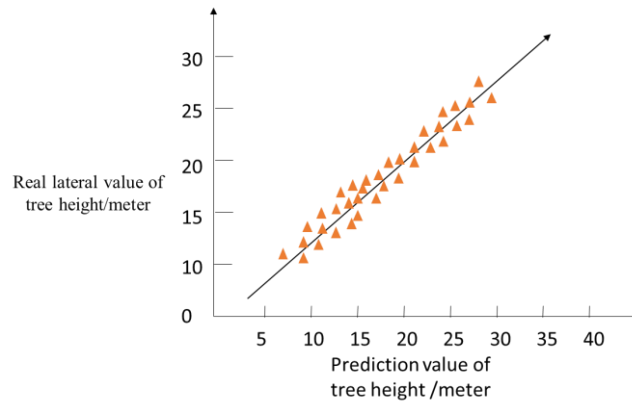


Fig.7 Predicted and measured average height of tree barriers

According to the actual measurements, the total proportion of "significant risk" and "high risk" categories for tree barriers is 30.62%, which is a key focus of the electricity grid company, as shown in Table 8.

Tab.8 Risk Level Distribution Table

Risk Level	Distribution of Tree Obstacle
	Risk Ratio
A (Significant risk)	11.24%
B (High risk)	19.38%
C (Moderate risk)	38.42%
D (General risk)	21.52%
E (Low risk)	10.33%

The prediction accuracy of the GA-BP model described in the article is evaluated by comparing it with conventional prediction models such as BP, CNN, ResNet, VGG16, etc. The article uses Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Accuracy as evaluation metrics [19], and their calculation formulas are as follows:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \tag{7}$$

$$Accuracy = \frac{T_{pn}}{T_{total}} \tag{8}$$

Explanation:

- MAE is the sum of the absolute differences between the target value and the predicted value, which represents the average error amplitude of the predicted value without considering the direction of the error.
- RMSE is the square root of the MSE, which eliminates the influence of stiffness.
- Accuracy is the percentage of correctly predicted results in the whole sample.
- In the above formula, m is the number of samples, y_i is the true value, \hat{y}_i is its predicted value, T_{pn} is the number of correctly predicted samples, and T_{total} is the total number of samples [11].

C. Experimental result

Following verification, the GA-BP model described in the article was found to exhibit a relatively small average absolute deviation and root-mean-square error in comparison to related prediction models. This indicates a significant improvement in the prediction accuracy of the GA-BP model, as evidenced in Table 9.

Tab.8 Risk Level Distribution Table

Prediction model	MAE	RMSE	Accuracy
GA-BP	0.187	0.243	0.985
BP	1.523	1.856	0.674

CNN	1.352	1.648	0.524
ResNet	0.985	0.937	0.775
VGG16	1.267	1.492	0.593

By integrating multi-source data such as laser point cloud data and multispectral data, a digital twin platform for transmission lines was constructed, and an optimized GA-BP model was used to predict the risk of tree obstacle growth. After verification and comparison with relevant models such as BP, CNN, ResNet, VGG16, etc., the prediction accuracy of the GA-BP model reached 98.5% (0.985), while the average accuracy of other models was 64.1% (0.641), which was about 30% higher overall. Through experimental data, the effectiveness of the GA-BP model in predicting the risk of tree obstacle growth has been strongly demonstrated.

IV. CONCLUSION

The hidden danger of transmission line tree faults is a typical problem in the operation and inspection of power grid companies. Conventional technical methods have problems such as insufficient generalization ability and low prediction accuracy. The article proposes a transmission line tree fault risk prediction method based on GA-BP neural network and digital twin. The transmission line digital twin platform constructed through the platform layer, model layer, and physical layer improves the monitoring and visualization capabilities of transmission lines and supports the development of transmission line operation and inspection businesses. A tree obstacle risk prediction model was constructed by optimizing the GA-BP neural network, which achieved the prediction and analysis of the growth height and risk level of various tree obstacles. The experimental results show that the GA-BP model described in the article has a prediction accuracy of 98.5%, which is superior to similar conventional methods. In the future, we plan to apply the research results to business scenarios such as sag, icing, and external damage prevention, providing lean and intelligent support for transmission line inspections.

