1052. Ensemble learning-based intelligent fault diagnosis method using feature partitioning

Yongsheng Zhu[1], Xiaoran Zhu[2], Jing Wang[3]
School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China
[3]Corresponding author
E-mail: [1]*yszhu@mail.xjtu.edu.cn*, [2]*lecijian@gmail.com*, [3]*wangpele@gmail.com*

Abstract. Diversity and fusion strategy are the key factors which affect the performance of the ensemble learning systems. In this paper, to tackle the neighborhood factor selecting difficulty of the traditional neighborhood rough set method, the wrapper feature selection algorithm based on kernel neighborhood rough set is introduced to find a set of feature subsets with high diversity, and then a base classifier selection method is proposed for constructing the ensemble learning systems. To increase the diversity, the heterogeneous ensemble learning algorithm based on the proposed base classifier selection method is designed and compared with the similar homogeneous ensemble learning algorithm. To study the effect of the fusion strategy on the final performance of the ensemble learning system, majority voting and D-S theory for fusing the outputs of base classifiers of ensemble learning system to get final decision are compared experimentally. The results on UCI data and the fault signals of rotor-bearing system show that the heterogeneous ensemble learning system with D-S fusion strategy can get the best classifying performance, and the ensemble learning system is superior to single classification system in most cases.

Keywords: intelligent fault diagnosis, homogeneous ensemble algorithm, heterogeneous ensemble algorithm.

1. Introduction

The essence of mechanical equipment fault diagnosis is the pattern recognition, therefore various intelligent classification methods have been developed and adopted to tackle this problem [1, 2]. Recently, the ensemble learning system, which is proved to have better performance than the single classifier, is attracted increasing attention and is applied successfully to the field of fault diagnosis [3, 4]. Although various classifiers for ensemble models have been proposed and studied, pursue on ensemble learning algorithm of higher performance is still a keep-on-going research topic.

The performance of the ensemble learning system is supposed to be better than the single classifier [5], since the ensemble learning system combines the outputs of the several independent base classifiers with a fusion strategy to make the final decision [6, 7] and therefore the mis-decision risk is reduced. Empirically, the ensemble learning system tends to yield better results when there is a significant diversity among the base classifiers. Two methods: data partitioning method and feature partitioning method are normally used to construct the single base classifiers for ensemble system. The data partitioning method, which is to train base classifiers with different sample subsets, mainly includes the bagging [8] and boosting [9, 10]. However, the performance of data partitioning method decreases significantly when the dataset contains small samples. Actually, this is the case in mechanical fault diagnosis field, where the fault samples are always insufficient. In contrast, various kinds of features can be obtained easily from each sample since the development of the signal processing technology, and therefore enough number of features can be prepared even though only limited numbers of samples are available. For this reason, the feature partitioning method is expected to achieve better results and is more preferable for fault diagnosis applications than data partitioning method.

Feature partitioning method divides the feature set into a set of feature subsets, and then combines the results of base classifiers that are trained with the feature subsets. For example, in 1998, the random subspace method was introduced by Ho [11]. In 2003, Robert Bryll et al.

proposed a feature bagging based on the random subsets of features [12]. In 2005, Oliveira et al. proposed an ensemble feature selection approach based on the hierarchical multi-objective genetic algorithm [13]. Intuitively, the feature partitioning method tries to produce diversity via using different feature subsets, so the key factor affecting the performance of feature partitioning method is how to generate a set of feature subsets which do not lose the distinguishing information [14] while keeping the good diversity properties. Normally, the subsets presented to multiple classifiers may be produced by employing the feature selection method [15]. A well-designed feature selection algorithm would significantly improve the performance of feature partitioning ensemble system. The rough set (RS, rough set), which was proposed by Pawlak [16], has attracted much attention from machine learning and data mining fields and recently introduced in ensemble learning system [17, 18]. In the view of RS, different reducts can be generated from the initial feature set by adjusting the controlling factor of RS. The reducts are in fact the feature subsets which keep the approximation ability of the initial total features but contain less number of features. So, the base classifiers trained with the obtained reducts, are expected to improve or maintain the similar classification performance as the classifier trained with the initial features. In addition, the difference among reducts can provide diversity among the obtained base classifiers. Therefore, feature partitioning ensemble system based on rough set is supposed to obtain better generalization ability.

Several ensemble learning systems based on RS have been proposed. For example, Suraj et al. proposed an approach of multiple classifier system using RS to construct classifier ensemble [19]. In the method, the different combinations of selected reducts but not the recudts themselves were used to train base classifiers, the results show that in most cases, the ensemble systems get better performance than the single classifier system, however, the selection algorithm of reducts is said by author to be complex. In reference [20], Qinghua Hu et al. proposed a FS-PP-EROS algorithm for selective ensemble of rough subspaces. In Hu's method, each obtained reduct was used to train a classifier and then the classifiers were sorted by the accuracy. To construct the ensemble system, the classifier was added into the ensemble system sequentially and the classification performance during the addition was recorded. Finally, the post-pruning is conducted by eliminating the base classifiers which are added after the peak accuracy. Since the performance curve fluctuates with the adding of the base classifiers, the proposed method only eliminated the base classifiers after the peak while kept the base classifiers which cause the performance fluctuation before the peak, and the final performance and efficiency of the obtained ensemble system are possibly affected strongly by these unwanted classifiers.

On the other hand, the traditional Pawlak's RS method was originally proposed to deal with categorical data, and the numerical data such as the features of vibration signals of the mechanical equipment must be reduced by using the Pawlak's RS method after the discretization. However, some important information may be missed from numerical features after discretization [21]. In order to directly deal with the numerical data, the neighborhood rough set was proposed [22, 23, 24]. In practice, it was discovered that the neighborhood value has serious influence on the reducts. In our previous work, the feature selection algorithm based on kernel neighborhood RS method was proposed to solve this problem [25]. In the method, the kernel method and neighbor rough set are combined to design the wrapper feature selection algorithm, and the neighborhood factor was calculated by mapping the data to a high-dimension feature space via the Gaussion kernel function and calculating the hypersphere radius as the neighborhood value. The experimental results shown that the proposed method can obtain reduct of better performance more easily than the traditional RS method. Also, by changing the hypersphere radius, a set of reducts of numerical features may be generated by the kernel neighborhood RS method, and the ensemble learning system based on the kernel neighborhood RS method could be constructed.

