*Mechatronics and Automation Technology*
*J.-Y. Xu (Ed.)*

*doi:10.3233/ATDE231089*

# Condition Monitoring of IGBT Devices in Internet of Things Based on Random Forest Model

Lidan ZHOU[a], Gang YAO[a,1], Hongyu WANG[b], Jian LI[c], Siyang LIU[d], Yongjie NIE[d]

[a] *School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[b] *Wiscom System Co., LTD, Nanjing, China*
[c] *State Key Laboratory of Power Transmission Equipment & System Security and New Technology Chongqing University, Chongqing 400044, China*
[d] *Electric Power Research Institute, Yunnan Power Grid Co.Ltd., Kunming, Yunnan, 650217, China*

**Abstract.** Aiming at the problem of frequent aging faults of insulated gate bipolar transistor modules, this paper takes IGBT aging fault characteristic parameter data collected from the intelligent sensing layer of the Internet of Things terminal as samples. In the computing fusion cooperation layer, multi-terminal cooperation training and establishment of an optimized random forest model is applied to the IGBT aging fault diagnosis system to realize the condition monitoring of IGBT devices. Firstly, the characteristic parameters of aging fault are selected from the data samples and preprocessed to establish the aging fault diagnosis data set. Secondly, the traditional random forest model is built and optimized by parameter optimization of base evaluator, parameter optimization of model frame and bagging method. The optimization model training was completed on the basis of cross-validation. Finally, the prediction effect of the model in this paper and other models on IGBT aging fault diagnosis data set was evaluated by various evaluation indexes. Finally, the optimized model fit well, the error between the training set curve and the test set curve was 1.19%, and the prediction accuracy on the test set could reach 98.81%. The feasibility and accuracy of the optimized random forest model applied to IGBT condition monitoring system in the Internet of Things environment are verified.

**Keywords.** Iot intelligent sensing terminal; Random forest model; Insulated gate bipolar transistor; Condition monitoring; Multi-terminal collaboration.

## 1. Introduction

With the rapid development of wireless sensor network and embedded technology, Internet of Things (IOT) realizes the connection[1], information exchange, intelligent identification, positioning and supervision between objects and networks through information sensing equipment and communication media, and has been widely used in

---

[1] Corresponding Author, Gang YAO, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; Email: yaogangth@sjtu.edu.cn.

power system. The Internet of Things terminal uses sensors and other terminal sensing devices to collect the electrical signals of power electronic devices in the power system, and completes the transmission, processing and transformation of information through effective transmission methods, so as to realize the status monitoring of power electronic devices in the system, so as to ensure the stable and reliable operation of the power system.

The Insulated Gate Bipolar Transistor (IGBT) is characterized by high voltage tolerance and high power density. The bad operating environment and the repeated high voltage and high current in the practical application are superimposed on each other, resulting in frequent aging failure. It is an important means to improve system operation reliability and safety by using IOT technology to obtain sample data of IGBT device condition monitoring and life prediction strategy and to predict before complete aging failure occurs.

Random Forest[2], as a non-parametric classification model, is widely used in multi-domain model prediction and risk assessment. In this paper, combined with the fault monitoring data obtained by IOT technology, the traditional random forest model was optimized in collaboration with multiple terminals, and the optimized random forest model was applied to the IGBT device aging fault system to verify the prediction accuracy and fitting. The results show that in the Internet of Things environment, compared with the traditional random forest model and other models, the optimized random forest model proposed in this paper has higher prediction accuracy and better fitting.

## 2. Collect and Process Fault Characteristics

The intelligent perception terminal of the Internet of Things includes a four-layer architecture[3] of intelligent perception, computing fusion and collaboration, local application and communication integration, and interacts with the cloud application layer. Figure 1 shows the system diagram of the intelligent sensing terminal of the Internet of Things.

