ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20230417000543
Original scientific paper

An Ensemble Learning Method for the Fault Multi-classification of Smart Meters

Shuhua LIANG, Changji CHEN, Dalei WU*, Longjin CHEN, Qingyao WU, Ting Ting GU

Abstract: With the rapid development of the power industry and the widespread adoption of smart meters, the occurrence of smart meter failures has also become more frequent. Consequently, the classification of smart meter faults has become a crucial task to ensure quality assurance in the power industry. Accurately determining the fault types of smart meters and improving maintenance efficiency are of utmost importance to ensure their safe and stable operation. Traditional methods for classifying smart meter faults primarily rely on manual inspection and testing, which suffer from issues such as low classification efficiency, high cost, susceptibility to missed detections, and false detections. In recent years, machine learning methods have demonstrated advantages in this field. This paper proposes an ensemble learning method for the multi-classification of smart meter faults to enhance the efficiency and accuracy of fault classification. Firstly, various data preprocessing techniques are employed to clean and extract features from a real-world dataset, thereby enhancing the data quality of the smart meter fault types. Secondly, a selection process is conducted to screen classical machine learning algorithms, resulting in the choice of three algorithms: K Nearest Neighbors (KNN), Random Forest (RF), and Xtreme Gradient Boosting (XGBoost). These algorithms are then utilized to classify the fault types of smart meters. Finally, a multi-classification ensemble learning method is introduced to combine the results from multiple classifiers, thereby improving the accuracy and robustness of smart meter fault classification. Experimental results demonstrate that the proposed method exhibits high accuracy and robustness in fault classification, offering promising applications and value for widespread adoption.

Keywords: ensemble learning; fault classification; multi-classification; smart meters

1 INTRODUCTION

The popularization and application of smart meters promotes the construction and development of smart grid, which is of great significance to maintaining the security of power grid [1-3]. As an important part of power consumption information collection system, fault classification of smart meters has become an important task to ensure the stable operation and quality assurance of power industry [4]. Faulty meters may not accurately measure energy consumption, making it difficult to identify areas for energy conservation. Accurate classification helps in promoting energy efficiency. Reliable meter data is essential for grid management and analytics, making accurate fault classification vital for maintaining data integrity [5]. However, the traditional fault classification methods of smart meters rely on manual inspection and manual test, which has some problems such as low classification efficiency, high cost, and proneness to miss detection [6]. When a fault occurs, the maintenance personnel are required to maintain it quickly. However, the operation and maintenance system cannot judge the specific fault in actual operation, which leads to the failure of the maintenance of the faulty smart meters. Therefore, it is of great theoretical significance and practical application value to study an efficient and accurate fault classification method on how to determine what kind of fault has occurred in the smart meters [7, 8].

At present, with the development of computer science and the rapid progress of artificial intelligence, researchers have applied machine learning to all aspects of production and life, and achieved good application results and technical improvement [9]. The concept of machine learning has also been widely concerned in the application research of smart meters. Through the relevant machine learning algorithm, the outstanding features in the fault dataset of smart meters can be extracted more comprehensively, and the dependence and interference of human factors can be reduced. At the same time, it effectively solves the problems of anomalies, errors and imbalanced distribution in data, greatly improves work efficiency and promotes the intelligent development of fault handling of smart meters [10, 11].

Cheng et al. [12] found that the failure rate of batch smart meters demonstrated a better fit solely in the time dimension, and they concluded that the linear regression model is unsuitable for predicting the failure rate of batch smart meters. Zhang et al. [13] proposed a cost-sensitive multi-classification ensemble tree model, which compresses the feature dimension of data to mitigate overfitting through preprocessing techniques such as hierarchical clustering. Via optimizing the cost-sensitive objective function designed based on the prior probability of categories, they effectively overcame the deviation caused by the imbalance of datasets. Du et al. [14] proposed a meter fault evaluation method based on Wei bull parameters model which can effectively integrate data from different areas and predicted the fault of power meters in multiple regions in the short term. Peng et al. [15] designed an improved fault classification and state prediction model for smart meters based on the improved Light GBM algorithm based on Random Forest.

However, the fault data structure of smart meters is complex, and the existing methods have some problems in solving the fault classification of smart meters: the sample data under different fault types are extremely uneven [16, 17], so it is necessary to adopt reasonable sampling strategies to eliminate the influence of data imbalance on the classification results. Li Ning et al. [18] proposed a fault prediction method for smart meters that integrates multi classification machine learning models to address the characteristics of large scale, high dimensionality, errors, and abnormal data in smart meter fault data. Using normal distribution completion and box plot methods to fill in missing values and replace outliers in the original dataset. By calculating the correlation coefficient between feature attributes and fault types, irrelevant features are eliminated and a feature subset is formed; building a mixed sampling strategy to solve the problem of imbalanced fault data. Gao Wenjun et al. [19] proposed a fault prediction method based on spatio-temporal convolutional neural network (ST-CNN) to address the sudden, complex, and multifaceted characteristics of smart meter faults. This method first incorporates time information into feature variables using sliding windows, constructs an input matrix with spatio-temporal characteristics, and then combines it with CNN to establish an intelligent meter fault prediction model. The model parameters are optimized using Adam algorithm.

Different machine learning algorithms have different advantages and disadvantages [20], and they perform differently on different types of faults. It is necessary to construct a reasonable multi-model hybrid integration algorithm to improve the accuracy of prediction results. However, the fault data structure of smart meters is complex, and the sample data size under different fault types is extremely uneven [21]. It is necessary to adopt a reasonable sampling strategy to eliminate the impact of data imbalance on the prediction results. In addition, different machine learning algorithms have their own advantages and disadvantages [22-24], and their recognition ability varies for different types of faults. It is necessary to construct a reasonable multi model hybrid ensemble algorithm to improve the accuracy of prediction results.