In this paper, a new selection method of the base classifiers is designed and applied to both heterogeneous ensemble and heterogeneous ensemble learning systems. At the same time, two fusion strategies of base classifiers (majority voting and D-S theory) are tested. First, the kernel

neighborhood RS method is adopted to obtain feature subsets. Then the homogeneous ensemble algorithm and heterogeneous ensemble algorithm, which are proved to have different classification performance, are designed based on the optimal feature subset and the obtained a set of sub-optimal feature subsets. In both algorithms, the subsets are added into the ensemble learning system one by one, and the performance of the ensemble learning system is evaluated for each addition. The base classifier which contributes to the performance increasing is kept while the one which causes the performance decreasing is ignored. Third, to improve the final classifying accuracy furthermore, the majority voting and D-S theory fusion strategies of ensemble system which are applicable to different types of classifier's outputs of label information and probability information are compared.

The rest of the paper is structured as follows. Basic information about feature selection based on kernel neighborhood rough set is given in Section 2. Section 3 presents the ensemble algorithm. The analysis results are shown in Section 4 and Section 5. The conclusion comes in Section 6.

2. Feature selection based on kernel neighborhood rough set

The basic idea of the feature selection based on kernel neighborhood rough set is mapping datasets into the high-dimensional space using the kernel function. The smallest hypersphere, which contains all datasets, is obtained. The hypersphere radius R may be considered the maximum neighborhood value for δ. Moreover, R is considered the upper bound for δ. The gaussian kernel function is adopted:

$$K(x,y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right).$$ (1)

The kernel parameter value σ affects the hypersphere. Thus, the wrapper feature selection algorithm is designed with $\sigma = 3$ to 4.5 with step 0.05 to determine a set of feature subsets. The feature selection algorithm is shown as Fig. 1. For more detailed information on this algorithm, please review [25].
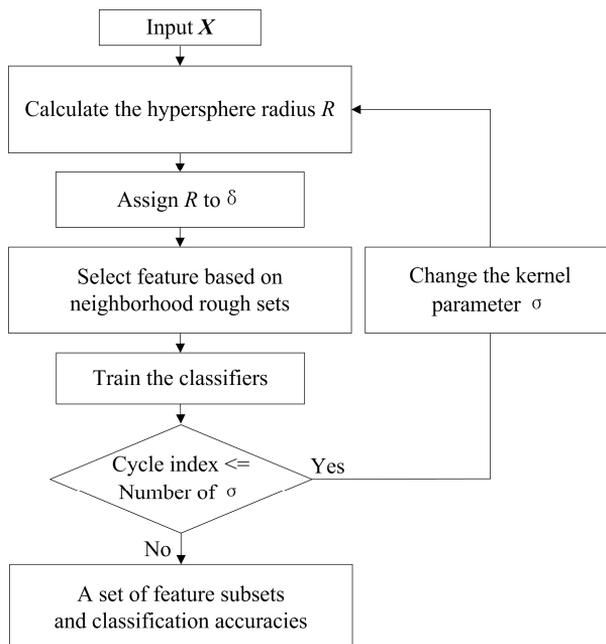
Fig. 1. The flow chart of the feature selection algorithm

3. Ensemble learning algorithms design

3.1. Homogeneous ensemble algorithm

According to the feature selection algorithm introduced in section 2, when the kernel parameter changed, the resultant feature subset might be changed. By using different kernel parameters, a set of feature subsets can be obtained. The feature subsets have two important characteristics. (1) Each of the feature subsets can improve or maintain the classification ability since some redundant or even noised features are ignored during feature selection process. (2) The feature subsets have diversity with each other because each of the feature subsets includes different features. Therefore, the homogeneous ensemble algorithm could be designed as follows. First of all, a series of base classifiers are trained with a set of feature subsets and the corresponding classification accuracies are recorded, then the feature subsets are sorted with descending order according to the classification accuracies. In the first ensemble step, the feature subset with the highest accuracy is selected, and then the next feature subset from the rest in each stage is selected and is added into the homogeneous ensemble algorithm. If the accuracy of the homogeneous ensemble algorithm is improved, the new added feature subset is retained; otherwise, the feature subset is deleted. Finally, a set of optimal feature subsets and the optimal classification accuracy of the homogeneous ensemble algorithm can be obtained. The homogeneous ensemble algorithm is shown as Fig. 2.
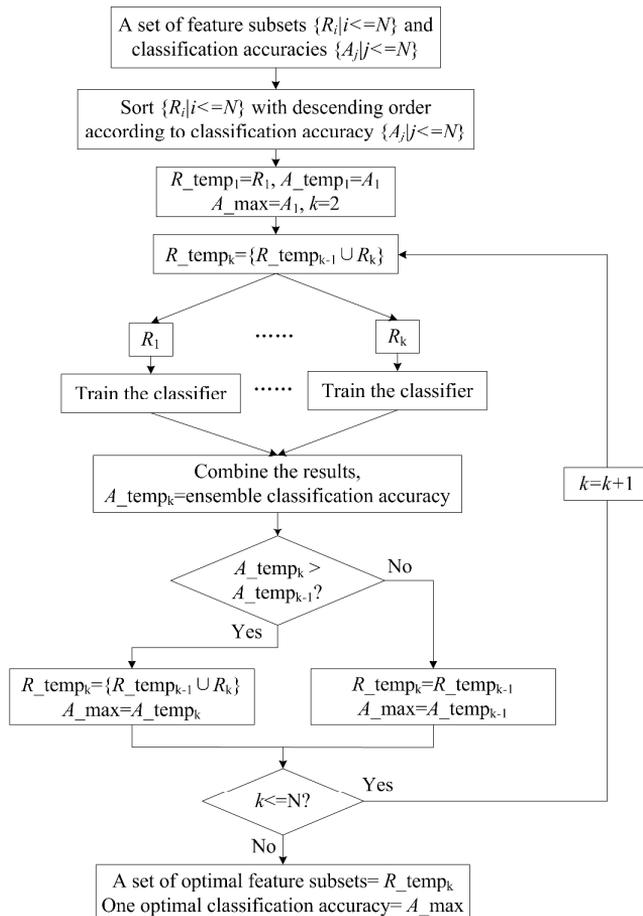
Fig. 2. The flow chart of the homogeneous ensemble algorithm