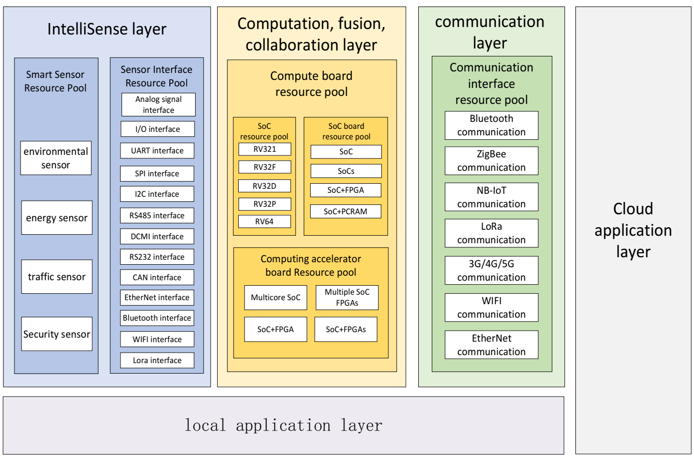



**Figure 1.** Iot intelligent sensing terminal system diagram

### *2.1. Intelligent Perception Layer*

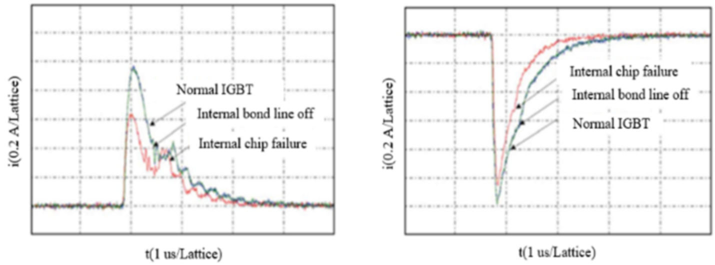
The intelligent sensing layer of the Internet of Things intelligent sensing terminal contains various types of intelligent sensors, intelligent sensor hardware interface and software interface, which is responsible for the collection and upload of various characteristic data of the monitored power electronic devices, providing data samples and sources for the IGBT device aging fault diagnosis strategy proposed in this paper. Through the collaborative work of multiple terminals, the subsequent model is established, optimized and applied in the collaborative layer of computing fusion.

Intelligent sensor hardware interface is the physical interface between intelligent sensing terminal and intelligent sensor, mainly for communication interface, including: UART interface, SPI interface, LoRa interface, etc. In order to be compatible with conventional sensors, intelligent sensor interfaces also include analog signal interfaces and I/O interfaces. In order to facilitate the call of the software system, the hardware interface of the intelligent sensor is programmed, and the corresponding software interface is formed. As the core device of the intelligent perception layer, the application type of intelligent sensor will be configured according to the environment and requirements of the intelligent perception terminal of the Internet of Things, and adopt a unified data format to upload data. Current, voltage and temperature sensors are installed in the intelligent sensing layer of the intelligent sensing terminal of the Internet of Things adopted in this paper to realize the monitoring of characteristic parameter data in the aging process of IGBT devices, and the data is transmitted to the computational fusion collaborative layer through hardware and software interfaces. The computational fusion collaborative layer allocates large tasks to parallel computing at multiple terminals. Complete input signal analysis and processing, machine learning modeling, etc. The computing fusion collaboration layer includes SoC chip, SoC board, computing acceleration board, memory chip and other hardware and collaborative software. Among them, SoC chip, SoC board and computing acceleration board are the core components of intelligent perception terminal computing in the Internet of Things, whose performance determines the computing capacity of the Internet of Things terminal. Finally, the original terminal, which performs task assignment by the computational fusion cooperation layer, collects the processing results of multi-terminal collaborative processing tasks and outputs the final status monitoring results of power electronic devices.

### *2.2. Selection of Fault Characteristic Parameters*

Due to the inconsistent expansion coefficients of different materials of IGBT devices, it is easy to cause bonding line shedding, solder layer cracking and other package failures or internal chip failures under the influence of junction temperature fluctuation[4-5]. The existing data show that the characteristic parameters of IGBT devices will gradually offset with the deepening of aging degree and performance decline. After the initial aging occurs and lasts for a period of time, the devices will completely age and fail. Therefore, it is necessary for fault diagnosis research to select suitable fault characteristic parameters and apply them to IGBT aging fault diagnosis system.

When different internal parameters of IGBT devices fail, the change of gate current when it is switched on and off is shown in figure 2. As can be seen from figure 2, the loss of the bond line inside the IGBT has little influence on the whole parasitic inductance and resistance of the IGBT module. However, when the chip aging failure occurs inside the IGBT, its parasitic parameters will change greatly, and the gate current opening and closing waveform will change obviously.




