

# Transformer graded fault diagnosis based on neighborhood rough set and XGBoost

RuYan Guo<sup>1,\*</sup>, MinFang Peng<sup>1</sup>, ZhenQi Cao<sup>2</sup> and RunFu Zhou<sup>3</sup>

<sup>1,2,3</sup>Hunan University, College of Electrical and Information Engineering, 410082 Changsha, China

**Abstract.** Aiming at the uncertainty of fault type reasoning based on fault data in transformer fault diagnosis model, this paper proposed a hierarchical diagnosis model based on neighborhood rough set and XGBoost. The model used arctangent transformation to preprocess the DGA data, which could reduce the distribution span of data features and the complexity of model training. Using 5 characteristic gases and 16 gas ratios as the input characteristic parameters of the XGBoost model at all levels, reduction was performed on these 21 input feature attributes, features that had a high contribution to fault classification were retained, and redundant features were removed to improve the accuracy and efficiency of model prediction. Taking advantage of XGBoost's strong ability to extract a few features, the output of the model was the superposition of leaf node scores for each type of fault, the maximum score was the type of failure the sample belonged to, and its value was also the probability value. The obtained probability was used as one of the evidence sources to use D-S evidence theory for information fusion to verify the reliability of the model. Experiments have proved that the XGBoost graded diagnosis model proposed in this article has the highest overall accuracy rate comparing with the traditional model, reaching 93.01%, the accuracy of XGBoost models at all levels has reached more than 90%, the average accuracy rate is higher than that of the traditional model by an average of more than 2.7%, and the average time-consuming is only 0.0695 s. After D-S multi-source information fusion, the reliability of the prediction results of the model proposed in this paper has been improved.

## 1 Introduction

Transformers are important equipment for transforming and transmitting electrical energy. Aiming at addressing the inaccuracy of the existing transformer fault diagnosis knowledge and in order to ensure the normal operation of the transformer, scholars have done research on transformer faults in multiple directions.

Reference [1] introduces a correction factor to the nearest neighbor component analysis algorithm, and maps the K nearest neighbors in combination with the training metric matrix, thereby improving the classification performance of the K nearest neighbor algorithm on unbalanced data sets; reference [2] inputs the DC transient excitation into the transformer winding, and takes the oscillating wave response at the end of the winding as the analysis object, and proposes a winding fault diagnosis technology; reference [3] quantifies the change characteristics of condition monitoring data over time, calculates the control limit of  $T^2$  and Q statistics, and determines the samples that exceed the control limit as fault samples, thus proposes unsupervised concept drift recognition and dynamic graph embedded transformer fault detection method; reference [4] uses deep belief network for unsupervised training, extracts features from DGA data and combines D-S evidence theory to solve the uncertainty problem of transformer fault diagnosis. However,

reference [1] requires Bayesian algorithm for hyperparameter tuning, and the overall structure of the model is complex and consumes lots of computing resources. reference [2] requires high signal-to-noise ratio of the excitation source and in high frequency bands, the sensitivity of the oscillatory wave method is low. Reference [3] only studies the boundary point between the fault data set and the normal data set, and does not involve the identification of specific fault types. Reference [4] uses deep belief network pre-training and parameter tuning for a long time.

In order to find early faults inside the transformer, combining with the characteristics of real-time, online, no electricity, and magnetic field interference based on DGA diagnosis<sup>[5]</sup>, This paper proposes XGBoost's multi-level transformer fault diagnosis based on neighborhood rough set. XGBoost<sup>[6]</sup> is an extreme gradient boosting algorithm that forms a strong classification model by integrating multiple CART<sup>[7]</sup> trees. This paper consequently uses D-S evidence theory<sup>[8]</sup> for information fusion to solve the uncertainty<sup>[9]</sup> and imprecision of diagnosis knowledge and methods, thereby improving the reliability of model diagnosis.

## 2 Related theories and concepts

\* Corresponding author: [609246209@qq.com](mailto:609246209@qq.com)

### 2.1 The concept of neighborhood rough set

Given neighborhood decision system  $NDS = \langle U, A \cup D \rangle$ ,  $B \subseteq A$ , if the condition attribute subset  $B$  satisfies the following conditions:

- a)  $\gamma_B(D) = \gamma_A(D)$ , which is  $Pos_B(D) = Pos_A(D)$ , conditional attribute subsets  $B$  and  $A$  have the same classification capabilities.
- b)  $\forall a \in B, \gamma_B(D) > \gamma_{B-\{a\}}(D)$ , that is, there are no extra attributes in condition subset  $B$ .

then  $B$  is a conditional reduction set of  $A$ .