3.2. Heterogeneous ensemble algorithm

The objective of the heterogeneous ensemble algorithm is to reduce the uncertainty and inaccuracy in classification problem by using the complementary of different base classifiers. In this paper, the basic idea of the heterogeneous ensemble algorithm is very similar to the homogeneous ensemble algorithm, except for that the heterogeneous ensemble system needs two or more base classifiers to integrate. The feature partitioning method refers to making use of the instability of the base classifier and the neural network is subject to the influence of the initialization. Therefore, the BP network and the RBF network are chosen as the base classifier. The heterogeneous ensemble algorithm is shown as Fig. 3.
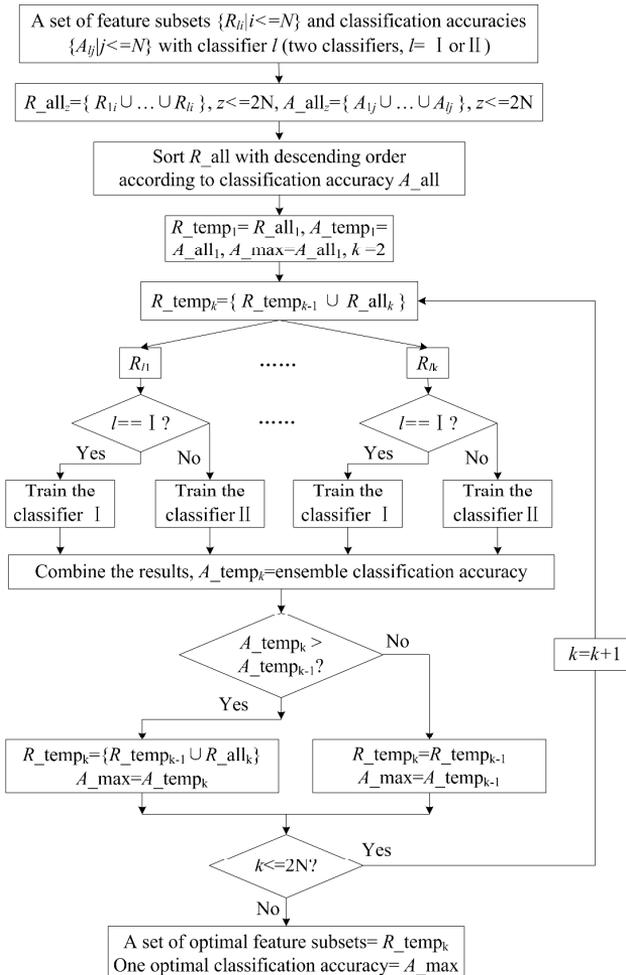
Fig. 3. The flow chart of the heterogeneous ensemble algorithm

3.3. Base classifier and fusion strategy

How to select fusion strategy according to the output type of the base classifier is an important study on ensemble learning system. Generally speaking, the output type of the base classifier is classified into two main categories: label information and probability information [26]. In this paper, two types of base classifiers, the BP network and the RBF network, are used. The label information or the probability information can be exported for neural network. With considering

the output type of the base classifier, selecting the suitable fusion strategy will be a good solution. In this paper, the majority voting is chosen as the fusion strategy for label information and the D-S theory is chosen as the fusion strategy for probability information.

4. Algorithm verification with UCI datasets

The datasets (Diab, Iono, Sonar, Wine, Wpbc) from University of California at Irvine (UCI) [27] are firstly chosen to test the proposed ensemble learning system. The UCI datasets information is shown in Table 1 and the features are normalized into the interval [0; 1]. The feature selection algorithm based on the kernel neighborhood rough sets, which is described in section 2, is applied to the UCI datasets to select features for each dataset. The number of selected features and the corresponding classification accuracy of two kinds of base classifiers are shown in Fig. 4.
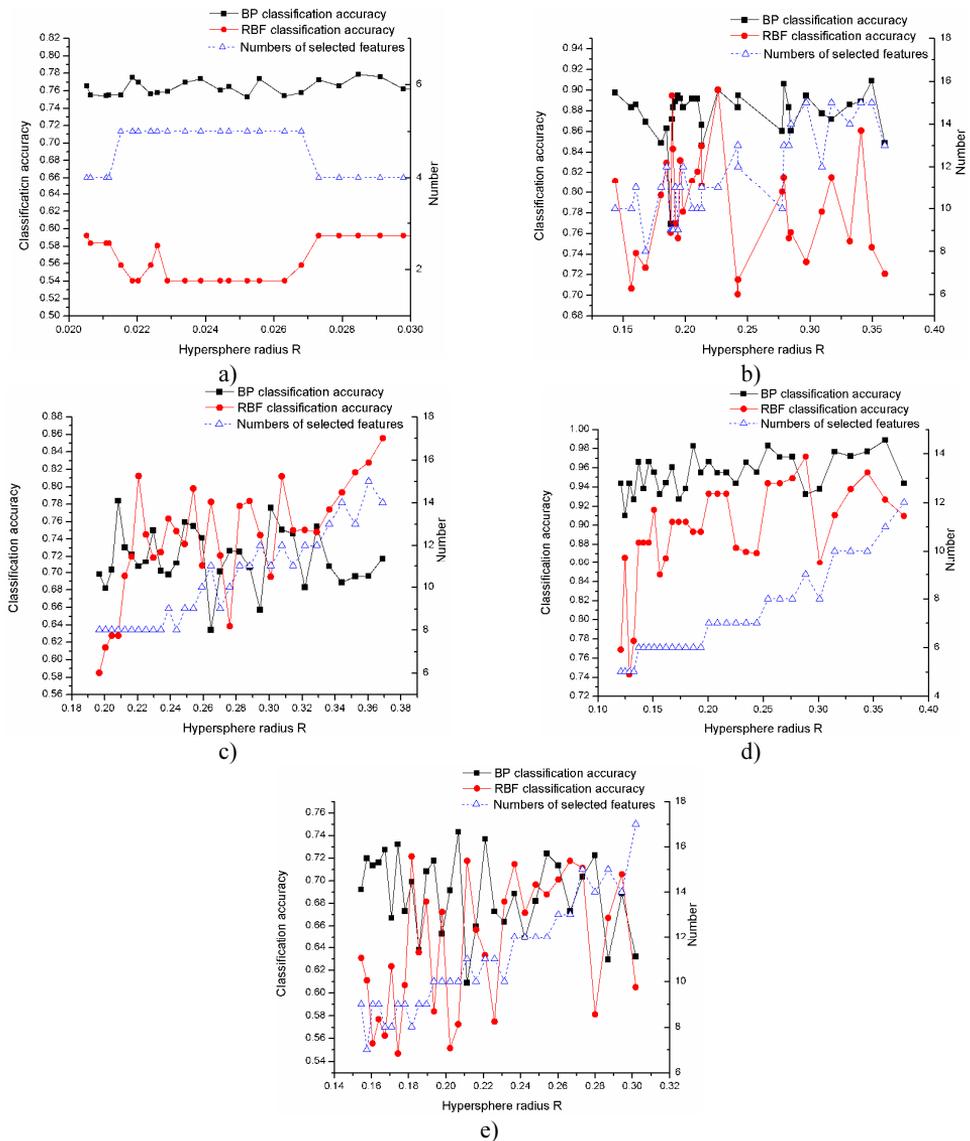
Fig. 4. Variation of classification accuracies and numbers of selected features with hypersphere radius R on (a) Diab, (b) Iono, (c) Sonar, (d) Wine and (e) Wpbc dataset