(a) IGBT gate current opening waveform under different fault states

(b) IGBT grid current turn-off waveform under different fault states

**Figure 2.** IGBT module gate current turn-on and turn-off waveform under different states

Considering the close relationship between various electrical parameters and the degree of device aging, the device package temperature T, collector current IC, collector voltage VC, grid current IG and grid voltage VG monitored by the intelligent sensing layer of the Internet of Things terminal during the aging process of IGBT devices are used as the characteristic parameters of the diagnosis model for the aging failure state of IGBT devices. Based on the above characteristic parameters, IGBT device aging fault diagnosis data set is established.

## 2.3. Data Preprocessing

According to the analysis of experimental data and the monitoring data change curve, the collected data were cut and defined as follows: The data between 140,000 and 146300 sampling points were taken as the characteristic parameters of IGBT health state, and the label was set as T0; The data between 146300 and 298000 sampling points were used as the characteristic parameters of the early aging stage of IGBT, and the label was set as T1. The data between 298000 sampling points and 301680 sampling points at the end of the experiment were taken as the characteristic parameters of IGBT aging fault state, and the label was set as T2. Part of the data set after standardization is shown in table 1.

**Table 1.** Partial Aging Fault Diagnosis Dataset

| Serial number | Collector current $I_C$/A | Collector voltage $V_C$/V | Grid current $I_G$/A | Grid voltage $V_G$/V | Package temperature $T$/°C | Aging status tag |
|---|---|---|---|---|---|---|
| 1 | 0.026671 | 0.986668 | 0.006957 | 0.03252 | 0.40146 | T0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 6300 | 0.807007 | 0.023287 | 0.885044 | 0.989315 | 0.058077 | T0 |
| 6301 | 0.807339 | 0.023199 | 0.878516 | 0.985618 | 0.057759 | T1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 158000 | 0.806842 | 0.005742 | 0.881586 | 0.984271 | 0.291336 | T1 |
| 158001 | 0.806345 | 0.005628 | 0.881222 | 0.987168 | 0.289749 | T2 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 161680 | 0.80651 | 0.163553 | 0.020059 | 0.03229 | 0.998413 | T2 |

## 3. Igbt Fault Diagnosis System

### 3.1. Establishment of Traditional Random Forest Model

In this paper, the decision tree based on classification tree algorithm is used as the base evaluator of the stochastic forest model to establish the traditional stochastic forest model. Assuming that data set D contains m categories of samples, $Gini$ coefficient[6] of data set D is defined as:

$$Gini(t) = \sum_{i=1}^{m} p(x_i \mid t)(1 - p(x_i \mid t)) \tag{1}$$

Among them, $p(x_i|t)$ said node t place some samples randomly selected data set and the probability that the sample that category for $x_i$; $Gini(t)$ is the probability that the categories of two samples randomly selected from the data set at node t are inconsistent. The decision tree establishment process is as follows:

1) Calculate $Gini$ coefficients of different categories in the data set, and take feature A with the smallest Gini coefficient as the root node of the decision tree;
2) Starting from the root node, data set D is divided into two sub-data sets according to feature A. Then, under the condition of feature A, Gini coefficient of data set D is defined as follows:

$$Gini(D, A) = \frac{D_1}{D} Gini(D_1) + \frac{D_2}{D} Gini(D_2) \tag{2}$$

3) Take the minimum $Gini(D, A)$ value as the optimal split node of the root node, and the root node continues to split downward until the conditions required for splitting at a node cannot be met or its Gini coefficient has reached the minimum value and cannot continue splitting, the decision tree stops growing, and the classification result at the leaf node is the final classification result of the decision tree.
4) After all decision trees in the model finish growing, the classification results of all decision trees in the random forest model are voted on, and the category with the largest number of votes in the classification results of decision trees is taken as the final output result of the model, as shown in Equation (3):

$$Y(x) = \underset{i=1,2,\cdots,z}{\arg\max} \sum_{n=1}^{N} \lambda\left(y_n(x) = i\right) \tag{3}$$

Where, $Y(x)$ is the output result of random forest model; $y_n(x)$ is the output result of the NTH decision tree in the forest. The expression in parentheses indicates that the final classification result of the decision tree is $i$. $\lambda(*)$ is the number of decision trees that satisfy the expression in parentheses; $z$ is the number of categories in the random forest model.

