Mathematical Biosciences and Engineering

Research article

Research on fault diagnosis technology of heat meter based on multi classifier fusion of pigeon swarm algorithm

Shuchun Yu[1], Jinjian Tao[1], Jun Liu[2] and Yanshu Miao[3,4,*]

[1] China Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentation of Heilongjiang Province, Harbin University of Science and Technology, Harbin 150080, China

[2] College of Civil and Architectural Engineering, Heilongjiang Institute of Technology, Harbin, Heilongjiang 150001, China

[3] School of Architecture, Harbin Institute of Technology, Harbin 150001, China

[4] Ministri Ind & Informat Technol, Key Lab Cold Reg Urban & Rural Human Settlement E, Harbin, Heilongjiang 150001, China

* Correspondence: Email: miaoyanshu2022@163.com.

Abstract: In order to improve the availability of fault data, the fault data of heat meters had been classified, and balances all kinds of fault data according to interpolation algorithms to meet the needs of fault diagnosis algorithms. Based on the voting mechanism, an integrated model of multi classifier fusion is established, and the weight of each classifier is optimally configured through pigeon swarm algorithm. In the experiment, three kinds of integration models are obtained according to the voting mechanism and pigeon swarm algorithm. The three integrated models are used to diagnose the fault of the heat meter, and the three indicators of precision, recall and F1 Core have achieved satisfactory results.

Keywords: heat meter; fault diagnosis; integrated model; pigeon swarm algorithm

1. Introduction

Heat supply is an important issue related to the national economy and people's livelihood, and the accuracy of heat energy consumption statistics is more related to the income of heating

companies [1]. In order to improve the accuracy of heat energy consumption statistics, smart heat meters have been widely used and play a very important role between users and heating companies.

Because the design and use of intelligent heat meters involve multiple functions such as heat conversion, metering, communication, etc., there are also many possible types of failures [2,3]. If the fault type of the intelligent heat meter cannot be identified quickly and accurately, it cannot be repaired in time, which will not only cause economic losses to the heating company, but also lead to conflicts between the heating company and users. Therefore, fast and accurate fault diagnosis of intelligent heat meters is of great significance in the field of thermal energy.

At home and abroad, extensive and in-depth research has been carried out on the fault diagnosis of various intelligent instruments, including heat meters. Machine learning concepts and common methods have been widely used. Koziy believes that the fault diagnosis of intelligent instruments can be made by statistically classifying the common fault types of intelligent instruments, establishing a fault database, and comparing the unknown faults with the fault types in the database [4]. Namaki uses Bayesian model to predict the fault type of intelligent instrument, but this method relies on expert experience, and the conditional probability of each node in the Bayesian model is difficult to calculate [5]. Lyashevska believes that the primary problem of fault diagnosis is to solve the problem of data imbalance, and accordingly he proposes an intelligent classification method based on CNN network [6]. Yagi uses the random forest model to build the fault diagnosis method of intelligent instrument respectively, obtaining high recognition accuracy, but there are few fault types that can be identified [7]. In order to further improve the accurate judgment of intelligent instrument fault types, a new idea of multi method fusion began to emerge [8–10]. Sonne fused the random forest model and support vector machine algorithm, and the new fusion method greatly improved the accuracy of fault identification of intelligent instruments [11]. Yy adopted the fusion method of Bayesian model and binary tree model to improve the accuracy of fault identification and expand the capacity of fault type library [12]. Zhou analyzed and modeled the possibility of fault diagnosis from the perspective of cognitive uncertainty and random uncertainty, and used the deep learning network based on probabilistic Bayesian estimation for fault diagnosis, which achieved good results [13]. Gompel diagnosed six types of faults in the photovoltaic system, extracted fault features according to the principle of machine learning, and then incorporated them into the recurrent neural network for analysis and decision-making, achieving a high accuracy of fault identification [14].

According to the analysis of existing research on fault diagnosis, uncertainty and randomness are the difficulties of fault diagnosis, and faults caused by different reasons will also bring large differences between fault data. For this reason, this paper first extracts the features of the fault data of the heat meter, and then achieves the fault data balance through the interpolation algorithm to avoid the impact of the difference of fault data on the subsequent fault diagnosis. Secondly, for a variety of possible faults, corresponding classifiers are established, and then the rationality of different classifier combinations is determined by voting mechanism. Finally, the weight of the classifier determined to be retained is set according to the pigeon swarm algorithm to enhance the rationality and accuracy of the fault diagnosis results. In this paper, a multi classifier fusion model based on voting mechanism and pigeon swarm algorithm is established to achieve more accurate and efficient fault diagnosis of heat meters.