Table 1. UCI dataset descriptions

Datasets	Samples	Features	Classes
Diab	768	8	2
Iono	351	34	2
Sonar	208	60	2
Wine	178	13	3
Wpbc	198	33	2

Fig. 4 shows that for all datasets, the classification accuracies and the numbers of the selected features vary with the different hypersphere radius. Meanwhile, the numbers of the feature subsets are much less than before, it means that the redundant features can be efficiently deleted by using the feature selection algorithm. Another interesting phenomenon is that even for the same number of feature subsets, the obtained classification accuracies are different apparently, this result further illustrated that the feature subset exerts a serious influence on classification accuracies, and the proposed feature selection algorithms can produce feature subsets of high diversity.

The diversity among the base classifiers also plays an important role in constructing an ensemble learning system. By observing the results shown in Fig. 4, one can easily find that for each dataset, the diversity exists among the same kind of base classifiers which are trained with the different feature subsets, so the homogeneous ensemble algorithm could be developed based on these observation and a better classification performance could be expected. On the other hand, the big differences of the classification accuracy between two types of base classifiers and the fluctuation with the subsets suggests the possibility of improving classification performance by applying the heterogeneous ensemble algorithm.

The ensemble algorithms are conducted on UCI datasets and the results are summarized in Fig. 5. For each dataset, to show the effects of feature selection, the BP network and the RBF network are firstly trained with the original dataset and the selected subsets separately; and then the homogeneous ensemble algorithm is applied to each kind of networks and the heterogeneous ensemble algorithm is applied to combine the results of BP and RBF networks. In each ensemble experiment, both the majority voting and the D-S fusion strategies are tested.

Fig. 5 shows that for the single base classifier, BP networks get better classification accuracy than RBF networks, and the selection process of features improves the classification ability of the base classifier apparently.

a)

b)

c)

d)

e)
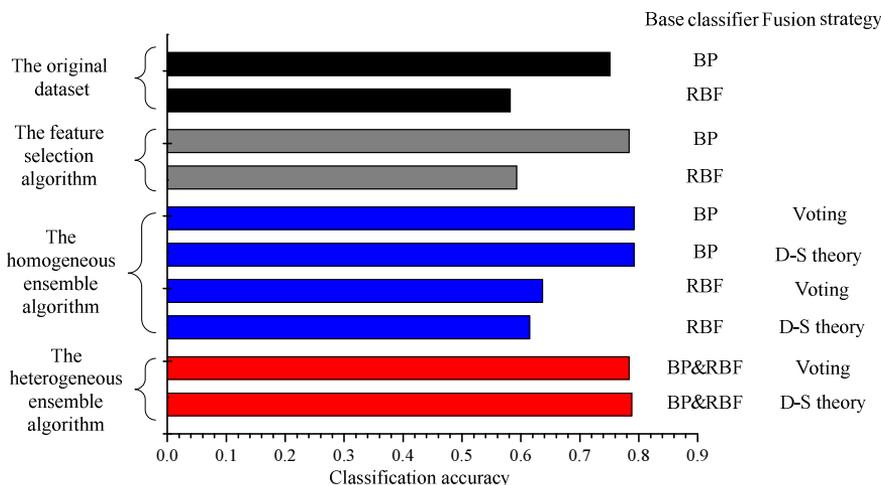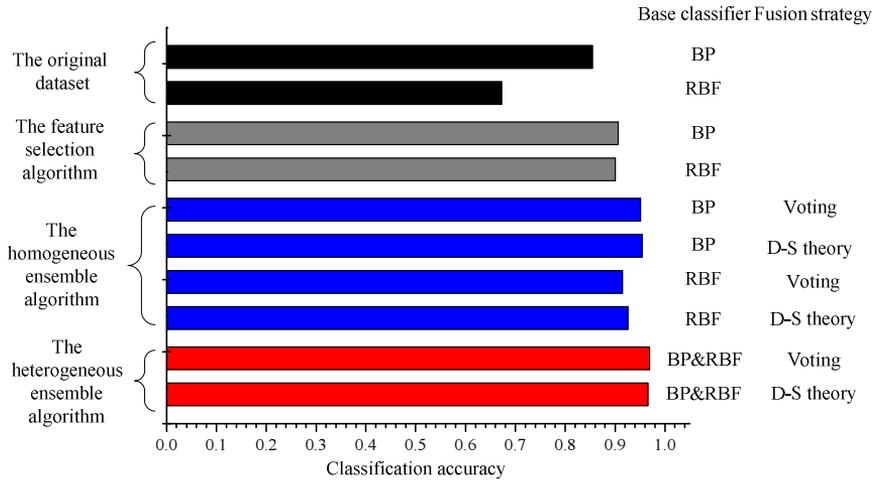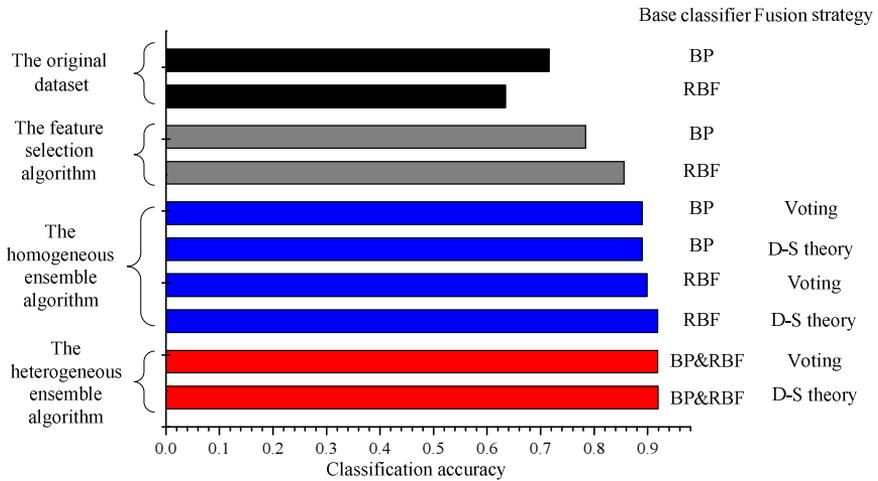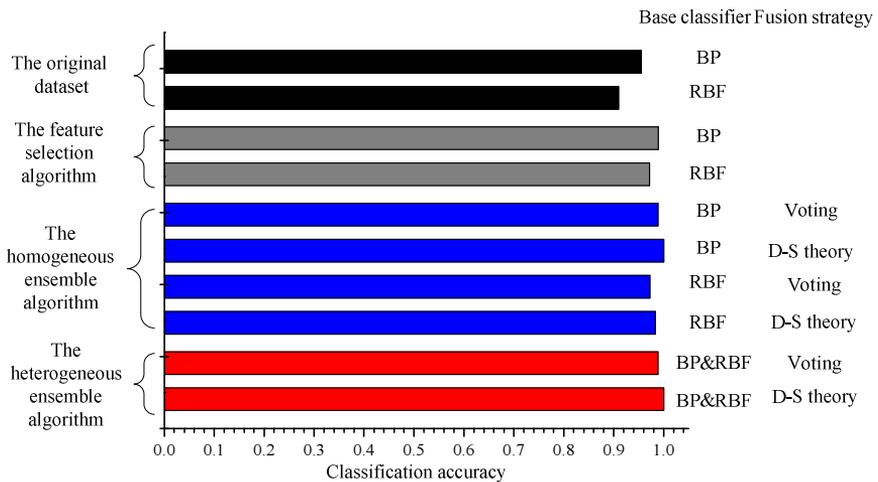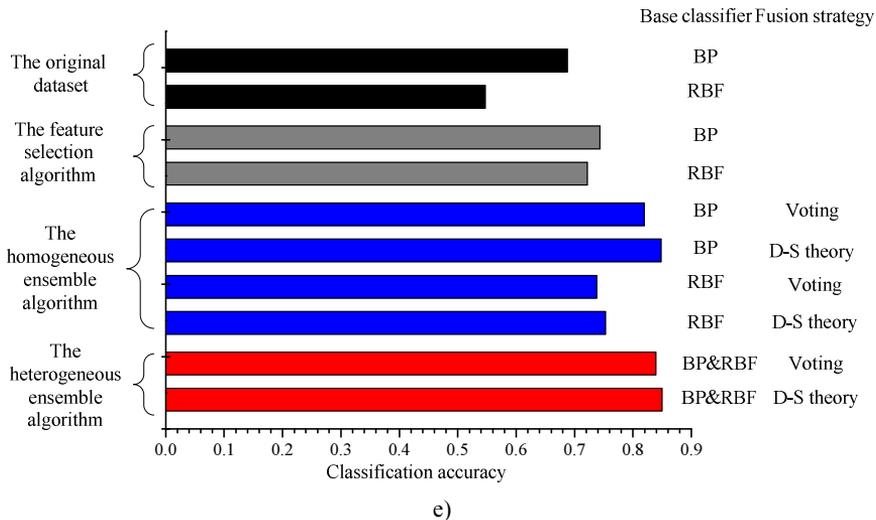
Fig. 5. Comparison of the classification accuracies for the UCI datasets
(a) Diab, (b) Iono, (c) Sonar, (d) Wine and (e) Wpbc dataset