Ensembles combine multiple models to enhance classification accuracy, reducing the risk of misclassifying meters as faulty or non-faulty. They are more robust to noise and variations in meter data, ensuring reliable fault detection even in challenging scenarios. Ensembles adapt well to changing fault patterns over time, providing sustainable fault classification performance. By incorporating diverse base models, ensembles capture various fault characteristics, enhancing fault detection capabilities.

In this paper, we propose an ensemble learning method to solve the problem of fault classification of smart meters, exploring the collected historical fault data information from a real-world dataset. In summary, the main contributions of this paper are as follows.

1. This paper cleans and standardizes the data collected by real-world smart meters through a variety of data preprocessing technologies, and provides a dataset through fault type selection, feature selection, and imbalanced data sampling.
2. This paper establishes an ensemble learning model by series of experiments to determine the most suitable classifier combination, the performance of which is also proved by comparison experiments.

2 RESEARCH METHODOLOGY

The original real-world dataset utilized in this paper is provided by the system data center of the Power Grid Corporation, encompassing a multitude of attributes. This dataset needs to be processed and analyzed for comprehensive examination. Initially, samples with missing attributes and anomalies are systematically eliminated, and the attributes of the remaining samples are referred to as features in the subsequent sections of the paper. Then, considering that there may be redundant or irrelevant features for fault classification, we calculate the correlation coefficient between features, select different feature subsets for comparative experiments, and finally determine the features that should be retained in the dataset. Furthermore, to address the issue of data imbalance among different fault types in the dataset, a hybrid sampling method is employed. This method combines both over-sampling and under-sampling techniques, and its viability is substantiated through comparative experiments.

2.1 Analysis and Features Selection of the Real-world Dataset

The purpose of feature selection in the real-world dataset of smart meter fault types is to find out the most important features for fault classification, so as to improve the accuracy and interpretability of the classification model. After preliminary feature selection, we exclude some features completely unrelated to the fault type from the dataset. Then we sort out the remaining features, and keep the features that affect the fault types in the dataset. Assuming that dataset is defined as:

$$D=\left\{\left(x_1,y_1\right),\ \left(x_2,y_2\right),\ ...,\ \left(x_M,y_M\right)\right\} \tag{1}$$

In D, the total number of samples is M, each sample in D is (x_i, y_i) $\left(i\in\left[1,M\right]\right), x_i\in X\left(\left|X\right|=K\right)$ and X is the set of features which represents the feature information of the i-th sample, $y_i\in Y\left(y\in\left[1,N\right]\right)$ represents the fault type of the i-th sample.

Without loss of generality, the following discussion is about a typical item in D. For the item, the correlation coefficient between each feature and the fault type is expressed as:

$$r=\left[\rho_1,\rho_2,...,\rho_k,...,\rho_K\right] \tag{2}$$

where ρ_k represents the relationship between the k-th feature and fault type, and r represents the correlation coefficient set.

The calculation of ρ_k is:

$$\rho_k=\frac{cov\left(\lambda_k,Y\right)}{\sqrt{D\left(\lambda_k\right)\cdot D\left(Y\right)}} \tag{3}$$

where $cov(\lambda_k, Y)$ represents the covariance between the feature λ_k and the fault type Y. $D(\lambda_k)$, and $D(Y)$, represent the variance of feature λ_k and fault type Y, respectively.

2.2 Mixed Sampling Method for the Imbalanced Dataset

As mentioned above, the numbers of different fault types of samples are quite different, so a certain mixed sampling method is needed to eliminate the influence of imbalanced data. Without loss of generality, we assume that there are N types in the dataset. The number of samples corresponding to each type in the dataset is S_i, (i = 1, 2, …, N). The average number of the samples corresponding to N types is S_{avg}:

$$S_{avg}=\frac{1}{N}\sum_{i=1}^{N}S_i \tag{4}$$

In this paper, a mixed sampling method combining over-sampling and under-sampling is adopted to deal with samples of large sizes as well as small sizes at the same time. Firstly, under-sampling is carried out to reduce the number of samples of large sizes. If the number of samples corresponding to the type i is larger than S_{avg}, namely, $S_i > S_{avg}$, we use the method of undersampling. Then, the samples of other types are oversampled. This can increase the diversity of data and improve the generalization of the model.

$$S_i^m = \begin{cases} S_i - \dfrac{S_i - S_{avg}}{2}, \text{undersampling} \\ S_i + \dfrac{S_{avg} - S_i}{2}, \text{oversampling} \end{cases} \tag{5}$$

where S_i is the fault data sample before sampling, S_i^m is the fault data sample after sampling, S_{avg} is the average of the number of samples corresponding to N types.

The most basic and simple method of oversampling is to randomly copy a portion of the original samples from the samples to be sampled and add them to this type of sample. The advantages of this method are simplicity and convenience, but the disadvantages are also obvious. After oversampling, certain samples will appear repeatedly in the dataset, resulting in overfitting in the trained model, Therefore, in order to solve this problem, we choose to add slight random perturbation every time a new sample point is generated, and ultimately decide to sample the SMOTE algorithm, which is the synthesis of minority class oversampling technology. The basic idea of this algorithm is based on interpolation method. Firstly, we analyze the set of minority class samples to be sampled, and then synthesize new sample points through a specific formula. The specific process is as follows:

Assuming the number of samples for a minority class is T, the SMOTE algorithm will synthesize NT new samples for this minority class. Here N is the sampling ratio, and the calculation formula is:

$$N = \left(Num_{new_i} - Num_i\right) / Num_i \tag{6}$$

Here, it is required that N must be a positive integer, so in the end, N obtained from Eq. (6) needs to be rounded. The new set of NT samples and the original T samples, totaling $(n + 1)\ T$ samples, is used for subsequent model training and learning.