ACKNOWLEDGMENT

This work is supported by the Southern Power Grid Technology Project (030100WS23220001).

REFERENCES

- [1] Jinqiu, Wu, et al. "Tree barrier prediction of power lines based on tree height growth model." *IOP Conference Series: Earth and Environmental Science* 645,no.1(2021):012008.
- [2] Wu, Shan, et al. "Research on risk prediction of tree growth in power transmission lines based on laser lidar." *2024 5th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)*,pp.524-527.IEEE, 2024.
- [3] Diamantopoulou, M. J., and R. Özçelik. "Evaluation of different modeling approaches for total tree-height estimation in Mediterranean Region of Turkey." *Forest Systems* 21.3 (2012): 383-397.
- [4] Özçelik, Ramazan, et al. "Estimating Crimean juniper tree height using nonlinear regression and artificial neural network models." *Forest ecology and management* 306 (2013): 52-60.
- [5] Long Tao, Qin LianHuan , and Ye ShaoMing. "Prediction for the growth of eucalyptus plantations with continuous-planting rotations based on BP neural network." (2012): 122-125.
- [6] Leslie, A. D., Maurizio Mencuccini, and Mike P. Perks. "Preliminary growth functions for Eucalyptus gunnii in the UK." *Biomass and Bioenergy* 108 (2018): 464-469.
- [7] Ghamisi, Pedram, et al. "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art." *IEEE Geoscience and Remote Sensing Magazine* 7,no.1(2019): 6-39.
- [8] Karatepe, Yasin, et al. "Total tree height predictions via parametric and artificial neural network modeling approaches." *iForest-Biogeosciences and Forestry* 15,no.2(2022): 95.
- [9] Chen, Xianda, et al. "The identification and ranging method of transmission line tree barrier based on 3D mapping of point cloud." *International Conference on Computer Vision, Application, and Algorithm (CVAA 2022)*,pp.53-57.SPIE,2023.
- [10] Ghamisi, Pedram, et al. "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art." *IEEE Geoscience and Remote Sensing Magazine* 7,no.1(2019):6-39.
- [11] Zhang, Jixian, and Xiangguo Lin. "Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing." *International Journal of Image and Data Fusion* 8,no.1(2017):1-31.
- [12] Wei M, Yang Z, Meng L, et al. Identification and prediction of transmission line tree obstacles based on multispectral and LiDAR data[C]//2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL),pp.305-309.IEEE, 2023.
- [13] Wu, Shan, et al. "Research on risk prediction of tree growth in power transmission lines based on laser lidar." *2024 5th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)*,pp.524-527.IEEE, 2024.
- [14] Zong, L. I., and Jingxing ZHENG. "Improved BP model and its application of earth dam monitoring data analysis." *Journal of Hydraulic Engineering In Chinese* (2003):111-114.
- [15] Li, Xiuting, et al. "Classification of Transmission Line Corridor Tree Species Based on Drone Data and Machine Learning." *Sustainability* 14,no.14(2022): 8273.

- [16] Liang, Haisheng, et al. "Application of digital twin in operation and maintenance of power transmission project." 2021 Power System and Green Energy Conference (PSGEC).pp.577-582.IEEE,2021.
- [17] K ksoy, Onur. "Multiresponse robust design: Mean square error (MSE) criterion." Applied Mathematics and Computation 175,no.2(2006): 1716-1729.
- [18] WANG, He, et al. "Magnetotelluric inversion based on BP neural network optimized by genetic algorithm." Chinese Journal of Geophysics 61,no.4(2018): 1563-1575.
- [19] Zhang, Lingling, et al. "Analysis of energy consumption prediction for office buildings based on GA-BP and BP algorithm." Case Studies in Thermal Engineering 50 (2023): 103445.
- [20] Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation." Peerj computer science 7 (2021): e623.
- [21] Hodson, Timothy O. "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not." Geoscientific Model Development Discussions 2022(2022): 1-10.
- [22] Jierula, Alipujang, et al. "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data." Applied Sciences 11, no.5(2021): 2314.