The ensemble learning method, no matter the homogeneous ensemble algorithm or the heterogeneous ensemble algorithm, can improve the classification performance further compared with the single classifier for most datasets. And the performance of base classifiers has strong effect on the final accuracy of ensemble system. For example, if the base BP network has superior performance, the ensemble of BP networks will give a better classification performance than RBF networks and vice versa. By comparing the performance of two types ensemble learning algorithms, one can conclude that as expected, the heterogeneous ensemble algorithm can gain a better classification performance than the homogeneous ensemble algorithm, since it use not only the diversity induced by different feature subsets, but also the diversity existing between the different classifiers.

As for the influence of the fusion strategy on the final classification accuracy, as shown in the Fig. 5, for the homogeneous ensemble, the results indicate that the D-S theory can gain a better classification performance than the majority voting except for the Diab dataset. For the heterogeneous ensemble, the results indicate that the D-S theory can gain a better classification performance except for the Iono dataset. Since D-S theory is applied to the probability information while the majority voting method is applied to the label information, the results suggest that combining the probability information should be better than combining the label information.

5. Experimental result

The test rig layout is shown in Fig. 6. The test rig is driven by a DC motor. A flexible coupling was used between the shaft and the motor to isolate the vibration transmitted from the motor. Some eddy current displacement sensors were mounted on different positions and acquired the vibrations information in horizontal and vertical directions, respectively. In the experiments, the vibration data was collected in four different states of the rotor system, including normal, rubbing, unbalance and misalignment (by adjusting the eccentric bearing) state. The rotational speed is set as 300 r/min.

Because the statistical features of vibration signal contain an amount of information and is very efficient in calculation, the intelligent fault diagnosis based on statistical features has received increasing attention [3, 30]. In this paper, six kinds of statistical features (as shown in Table 2) are extracted from both time domain and frequency domain. To extract features in time domain, raw vibration signals are filtered into low frequency band ($\leq 0.49 f_0$, where the f_0 is rotation

frequency), medium frequency band ($0.5f_0 \sim 1.5f_0$) and high frequency band ($\geq 1.51f_0$) by using filters, and are decomposed into 3 levels with wavelet packet transform (WPT, and the wavelet basis function is db3), and then six statistical features are extracted for each band, each decomposed signal and the raw signal respectively. To extract the features in frequency domain, the power spectrums are calculated for above-mentioned three filtered signals and the original signals, and then six statistical features are extracted from every power spectrum respectively. Finally total 96 features can be obtained. The information about the fault dataset (the dataset name is DALL) is shown in Table 3.