### 2.2 Introduction to XGBoost

XGBoost uses the CART tree as the basic model and is developed from the gradient boosting decision tree. The objective function is:

$$Obj^n = \sum_{j=1}^T \left[ \omega_j G_j + \frac{1}{2} \omega_j^2 (H_j + \lambda) \right] + \gamma T \quad (1)$$

$\omega_j$  is the weight of leaf node  $j$ ,  $G_j = \sum_{i \in I_j} g_i$ ,  $H_j = \sum_{i \in I_j} h_i$ ,  $g_i$  and  $h_i$  are respectively the first and second derivatives of  $\hat{y}_i^{(t-1)}$  on the loss function  $l(y_i^t, \hat{y}_i^{(t-1)})$ ,  $I_j$  is the set of samples contained on the leaf with index  $j$ . When each leaf  $j$  takes the value of the quadratic function to take the minimum value, the objective function takes the minimum value, so that the derivative of  $\omega_j$  is obtained on the quadratic function, and the derivative is 0 for the extreme value:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (2)$$

The structure score is:

$$Obj^n = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (3)$$

To split a leaf node, the gain after splitted is:

$$Gain = \frac{1}{2} \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (4)$$

Perform gain calculation on all branch points of all features, and select the node with the largest gain, that is, the node with the fastest decline in the objective function for branching, when the gain is lower than the setted threshold, stop the growth of the tree and find the best tree structure.

### 2.3 D-S evidence theory

1) Confidence function:

If the proposition  $A$  is a subset of  $\Theta$ , the sum of the probability distributions of all the subsets in  $A$  is the confidence function of  $A$ ,

$$B_{el}(A) = \sum_{B \subseteq A} m(B) \quad (5)$$

(5) shows the minimum degree of trust in proposition  $A$ .

2) Likelihood function:

$$P_i(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (6)$$

(6) indicates the maximum confidence in proposition  $A$ . The confidence interval is  $[B_{el}(A), P_i(A)]$ , which indicates the degree of confidence in a certain proposition.

3) Synthesis rule:

Assuming that there are  $n$  information sources, the mass function value fused on  $\Theta$  can be obtained through the synthesis rule.  $\forall A \subseteq \Theta$ , the synthesis rule of  $n$  mass functions  $m_1, m_2, \dots, m_n$  on  $\Theta$  is:

$$(m_1 \oplus m_2 \dots \oplus m_n)(A) = \frac{1}{K} \sum_{A_1 \cap A_2 \cap \dots \cap A_n = A} \prod_{1 \leq i \leq n} m_i(A_i) \quad (7)$$

where  $A_i \subseteq \Theta$ ,  $K$  represent the conflict factor.

$$K = 1 - \sum_{A_1 \cap A_2 \cap \dots \cap A_n = \emptyset} \prod_{1 \leq i \leq n} m_i(A_i) \quad (8)$$

$K \in [0, 1]$ , the larger the  $K$ , the more intense the conflict between the evidence.

## 3 Input attribute reduction and model establishment

### 3.1 Choice of Input feature vector and fault type

The first level (XGBoost1) diagnoses normal or fault, the second level diagnosis (XGBoost2) diagnoses overheating(H), complex(M) or discharge(C), and the third level includes 3 models of XGBoost3-XGBoost5, which respectively diagnose 9 types of faults proposed by the uncoded ratio method. The volume fractions of five characteristic gases are selected as the reduction object of the first-level model. And take the characteristic gas ratio in Table 1 as the reduction object of the second-level and third-level models, where TH represents total hydrocarbons,  $D = CH_4 + C_2H_2 + C_2H_4$ , select 9 types of faults proposed by the non-code ratio method as output: low energy discharge and overheating (MF1), high energy discharge and overheating (MF2), partial discharge (PD), low energy discharge (D1), high energy discharge (D2), low temperature overheating (T1), medium temperature overheating (T2), high temperature overheating (T3), normal (N).

Table 1. Characteristic gas ratio.

