A Power Transformer Fault Diagnosis Method Based on Hierarchical Classification and Ensemble Learning

L Su¹, Y Cui²*, Z Yu¹, L Chen¹, P Xu¹, H Hou² and G Sheng²

¹State Grid Shanghai Electric Power Research Institute, Shanghai 200437, China
²Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

E-mail: adacui523@sjtu.edu.cn

Abstract. Since conventional machine learning methods often result in low diagnostic accuracy and non-negligible recognition difference among fault types with imbalanced class distribution among transformer fault types, a multi-level hierarchical power transformer fault diagnostic model is proposed based on hierarchical classification and ensemble learning, where classifiers are hierarchically constructed for level-by-level diagnosis according to the imbalance extent on each level. The Level I neural network classifier extracts 3 generalized feature labels of normal, discharge fault and thermal fault for feature fusion with original data input, to guide classification among 9 detailed operation status under DL/T 722-2014 standard. The Level II classifier adopts EasyEnsemble, generating balanced training subsets by undersampling majority classes and training sub-classifiers in parallel for parameter synthesis in ultimate classifier, to balance information between major fault types and minor ones. Experimental result shows that: compared to traditional methods, our proposed method improves the generalization ability on minority class faults and the overall accuracy by 7%.

1. Introduction
Power transformer plays an important role in the power system, whose normal running operation is closely related to reliable supply from the power grid. Although the electrical equipment development and manufacture technologies are getting mature in recent years, adverse factors such as natural aging, severe environmental conditions and operating overload could still trigger power transformer failure.

Researchers initially established basic methods like Rogers ratio [1] to evaluate the transformer operation status through analysing the difference of constituent contents and gas generation rate of characteristic gases in transformer oil [2]. However, it has limits of invalid embedding and threshold absoluteness [3][4]. AI-based fault diagnostic method has become popular with high accuracy in classifying transformer operation status. Due to low fault occurrence in transformer operation and the big difference between samples of various fault types, the collected dissolved gas-in-oil fault database is a small-scale imbalanced dataset. Because of this, most basic models such as support vector machine [5] and decision tree [6] tend to focus on updating parameters of major fault types while ignore the correct classification of minor fault class during maximizing overall accuracy [7].

In practice, the oversampling and the ensemble learning are applied to solve the imbalanced data set training problem from data acquisition and algorithm optimization respectively. Reference [8] uses SMOTE oversampling to generate new minority class faults to balance the number of the samples of various classes. However, it may results in overfitting and overlapping. Reference [9] uses random
forest, deriving from bagging ensemble learning to generate different training subsets through random sampling and diversified multiple decision trees. This ensemble learning method only proceeds from differentiation of training subsets but without considering the balance of samples in training subsets.

To mitigate the adverse effect of the data imbalance among fault types, we construct a transformer fault diagnostic model from hierarchical classification and ensemble learning, to improve the generalization ability of various fault types in a comprehensive way. First, hierarchical classification [10] is established from general classification to detailed classification level by level. Meanwhile, the EasyEnsemble algorithm [11] is introduced where multiple balanced training subsets are generated from undersampling to conduct parallel training among sub-classifiers to enhance the information of minor class in each sub-classifier. Through case study, our proposed method not only improves the accuracy of all samples with 7% increase compared to traditional methods, but also ensures the proximity of recognition rate on minority fault class to majority fault classes by using the relationship among multi-level classes and balancing the information of each class.

2. Related theory

2.1. Hierarchical Classification

According to the inter-class relationship among 9 operation status under standard DL/T 722-2014 [12], use normal state(NS), discharge fault and thermal fault as 3 generalized feature labels for Level I classification. In Level II, discharge fault is further classified based on different discharge energy densities. Thermal fault is further classified based on different hot-spot temperatures. Discharge of low energy with thermal fault(LDT) and discharge of high energy with thermal fault(HDT) are both tagged with discharge fault label and thermal fault label in Level I as shown in figure 1.

Hierarchical classifier per level is selected as the solution to hierarchical classification in figure 1, which establishes a black box classifier at each class level respectively to undertake the decision-making strategy for all node labels of that level. Moreover, the class labels and primitive input of the upper level are fused to be used as input for the classifier of the next level [13], to provide guidance information through the class hierarchy in a node-to-node way, as shown in figure 2.