The DC motor The flexible coupling The eddy current displacement sensor The mass disk The simulation of misalignment fault The simulation of unbalance fault The simulation of rubbing fault
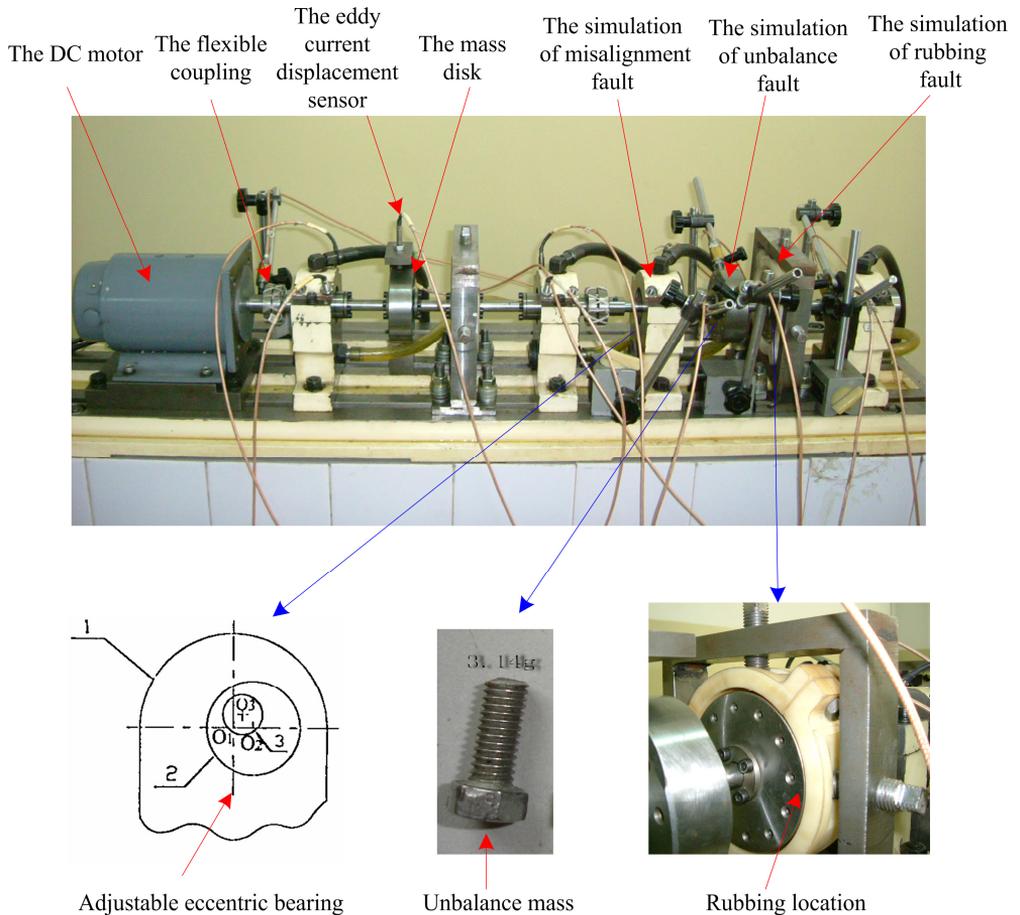
Adjustable eccentric bearing   Unbalance mass   Rubbing location
Fig. 6. The test rig

Table 2. Statistical features

Feature	Equation	Feature	Equation
Skewness	$x_{ske} = \dfrac{1}{n}\sum_{t=1}^{n}(x(t)-\bar{x})^3$	Kurtosis	$x_{kur} = \dfrac{1}{n}\sum_{t=1}^{n}\left(x(t)-\dfrac{1}{n}\sum_{t=1}^{n}x(t)\right)^4$
I factor	$I = E\{\max(|x(t)|)\}\Big/\dfrac{1}{n}\sum_{t=1}^{n}x(t)$	S factor	$S = \left(\dfrac{1}{n}\sum_{t=1}^{n}x^2\right)^{1/2}\Big/\dfrac{1}{n}\sum_{t=1}^{n}x(t)$
C factor	$C = E\{\max(|x(t)|)\}\Big/\left(\dfrac{1}{n}\sum_{t=1}^{n}x^2\right)^{1/2}$	L factor	$L = E\{\max(|x(t)|)\}\Big/\left(\dfrac{1}{n}\sum_{t=1}^{n}|x|^{1/2}\right)^2$
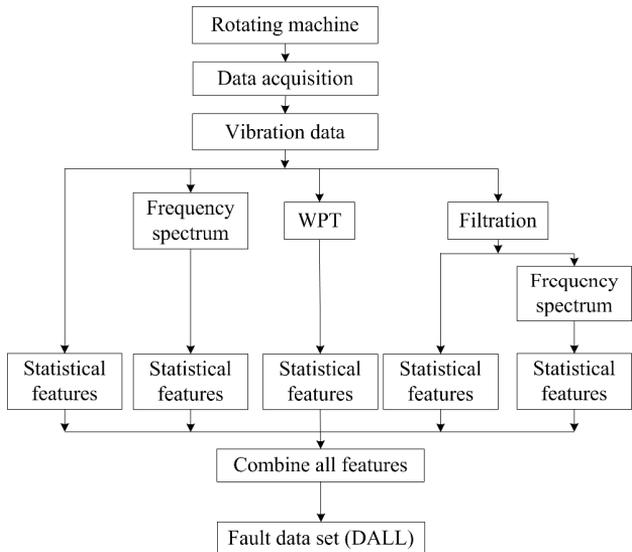
Fig. 7. The flow chart of the fault dataset

Table 3. The fault dataset DALL description

Dataset	Samples	Features	Fault type	Classification label
DALL	30	96	Normal	1
	30	96	Rubbing	2
	30	96	Unbalance	3
	30	96	Misalignment	4

The feature selection algorithm is applied to the DALL datasets. The result is shown in Fig. 8. The trend of classification accuracy changing is similar to UCI datasets. It is found that unrelated features are substantively deleted.
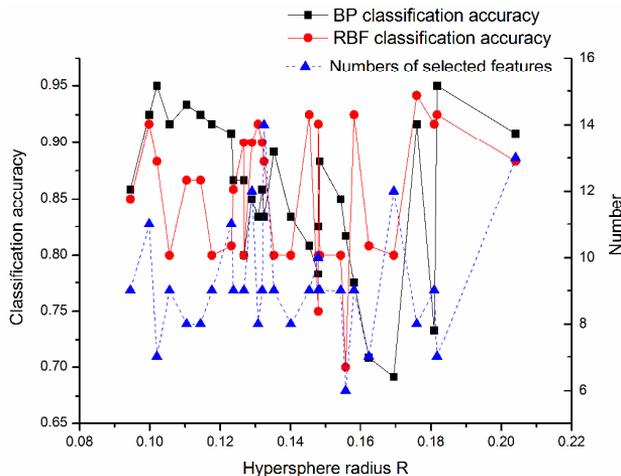
Fig. 8. Variation of classification accuracies and numbers of selected features
with hypersphere radius R on the DALL dataset

When FFT, WPT and filters are used to process the vibration signal, the obtained 96 features include not only the time-domain information, but also the frequency-domain information. Obviously, it is possible that some unrelated features are also acquired during the processing.