### 3.2. Optimization of Traditional Random Forest Model

The formula for calculating the upper bound of generalization error of random forest model is as follows:

$$err \le \frac{\overline{\rho}(1-s^2)}{s^2} \tag{4}$$

Therefore, optimization of random forest model from improve the prediction accuracy of a single decision tree s and reduce the average correlation coefficient between the decision tree $\bar{\rho}$ two aspects, on the basis of both model efficiency: first of all to parameter optimization of matrix estimator, improve the prediction accuracy of a single base assessment s efficiency and model; Secondly, frame parameters of the model are optimized to achieve a balance between model accuracy s and model efficiency. Finally, using the method of bags to reduce the average correlation coefficient between the estimator rho¯, further enhance the model prediction accuracy.

#### 3.2.1. Base Evaluator Parameter Optimization

The "pre-pruning" method limits the complete growth of the decision tree by setting the generation parameters of the decision tree in the tree building process[7-8], so that the complexity of the random forest model can be effectively controlled. On the basis of "pre-pruning", this paper adopts the method of grid search to optimize the parameters of the maximum depth, minimum number of branch nodes and minimum number of branch samples during the establishment of the decision tree. The value range of the depth of the decision tree is set as [1, 50], and the step size is 1. The value range of the minimum number of branch nodes is [2, 25], and the step size is 1. The range of minimum branch sample number is [2, 25], and the step size is 1. Grid search results of prediction accuracy and running time of a single decision tree are output, as shown in figure 3. In order to ensure both prediction accuracy and efficiency, the difference between value 1 and prediction accuracy is taken, and the value of decision tree running time is superimposed to output grid search results. As shown in figure 4, the 3D surface graph formed by the normalization of the numerical lowest point (optimal parameter) of all the section data in figure 4 shows a trend of first decreasing and then increasing. The numerical lowest point is obtained at the lowest point of the surface depression, that is, when the maximum depth of the decision tree is 7, the minimum number of branch nodes is 15, and the minimum number of branch samples is 3, the surface reaches the lowest point. That is, the model has the highest prediction accuracy and the shortest time.

#### 3.2.2. Model Frame Parameter Optimization

During the modeling process, the value of n_estimators defined the number of base estimators and determined the complexity of the stochastic forest model. After the optimization of base estimators parameters is completed, all base estimators in the model are set as the optimal parameters, the value range of n_estimators is set as [1, 200], and the step is set as 1. When the number of base estimators increases, the learning curve of forecasting accuracy of the output model is shown in figure 5. It can be concluded that: When the parameters of base estimators have been adjusted to the optimal level, we can see: When the n_estimators value is 24, the stochastic forest model has the best performance.

