



A New Multi-classifier Ensemble Algorithm Based on D-S Evidence Theory

Kaiyi Zhao¹ · Li Li^{2,3} · Zeqiu Chen¹ · Ruizhi Sun^{1,4}  · Gang Yuan¹ · Jiayao Li¹

Accepted: 11 April 2022 / Published online: 9 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Classifier ensemble is an important research content of ensemble learning, which combines several base classifiers to achieve better performance. However, the ensemble strategy always brings difficulties to integrate multiple classifiers. To address this issue, this paper proposes a multi-classifier ensemble algorithm based on D-S evidence theory. The principle of the proposed algorithm adheres to two primary aspects. (a) Four probability classifiers are developed to provide redundant and complementary decision information, which is regarded as independent evidence. (b) The distinguishing fusion strategy based on D-S evidence theory is proposed to combine the evidence of multiple classifiers to avoid the mis-classification caused by conflicting evidence. The performance of the proposed algorithm has been tested on eight different public datasets, and the results show higher performance than other methods.

Keywords Evidence theory · Combination · Machine learning · Neural networks

1 Introduction

A classifier is a system that uses known knowledge to assign the unknown object to one class or category. The most common way of building the mapping function is to learn the previous classification instance by a specific learning algorithm. Although there are many classification algorithms, such as Decision Tree [4], Naïve Bayes (NB) [31], Artificial Neural Networks (ANN) [26], K-Nearest Neighbors (KNN) [13], Support Vector Machine (SVM) [33], Class Association Rules (CARS) [15] and others [12, 19, 25], there is not a single

✉ Ruizhi Sun
sunruizhi@cau.edu.cn

✉ Gang Yuan
yuangang_cau@163.com

¹ College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

² Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin 541004, China

³ Computer School, Beijing Information Science and Technology University, Beijing 100101, China

⁴ Scientific Research Base for Integrated Technologies of Precision Agriculture (Animal Husbandry), The Ministry of Agriculture, Beijing 100083, China

method that is superior than any other. So, the strategy of integrating different classification algorithms has attracted the interest of more and more researchers [23, 29, 34, 35].

The main idea of a multi-classifier ensemble algorithm is to use some complementary predictions to obtain a better performance than base classifiers [9]. Therefore, to obtain the final ensemble decision, it is necessary to establish a process of combining base decisions. There are two kinds of ensemble strategies for combining base classifiers: selection and fusion [20]. In classifier selection, each base classifier has its domain, where they are the most reliable classifiers. Ensemble decision is to select the corresponding base classifier decision as the final decision based on the domain of the unknown object. In classifier fusion, ensemble decision is established by combining some complementary base classifiers. The existing classifier fusion methods are the average method, voting method, weighting method, and others, such as meta-learning methods [10, 30].

In the ensemble method based on classifier fusion, the main premise to obtain best ensemble performance is that the error rate of base classifiers must be very low, and the error made by one classifier must be compensated by the correct prediction of other base classifiers. That is, the basic classifiers must be as accurate and complementary as possible [14]. However, The more accurate the basic classifiers are, the more similar they are. And the better the complementarity of classifiers, the more uncertain the prediction of classifiers. So, the main challenge for classifier ensemble is to achieve the balance between accuracy and complementarity.

Faced with this challenge, a multi-classifier ensemble algorithm based on probabilistic classifier and distinguish fusion strategy is proposed in this paper. Firstly, considering that classifiers with different principles are more likely to complement each other, classifier training should be as easy as possible. Support Vector Machine(SVM), Echo State Network(ESN), k-Nearest Neighbor(KNN), and Extreme Learning Machine(ELM) are selected as the base classifiers. The posterior probability estimation methods are proposed based on probability theory to obtain four probability classifiers to output the possible probability of each category in different basic classifiers. Then, the prediction of each base probability classifier is regarded as a piece of evidence to form the body of evidence. In order to avoid recognition errors caused by highly conflicting evidence, the distinguish fusion strategy based on conflict coefficient is proposed, i.e., the body of evidence with conflicting evidence is preprocessed first, and then fused by utilizing Dempster's rule, otherwise, Dempster's rule can be directly used for fusion. Finally, the recognition results of the proposed multi-classifier ensemble algorithm can be obtained by converting the fusion results into labels. In the experimental analysis, eight public datasets are used to verify the proposed ensemble algorithm, and the analytical conclusion shows that the proposed algorithm has better classification performance in different scenarios than single models, the voting-based method, and the Multi-modal method [38].

The rest of the paper is organized as follows. Section 2 introduces some preliminaries related to this paper. Section 3 shows the detailed process of the proposed multi-classifier ensemble algorithm. Section 4 gives the analysis and discussion of comparative experiments. Section 5 is the conclusion and outlook for our works.