![Figure 1. Hierarchical structure of operation status of power transformer.](image-url)
2.2. EasyEnsemble Learning Algorithm

We introduce EasyEnsemble learning method [11], which conducts undersampling to the majority class samples to build multiple balanced training subsets and integrate differentiated sub-classifiers into the final strong classifier, to fully dig the potential useful information from the imbalance dataset. Given a multi-class classification problem for $K$ classes, regard the class with the least number of samples in the imbalanced dataset $C_1, C_2, ..., C_K$ as minority class $P$, of which the number is $|P|$, and the rest as majority classes $N_1, N_2, ..., N_{K-1}$ . Repeat $T$ times of random undersampling to all majority classes, transferring each majority class $N_i$ into a couple of subsets $N_{it}$ . Denote the set of all majority classes obtained from each undersampling as $N_t = \{N_{it}, N_{it+1}, ..., N_{it+|P|-1}\}$, combined with the minority class $P$ to form the balanced training subsets $D_t = N_t \cup P$ for each weak classifier.

Feed the training subset as the input for each weak classifier $h_i$ with $s_i$ times iterative training. After parallel training, all parameters of weak classifiers $h_i$ are integrated to establish the final classifier $H$, of which the output is shown in equation (1), where, $y$ is the true label of input $x$, $h_{id}(x)$ is the output of weak classifier at training epoch $d$, $\varepsilon_i$ is the error function of the weak classifier $h_i$, $k \in \{1,2,\ldots,K\}$ is the subscript of $C_1, C_2, ..., C_K$.

$$H(x) = \arg \max_{k} \sum_{i=1}^{T} \sum_{d=1}^{s_i} \frac{1-\varepsilon_i}{\varepsilon_i} \| h_{id}(x) = y \|$$

3. Proposed Method

3.1. Dissolved gas-in-oil data pre-processing

To reduce the adverse effect of the value fluctuation of the concentration of the 5 characteristic gases of hydrogen, methane, ethylene, acetylene and ethane in different cases, the normalization method in equation 2 is adopted to process each primordial gas sample $x_{ij} = (g_{ij}^1, g_{ij}^2, g_{ij}^3, g_{ij}^4, g_{ij}^5)$, where, $j \in \{1,2,3,4,5\}$ respectively represents 5 gases, $i = 1,2,\ldots,M$ is the sample index. The normalized gas data are recorded as $X = \{x_{ij}\}_{i=1}^{M}$.

$$x'_{ij} = \frac{g_{ij}^l - \min(g_{ij}^l)}{\max(g_{ij}^l) - \min(g_{ij}^l)} \cdot (\max(g_{ij}^l) - \min(g_{ij}^l)) + \min(g_{ij}^l)$$

3.2. Code embedding for operation status and sample distribution

Based on the class hierarchy of operation status in figure 1, we embed codes for normal state, discharge fault and Thermal fault as $Y_1$ at Level I, and codes of 9 subdivided operation status as $Y_2$ at Level II in table 1. Samples in our dissolved gas-in-oil fault database consist of data from power grid
companies and data from published literatures. For the modelling process, 80% and 20% of the samples are used to conduct training dataset and testing dataset respectively. The detailed sample distribution among each class is shown in Table 2.

### Table 1. Code embeddings for operation status.

| Operation Status                             | Abbr. | Level I               | Level II              |
|----------------------------------------------|-------|-----------------------|-----------------------|
| Normal State                                 | NS    | 1,0,0                 | 0: 1,0,0,0,0,0,0,0,0  |
| Partial Discharge                            | PD    | 0,1,0                 | 1: 0,1,0,0,0,0,0,0,0  |
| Discharge of Low Energy                      | LD    | 0,1,0                 | 2: 0,0,1,0,0,0,0,0,0  |
| Discharge of High Energy                     | HD    | 0,1,0                 | 3: 0,0,0,1,0,0,0,0,0  |
| Thermal Fault of Low Temperature             | LT    | 0,0,1                 | 4: 0,0,0,0,1,0,0,0,0  |
| Thermal Fault of Medium Temperature          | MT    | 0,0,1                 | 5: 0,0,0,0,0,1,0,0,0  |
| Thermal Fault of High Temperature            | HT    | 0,0,1                 | 6: 0,0,0,0,0,0,1,0,0  |
| Discharge of Low Energy with Thermal Fault   | LDT   | 0,1,1                 | 7: 0,0,0,0,0,0,0,1,0  |
| Discharge of High Energy with Thermal Fault  | HDT   | 0,1,1                 | 8: 0,0,0,0,0,0,0,0,1  |