The oversampling mainly uses the SMOTE algorithm, whose main principles are as follows:

Consider a sample i of this minority class, whose eigenvector is i □ $\{1, 2, ..., T\}$:

1. Based on the KNN algorithm, find k nearest neighbor points of the sample from all samples of the minority class, denoted as, $near$ □ $\{1, 2, ..., k\}$.
2. Then randomly select a sample point from these k nearest neighbors, and add slight interference to generate a random number between 0 and 1. According to Eq. (7), synthesize a new sample point:

$$X_{i1} = X_i + \xi_1 \cdot \left(X_{i(nn)} - X_i\right) \tag{7}$$

3. Repeat step 2 N times to synthesize N new sample points: $X_{i(\text{new})}$, new $\in \{1, ..., N\}$. If the above operation is performed on all T minority class samples, a total of NT new samples can be synthesized for that minority class.

Under sampling adopts a random sampling method, which directly selects a portion of samples from the samples to be sampled. The advantage of this method is that it is simple and fast.

2.3 The Ensemble Learning Method for the Multi-classification of Smart Meter Fault Types

After preprocessing the dataset of smart meter fault types to mitigate the influence of data imbalance on the prediction results, we perform the task of classifying the fault types. Given the varied strengths and weaknesses of different machine learning algorithms, their performance can differ across distinct fault types. To make full use of base classifiers, we propose an ensemble learning method to construct a reasonable multi classifier hybrid ensemble algorithm to improve the accuracy of the prediction results. The ensemble learning method for the multi-classification of smart meter fault types is illustrated in Fig. 2.

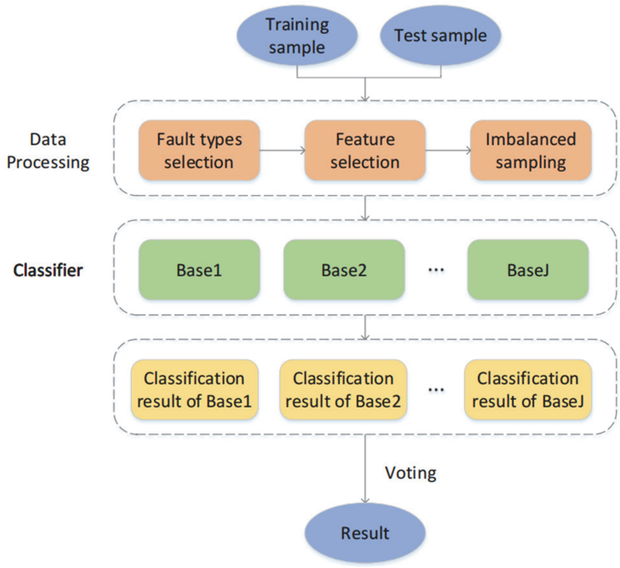

Figure 1 The flowsheet of the ensemble learning method for the multi-classification of smart meter fault types

The purpose of the ensemble learning is to integrate the advantages of base classifiers, and then improve the fault classification accuracy of the model. In addition, it can reduce the risk of over-fitting, improve the generalization and stability of the model.

As shown in Fig. 2, assuming that there are J base classifiers used to build the ensemble learning model, the sample dataset is as shown in Eq. (1) as $D = \{(x_1, y_1), (x_2, y_2), ..., (x_M, y_M)\}$.

The classification accuracy of each type in the training results of the j-th model a_j is:

$$a_j = \left[a_j^1, a_j^2, ..., a_j^N\right] \tag{8}$$

Then the prediction results of the multi-classification ensemble learning model for the n-th fault type $\hat{y}^n$ is:

$$\hat{y}^n = arg\,max\left(\sum_{j=1}^{J} y_j^n\right) \quad (9)$$

where *argmax* represents the type with the highest score as the final prediction result; y_j^n represents the classification result of the j-th base classifier for the n-th fault type. Soft Voting Classifier is used in our method by taking the averages produced by all the base classifiers as the evaluation values, and the corresponding type with the highest value is the final prediction result.

3 DATA PROCESS AND EXPERIMENT

In this part, we first introduce the evaluation metrics that will be used in the experiment. Then, we set up experiments on 5 public datasets and the Actual Dataset of Smart Meter Fault Types in terms of ACC to select the base classifiers, and then we combine the base classifiers with better performance by ensemble learning, and choose the best composition from them. Then, we evaluate the accuracy of ensemble learning classifier and base classifier for various fault types on the Actual Dataset, and draw the confusion matrix corresponding to each classifier. Finally, we compare the *Precision*, *Recall* and *F*1 of each basic classifier and the Ensemble Learning classifier.

3.1 Overall Statistical Analysis and Preliminary Selection of Fault Types

We preliminarily process the real-world dataset, and then statistically analyze the fault types of samples. According to statistics, in the above dataset, there are 17467 normal smart meter samples (without faults), and the other 45015 samples contain 28 fault types. The name of each fault type and the corresponding sample number are shown in Tab. 1, and the distribution of fault types is shown in Fig. 1.