Input	Ratio	Input	Ratio
$x_1$	$CH_4 / H_2$	$x_9$	$(CH_4 + C_2H_4) / TH$
$x_2$	$C_2H_2 / C_2H_4$	$x_{10}$	$C_2H_2 / H_2$
$x_3$	$C_2H_4 / C_2H_6$	$x_{11}$	$C_2H_2 / C_2H_6$
$x_4$	$H_2 / (TH + H_2)$	$x_{12}$	$C_2H_2 / CH_4$
$x_5$	$CH_4 / TH$	$x_{13}$	$C_2H_6 / CH_4$
$x_6$	$C_2H_6 / TH$	$x_{14}$	$CH_4 / D$
$x_7$	$C_2H_4 / TH$	$x_{15}$	$C_2H_2 / D$
$x_8$	$C_2H_2 / TH$	$x_{16}$	$C_2H_4 / D$

### 3.2 Data preprocessing

This article uses 2093 pieces of data from a 330kV oil-immersed transformer as samples, 1079 pieces of data as the test set and 1014 pieces as inspection data. In order to avoid the highly skewed distribution of DGA data and the large gap between the characteristic gas ratios, the paper uses arctangent transformation processing on 5 characteristic gases and 16 characteristic gas ratios, in order to reduce the impact of initialization, normalize the data again:

$$p_i^* = \frac{p_i - e(p_i)}{\sqrt{\sigma(p_i)}} \quad (9)$$

$p_i^*$  is the normalized feature vector,  $e(p_i)$  is the expectation of feature vector  $p_i$ , and  $\sqrt{\sigma(p_i)}$  is the standard deviation of feature vector  $p_i$ .

### 3.3 Attribute reduction

In order to enable the model to identify the mapping between DGA data and fault types more efficiently and accurately, Using 1079 test set as the reduction object, this paper applies neighborhood rough set to reduce attributes, keep important attributes, and eliminate redundant attributes. It adopts forward greedy algorithm and does not need to discretize DGA data, thus maintaining the integrity of feature data, the lower limit of the importance of this article is 0.001, so that the reduction is accurate.

### 3.4 The definition of probability distribution function in D-S evidence theory

This paper uses the output of the XGBoost model as one of the evidences, and its basic probability value is calculated as follows:

$$m(R_i) = y(R_i) / \left( \sum_{i=1}^{N_1} y(R_i) + E_n \right) \quad (10)$$

$y(R_i)$  is the output probability value of the XGBoost model,  $E_n$  is the error,

$$E_n = \frac{1}{2} \sum (t_j - y_j)^2 \quad (11)$$

$t_j$  is the actual one hot label,  $y_j$  is the output probability value of the XGBoost model,  $R_i (i = 1, 2, \dots, N_1)$  is the fault category;  $N_1$  is the total number of fault categories. Other evidence uses statistical methods to determine its basic probability value.

### 3.5 The algorithm flow of XGBoost

- 1) Divide 1014 inspection data into 811 training set and 203 test set according to 4:1, and use the reduced attributes as input to complete the training of the XGBoost model, and use 203 test set data for prediction to verify the performance of the XGBoost model.
- 2) Input the data to be diagnosed into the trained XGBoost model according to the reduced attributes to obtain the output probability value.

3) According to equations (10) and (11), calculate the basic probability value of the network output, and use statistical analysis to obtain the basic probability value of other evidence.

4) Use equations (5)—(8) to calculate the confidence function, likelihood function, and conflict factor  $K$  of each focal element.

5) Obtain the confidence interval  $(B_{le}, P_1)$ .

6) According to the fault type, select the next-level XGBoost diagnostic model, and follow steps 2) —5) to further find the cause of the fault.

7) Get the final conclusion.