2 Preliminaries

2.1 Dempster/Shafer Evidence Theory

Evidence theory is an inexact reasoning theory first proposed by Dempster [6] in 1967 and further developed by Shafer [32] in 1976, also known as Dempster/Shafer evidence theory (D-S evidence theory). Its basic definition and combination rules are as follows:

Discernment Frame The discernment frame Ω is defined as a non-empty set, $\Omega = \{x_1, x_2, \dots, x_n\}$, which contains mutually exclusive events. The power set of discernment frame 2^Ω contains 2^n elements, which are represented as follows:

$$2^\Omega = \{\emptyset, \{x_1\}, \{x_2\}, \{x_3\}, \dots, \{x_n\}, \{x_1, x_2\}, \dots, \{x_1, x_2, \dots\}, \dots, \Omega\} \tag{1}$$

Mass Function It is also called Basic Probability Assignment(BPA). In the discernment frame, the basic probability assignment function is defined to represent uncertain information. The mass function m is the mapping of power set 2^Ω on the interval $[0, 1]$, which satisfies the following conditions:

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{x \in 2^\Omega} m(x) = 1 \end{cases} \tag{2}$$

If the above conditions are true, the subset of the power set 2^Ω is called a focal element. $m(x)$ is the value of mass function.

Dempster's Combination Rule In the framework of evidence theory, two independent mass functions m_1 and m_2 can be fused through the following Dempster's combination rule:

$$m(x) = (m_1 \oplus m_2)(x) = \frac{\sum_{B \cap C = x} m_1(B)m_2(C)}{1 - K} \tag{3}$$

Here, B and C is a subset of 2^Ω . K is the normalization factor, defined as follows:

$$K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C) \tag{4}$$

or

$$K = 1 - \sum_{B \cap C \neq \emptyset} m_1(B)m_2(C) \tag{5}$$

2.2 Single Classifier

For each ensemble algorithm, it is very important to choose a suitable single model. The principle of classifier determines the analysis perspective of the training model in judging the category of new instances. In other words, classifiers developed based on different principles are more likely to be trained into complementary models. Based on this idea, SVM, ESN, KNN, and ELM are selected as the base classifiers of the ensemble algorithm, and their training process is very simple, which can quickly get the base classifier. The basic principles of these four classifiers are introduced as follows.

Support Vector Machine Support Vector Machine(SVM), which is proposed by Vapnik with colleagues [5], is a kind of generalized linear classifier, it can separate data in a supervised learning manner, and its decision boundary is the maximum margin hyperplane. When there are more than two categories to be divided, SVM will convert this problem into multiple binary classification problems. In target recognition and feature-based classification applications, SVM has a pretty good performance in classifying categories based on the maximum margin hyperplane [7, 11, 22, 33]. Therefore, we choose it as a single model to verify the effectiveness of the ensemble algorithm, hoping to find its unique evidence from the perspective of SVM.

Echo State Network Echo State Network (ESN), proposed by Jaeger [17, 18], is a recurrent neural network. Its hidden layer is generally sparsely connected, usually only 1% connectivity. The connectivity and weights of hidden neurons are randomly assigned as constants. The main interest for ESN is that its behavior is non-linear, but the only weights modified during

training are the mapping functions that connect hidden neurons to output neurons. And it is widely used in regression analysis [3, 21, 37] due to its simple network structure.

k-Nearest Neighbor *k*-Nearest Neighbor (KNN) [1] was proposed by Thomas Cover and is a non-parametric method for classification and regression. KNN, which relies on distance for classification, is an algorithm based on instance learning or lazy learning. It is very simple to implement and the error is easy to control. And it can handle a series of non-linear problems and is active in various regression or classification applications [2, 8, 13, 39]. Therefore, KNN can also be selected as a base classifier to generate characteristic evidence for the ensemble classification algorithm.

Extreme Learning Machine Extreme Learning Machine (ELM), proposed by Huang Guangbin [16], is a feed-forward neural network. In ELM, the neurons in the hidden layer are randomly allocated and never updated, that is, ELM is a projection having the nonlinear transformation. Since ELM has good generalization performance, it has achieved good results in classification and regression applications [24, 36, 40].

3 Proposed Multi-classifier Ensemble Algorithm

Using the idea of classifier fusion, four classifiers SVM, ESN, KNN, and ELM are selected as base classifiers, and these classifiers output the probability of each prediction category instead of the label. On the basis of previous research on solving the issue that highly conflict evidence is easy to lead to counterintuitive conclusions [42, 43], the novel multi-classifier ensemble algorithm is proposed based on the D-S evidence theory. The schematic diagram of the algorithm is shown in Fig. 1, which is described as follows:

1. Preprocessing data. There may be null values and characters in the original data that are difficult for the classifier to handle directly, so data preprocessing is required. Here, the real numbers will be used as labels to replace characters and nulls, and all records will be normalized to reduce the impact of the range of values on classification performance.
2. Generating BPAs for each classifier. Different from general classifiers, each classifier in the proposed ensemble algorithm outputs the probability of each category instead of the category label. The prediction of each classifier will be regarded as a piece of evidence, and they will together form a complete body of evidence. Therefore, the most