Namely, not all features make contributions to the faulty identification; some of them are insensitive to distinguish the faults. Therefore, the classifiers trained with all features would be confused and the classification accuracy is possibly low. However, when the sensitive features are selected by using the feature selection algorithm, the classification accuracy is improved greatly. To further improve the classification accuracy, the ensemble learning systems that are proposed in this paper are conducted on the DALL dataset, the results are shown in Fig. 9.
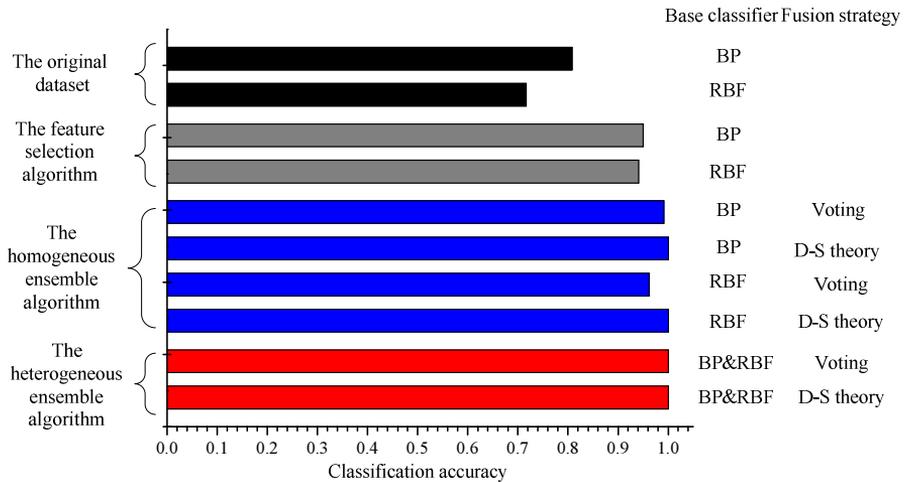
Fig. 9. Comparison of the classification accuracies for the DALL dataset

The classification accuracy is improved from 0.7624 (average) to 1 (average) by adopting the ensemble learning algorithms, it means that the proposed ensemble intelligent fault diagnosis approach can diagnose different faults of the rotor-bearing system accurately. From the results, we get the similar conclusion as the above: ensemble learning can improve the classification accuracy; the D-S theory seems to be the powerful fusion tool.

6. Conclusions

In order to improve the classification accuracy, the ensemble learning systems based on the feature partitioning are proposed. First, the homogeneous ensemble algorithm and the heterogeneous ensemble algorithm that both are based on the feature selection of the kernel neighborhood rough set are designed. Second, the majority voting is chosen as the fusion strategy for label information and the D-S theory is chosen as the fusion strategy for probability information. Third, the UCI datasets and the fault dataset DALL are used to test the classification performance of the ensemble learning systems. The results show that: (1) the ensemble learning systems can improve the classification performance and the heterogeneous ensemble algorithm can gain a better classification performance than the homogeneous ensemble algorithm. (2) Because probability information can provide more complementary information, the fusion strategy of the D-S theory can gain a better classification performance than the majority voting. (3) The application results of the rotating machinery demonstrate the effectiveness of the presented fault approach.

Automatic identifying the running condition of machine is supposed to be the developing trend of fault diagnosis technology, and it is also found by many researchers to be difficult to find one feature superior to the others in many cases. In this paper, a feature selection method is introduced and the fault identifying accuracy is guaranteed by designing homogeneous and heterogeneous ensemble learning systems. The improvement on the classification accuracy is attributed to the

diversity among subsets of features and base classifiers, as well as the good fusion strategy.

Acknowledgements

References

[1] **A. K. S. Jardine, Daming Lin, D. Banjevic** A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, Vol. 20, Issue 7, 2006, p. 1483-1510.

[2] **A. Widodo, Bo-Suk Yang** Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, Vol. 21, Issue 6, 2007, p. 2560-2574.

[3] **Yaguo Lei, Zhengjia He, Yanyang Zi, Qiao Hu** Fault diagnosis of rotating machinery based on multiple ANFIS combination with Gas. Mechanical Systems and Signal Processing, Vol. 21, Issue 5, 2007, p. 2280-2294.

[4] **Qinghua Wang, Youyun Zhang, Lei Cai, Yongsheng Zhu** Fault diagnosis for diesel valve trains based on non-negative matrix factorization and neural network ensemble. Mechanical Systems and Signal Processing, Vol. 23, Issue 5, 2009, p. 1683-1695.

[5] **Zhi-Hua Zhou, Jianxin Wu, Wei Tang** Ensembling neural networks: many could be better than all. Artificial Intelligence, Vol. 37, Issue 1, 2002, p. 239-263.

[6] **Thomas G. Dietterich** Ensemble methods in machine learning. Multiple classifier systems, Lecture Notes in Computer Science, Calgari, Italy, 2000.

[7] **L. K. Hansen, P. Salamon** Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, Issue 10, 1990, p. 993-1001.

[8] **L. Breiman** Bagging predictors. Machine Learning, Vol. 24, Issue 2, 1996, p. 123-140.

[9] **R. E. Schapire** The strength of weak learnability. Machine Learning, Vol. 5, Issue 2, 1990, p. 197-227.

[10] **Y. Freund** Boosting a weak algorithm by majority. Information and Computation, Vol. 121, Issue 2, 1995, p. 256-285.

[11] **Tin Kam Ho** The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, Issue 8, 1998, p. 832-844.

[12] **R. Bryll, R. Gutierrez-Osuna, F. Quek** Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. Pattern Recognition, Vol. 36, Issue 6, 2003, p. 1291-1302.

[13] **L. S. Oliveira, M. Morita, R. Sabourin, F. Bortolozzi** Multi-objective genetic algorithms to create ensemble of classifiers. International Conference on EMO 2005, Guanajuato, Mexico, 2005, p. 592-606.

[14] **D. W. Opitz** Feature selection for ensembles. Proceedings of 16th National Conference on Artificial Intelligence, Orlando, 1999, p. 379-384.

[15] **J. Kittler, M. Hatef, R. P. W. Duin, J. Matas** On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, Issue 3, 1998, p. 226-239.

[16] **Z. Pawlak** Rough Sets – Theoretical Aspects of Reasoning About Data. Kluwer Academic, Dordrecht, 1991.

[17] **W. Roman, S. Skowron** Rough set mehods in feature selection and recognition. Pattern Recognition Letters, Vol. 24, Issue 6, 2003, p. 833-849.

[18] **R. Jensen, Q. Shen** Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. IEEE Transactions on Knowledge Data Engineering, Vol. 16, Issue 12, 2004, p. 1457-1471.

[19] **Z. Suraj, N. E. Gayar, P. Delimata** A rough set approach to multiple classifier systems. Fundamenta Informaticae, Vol. 72, Issue 1, 2006, p. 393-406.

[20] **Qinghua Hu, Daren Yua, Zongxia Xiea, Xiaodong Lia** EROS: Ensemble rough subspaces. Pattern Recognition, Vol. 40, Issue 12, 2007, p. 3728-3739.