## 4 The evaluation of XGBoost performance

### 4.1 The contrast of before and after actangent p-processing

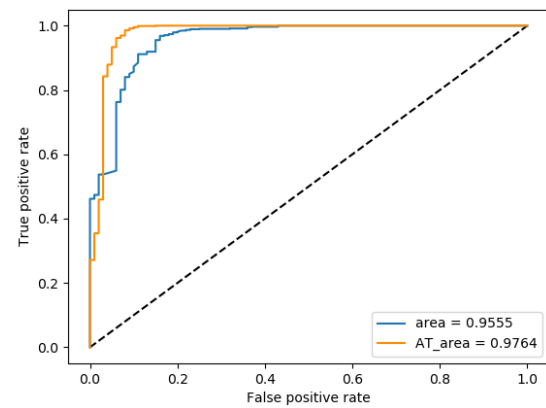


Figure.1 ROC curve of level I diagnostic layer

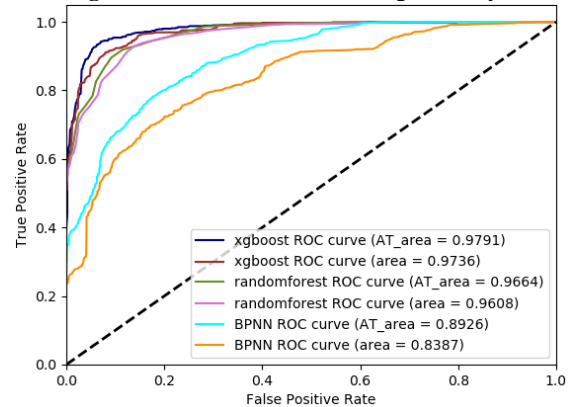


Figure.2 ROC curve of level II diagnostic layer

Take the I and II diagnostic models as an example, the ROC<sup>[10]</sup> curves are shown in Figures 1 and 2 respectively. The first level only uses the XGBoost model. It can be seen from Figure 1 that after arctangent processing, the AUC area of the level I diagnostic model has increased by 0.0209 to 0.9764. For the same reason, Figure 2 shows that the AUC area of the three models XGboost, RandomForest, and BPNN after arctangent processing increased by 0.0217 on average, and the AUC area stratification occurred: the AUC area of XGboost is larger than RandomForest, and the AUC area of RandomForest is larger than BPNN. It shows that through arctangent transformation, the input data is mapped to  $[0, \pi/2]$ , and

then normalized processing can effectively solve the long-tailed<sup>[11]</sup> distribution problem of DGA data, so that the model can better express the nonlinear relationship between input and output .

#### 4.2 Comparison of accuracy after feature reduction

Supposing five characteristic gases are input characteristic parameters  $S_1$ , and the ratio of 9 characteristic gases of the non-code ratio method is input characteristic parameters  $S_2$ , five characteristic gases and the ratios of 16 characteristic gases are input characteristic parameters  $S_3$ , The input feature parameter after reducing  $S_3$  by the neighborhood rough set theory is  $S_4$ , Input the above four characteristic parameters into the XGBoost graded diagnosis model, and get Tables 2 and 3.

**Table 2.** Diagnosis accuracy rates of different input feature parameters at all levels(%).

Diagnostic Layer	$S_1$	$S_2$	$S_3$	$S_4$
XGBoost1	90.14	90.96	93.86	95.40
XGBoost2	87.95	89.96	90.99	92.62
XGBoost3	86.67	87.53	91.38	92.02
XGBoost4	85.04	85.41	89.29	91.36
XGBoost5	86.91	90.34	89.34	90.95

Observing Table 2, we can see that the diagnostic accuracy of level I of the four input feature parameters is the highest compared to level II and III, all of which are above 90%, and the highest is 95.40%. This is because the I-level XGBoost model can easily grasp the difference between the characteristics of normal and fault data through learning. The accuracy of the four feature quantities as the input of the II and III models has decreased, indicating that when the fault type needs to be specifically subdivided, the generalization ability of the model is more demanding. In level II diagnosis, due to the large sample size, the generalization ability of the model is effectively improved through learning, so the accuracy rate is high. The average accuracy of level II is 90.38%, and the highest is 92.62% for  $S_4$ .

Since the five characteristic gases are directly used to distinguish fault types from DGA data, they contain more redundant information, so the level III diagnosis accuracy of the five characteristic gases is low, with an average accuracy rate of only 86.21%. The non-code ratio method is used as the input feature parameter. Since the ratio selection is only 9 kinds, it is not comprehensive and detailed, and it is easy to cause uncertainty in reasoning. Therefore, the average accuracy rate of the third level of the non-code ratio method is only 87.76%.