### Table 2. Sample Distribution.

| Operation Status                             | Total Samples | Training Samples | Testing Samples |
|----------------------------------------------|---------------|------------------|-----------------|
| Normal State                                 | 652           | 519              | 133             |
| Partial Discharge                            | 161           | 129              | 32              |
| Discharge of Low Energy                      | 216           | 173              | 43              |
| Discharge of High Energy                     | 460           | 366              | 94              |
| Thermal Fault of Low Temperature             | 119           | 95               | 24              |
| Thermal Fault of Medium Temperature          | 203           | 162              | 41              |
| Thermal Fault of High Temperature            | 310           | 247              | 63              |
| Discharge of Low Energy with Thermal Fault   | 90            | 72               | 18              |
| Discharge of High Energy with Thermal Fault  | 46            | 37               | 9               |
| Overall                                      | 2257          | 1800             | 457             |

3.3. Fault diagnosis model

In the hierarchical classification model shown in figure 2, corresponding classifiers are established based on the imbalance degree of class distribution of each level for fault diagnosis level by level. At Level I, the number of normal state, discharge fault and thermal fault is relatively balanced. Therefore, no special balancing technique is required. For Level I classifier, a simple structure neural network is selected to solve the problem of multi-label classification, and feature fusion method is adopted to strengthen the guidance of Level I labels in classification of Level II. When further subdividing the fault types from Level I to Level II, as shown in Table 2, the number of discharge fault of high energy samples is up to 10 times of the number of discharge fault of high energy with thermal fault, indicating the high imbalance under Level II. Therefore, Level II classifier adopts EasyEnsemble learning for multi-class classification, where multiple training subsets with balanced data are generated from undersampling for parallel training, to improve quality of classification on minority fault class.

3.3.1 Level I classifier. To realize multi-label classification with Level I classifier $H_1$, sigmoid activation function is used for interval mapping for the output value $H_1(X)$. Then the mapped result is compared to the threshold value $\rho$ to determine whether it belongs to the operation by step function at
each label position. The final output label $\hat{Y}_1$ of Level I classifier is expressed as equation (3), where, 

$$\rho$$ is pre-set classification threshold value, $\sigma_i(x) = \text{sigmoid}(x) = \frac{1}{1 + e^{-x}}, \quad \text{sgn}(x) = \begin{cases} 0 & \text{if } x \geq 0 \\ 1 & \text{if } x < 0 \end{cases}.$$

$$\hat{Y}_1 = \text{sgn}(\sigma_1(H_1(X)) - \rho) \quad (3)$$

Loss function in equation (4) is adopted for gradient computation and back propagation to update the relevant weight and bias of Level I classifier $H_1$. The output feature $\hat{Y}_1$ of Level I goes through feature fusion with the primitive dissolved gas-in-oil input $X$, combined as the input for Level II classifier.

$$\hat{Y}_1 = \text{sgn}(\sigma_1(H_1(X)) - \rho) \text{Loss}_1 = -(Y_i \log \sigma_i(H_1(X)) + (1 - Y_i) \log(1 - \sigma_i(H_1(X))) \quad (4)$$

$$X_2 = [X, \hat{Y}_1] \quad (5)$$

3.3.2 Level II classifier. While implementing EasyEnsemble learning method, the discharge fault of high energy with thermal fault with least number among all fault types is set as the minority class, the rest classes are treated as majority fault classes to conduct undersampling for $T$ balanced training subsets $D_t = \{(X_{2,t}, Y_{1,t})\}_{t=1}^T$ in Level II with input $X_2$. Parallel training is introduced to train $T$ sub-classifiers $h_{2,t}$, where softmax $\sigma_2(x) = \text{softmax}(x) = \frac{e^{x_j}}{\sum_{j=1}^k e^{x_j}}$ is used as activation function for probabilistic calculation in multi-class classification and the class label $k \in \{0,1,2,3,4,5,6,7,8\}$ with the highest probability as the diagnostic operation status. The loss function in equation (7) with cross entropy is the core to update parameters of sub-classifiers. After fine-tuning, the weight and bias parameters of sub-classifiers are weighted with the pre-set $\alpha_t$, and combined into the ultimate Level II classifier $H_2$. The final output is interpreted as equation (8).