2. Fault data analysis and balance processing of heat meter

2.1. Fault data analysis of heat meter

In the heating system, the intelligent heat meter transmits data back to the heat energy control system through remote means. When the heat meter fails, it mainly shows data error or abnormality. From the perspective of fault data, the common faults of heat meters include the following types: faults caused by electrical hardware, faults caused by mechanical hardware, faults caused by communication problems, faults caused by storage problems, faults caused by billing software, faults caused by metering software, and faults caused by other reasons.

According to the statistics in the industry, the percentage of these seven types of failures is shown in Table 1.

**Table 1.** Proportion of type 7 faults of heat meter.

Order number	Fault type	Proportion
1	Faults caused by electrical hardware	7.7%
2	Faults caused by mechanical hardware	2.3%
3	Faults caused by communication problems	33.6%
4	Faults caused by storage problems	9.8%
5	Faults caused by billing software	13.2%
6	Faults caused by metering software	17.0%
7	Faults caused by other reasons	16.4%

It can be seen from the data in Table 1 that among the reasons for the failure of the heat meter, the failure caused by communication problems accounts for the highest proportion, reaching 33.6%; The faults caused by metering and billing software ranked second and fourth respectively, accounting for 17% and 13.2% respectively, which are also the faults most likely to cause disputes between users and thermal energy companies; The failure caused by other reasons accounted for 16.4%, which also shows the complexity of the failure causes of the heat meter.

2.2. Balance processing of fault data of heat meter

In order to achieve the accurate diagnosis of the heat meter fault, it is necessary to train the diagnosis model according to the heat meter fault data. However, due to the different probabilities of the seven types of faults, the amount of data for some fault types is too small, resulting in data imbalance between the types of faults with high incidence. Therefore, interpolation algorithm is used to increase the balance between different types of fault data.

The principle of data balance algorithm based on interpolation is shown in Figure 1.
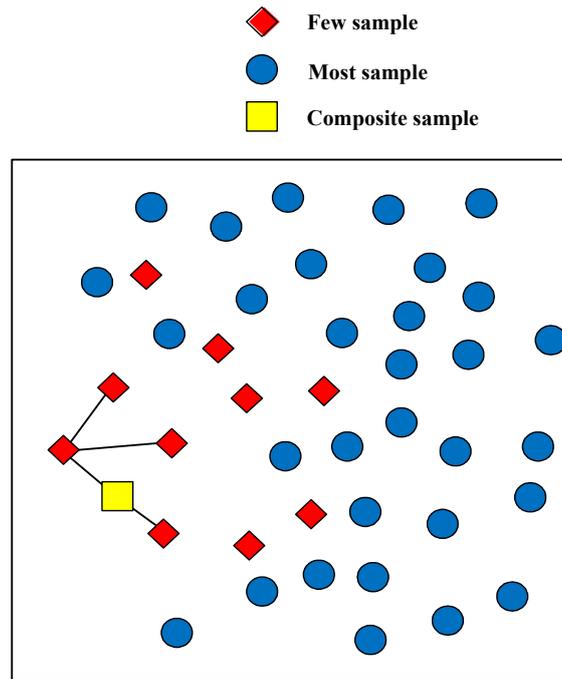
Figure 1. Principle of interpolation based data balancing algorithm.

As shown in Figure 1, the implementation process of the interpolation based data balancing algorithm is as follows:

Firstly, the set of a few data is set as *X*, and there are *n* samples in this set.

Secondly, sample $x_i(i \in [1, n])$ is selected from a small number of data sets as the root sample for synthesizing new samples.

Thirdly, according to the upward sampling magnification *n*, select an odd number *k* (for example, *k* = 3), and use k neighborhood samples x_{ij} as auxiliary samples for synthesizing new samples, where $x_{ij} \in X$, *j*=1, 2, ..., *k*.

Finally, a new sample is generated by interpolation between the root sample x_i and the auxiliary sample x_{ij}. The execution formula of interpolation processing is as follows:

$$x_{new} = x_i + \gamma * \left| x_{ij} - x_i \right| \tag{1}$$

Here, x_{new} represents a new sample, γ Represents a random number that takes a value on the 0–1 interval.

3. Fault diagnosis method of heat meter based on voting mechanism and pigeon swarm algorithm

3.1. Integrated learning based on voting mechanism

The intelligent diagnosis of heat meter faults is generally solved from the perspective of machine learning. In order to improve the accuracy of fault diagnosis, integrated learning is a better

choice. The basic principle of ensemble learning is to fuse multiple classifiers to obtain a better ensemble model than a single classifier.