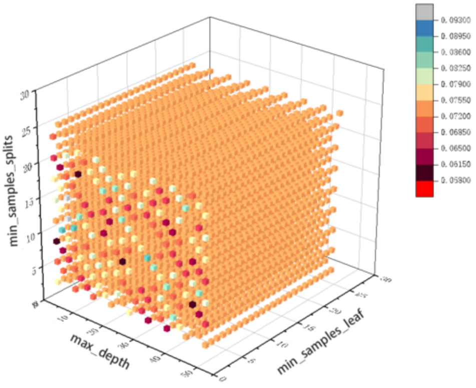



**Figure 3.** Base Evaluator Runtime Grid Search Results

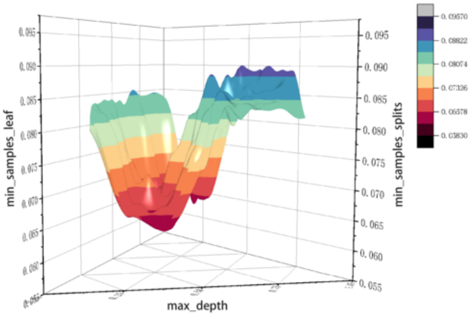



**Figure 4.** Base Evaluator for Comprehensive Evaluation of 3D Surface Maps

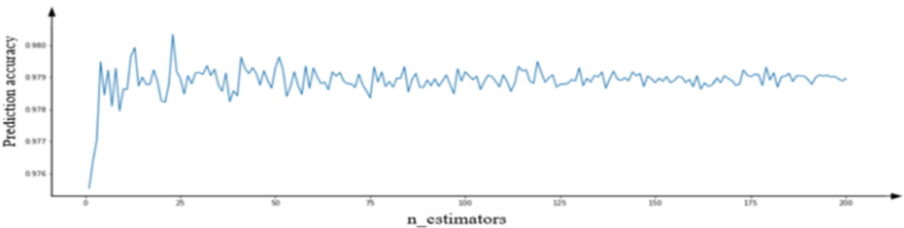



**Figure 5.** Number of base estimators optimization curve

### 3.2.3. Bagging

In essence, the bagging method[9] re-samples the random samples that have been returned to the training set to form multiple new data sets with similar scale but different from the original training set. Due to the randomness and independence of the resampling of the training sample set, the difference of several base evaluators formed on this basis increases, which significantly reduces the correlation between any two base evaluators.

## 4. Experimental Analysis

In order to verify the prediction accuracy of the optimized random forest model in the IGBT device fault diagnosis model, this paper verifies the IGBT aging fault diagnosis data collected by the intelligent sensing terminal of the Internet of Things, and compares and analyzes the prediction accuracy and fitting effect of different models.

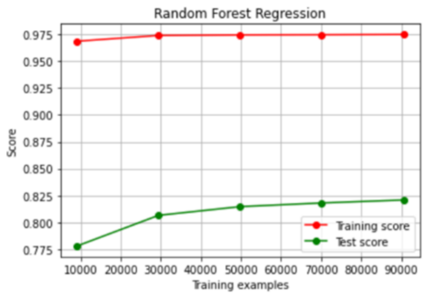
### *4.1. Cross-validation of Models*

In the process of model training, the K-fold cross-validation method[10] is adopted to divide the eigenmatrix into K equal parts, each part is taken as the test set in turn, and the remaining K-1 part is taken as the training set. After K training, the mean value of the obtained test set results is output as the final model prediction result.

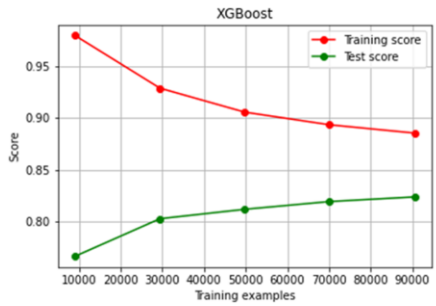
### *4.2. Model Evaluation Index*

The learning curve of the optimized random forest model and other models after multiple cross-validation training is shown in figure 6. Where, the red line and the blue line are the variation trend of the model's prediction accuracy in the training set and the test set respectively. The vertical axis is the prediction accuracy value, and the horizontal axis is the number of samples in the training set. As can be seen from Figure 6, the prediction accuracy of the random forest regression model is higher in the training set but lower in the test set, which is manifested as overfitting. The prediction accuracy of XGBoost model is poor in both the training set and the test set. The traditional random forest model has a general fitting degree, but the prediction accuracy of the training set and the test set is lower than that of the optimal random forest model. The optimized random forest model has high prediction accuracy on the training set and the test set, and can achieve complete fitting on the training set. The difference between the two prediction curves is only 1.19%.

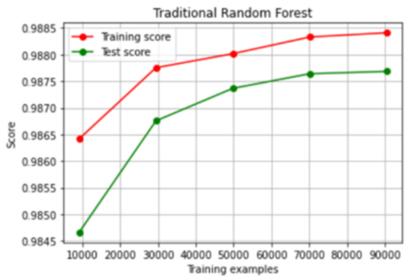
Finally, the mean square error (MSE), mean absolute error (MAE), model decision coefficient ($R^2$), and the mean value of the model's prediction accuracy on the training set and the test set under multiple training are taken as evaluation indexes, and the performance of the model established in this paper and other models in IGBT fault diagnosis system is compared and analyzed. The evaluation data of the output model are shown in table 2, where: the smaller the value of MSE and MAE, the larger the value of R2, and the higher the correlation between the model prediction accuracy and the test data. As can be seen from the data in table 2, XGBoost model has the worst prediction effect; The optimized random forest model has the smallest mean square error and the smallest mean absolute error, the highest prediction accuracy and the strongest correlation with the test data.