$$\hat{Y}_{2,t} = \text{argmax}_{k} \sigma_2(h_{2,t}(X_{2,t})) \quad (6)$$

$$\text{Loss}_2 = -(Y_{2,t} \log \hat{Y}_{2,t} + (1 - Y_{2,t}) \log(1 - \hat{Y}_{2,t})) \quad (7)$$

$$\hat{Y}_2 = \text{argmax}_{k} \sigma_2((\sum_{t=1}^T \alpha_t h_{2,t})(X_2)) \quad (8)$$

4. Case Study

Samples in Table 2 are used for building our data-driven model, where the input is the set of concentrations of 5 gases and the final target labels for analysis are 9 operation status in Level II. The precision and recall rate of each class are used to assess the exactness and completeness of the model reported failure. The generalization ability is measured by overall accuracy on testing dataset.

4.1. Verification of hierarchical classification

The configuration of Level I back-propagation neural network (BPNN) classifier is set as 5 neurons for input, [15, 30, 15] neurons for 3 hidden layers and 3 for output, with learning rate as 0.001 and training epochs as 50000. By grid searching pre-set threshold value $\rho$, the highest classification accuracy of Level I is 97.37% when $\rho = 0.6$.

To verify the effectiveness of level-by-level classification supported by our hierarchical classification model, whether the feature fusion of Level I classifier plays a guiding role in Level II classification or not, 2 groups are assigned with (proposed model and controlled trial1) and (controlled trial2 and controlled trial3) for comparison. In each group, the configuration of Level I classifier is set as the control variable, one is applied with hierarchical classification while the other omits Level I and classifies operation status directly. The first group adopts EasyEnsemble as the Level II classifier, while the second group adopts BPNN.
In table 3, the overall accuracy is boosted by the introduction of Level I classifier in both groups that 1.8% increase is made in proposed method compared to controlled trial1, and a more obvious one of 3.8% is achieved in controlled trial2 by taking the advantage of constraint relationship between labels among 2 levels provided by class hierarchy. Through feature fusion of the output label of Level I classifier and the input gas, the class information implied in the generalized 3 labels could provide instructions of next level classification, improving fault diagnostic result.

| Table 3. Diagnostic Accuracy of Proposed Model and Controlled Trials. |
|---------------------------------------------------------------|
| Level I Classifier      | Proposed Method | Controlled Trial 1 | Controlled Trial 2 | Controlled Trial 3 |
|-------------------------|-----------------|-------------------|-------------------|-------------------|
| BPNN                    | EasyEnsemble    | BPNN              | None              | BPNN              |
| Accuracy                | 90.37%          | 88.62%            | 87.31%            | 83.59%            |

| Table 4. Diagnostic Result on Testing Dataset. |
|-----------------------------------------------|
| Operation Status | BPNN  | SVM  | Random Forest | Proposed Method |
|------------------|-------|------|---------------|-----------------|
|                  | Precision | Recall | Precision | Recall | Precision | Recall | Precision | Recall |
| NS               | 99.3%   | 100.0% | 100.0%      | 97.7%  | 100.0%    | 100.0% | 100.0%    | 100.0% |
| PD               | 75.8%   | 78.1%  | 71.4%       | 78.1%  | 59.3%     | 100.0% | 76.2%     | 100.0% |
| LD               | 87.9%   | 67.4%  | 73.5%       | 58.1%  | 55.6%     | 46.5%  | 83.3%     | 69.7%  |
| HD               | 77.0%   | 71.3%  | 85.3%       | 98.9%  | 84.2%     | 73.4%  | 93.0%     | 85.1%  |
| LT               | 66.7%   | 83.3%  | 51.3%       | 83.3%  | 54.1%     | 83.3%  | 88.0%     | 91.7%  |
| MT               | 81.8%   | 87.8%  | 62.5%       | 61.0%  | 53.6%     | 36.6%  | 82.2%     | 90.2%  |
| HT               | 88.0%   | 81.0%  | 81.3%       | 41.3%  | 71.2%     | 66.7%  | 93.1%     | 85.7%  |
| LDT              | 89.5%   | 94.4%  | 58.3%       | 77.8%  | 82.4%     | 77.8%  | 88.5%     | 94.4%  |
| HDT              | 21.1%   | 44.4%  | 42.9%       | 66.7%  | 54.6%     | 66.7%  | 61.5%     | 88.9%  |
| Accuracy         | 83.59%  | 79.65% | 76.81%      | 90.37% | 76.81%    | 90.37% | 76.81%    | 90.37% |

4.2. Verification of EasyEnsemble algorithm
To verify the effectiveness of EasyEnsemble learning algorithm in imbalanced dataset, we assign 2 control groups, one with proposed method and controlled trial2, another with controlled trial1 and controlled trial2. Control variable is set as the type of classifier applied in Level II. EasyEnsemble learning and BPNN is respectively used in each group, and the first group is in hierarchical classification process while the other takes the simple classification.