The integrated learning framework designed in this paper is based on the voting mechanism. The training set is trained through multiple base classifiers, and the test set results of each base classifier are fused according to the voting mechanism to determine the final output. The voting based integrated learning framework is shown as Figure 2.
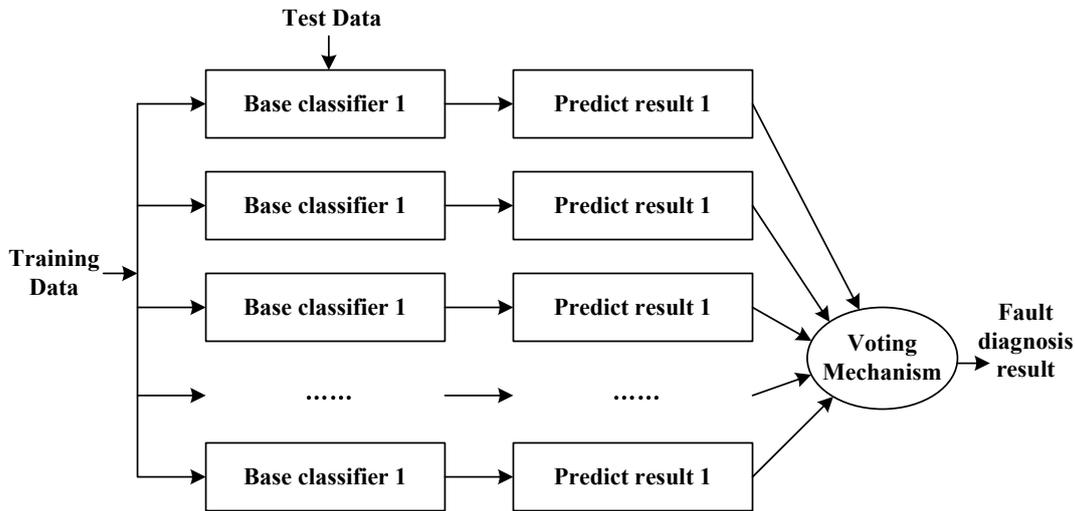
Figure 2. Integrated learning framework based on voting mechanism.

The voting mechanisms used in this paper are divided into two categories: hard voting mechanisms and soft voting mechanisms. Taking the binary classification problem as an example, the prediction target is $y \in \{-1, 1\}$, There are three models A, B, and C to predict the sample (x_i, y_i). The hard voting mechanism is to vote on the model prediction label. Assume that the prediction results of A, B, and C models are 1, 1, and 1, so the final prediction result is 1. The soft voting mechanism refers to weighted voting on the probabilities of each category predicted by each base model. The highest category probability is the final result, and its calculation is list in Eq (2):

$$\begin{cases} E_k = \dfrac{\sum p_k w_i}{\sum w_i} \\ y_i = \max(E_k) \end{cases} \tag{2}$$

Here, P_k represents the probability that the fault predicted by the base classifier is category k, $k = 1, 2, ..., m$, m represents the total number of fault categories, w_i represents the weight of each base classifier, $i = 1, 2, ..., n$, n represents the total number of base classifiers, and E_k represents the probability that the predicted fault category is k after weighted voting, and y_i is the final result.

3.2. Combination of base classifiers based on diversity criteria

The integrated model for fault diagnosis of heat meters is the result of the combination of

multiple base classifiers. The standard for building an integrated model is that the recognition accuracy of fault types is high, and the diversified combination of base classifiers should be achieved, that is, multiple base classifiers should be used as much as possible.

Taking a binary classification problem as an example, this paper analyzes the difference of the integration effect when multiple base classifiers are combined. For the binary classification problem, the prediction target is $y \in \{-1,1\}$, the mapping relationship of the function is f, and the error of the classifier is ε, Then the error relationship of all classifiers is:

$$P(h_i(x) \neq f(x)) = \varepsilon \tag{3}$$

The T classifiers are processed according to the voting mechanism, and the error relationship is as follows:

$$P(H(x) \neq f(x)) = \sum_{k=0}^{T/2} \binom{T}{K}(1-\varepsilon)^k \varepsilon^{T-k} \leq \exp(-\frac{1}{2}T(1-2\varepsilon)^2) \tag{4}$$

Here, $h_i(x)$ represents the ith classifier, and $i = 1, 2, ..., T$, T represents the total number of classifiers.

It can be seen from the above formula that the error of integrated learning decreases with the increase of T. The contrast effect of different base classifiers after integration is shown in Table 2.

Table 2. Comparison effect of different base classifiers after integration.

	Test case 1	Test case 2	Test case 3
Base classifier 1	True	True	False
Base classifier 2	True	False	True
Base classifier 3	False	True	True
Integration model A	True	True	True
Base classifier 4	True	True	False
Base classifier 5	True	True	False
Base classifier 6	True	True	False
Integration model B	True	True	False
Base classifier 7	True	False	False
Base classifier 8	False	True	False
Base classifier 9	False	False	True
Integration model C	False	False	False

From the comparison results in Table 2, we can see that the prediction rate of base classifier 1, base classifier 2 and base classifier 3 is 66.7% respectively, and they have different performance in different test cases, so the effect of integrated model A reaches 100%; Although the prediction rate of base classifier 4, base classifier 5 and base classifier 6 is 66.7%, the performance of the three is the same and they do not meet the diversification conditions, which leads to the failure of integrated model B; The performance of base classifier 7, base classifier 8 and base classifier 9 is poor, and the integrated model C is discarded directly.