[21] **Randall Wilson, Tony R. Martinez** Improved heterogeneous distance functions. Journal of Artificial Intelligence Research, Vol. 6, 1997, p. 1-34.

[22] T. Y. Lin, K. J. Huang, Q. Liu, W. Chen Rough sets, neighborhood systems and approximation. Proceedings of the Fifth International Symposium on Methodologies of Intelligent Systems, Knoxville, Tennessee, 1990.

[23] Qinghua Hu, Daren Yu, Zongxia Xie Neighborhood classifiers. Expert Systems with Applications, Vol. 34, Issue 2, 2008, p. 866-876.

[24] Qinghua Hu, Jinfu Liu, Daren Yu Mixed feature selection based on granulation and approximation. Knowledge-Based Systems, Vol. 21, Issue 4, 2008, p. 294-304.

[25] Xiaoran Zhu, Youyun Zhang, Yongsheng Zhu Intelligent fault diagnosis of rolling bearing based on kernel neighborhood rough sets and statistical features. Journal of Mechanical Science and Technology, Vol. 26, Issue 9, 2012, p. 2649-2657.

[26] Xu L., Krzyzak C., Suen C. Y. Methods of combining multiple classifiers and their application to handwriting recognition. IEEE Transactions on Systems, Man and Cybernetics, Vol. 22, Issue 3, 1992, p. 418-435.

[27] Frank A., A. Asuncion UCI Machine Learning Repository 2010. University of California, Irvine, School of Information and Computer Sciences.

[28] Shun Bian, Wenjia Wang On diversity and accuracy of homogeneous and heterogeneous ensembles. International Journal of Hybrid Intelligent Systems, Vol. 4, Issue 2, 2007, p. 103-128.

[29] A. L. V. Coelho, D. S. C. Nascimentoa On the evolutionary design of heterogeneous Bagging models. Neurocomputing, Vol. 73, Issue 16, 2010, p. 3319-3322.

[30] B. Samanta, K. R. Al-Balushi Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mechanical Systems and Signal Processing, Vol. 17, Issue 2, 2003, p. 317-328.

Appendix

The ensemble learning systems that are proposed in this paper are conducted on the UCI datasets and fault dataset DALL. The results are shown in Table 4 – Table 9.

Table 4. The results of classification accuracies for the Diab dataset

	Base classifier	Fusion strategy	Classification accuracies
The original dataset	BP		0.7513 ± 0.0232
	RBF		0.5819 ± 0.0622
The feature selection algorithm	BP		0.7839 ± 0.0241
	RBF		0.5923 ± 0.0353
The homogeneous ensemble algorithm	BP	majority voting	**0.7917 ± 0.0279**
		D-S theory	**0.7917 ± 0.0215**
	RBF	majority voting	0.6367 ± 0.0130
		D-S theory	0.6145 ± 0.0249
The heterogeneous ensemble algorithm	BP& RBF	majority voting	0.7839 ± 0.0261
		D-S theory	0.7878 ± 0.0277

Table 5. The results of classification accuracies for the Iono dataset

	Base classifier	Fusion strategy	Classification accuracies
The original dataset	BP		0.8546 ± 0.0824
	RBF		0.6724 ± 0.0374
The feature selection algorithm	BP		0.9060 ± 0.0459
	RBF		0.9002 ± 0.0545
The homogeneous ensemble algorithm	BP	majority voting	0.9515 ± 0.0240
		D-S theory	0.9544 ± 0.0275
	RBF	majority voting	0.9145 ± 0.0484
		D-S theory	0.9259 ± 0.0384
The heterogeneous ensemble algorithm	BP& RBF	majority voting	**0.9687 ± 0.0186**
		D-S theory	0.9658 ± 0.0279

Table 6. The results of classification accuracies for the Sonar dataset

	Base classifier	Fusion strategy	Classification accuracies
The original dataset	BP		0.7159 ± 0.0495
	RBF		0.6343 ± 0.0300
The feature selection algorithm	BP		0.7838 ± 0.0584
	RBF		0.8557 ± 0.0388
The homogeneous ensemble algorithm	BP	majority voting	0.8898 ± 0.0250
		D-S theory	0.8898 ± 0.0303
	RBF	majority voting	0.8986 ± 0.0678
		D-S theory	0.9184 ± 0.0683
The heterogeneous ensemble algorithm	BP& RBF	majority voting	0.9184 ± 0.0700
		D-S theory	**0.9187** ± 0.0426

Table 7. The results of classification accuracies for the Wine dataset

	Base classifier	Fusion strategy	Classification accuracies
The original dataset	BP		0.9552 ± 0.0246
	RBF		0.9095 ± 0.0785
The feature selection algorithm	BP		0.9889 ± 0.0152
	RBF		0.9716 ± 0.0208
The homogeneous ensemble algorithm	BP	majority voting	0.9886 ± 0.0157
		D-S theory	**1** ± 0.0000
	RBF	majority voting	0.9719 ± 0.0197
		D-S theory	0.9830 ± 0.0250
The heterogeneous ensemble algorithm	BP& RBF	majority voting	0.9886 ± 0.0157
		D-S theory	**1** ± 0.0000

Table 8. The results of classification accuracies for the Wpbc dataset

	Base classifier	Fusion strategy	Classification accuracies
The original dataset	BP		0.6875 ± 0.0609
	RBF		0.5465 ± 0.0393
The feature selection algorithm	BP		0.7432 ± 0.0369
	RBF		0.7216 ± 0.0867
The homogeneous ensemble algorithm	BP	majority voting	0.8190 ± 0.0389
		D-S theory	0.8480 ± 0.0215
	RBF	majority voting	0.7374 ± 0.0298
		D-S theory	0.7527 ± 0.0420
The heterogeneous ensemble algorithm	BP& RBF	majority voting	0.8392 ± 0.0409
		D-S theory	**0.8495** ± 0.0488

Table 9. The results of classification accuracies for the DALL dataset

	Base classifier	Fusion strategy	Classification accuracies
The original dataset	BP		0.8083 ± 0.0547
	RBF		0.7166 ± 0.0494
The feature selection algorithm	BP		0.95 ± 0.0456
	RBF		0.9416 ± 0.0364
The homogeneous ensemble algorithm	BP	majority voting	0.9916 ± 0.0186
		D-S theory	**1** ± 0
	RBF	majority voting	0.9616 ± 0.0166
		D-S theory	**1** ± 0
The heterogeneous ensemble algorithm	BP& RBF	majority voting	**1** ± 0
		D-S theory	**1** ± 0