According to table 3, the overall accuracy of model with EasyEnsemble as Level II classifier is higher than the model embedded with BPNN by over 3% increase in the hierarchical group, and about 5% boost in simple classification group, showing that training balanced training subsets in parallel to form the ultimate Level II classifier could achieve a better generalization ability in fault diagnosis.

4.3. Comparison with Traditional Methods
To validate proposed approach, 2 common machine learning methods BPNN and support vector machine (SVM) are also modelled for power transformer fault diagnosis. The configuration of BPNN is the same with Level II classifier with 3 hidden layers; SVM is embedded with radial basis function (RBF) and inter-class imbalance weight adjustment in training.

In terms of generalization ability on minority fault class, the precision and recall rate of our proposed method for all 9 operation status is no less than 60% as shown in table 4, achieving a stable performance in exactness and completeness of reported failures compared to other 2 models. In particular, for the minority fault class, discharge of high energy with thermal fault, BPNN ranks the lowest at 21.1% precision and 44.4% recall rate without adopting any technology against imbalanced
data training. Although SVM adopts inter-class imbalance weight adjustment, it slightly reduces the performance difference between majority fault classes and minority fault class but is not ideal enough.

For overall diagnostic accuracy, our proposed method outperforms other 2 methods. During training, BPNN and SVM conduct supervised learning only with 9 detail-classified labels, whereas the proposed model is supervised both by labels at Level II and Level I, which improves accuracy by 7% than BPNN and 11% than SVM.

5. Conclusions
By establishing hierarchical classification model to extract the hidden label information between Level I and Level II, and taking EasyEnsemble learning to balance information among all operation status types, the proposed method shows its competitiveness than the traditional machine learning fault diagnostic method. In contrast to other machine learning diagnostic methods, our multi-level hierarchical fault diagnosis model could improve the overall classification accuracy by 7% and is of better generalization ability in minority fault class, obtaining a more accurate and more balanced fault diagnostic result on imbalanced samples.

Acknowledgments
This work was supported Project supported by National Natural Science Foundation of China (51477100), Startup Program of Young Faculties of SJTU (Fault diagnosis of power equipment based on artificial intelligence) and Science and Technology Project of SGCC (Research on Transformer State Recognition Technology Based on Deep Learning).

References
[1] Rogers, R 1978 IEEE Trans on Electrical Insulation 5 349–354
[2] Duval, M 1989 IEEE Electrical Insulation Magazine 5 22–27
[3] Dai J, Song H, Sheng G and Jiang X 2017 IEEE Trans on Dielectrics and Electrical Insulation 24 2828–2835
[4] Lee L, Cheng Y, Xie L, Jiang L, Ma N and Lu M 2015 IEEE Trans on Dielectric and Electrical Insulation 22 2368–2378
[5] Li J, Zhang, Q, Wang K, Wang J, Zhou T and Zhang, Y 2016 IEEE Trans on Dielectrics and Electrical Insulation 23 1198–1206
[6] Samantaray S and Dash P 2011 Int. J. Electrical Power and Energy Systems 33 1043–1048
[7] He H and Garcia E 2008 IEEE Trans on Knowledge and Data Engineering 9 1263–1284
[8] Cui Y, Ma H and Saha T 2014 IEEE Trans on Dielectrics and Electrical Insulation 21 2363–2373
[9] Chen X, Cui H and Luo L 2011 Fault diagnosis of transformer based on random forest Proc. Int. Conf. on Intelligent Computation Technology and Automation vol 1 (IEEE) p 132–134
[10] Gordon A 1987 A review of hierarchical classification J. Royal Statistical Society: Series A 150 119–137
[11] Liu X, Wu J and Zhou Z 2009 IEEE Transactions on Systems, Man, and Cybernetics, Part C 39 539–550
[12] Guide to the analysis and the diagnosis of gasses dissolved in transformer oil: DL/T 722-2014 2014 (Beijing: National Energy Administration)
[13] Cerri R, Barros R and De Carvalho 2014 J. Computer and System Sciences 80 39–56