3.3. Weight setting of base model based on pigeon swarm algorithm

The accuracy of the diagnosis results of the fault diagnosis model is closely related to the weight of each base classifier in decision-making. In order to obtain the reasonable weight allocation of each base classifier, this paper uses pigeon swarm algorithm to optimize the weight.

Pigeon colony algorithm is different from genetic algorithm and ant colony algorithm in that it performs global optimization with relatively clear objectives. For the fault diagnosis model proposed in this paper, the possible fault causes of the heat meter and the classifier determined in Section 3.2 are clear, so the pigeon swarm algorithm is more appropriate.

Pigeon swarm optimization algorithm is inspired by the special navigation behavior in the homing process of pigeons.

Set the weight of each base classifier as a pigeon to form a pigeon group. During initialization, the position and speed of pigeons are set randomly. After the implementation of the pigeon swarm algorithm, each pigeon in the pigeon swarm obtains the fitness value through the optimization function. In the iterative process of the pigeon swarm algorithm, each pigeon searches its own individual optimal value P_{best}, and also searches the global optimal value G_{best} of the pigeon swarm through sharing. In the pigeon group, all pigeons constantly adjust their position and speed according to the individual optimal value and the global optimal value until all pigeons reach or approach the optimal result.

In a pigeon group consisting of n pigeons, the position of a single pigeon (the weight of the base classifier in this paper) w_i can be expressed as a D dimension vector, namely:

$$w_i = (w_{i1}, w_{i2}, \Lambda, w_{iD}), i = 1, 2, \Lambda, n \tag{5}$$

The flight speed of each pigeon can be expressed as:

$$V_i = (v_{i1}, v_{i2}, \Lambda, v_{iD}), i = 1, 2, \Lambda, n \tag{6}$$

The current individual optimal value of each pigeon can be expressed as:

$$p_{best} = (p_{i1}, p_{i2}, \Lambda, p_{iD}), i = 1, 2, \Lambda, n \tag{7}$$

The current global optimal value of the whole pigeon group can be expressed as:

$$g_{best} = (p_{g1}, p_{g2}, \Lambda, p_{gD}) \tag{8}$$

In the iterative process of pigeon swarm algorithm, the velocity update formula of each pigeon can be expressed as:

$$v_{id} = v_{id} \times e^{-R} + rand \times (g_d - w_{id}) \tag{9}$$

Here, R represents the map parameter, $rand$ is a random value within the range of 0–1, and g_d represents the dth dimensional component of the global optimal value.

In the iterative process of pigeon swarm algorithm, the position update formula of each pigeon can be expressed as:

$$w_{id} = w_{id} + rand \times \frac{1}{1 + e^{v_{id}}}$$ (10)

3.4. Pseudocode of proposed algorithm

In order to realize the fault diagnosis of the heat meter data, the proposed algorithm involves data equalization, classifier selection, classifier weight determination and other aspects. Here, the pseudo code of the entire algorithm is given as follows:

///

The Algorithm for Fault Diagnosis
1: Data equalization based on interpolation by Eq (1);
2: Classifier selection based on voting mechanism by Eq (4);
3: Generate Random Pigeon Swarm by Eq (5);
4: Compute Fitness;
5: Compute Local Optimal Position by Eq (7):
6: Compute Global Optimal Position by Eq (8);
7: While Iterative_error > Threshold
8:          Update Position Particle;
9:          Update Local Optimal Position;
10:         Update Global Optimal Position;
11: Output Weight of Each Classifier;
12: Calculate the judgment result of the fault type;

///

4. Experimental results and analysis

In order to verify the effectiveness of the heat meter fault diagnosis method proposed in this paper, the fault data set of heat meters in Heilongjiang Province is taken as the processing object, and 10,000 of them are selected as sample data. Among the 10,000 samples data, 66% of them are used as the training data set, and 34% of them are used as the test data set.

4.1. Selection experiment of base classifier

At present, in the field of fault diagnosis of intelligent instruments, many mature methods can be used as the base classifier, which needs to be selected through experiments. In this paper, a total of eight common fault diagnosis methods are selected as the candidates for the base classifier, namely: NB (Naive Bayes) algorithm, SVM (Support Vector Machines) algorithm, LR (Logistic Regression) algorithm, DT (Decision Tree) algorithm, BP (Backpro Pogation) algorithm, RF (Random Forest) algorithm, KNN (K-Nearest Neighbor) algorithm, GBDT (Gradient Boosting Decision Tree) algorithm. According to the classification criteria for seven types of common faults of heat meters given in this paper, the diagnostic accuracy of the above eight algorithms for each type of fault is shown in Table 3.