Table 1 Fault type and the corresponding sample number

FaultType	Sample Number	FaultType	Sample Number
Mechanical fault	15149	Crash	94
Burning fault	9228	Wiring fault	84
Screen fault	6596	Multiplication fault	64
Communication fault	3880	Fire causing destruction	60
Other faults	3644	Watch box destruction	29
Data exception	2102	Time disorder	27
Lightning causing destruction	1109	Programming fault	17
Clock disorder	865	TV fuse breaker	11
Typhoon causing destruction	583	Anthropogenic destruction	7
Flood causing destruction	451	TA overload fault	2
Pulse sampling fault	330	TA turn-to-turn short circuit	2
Battery fault	307	TV turn-to-turn short circuit	1
Overload burning fault	267	TA broken circuit	1
Lose	104	Measuring cabinet fault	1

As can be seen from Fig. 1, the number of samples of various fault types is imbalanced. Among them, the types of Mechanical fault, Burning fault, Screen fault, Communication fault, Other faults, Data exception, Lightning causing destruction, Clock disorder, and Typhoon causing destruction, account for nearly 98.57% in total, while the other fault types account for a relatively small proportion, only less than 5%. Therefore, in order to avoid the influence of fault types with small sample sizes and other faults (the type which contains many faults that cannot be distinguished or have no obvious characteristics), we eliminate the fault types with the sample size less than 500 in the dataset and the type of Other faults. In addition, we also adopt a mixed sampling method to solve the problem of data imbalance in the subsequent part of this paper.

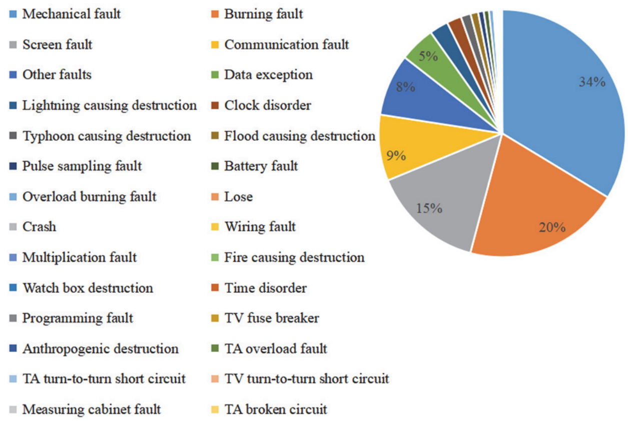

Figure 2 The number of various fault types in the real-world dataset of smart meter fault types

The correlation coefficient results obtained from our dataset are sorted and shown in Tab. 2.

Table 2 The rank of correlation coefficient between features and fault type

Feature	Rank	Feature	Rank
Manufacturer	1	Equipment type	5
Fault discovery date	2	Equipment state	6
Fault recovery date	3	Commissioning date	7
Fault source	4		

According to our ranking of correlation coefficient results in Tab. 2, we pass a judgement to the validity of features further by constructing feature subsets and classifying them by decision tree.

We construct the following two feature subsets from the original features:

(1) Equipment type, Manufacturer, Equipment state, Fault discovery date, Fault recovery date, Fault source and Commissioning date.

(2) Equipment type, Manufacturer, Equipment state, Fault discovery date, Fault recovery date, and Fault source.

The accuracy of the subsets calculated by the decision tree is as follows: (1) 51.01%; (2) 52.50%.

By analyzing the above data, we can see that the classification accuracy of the model trained by the feature subset (2) is higher, so the feature subset (2) is finally selected, that is, six features are reserved: manufacturer, fault discovery date, fault recovery date, fault source, equipment type, and equipment state.

After a series of data preprocessing operations, the number of remaining sample is 39512. We name the real-world dataset provided by the Power Grid Data Center after a series of data processing as the Actual Dataset of Smart Meter Fault Types (in the following part, we will refer to this dataset as the Actual Dataset).

3.2 Experimental Settings and Evaluation Metrics

Our experiments are carried out on a 64-bit Windows 10 operating system and a 2.5 GHz Intel(R) Core(TM) i5-7300HQ CPU. In our experiment, each dataset is randomly divided into a training set and a test set in a 7:3 ratio. Ten base models commonly selected for fault classification are trained, including K Nearest Neighbors (KNN), Random Forest (RF), Light Gradient Boosting Machine (LGBM), Gradient Boosting Decision Tree (GBDT), Adaptive Boosting (AdaBoost), Xtreme Gradient Boosting (XGBoost), Decision Tree (DT), Gaussian Naive Bayes (GaussianNB), Logistic Regression (LR), and Support Vector Machines (SVM), and the prediction accuracy of each model on the test dataset is obtained. For SVM, LR, DT, XGBoost, AdaBoost, GBDT and RF, we set random_state to 42. For XGBoost, AdaBoost, GBDT and RF, we set n_estimators to 100. In addition, we set n_estimators to 20 and 5 for LGBM and KNN, respectively. We use LabelEncoder in sklearn library to encode the classification features in the dataset, and all the classification features are converted into numerical features.

In order to comprehensively evaluate the performance of various base classifiers as well as the ensemble learning model, we adopt four widely used evaluation metrics. We use a one-to-one approach to extend these binary classification evaluation metrics to the multi-classification in discussion.

The confusion matrix is a table for evaluating the performance of the classification model. Based on the real categories and the classifier's prediction results, the samples are divided into four groups: True Positive (*TP*), False Positive (*FP*), True Negative (*TN*), and False Negative (*FN*).