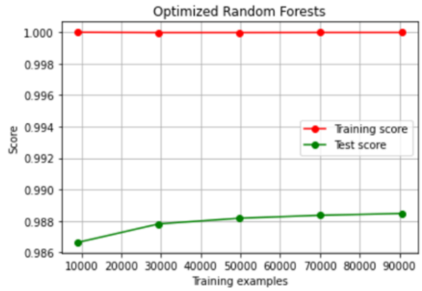

(a)Random forest regression model




(b)XG Boost model




(c)Tradition random forest model




(d)Optimize the random forest model

**Figure 6.** Diagram of learning curve of each model

**Table 2.** Model Evaluation Indicator Results

| Fault diagnosis model | MSE | MAE | $R^2$ | Prediction accuracy | |
|---|---|---|---|---|---|
| | | | | Training set | Test set |
| Random forest regression model | 0.0113 | 0.0224 | 0.8191 | 0.9752 | 0.8154 |
| XGboost model | 0.1063 | 0.0233 | 0.817 | 0.8815 | 0.8146 |
| Traditional random forest model | 0.0127 | 0.0121 | 0.7970 | 0.9884 | 0.9772 |
| Optimize the random forest model | 0.0119 | 0.0119 | 0.8095 | 1.0 | 0.9881 |

## 5. Conclusion

In this paper, the IGBT aging fault characteristic parameter data collected by the intelligent sensing terminal of the Internet of Things is used to establish the IGBT aging fault diagnosis data set by the collaboration of multiple terminals, and the optimized random forest model is applied to the IGBT aging fault diagnosis system. The results show that: In the Internet of Things environment, the optimized random forest model achieved the highest prediction accuracy on both the training set and the test set of IGBT aging fault diagnosis data. Compared with XGboost model, random forest regression model and traditional random forest classification model under the same conditions, the prediction accuracy of the test set increased by 17.35%, 17.27% and 1.09%, respectively. From the error distribution of multiple tests, the mean square error and mean absolute error of the optimized random forest model are low and stable, and have the highest correlation with the data set. The IGBT device aging fault diagnosis system based on the Internet of Things environment has high prediction accuracy, good fitting degree, and good practical value and application prospect.

## Acknowledgments

This work was supported by the National Key Research and Development Program "Intelligent Sensing Terminal Platform System and Application Verification of Internet of Things" (2018YFB2100100).

## References

[1] G. Thippa Reddy, et al. A deep neural networks based model for uninterrupted marine environment monitoring[J]. Computer Communications, 2020, (157): 64-75.
[2] Dong L J, LI X B, Peng K. Prediction of rockburst classification using Random Forest[J]. Transactions of Nonferrous Metals Society of China, 2013, 23(2): 472-477.
[3] Zhou L D, Li J, Yao G, et al. New Intelligent Sensing Terminal System Architecture of Internet of Things[C]// Internet of Things, Communication and Intelligent Technology Conference. August 21-23, 2022, Changsha, China.
[4] Oh H., Han B. , McCluskey P., Han C., and Youn B. D. "Physics-of-failure, condition monitoring, and prognostics of insulated gate bipolar transistor modules: A review," IEEE Trans. Power Electron., May 2015,30(5): 2413–2426, .
[5] Yang S., Xiang D., Bryant A., Mawby P., Ran L. and Tavner P. "Condition monitoring for device reliability in power electronic converters: A review", IEEE Trans. Power Electron., Nov. 2010, 25( 11) :2734–2752, .
[6] Yutaka Iwakami, Hironori Takuma, Motoi Iwashita. Properly initialized Bayesian Network for decision making leveraging random forest[J]. Artificial Intelligence Research, 2020, 9(1): 36-44.
[7] Vapnik V. The nature of statistical learning theory [J]. IEEE Transactions on Neural Networks, 1997, 8(16): 22-31.
[8] Doreen Y, Chee S, Ahmad I. Improved Boosting Algorithms by Pre-Pruning and Associative Rule Mining on Decision Trees for Predicting Obstructive Sleep Apnea[J]. Advanced Science Letters, 2017, 23(11): 11593-11598.
[9] Xu W, Tang J, Xia H, et al. Soft measurement of dioxin emission concentration based on Bagging semi-supervised deep forest regression [J]. Chinese Journal of Scientific Instrument, 2022, 43(6): 251-259.
[10] Sun Z X, Sun H X, Zhang J X.Multistep Wind Speed and Wind Power Prediction Based on a Predictive Deep Belief Network and an Optimized Random Forest[J]. Mathematical Problems in Engineering: Theory, Methods and Applications, 2018, 7: 1-15.