Table 3. Accuracy of eight kinds of algorithms for common faults of heat meters (%).

Algorithm	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5	Fault 6	Fault 7	Mean
NB	1.25	1.43	20.31	22.37	3.25	0.13	50.72	14.21
SVM	0.18	0.72	6.93	93.22	9.17	47.38	50.53	29.73
LR	0.09	48.90	21.22	76.33	0.15	31.24	49.64	32.51
DT	19.02	63.26	71.93	79.08	5.32	29.31	66.49	47.77
BP	30.44	73.21	69.53	61.30	19.83	63.44	76.52	56.32
RF	17.54	92.20	73.43	81.69	9.04	68.31	72.04	59.18
KNN	53.66	88.31	62.55	73.64	50.61	63.78	83.09	67.95
GBDT	43.26	55.68	90.35	48.53	81.22	72.69	87.51	68.46

In order to intuitively observe the diagnosis results of 8 types of methods for 7 types of faults of the heat meter, the data in Table 3 are drawn in the form of curves, as shown in Figure 3.
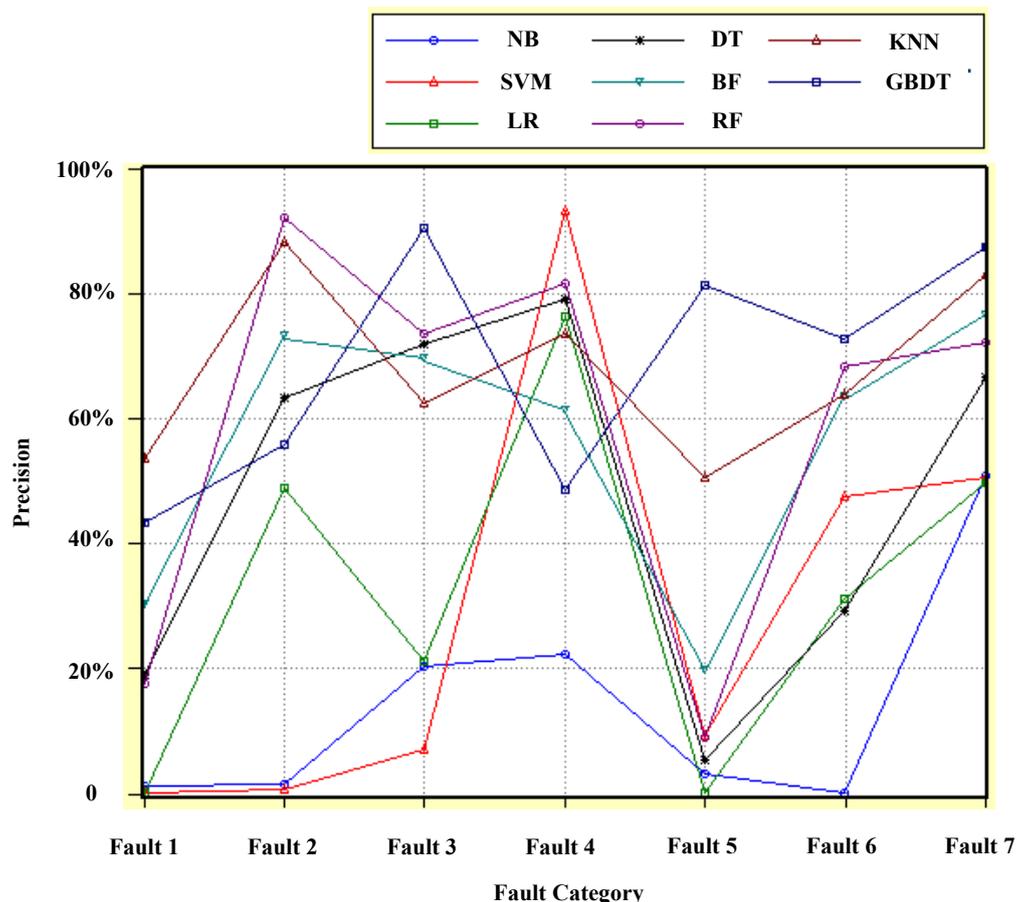
Figure 3. Curve of diagnostic accuracy of eight kinds of algorithms for common faults of heat meters.

From the results in Table 3 and Figure 3, it can be seen that the GBDT method has the best average diagnostic accuracy rate for seven types of faults of the heat meter, reaching 68.46%. Next is KNN method. The average diagnostic accuracy of seven types of faults of the heat meter reaches 67.95. Therefore, these two methods are selected as the first choice and the second choice of the three base classifiers. Then DT method, BP method and RF method are used as the three choices of the base classifier respectively, thus forming three integrated models for the failure diagnosis of heat meters, namely:

Integrated model A for fault diagnosis of heat meter: GBDT + KNN + DT;

Integrated model B of heat meter fault diagnosis: GBDT + KNN + BP;

Thermal energy meter fault diagnosis integrated model C: GBDT + KNN + RF.