When the five characteristic gases and the ratios of 16 characteristic gases are used as input parameters, because the included fault characteristic information data is relatively complete and accurate, the average accuracy rate of level III reaches 90%. When the feature parameters after attribute reduction are used as input, the diagnostic model of each level has selected features, so the accuracy of each level is the highest, and the average accuracy of level III is 91.44%. Through the comparison

of accuracy rate, it can be obtained that the predictive ability of XGBoost diagnostic models at all levels is the best after being reduced by the neighborhood rough set theory.

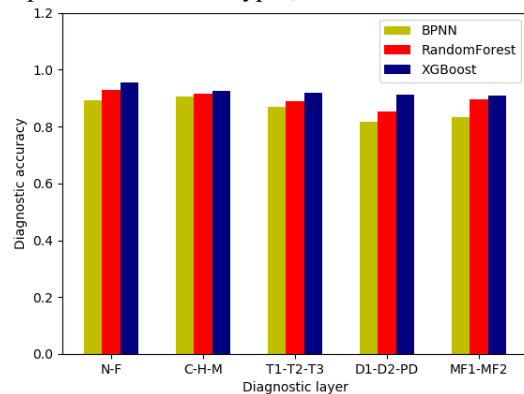
**Table 3.** Time-consuming diagnosis of different input characteristic parameters at all levels(s).

Diagnostic Layer	$S_1$	$S_2$	$S_3$	$S_4$
XGBoost1	0.0279	0.1165	0.0654	0.0569
XGBoost2	0.0863	0.0469	0.1178	0.1374
XGBoost3	0.0754	0.0359	0.0517	0.0489
XGBoost4	0.0614	0.0437	0.0569	0.0455
XGBoost5	0.0864	0.0325	0.0629	0.0588

Calculate the average time for each of the five models of the four characteristic parameters. The longest average time is the five characteristic gases and the ratios of the 16 characteristic gases ( $S_3$ ), the value is 0.07094s. The shortest average time is non-code ratio method( $S_2$ ), which has a value of 0.0551s, and the average time it takes to input the features to all levels of models after attribute reduction( $S_4$ ) is 0.0695s, which is shorter than  $S_3$  and meets actual requirements.

#### 4.3 Comparison of different diagnostic methods

Taking the hierarchical diagnosis proposed in this paper as a model, using the reduced feature parameters in as the input of BPNN, RandomForest and XGBoost, respectively, the accuracy comparison histogram of each classifier under different methods is obtained. In the Figure 3, N-F means normal and fault, C-H-M means discharge, overheating and complex faults. For the description of other fault types, see section 3.1 .



**Figure.3** Comparison of accuracy rates between various classifiers

Figure 3 shows that the XGBoost model has the highest accuracy rate compared to BPNN and RandomForest at the same level of diagnosis, which is 0.06046 higher than BPNN on average and 0.02774 higher than RandomForest on average. From the perspective of the AUC area under the ROC curve, the XGBoost model has the largest AUC area with a value of 0.9791, which is 0.0865 and 0.0127 higher than BPNN and RandomForest, respectively. It can be seen that the XGBoost model proposed in this paper has the best effect among the several classifiers. At the same time, it can be seen from Figure 3 that when the fault types are subdivided level by level, the accuracy rate decreases. This is due to the limited sample size,

unbalanced data, and the generalization ability of the model is limited.

The sample labels are set to the nine fault types proposed by the non-code ratio method, and the nine characteristic gas ratios proposed by the non-code ratio method are used as the input of BPNN and RandomForest, which do not use a graded model, and compared with the prediction accuracy of the XGBoost graded diagnosis model proposed in this paper. Table 4 can be obtained.

**Table 4.** Fault diagnosis accuracy rate of different methods(%).

Diagnosis Method	Non-code Ratio	BPNN	Random Forest	XGBoost Graded
Accuracy	88.46	90.14	91.91	93.01

Table 4 indicates that due to the limited fault feature information contained in the non-code ratio method, the accuracy rate is only 88.46%, BPNN is easy to fall into the local optimum, and the accuracy rate is only 90.14%. XGBoost is more suitable for processing data with a few features than RandomForest, and considering the second derivative, adding regular term coefficients to the loss function, and pre-pruning the decision tree, its accuracy is 1.1% higher than RandomForest, reaching 93.01%.

## 5 Case of study

Taking the DGA data in Table 5 as an example, the fault type is high-energy discharge. This article uses D-S evidence theory to conduct information fusion<sup>[12]</sup> analysis.