Accuracy represents the ratio of the number of samples correctly predicted by the classifier to the total number of samples, and the calculation formula is:

$$Accurancy = \frac{TP + TN}{TP + FP + TN + FN} \tag{10}$$

Precision indicates the ratio of the number of samples correctly predicted by the classifier to the number of all samples predicted as positive examples, and the calculation formula is:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Recall is an index used to measure the performance of classification model, which reflects how many positive samples the model can correctly identify. The calculation formula is:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

$F1$ is the harmonic average of *Precision* and *Recall*, which represents the comprehensive performance of the classifier. The calculation formula is:

$$F1 = \frac{(2 \cdot Precision \cdot Recall)}{Precision + Recall} \tag{13}$$

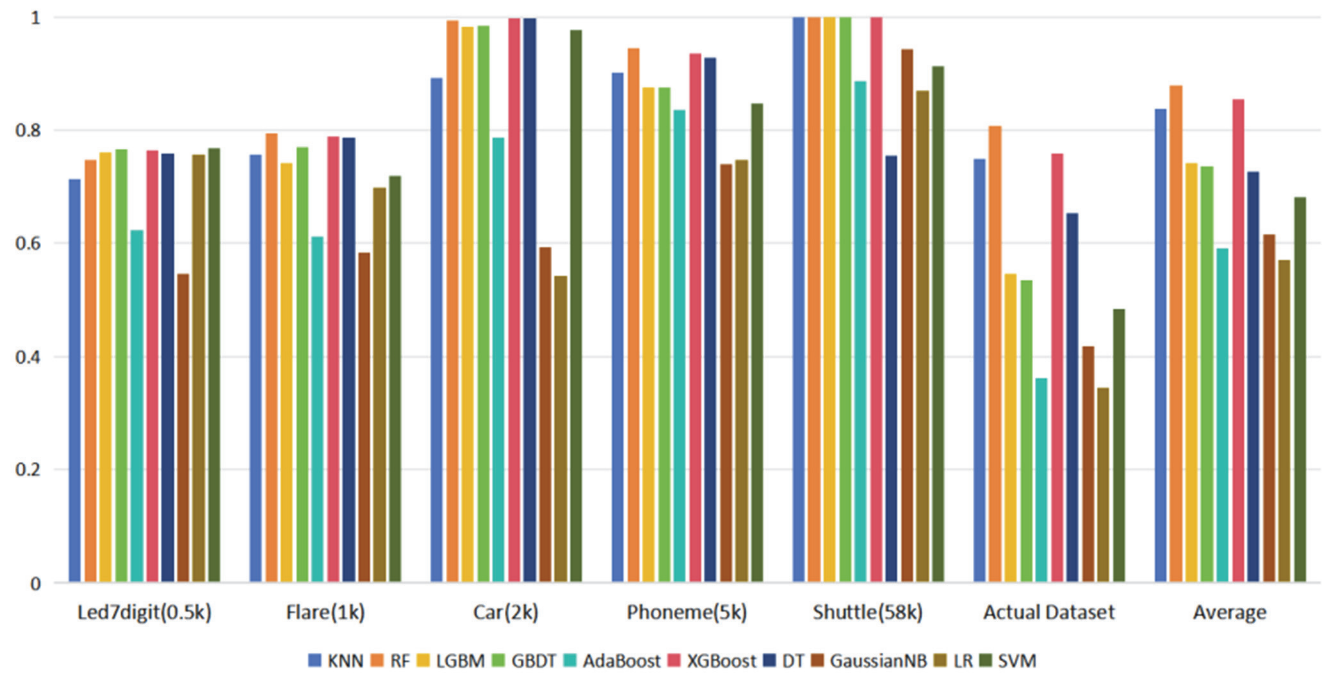

Figure 3 Visual results of the comparison of classification accuracy of 10 base classifiers

Table 3 Comparison of classification accuracy of 10 base classifiers

	Led7digit (0.5k)	Flare (1k)	Car (2k)	Phoneme (5k)	Shuttle (58k)	Actual Dataset	Average
KNN	0.712899	0.756285	0.892053	0.901273	0.999577	**0.748363**	**0.8371801**
RF	0.747826	**0.793460**	**0.994061**	**0.945481**	**0.999988**	**0.807986**	**0.8780308**
LGBM	0.761159	0.741155	0.981924	0.875405	**0.999988**	0.545883	**0.74092005**
GBDT	**0.765700**	0.770150	0.985023	0.874096	0.999392	0.534290	**0.73671105**
AdaBoost	0.623140	0.612109	0.786929	0.834641	0.887068	0.362557	0.59023985
XGBoost	**0.763478**	**0.787792**	**0.997676**	**0.935819**	**0.999984**	**0.758646**	**0.85355645**
DT	0.759034	**0.785897**	**0.997159**	**0.928617**	0.755422	0.653949	0.72713645
GaussianNB	0.546087	0.583727	0.592211	0.739844	0.942150	0.419010	0.61524345
LR	0.756715	0.698352	0.542361	0.746395	0.870251	0.343953	0.57024295
SVM	**0.767874**	0.718518	0.976506	0.846759	0.912302	0.483223	0.68078495

3.3 Experiments to Select Base Classifiers

In order to select the base classifiers for the ensemble learning method proposed in this paper, this section compares the classification accuracy of the base classifiers on 6 datasets, including the Actual Dataset and five public datasets from the authoritative KEEL of machine learning. For the sake of fairness, each dataset is given a weight, and we calculate the weighted ACC to evaluate the performance of the base classifiers, in which the weight of the shuttle dataset is 0.3, the weight of the Actual Dataset is 0.5, and the weight of each of the other datasets is 0.05. The weights of the public dataset are set according to the data volume of each dataset as shown in the bracket after the name of the dataset in the first row of Tab. 3.

In this part of the experiment, the ten base models, and the prediction accuracy of each model on the test dataset is obtained. The results are shown in Tab. 3. The three base classifiers with the highest classification accuracy in each dataset are displayed in bold, and the top five (a half of the base classifiers) weighted ACC values are also shown in bold.

From the result shown by Tab. 3, we compare the ACC of 10-base classifier on the six datasets. The performance of RF and XGBoost is steadily ahead of other classification methods in most cases, and their weighted average ACC values also rank the top two of the 10 classifiers. In addition, KNN, LGBM and GBDT can also achieve good results. Considering the above results and the scale and time constraints of the experiment, we conduct a combinations experiment on five classifiers, namely, RF, XGBoost, KNN, LGBM, and GBDT. Based on the combination of RF and XGBoost, we add KNN, LGBM, and GBDT one by one to build an ensemble learning classifier, and compare the performance of the ensemble learning classifiers of different combinations of base classifiers on the dataset with the two highest weights (shuttle dataset and Actual Dataset) to select base classifiers. The results are shown in Tab. 4.