The above three fault diagnosis models will be used in subsequent fault diagnosis experiments for further fault diagnosis effect evaluation. It should be pointed out here that if KNN algorithm or GBDT algorithm is selected separately, because the classifier is single, for some types of heat meter faults, the detection results will be obviously misjudged. The integration of the three classifiers according to the method in this paper will greatly reduce the possibility of miscalculation.

4.2. Evaluation index of fault diagnosis performance

The effect of fault diagnosis is closely related to the fault classification. Generally, confusion matrix method is used to judge whether the classification is accurate. The composition of confusion matrix is listed in Table 4.

Table 4. Composition of confusion matrix.

Confusion matrix		True value	
		P	N
Predictive value	P	TP	FP
	N	FN	TN

In Table 4, TP represents the number of samples that are both actually positive and predicted; FP represents the number of samples that are actually negative but are predicted to be positive; FN represents the number of samples that are actually positive but are predicted to be negative; TN indicates the number of samples that are actually negative and the prediction is also negative.

According to the confusion matrix, we can further build three evaluation indicators, namely Precision, Recall and F1 Core.

Precision reflects the model's ability to distinguish negative samples. The higher the value, the stronger the model's ability to distinguish negative samples, which can be expressed as:

$$\mathrm{Pr\,ecision} = \frac{1}{m} \sum_{i=1}^{m} P_i = \frac{1}{m} \sum_{i=1}^{m} \frac{TP_i}{TP_i + FP_i} \tag{11}$$

Recall reflects the model's ability to distinguish positive samples. The higher the value, the stronger the model's ability to recognize positive samples. It can be expressed as:

$$\text{Recall} = \frac{1}{m}\sum_{i=1}^{m} R_i = \frac{1}{m}\sum_{i=1}^{m} \frac{TP_i}{TP_i + FN_i} \tag{12}$$

F1 Score is the comprehensive result of Precision and Recall. The higher the value, the more robust the model is. It can be expressed as:

$$\text{F1} = \frac{1}{m}\sum_{i=1}^{m} \frac{2 * P_i * R_i}{P_i + R_i} \tag{13}$$

Here, *m* represents the total number of fault types; *i* indicates the fault type.

4.3. Experimental results of fault diagnosis for integrated model based on voting mechanism

In Section 4.1, three heat meter based fault diagnosis models are obtained, which integrate three base classifier algorithms respectively. The weight allocation results of the base classifier under each integrated model is shown in Table 5.

Table 5. Weight allocation of base classifier under three integration models.

Method	Base classifier and its weight		
Integration model A	GBDT	KNN	DT
	0.45	0.45	0.10
Integration model B	GBDT	KNN	BP
	0.60	0.30	0.10
Integration model C	GBDT	KNN	RF
	0.50	0.25	0.25

The above three types of integrated models are used to further diagnose the seven common faults of the heat meter, and the accuracy is shown in Table 6.

Table 6. Diagnosis accuracy of three types of integrated models for common faults of heat meters (%).

Algorithm	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5	Fault 6	Fault 7	Mean
Integration model A	80.52	82.36	85.38	84.21	85.23	86.84	87.11	84.52
Integration model B	71.32	86.53	78.61	82.35	88.53	75.63	91.25	82.03
Integration model C	72.19	71.23	74.28	78.51	76.36	77.07	81.28	75.85