**Table 5.** Transformer DGA data(ul/L).

H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>
164.56	36.65	9.88	85.97	193.83

Input the reduced characteristic parameters into the hierarchical XGBoost diagnosis model in turn. The actual output of the I-level XGBoost1 model is [0.0221, 0.9779], and the expected output is [0, 1]. According to equations (10) and (11), the basic probability distribution of evidence source  $e_1$  is [0.0220, 0.9774]. The analysis of transformer fault statistics data shows that the basic probability distribution of another source of evidence  $e_2$  is [0.0986,0.9000], then Table 6 can be obtained:

**Table 6.** Probability distribution of level I diagnostic model.

Evidence	Basic Probability		Uncertainty
	$N$	$F$	
$e_1$	0.0220	0.9774	0.0006
$e_2$	0.0986	0.9000	0.0014

According to formulas(5)—(8), carry out evidence fusion and calculate the conflict factor  $K$  to be 0.8838. The confidence and likelihood of various operating conditions are shown in Table 7:

**Table 7.** Confidence and likelihood of various operating conditions.

Fault Type	$B_{el}$	$P_l$
N	0.0025	0.00250095
F	0.9974	0.99740095

Through the level I decision fusion, the conclusion can be drawn: the operating condition is a fault, the uncertainty is  $9.5 \times 10^{-7}$ , the confidence interval is (0.9974, 0.99740095), and the reliability of the level I model for diagnosing the operating condition as fault F is improved.

Further analysis, there are three possible fault types: discharge (C), overheat (H), and complex (M). input the reduced feature parameters into the class II XGBoost2 model, the actual output of the class II model is [0.9389, 0.0359, 0.0252], and the expected output is [1, 0, 0]. According to equations (10) and (11), the basic probability distribution of evidence source  $e_3$  is calculated as [0.9363, 0.0358, 0.0251], and the basic probability distribution of another evidence source  $e_4$  is calculated as [0.3829, 0.3160, 0.2009] based on the analysis of failure statistical data, then available Table 8:

**Table 8.** Probability distribution of level II diagnostic model.

Evidence	Basic Probability			Uncertainty
	$C$	$H$	$M$	
$e_3$	0.9363	0.0358	0.0251	0.0028
$e_4$	0.3829	0.3160	0.2009	0.1002

Using formulas (5)—(8) in the D-S evidence theory for fusion, the conflict factor  $K$  can be obtained as 0.4775. The confidence and likelihood of various failure types are shown in Table 9. The conclusion can be drawn from the second-level decision: the fault type is discharge, the uncertainty is 0.0005, and the confidence interval is (0.9495, 0.95), and the reliability of the second-level model for diagnosing the fault type as discharge is improved.

**Table 9.** Confidence and likelihood of various failure types

Fault Type	$B_{el}$	$P_l$
C	0.9495	0.9500
H	0.0320	0.0325
M	0.0170	0.0175

There are three possibilities for this level of diagnosis from the diagnosis result of the upper level as discharge: low-energy discharge ( $D_1$ ), high-energy discharge ( $D_2$ ) and partial discharge (PD), input the reduced features into the III-level XGBoost5 model, the actual output is [0.0281, 0.9507, 0.0212], and the expected output is [0, 1, 0]. According to formulas (10) and (11), the basic probability distribution of evidence  $e_5$  is calculated as [0.0280, 0.949, 0.0211], and the basic probability distribution of evidence  $e_6$  as which the non-code ratio method act is [0.27, 0.51, 0.21], then Table 10 :

**Table 10.** Probability distribution of level III diagnostic model.

Evidence	Basic Probability			Uncertainty
	$D_1$	$D_2$	PD	
$e_5$	0.0280	0.9490	0.0211	0.0019
$e_6$	0.2700	0.5100	0.2100	0.0100

In the same way, the conflict factor  $K$  is 0.5079, and the confidence and likelihood of various failure types are shown in Table 11:

**Table 11.** Confidence and likelihood of various failure types

Fault Type	$B_{el}$	$P_l$
$D_1$	0.0164	0.016437
$D_2$	0.9735	0.973537
PD	0.0099	0.009937

The conclusion can be drawn from the third-level decision: the fault type is high-energy discharge, the uncertainty is  $3.7 \times 10^{-5}$ , the confidence interval is (0.9735, 0.973537), and the reliability of the third-level model in diagnosing the fault type as high-energy discharge is improved.