Table 4 Comparison of classification accuracy of ensemble learning classifiers of different combinations of base classifiers on the shuttle dataset and the actual dataset

	Shuttle	Actual Dataset
RF + XGBoost	0.999992	0.842749
RF + XGBoost + KNN	**0.999996**	**0.848225**
RF + XGBoost + KNN + LGBM	0.999992	0.819614
RF + XGBoost + KNN + LGBM + GB	0.999992	0.770683

As can be seen from Tab. 4, when KNN, RF and XGBoost are used as base classifiers to construct ensemble learning classifier, the best accuracy can be achieved on both shuttle dataset and the Actual Dataset. Therefore, K Nearest Neighbors (KNN), Random Forest (RF) and Xtreme Gradient Boosting (XGBoost) are selected as the base classifiers for use in our ensemble learning method. We also give the following theoretical explanations. KNN is simple and easy to use, and there is no need to train data. Random Forest has a high tolerance for outliers and noise, and over-fitting is not easy to appear. It can select features to improve the generalization of the model. XGBoost has obvious advantages in training speed and memory usage, and can adaptively extract features, and can handle large datasets. Therefore, it is the most effective to select these three basic classifiers to construct our ensemble learning classifier (short for EL in the following part).

3.4 Experiments on the Actual Dataset of Smart Meter Fault Type in terms of ACC

For the top three base classifiers and the proposed EL, we conduct experiments on the Actual Dataset, and obtain the accuracy value for each fault type to evaluate the effect of the ensemble learning technique. It can be seen from Tab. 5 that EL shows advantage over the three basic classifiers when dealing with eight different types of smart meter faults, which proves the effectiveness of this method. And we draw the confusion matrix of this experiment in Fig. 3, where (a), (b), (c) and (d) represent the confusion matrixes of KNN, RF, XGBoost and EL, respectively. As shown in Fig. 3, the EL Classifier that we build can achieve the best classification results on all fault types.

In the case of experimental results, we conducted a significance test. In this test, we used the F-statistic to measure the variance differences between different groups. A larger F-statistic indicates a larger variance difference between groups, suggesting the presence of significant differences. The p-value is a probability value used to assess the significance of the F-statistic. It represents the probability of observing the F-statistic value under the null hypothesis. A smaller p-value indicates that the observed differences are more likely to be real rather than resulting from random factors. We typically set the significance level to 0.05, which means that we want to conduct significance tests at a 95% confidence level. If the p-value is less than the significance level, differences are generally considered significant. The test results show F-statistic: 0.9993562219903013, p-value: 0.4076432129126476, and we believe we can reject the null hypothesis, indicating the presence of significant differences between the models.

Table 5 Accuracy on the actual dataset

Fault Type	Label	KNN	RF	XGBoost	EL
Mechanical fault	0	1.000000	1.000000	1.000000	**1.000000**
Burning fault	1	0.970666	0.990112	0.756098	**0.998682**
Screen fault	2	0.998369	1.000000	0.990865	**1.000000**
Communication fault	3	0.613285	0.834265	0.628412	**0.912200**
Data exception	4	0.442899	0.642187	0.507603	**0.808476**
Lightning causing destruction	5	0.266871	0.461827	0.442059	**0.641104**
Clock disorder	6	0.819596	0.946044	0.844422	**0.987090**
Typhoon causing destruction	7	1.000000	1.000000	0.921363	**1.000000**

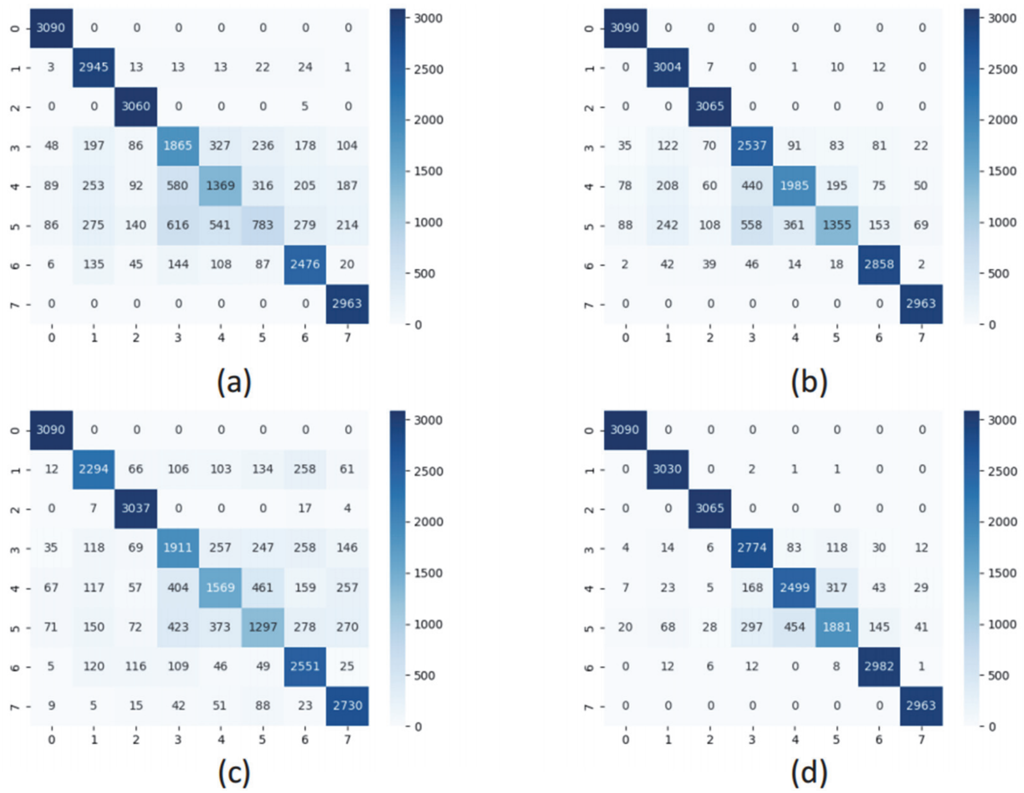

Figure 3 Confusion matrixes of KNN, RF, XGBoost classifier and the EL on the actual dataset