In order to intuitively observe the diagnosis results of the three types of integrated models on the seven types of faults of the heat meter, the data in Table 6 are drawn in the form of curves, as shown in Figure 4.
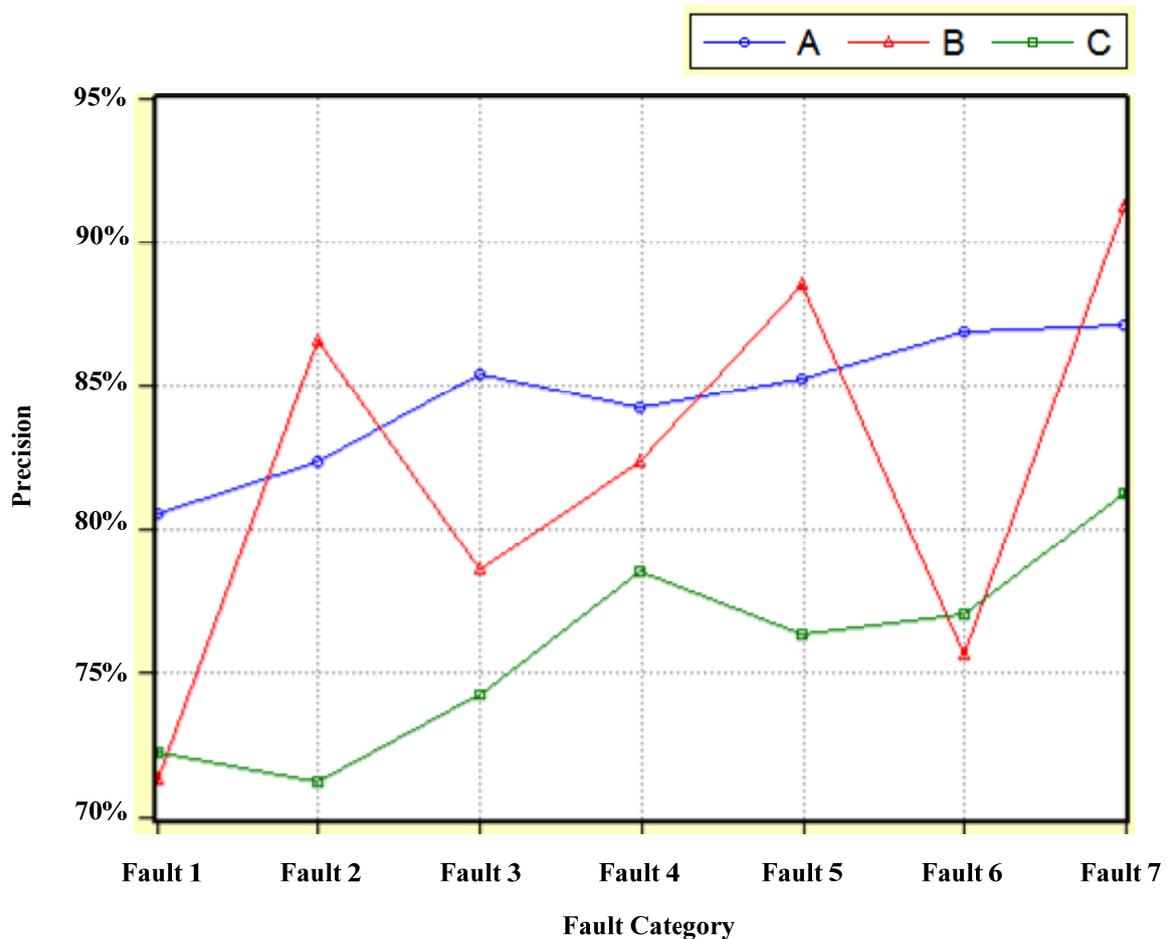
Figure 4. Diagnosis accuracy curve of three types of integrated models for common faults of heat meters.

It can be seen from the results in Table 6 and Figure 4 that the average diagnostic accuracy of integrated model A for seven types of faults of the heat meter is the best, reaching 84.52%. The second is the integrated model B. The average diagnostic accuracy rate of seven types of faults of the heat meter reaches 82.03. Although the diagnostic accuracy of the integrated model C for the seven types of faults of the heat meter is lower than that of the other two types of models, it is also significantly higher than that of the eight single methods in Table 3. This group of experimental results fully proves that the integrated model constructed by the diversified basis classifier is effective for the fault diagnosis of heat meters.

According to the method given in Section 4.2, further compare the Precision, Recall and F1 Core indicators of the three integration models. The results are shown in Table 7.

It can be seen from the results in Table 7 that integration model A shows better results in the three indicators of Precision, Recall and F1 Core. Although the results of integration models B and C are slightly worse than that of integration model A, they also achieve relatively good results.

Table 7. Comparison of fault diagnosis evaluation indicators of three integrated models.

Method	Precision	Recall	F1
Integration model A	0.8452	0.8129	0.7723
Integration model B	0.8203	0.8053	0.7541
Integration model C	0.7585	0.7764	0.7319

4.4. Comparison with other methods

Based on the integrated model A, the method in this paper is applied to the fault diagnosis of heat meters, and probabilistic bayesian network method (PBNM) and recursive neural network method (RNNM) are selected as reference methods to form the fault diagnosis results of the three methods, are shown in Table 8.

Table 8. Accuracy of three kinds of methods for fault diagnosis of heat meters (%).

Algorithm	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5	Fault 6	Fault 7	Mean
Our Method	80.52	82.36	85.38	84.21	85.23	86.84	87.11	84.52
PBNM	76.34	80.21	81.53	81.48	83.22	79.64	84.13	81.39
RNNM	75.88	79.27	81.36	80.48	81.29	80.57	80.36	80.31

It can be seen from the comparison results in Table 8 that the accuracy of this method for the diagnosis of seven types of common faults of heat meters is significantly higher than that of probabilistic bayesian network method (PBNM) and recursive neural network method (RNNM).