## 6 Conclusion

Based on the XGBoost model, this paper uses the neighborhood rough set theory to reduce its input, and uses its output as one of the evidence sources to employ D-S evidence theory for information fusion, and establish the XGBoost graded diagnosis model, which is compared with the non-code ratio method, BPNN and RandomForest, the prediction accuracy of the transformer graded diagnosis model proposed in this paper is higher than 2.7% on average, and the average time from training model to prediction is less than 0.07 s. In the analysis of the examples, the D-S evidence theory was used to verify the efficiency and reliability of the diagnosis model at all levels proposed in this paper. After summarizing, the following conclusions are drawn:

1) By preprocessing the data with arctangent transformation, and then normalizing, the AUC area of the model can be increased, and the long-tailed distribution problem of DGA data can be solved.

2) The feature attribute reduction using the neighborhood rough set theory can retain the features that have a greater contribution to distinguishing fault types and remove redundant features, which can improve the prediction accuracy and shorten the model training time.

3) Through the use of D-S evidence fusion theory, the ambiguity is resolved. In the future, more sufficient data will be used to establish the XGBoost hierarchical diagnosis model, which can solve the problem of data imbalance, so that all levels of models have a high diagnostic accuracy rate, and there will be no decline.

## References

1. Li Yaxin, Hou Huijuan, Zhang Lijing, Xu Mingkai, Sheng Gehao, and Jiang Xiuchen, "Transformer fault

diagnosis with unbalanced samples based on neighborhood component analysis and k-nearest neighbors," (2020)

- Wu Zhenyu, Zhou Lijun, Zhou Xiangyu, Lin Tong, Guo Lei, Liu Hongwen and Jiang Junfei, "Research on fault diagnosis method of transformer winding based on oscillatory wave," Proceedings of the CSEE, vol. 40, pp. 348–357, (2020)
- Liu Hang, Wang Youyuan, Chen Weigen, Liu Lifeng, and Zhang Jianguang, "Fault detection for power transformer based on unsupervised concept drift recognition and dynamic graph embedding," Proceedings of the CSEE, vol. 40, pp. 4358–4370, (2020)
- Li Gang, Yu Changhai, Fan Hui, Liu Yunpeng and Song Yu, "Deep fault diagnosis of power transformer based on multilevel decision fusion model," Electric Power Automation Equipment, vol. 37, pp.138–144, (2017)
- Wu Xiaoxin, He Yigang, Duan Jiajun, Zhang Hui, Zeng Zhaorong, "Bi-LSTM-based transformer fault diagnosis method based on DGA considering complex correlation characteristics of time sequence", Electric Power Automation Equipment, vol. 40, pp. 184–190,(2020)
- Yin Shi, Hou Guolian, Hu Xiaodong, Zhou Jiwei, Gong Linjuan, "Fault warning and identification of front bearing of wind turbine generator," Chinese Journal of Scientific Instrument, vol. 41, pp. 242–251, (2020)
- Pang Mengyang, Suo Zhongying, Zheng Wanze, Xu YuHeng, Bao Zhuangzhuang and Huang Lin, "Small sample fault diagnosis of aeroengine based on RS-CART decision tree," Journal of Aerospace Power, vol. 35, pp. 1559–1568, (2020)
- Ma Suliang, Jia Bowen, Wu Jianwen, Yuan Yang, Ji-ang Yuan and Li Weixin, "Multi-vibration information fusion for detection of HVCB faults using CART and D-S evidence theory," (2020)
- Gao Rong, Kou Peng, Liang Deliang, Liu Yibin, Wu Zihao, "Robust model predictive control for the voltage regulation in active distribution networks with Hybrid distribution transformers," Proceedings of the CSEE, vol. 40, pp. 2081–2090, (2020)
- Fawcett Tom, "An introduction to ROC analysis," Pattern Recognition Letter, vol.27, pp.861–874, (2006)
- Chu P, Bian X, Liu S and Ling H, "Feature space augmentation for long-tailed data," (2020)
- Li Yonggang, Wang Luo, Li Junqing and Ma Minghan, "Identification of inter-turn short-circuit fault in rotor windings of synchronous generator based on multi-source information fusion," Automation of Electric Power Systems, vol.43,pp.162–167,191,(2019)