3.5 Experiments on the Actual Dataset of Smart Meter Fault Types in Terms of *Precision*, *Recall* and *F*1

When the EL method proposed in this paper is constructed by KNN, RF and XGBoost, the *Precision* and *Recall* are used as the evaluation metrics of the classifiers, and usually the concerned type is the positive and other classes are the negative.

It can be seen from Eq. (11) that when the *Precision* and *Recall* are both high, the value of *F*1 will also be high. In the experiment, the dataset is randomly divided and trained for 100 times, and the average value of the results is taken as the final experimental result for comparison, so as to avoid the contingency of experimental results caused by the randomness of dataset division as well as increase the credibility of the experimental results.

Table 6 *Precision*, *recall* rate and *F*1 of each classifier on the actual dataset

Classifier	*Precision*	*Recall*	*F*1
KNN	0.741018	0.763961	0.742662
RF	0.858826	0.859304	0.849713
LGBM	0.750475	0.761353	0.751774
EL	**0.916249**	**0.919345**	**0.916314**

As depicted in Tab. 6, the Ensemble Learning (EL) method, a combination of KNN, RF, and XGBoost, exhibits noteworthy enhancements in *Precision*, *Recall*, and *F*1 when compared to the individual base classifiers. In comparison with the second best classifier RF, our EL classifier is improved by about 0.06 on *Precision*, 0.06 on *Recall* and 0.07 on *F*1. These findings underscore that the ensemble learning method proposed in this paper can effectively improve the performance and the reliability of fault type classification of smart meters, and proves the effectiveness of this EL method.

4 CONCLUSION

In this paper, we have conducted a comprehensive analysis and processing of a real-world smart meter fault dataset. Our approach encompasses several crucial steps to enhance the effectiveness of fault classification. Firstly, we have meticulously filtered out invalid samples from the dataset, ensuring that our analysis is based on reliable data. This step is fundamental for the accuracy of our results. Next, we employed feature selection techniques to eliminate redundant or irrelevant features. This not only streamlines the dataset but also significantly boosts classification efficiency by focusing on the most informative attributes. Addressing the challenge of data imbalance, we have introduced a mixed sampling method that combines over-sampling and under-sampling. This approach helps ensure that our model can effectively handle varying fault types and maintain robust predictive capabilities. Furthermore, we have carefully selected the base classifier for our ensemble learning model through a series of rigorous experiments. This step is pivotal as the choice of base classifier profoundly impacts the overall performance of the ensemble model.

The performance evaluation of our ensemble learning method demonstrates its effectiveness. The results validate that the model proposed in this paper enhances prediction performance significantly. This improvement holds valuable implications for real-world smart meter fault diagnosis applications, promising more accurate and reliable results. Looking ahead, our future research will expand in several directions. We plan to test our methodology on a more diverse set of datasets to ensure its generalizability. Additionally, we will explore the integration of deep learning techniques into our approach, as they hold great potential for further enhancing fault diagnosis and classification accuracy. These advancements will contribute to the ongoing development of intelligent metering systems, offering more reliable and precise fault detection capabilities.

Acknowledgement

This work was supported by Science and Technology Project of China Southern Power Grid Limited Liability Company (070000KK52200021).

5 REFERENCES

[1] Wang, Y., Chen, Q., Hong, T., & Kang, C. (2019). Review of smart meter data analytics: applications, methodologies and challenges. *IEEE Transactions on Smart Grid, 10*(3), 3125-3148. https://doi.org/10.1109/TSG.2018.2818167

[2] Wen, L., Zhou, K., Yang, S., & Li, L. (2018). Compression of smart meter big data: a survey. *Renewable and Sustainable Energy Reviews, 91*, 59-69. https://doi.org/10.1016/j.rser.2018.03.088

[3] Sovacool, B. K., Kivimaa, P., Hielscher, S., & Jenkins, K. (2017). Vulnerability and resistance in the United Kingdom's smart meter transition. *Energy Policy, 109*, 767-781. https://doi.org/10.1016/j.enpol.2017.07.037

[4] Yao, H., Wang, X., Wu, L., Jiang, D., Luo, T., & Liang, D. (2018). Prediction method for smart meter life based on big data. *Procedia engineering,* 15(8), 2810. https://doi.org/10.3390/en15082810

[5] Chakraborty, S. & Das, S. (2019). Application of smart meters in high impedance fault detection on distribution systems. *IEEE Transactions on Smart Grid, 10*(3), 3465–3473. https://doi.org/10.1109/TSG.2018.2828414

[6] Siddiqui, I. F., Qureshi, N. M. F., & Shaikh, M. A. (2019). Stuck-at fault analytics of IoT devices using knowledge-based data processing strategy in smart grid. *Wireless Personal Communications, 106*(4), 1969–1983. https://doi.org/10.1007/s11277-018-5739-9