5. Conclusions

This paper focuses on the fault diagnosis of heat meters. First of all, the common faults of the heat meter are classified, including those caused by electrical hardware, mechanical hardware, communication, storage, billing software, metering software, and other reasons. On this basis, the method based on interpolation is adopted to eliminate the imbalance of various fault data. An integrated model of multi classifier fusion is established based on voting mechanism. The weight of each classifier is optimally configured through pigeon swarm algorithm to realize the fault diagnosis of the heat meter. During the experiment, according to the diagnostic effects of eight classifier algorithms, three integration models are constructed, namely: Model A, GBDT + KNN + DT; Model B, GBDT + KNN + BP; Model C, GBDT + KNN + RF. The experimental results show that the integrated model proposed in this paper has greatly improved the accuracy of fault diagnosis of the heat meter, and has also received higher evaluation in the three indicators of Precision, Recall and F1 Core.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. M. H. Abokersh, M Vallès, L. F. Cabeza, D. Boer, A framework for the optimal integration of solar assisted district heating in different urban sized communities: A robust machine learning approach incorporating global sensitivity analysis, *Appl. Energy*, **267** (2020), 114903. https://doi.org/10.1016/j.apenergy.2020.114903

2. I. Mitiche, A. Nesbitt, S. Conner, P. Boreham, G. Morison, 1D-CNN based real-time fault detection system for power asset diagnostics, *IET Gener. Transm. Distrib.* **14** (2020), 5766–5773. https://doi.org/10.1049/iet-gtd.2020.0773

3. M. Biswal, S. Ghore, O. P. Malik, R. C. Bansal, Development of time-frequency based approach to detect high impedance fault in an inverter interfaced distribution system, *IEEE Trans. Power Delivery*, **36** (2021), 3825–3833. https://doi.org/10.1109/TPWRD.2021.3049572

4. K. Koziy, B. Gou, J. Aslakson, A low-cost power-quality meter with series arc-fault detection capability for smart grid, *IEEE Trans. Power Delivery*, **28** (2013), 1584–1591. https://doi.org/10.1109/TPWRD.2013.2251753

5. O. Namaki-Shoushtari, B. Huang, Bayesian control loop diagnosis by combining historical data and process knowledge of fault signatures, *IEEE Trans. Ind. Electron.*, **62** (2015), 3696–3704. https://doi.org/10.1109/TIE.2014.2375253

6. O. Lyashevska, F. Malone, E. Maccarthy, J. Fiehler, J. Buhk, L. Morris, Class imbalance in gradient boosting classification algorithms: Application to experimental stroke data, *Stat. Methods Med. Res.*, **30** (2021), 916–925. https://doi.org/10.1177/0962280220980484

7. A. Yagi, L. Nguyen, T. V. Ta, A sustainability condition for stochastic forest model, *Commun. Pure Appl. Anal.*, **16** (2017), 699–718. https://doi.org/10.3934/cpaa.2017034

8. B. de Silva, J. Callaham, J. Jonker, N. Goebel, J. Klemisch, D. McDonald, et al., Hybrid learning approach to sensor fault detection with flight test data, *AIAA J.*, **59** (2021), 3490–3503. https://doi.org/10.2514/1.J059943

9. J. Hage, P. Xu, P. Bonnifait, J. Ibanez-Guzman, Localization integrity for intelligent vehicles through fault detection and position error characterization, *IEEE Trans. Intell. Transp. Syst.*, **23** (2020), 2978–2990. https://doi.org/10.1109/TITS.2020.3027433

10. M. Ma, P. He, Y. Li, H. Li, M. Jiang, Y. Wu, Fault diagnosis method based on multi-source information fusion for weak interturn short circuit in synchronous condensers, *IET Electr. Power Appl.*, **15** (2021), 1245–1260. https://doi.org/10.1049/elp2.12094

11. J. Sonne, C. Seavey, J. Groshong, Rapid immunohistological measurement of tyrosine hydroxylase in rat midbrain by near-infrared instrument-based detection, *arXiv preprint*, (2021). https://doi.org/10.1016/j.jchemneu.2021.101992

12. Y. Yamazaki, S. Kanaji, T. Matsuda, T. Oshikiri, T. Nakamuraet, S. Suzuki, et al., Automated surgical instrument detection from laparoscopic gastrectomy video images using an open source convolutional neural network platform, *J. Am. Coll. Surg.*, **230** (2020), 725–732. https://doi.org/10.1016/j.jamcollsurg.2020.01.037

13. T. Zhou, T. Han, E. Droguett, Towards trustworthy machine fault diagnosis: A probabilistic Bayesian deep learning framework, *Reliab. Eng. Syst. Saf.*, **224** (2022), 1065–1077. https://doi.org/10.1016/j.ress.2022.108525

14. J. Gompel, D. Spina, C. Develder, Satellite based fault diagnosis of photovoltaic systems using recurrent neural networks, *Appl. Energy*, **305** (2022), 117874. https://doi.org/10.1016/j.apenergy.2021.117874