[7] Gilbert, E. P. K., Lydia, M., Baskaran, K., & Rajsingh, E. B. (2019). Trust aware fault tolerant prediction model for wireless sensor network based measurements in smart grid environment. *Sustainable Computing: Informatics and Systems, 23*, 1111-1114. https://doi.org/10.1016/j.suscom.2019.06.001

[8] Gao, X., Ji, W., Zhao, B., Jia, X., Huang, Z., & Ren, B. (2021). Multi-classification Method of Smart Meter Fault Types Based on CVAE-CNN Model Under Imbalanced Dataset. *Dianwang Jishu/Power System Technology, 45*, 3052-3060

[9] Diez-Olivan, A., Ser, J., Galar, D., & Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: trends and perspectives towards Industry 4.0. *Information Fusion, 50*, 92–111. https://doi.org/10.1016/j.inffus.2018.10.005

[10] Fazel, A., Selim, E., & Wilfried, S. (2018). Rethinking human-machine learning in Industry 4.0: how does the paradigm shift treat the role of human learning? *Procedia Manufacturing*, *23*, 117–122. https://doi.org/10.1016/j.promfg.2018.04.003

[11] Diogo Narciso, A. C. & Fernando, G. M. (2020). Application of machine learning tools for energy efficiency in industry: a review. *Energy Reports, 6*, 1181-1199. https://doi.org/10.1016/j.egyr.2020.04.035

[12] Cheng, S. & Zhu, J. (2022). Study on Failure Rate Prediction Method of Batch Smart Meters. Instrumentation. *Analysis. Monitoring, 150*(02), 11-16.

[13] Zhang, C., Zhao D., & He, X. (2021). Research on cost-sensitive fault prediction model for smart meters. *Electrical Measurement and Instrumentation, 58*(3), 180-184.

[14] Du, S., Cheng, D., & Qiu, W. (2018). Short-term fault evaluation and prediction of energy meter based on hierarchical Bayesian. *Journal of Electronic Measurement and Instrumentation, 32*(12), 178-184.

[15] Peng, T., Shen, H., Zhang, Y., Ren, P., Zhao, J., & Jia, Y. (2022). Status Forecast and Fault Classification of Smart Meters Using LightGBM Algorithm Improved by Random Forest. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2022/3846637

[16] Yuan, Y. & Jia, K. (2015). A Distributed Anomaly Detection Method of Operation Energy Consumption Using Smart Meter Data. *International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 310-313. https://doi.org/10.1109/IIH-MSP.2015.38

[17] Tureczek, A., Nielsen, P. S., & Madsen, H. (2018). Electricity Consumption Clustering Using Smart Meter Data. *Energies, 11*(4), 859. https://doi.org/10.3390/en11040859

[18] Li, N., Zhang, W., Guo, Z., Yuan, T., & Han, L. (2022). Intelligent meter fault prediction based on multi classification machine learning mode. *Electrical and energy efficiency management technology, 2022*(02), 6-11.

[19] Wang, Y., Hou, H., & Yao, Q. (2022). Research on Quality Analysis and Prediction Methods of Intelligent Energy Meters in Operation. *Electrical Measurement and Instrumentation, 59*(04), 34-40.

[20] Chen, Y., Han, T., & Wei, L. (2022). Intelligent Energy Meter Fault Prediction Based on Multi classification Fusion Model. *Electrical Measurement and Instrumentation, 59*(11), 162-168.

[21] Shang, X., Ji, Y., Zhang, J. (2020). Overview of Research on Operational Status Evaluation Technology for Intelligent Energy Meters. *Technical Research Review Electrical Measurement and Instrumentation, 57*(3), 134-141.

[22] Gao, X., Diao, X., Liu, J., Zhang, M., & He, Y. (2019). Model based adaptive selection fusion Multi classification method for intelligent electricity meter faults. *Power Grid Technology, 43*(6), 1955-1961.

[23] He, X., Ai, Q., & Qiu, C. (2019). Data mining for ubiquitous power IoT Overview of System Construction and Exploration of Data-driven Cognitive Framework. *Electrical Appliances Energy Efficiency Management Technology,* (19), 1-14.

[24] Wang, R., Yu, L., & Zhu, B. (2020). Big data and machine learning technology in Application Research in Cable Operation and Maintenance. *Electrical Appliances and Energy Efficiency Management Technology Shu, 12*, 114-118.

Contact information:

Shuhua LIANG, Engineer
Electric Energy Metering Center, Hainan Power Grid Co., Ltd.,
199, Haiyu Middle Line, Longhua District, Haikou City, Hainan Province, China
E-mail: 753378106@qq.com

Changji CHEN, Engineer
Wuzhishan Power Supply Bureau of Hainan Power Grid Co., Ltd.,
WuzhishanYingbin Avenue Hainan Power Grid Wuzhishan Power Supply
Company Building, Wuzhishan City, Hainan Province, China
E-mail: 285264@qq.com

Dalei WU, SN ENGR
(Corresponding author)
Electric Energy Metering Center, Hainan Power Grid Co., Ltd.,
199, Haiyu Middle Line, Longhua District, Haikou City, Hainan Province, China
E-mail: wdl198609@126.com

Longjin CHEN, SN ENGR
Electric Energy Metering Center, Hainan Power Grid Co., Ltd.,
199, Haiyu Middle Line, Longhua District, Haikou City, Hainan Province, China
E-mail: 289208474@qq.com

Qingyao WU, Engineer
Electric Energy Metering Center, Hainan Power Grid Co., Ltd.,
199, Haiyu Middle Line, Longhua District, Haikou City, Hainan Province, China
E-mail: 378700206@qq.com

Ting Ting GU, Engineer
Electric Energy Metering Center, Hainan Power Grid Co., Ltd.,
199, Haiyu Middle Line, Longhua District, Haikou City, Hainan Province, China
E-mail: 281509087@qq.com