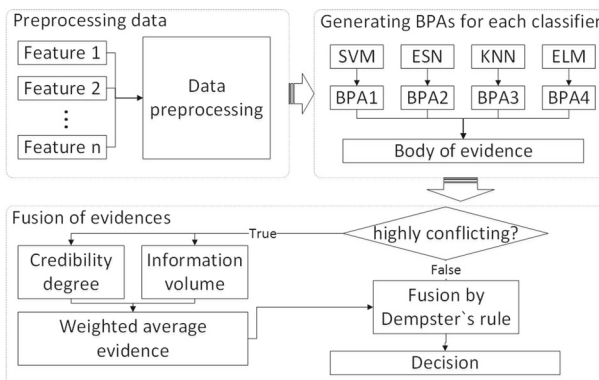


Fig. 1 Schematic diagram of the proposed multi-classifier ensemble algorithm

important part of this module is to discuss how making the classifier generate the possible probability of each category (Sect. 3.1).

3. Fusion of evidences. Since D-S evidence theory may produce counter-intuitive results when dealing with highly conflicting evidence, we will first judge whether there are conflicting evidence pairs in the body of evidence. If not, the evidence will be directly fused by using Dempster’s combination rule. Otherwise, we will generate weighted average evidence based on the credibility degree and information volume of each piece of evidence in the body of evidence, and then fuse the weighted average evidence by Dempster’s combination rule (Sect. 3.2).

3.1 Generating BPAs for Each Classifier

According to the principle of each classifier, their basic probability assignment for the prediction category will be obtained. To obtain the classification probability of SVM, Platt [28] proposes a method of estimating the posterior probability by using a sigmoid function to map the output of SVM to the interval [0, 1]. To obtain the classification probability of KNN, the probability of KNN recognition category is very easy to obtain according to its classification principle which takes the category of the majority item in the nearest k neighbors as the classification result. To obtain the classification probability of ESN and ELM, since the neural network converts the classification problem into regression for analysis, then converts the regression results into categories to output, we estimate the posterior probability of the category based on the distance between the forecast value and the target value.

Probability SVM Platt uses the sigmoid-fitting method to post-process the output of standard SVM and convert it into posterior probability, and its definition is as follows:

$$P(y = 1|f) = \frac{1}{1 + e^{Af+B}} \tag{6}$$

Here, f is the unthreshold output of sample input, and A, B are the values estimated with maximum likelihood estimate which is as follows:

$$\min - \sum_i t(f_i) \log(P(f_i)) + (1 - t(f_i)) \log(1 - P(f_i)) \tag{7}$$

with

$$t_i = \begin{cases} \frac{N_++1}{N_++2}, y_i = +1 \\ \frac{1}{N_++2}, y_i = -1 \end{cases} \quad i = 1, 2, 3, \dots, n \tag{8}$$

where $P(f_i) = P(y = 1|f_i)$, N_+ is the number of positive samples($y_i=+1$) and N_- is the number of negative samples($y_i=-1$).

Probability KNN In the absolute classification of KNN, judging the category of the new instance is to find the k instances that are most similar to the new instance based on the known data. If most of these k instances belong to a certain category, then the new instance will be the category.

However, when the new instance is at the boundary of two or more categories, no matter which category it belongs to, it is likely to be wrong. At this time, it will be more meaningful to output the possible probability of each category. For this purpose, the k instances of known categories closest to the new instance will be found and counted how many of them are in each category and divide by k to get the probability value of the new instance belonging to

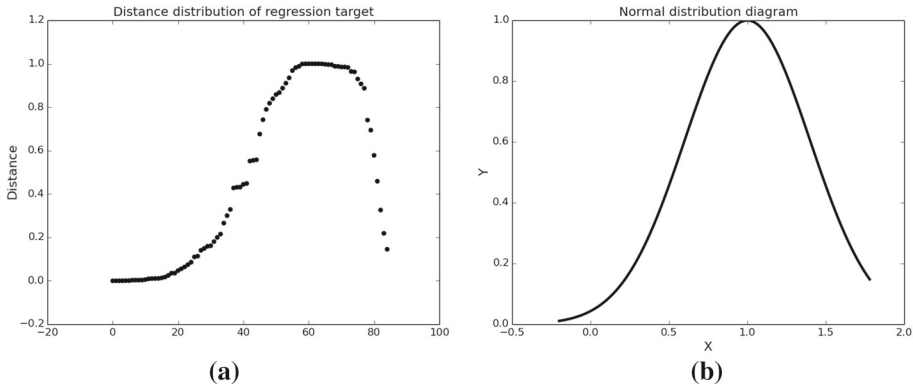


Fig. 2 a Distribution of absolute distance between the expected value and forecasting value; b Schematic diagram of normal distribution

each category. It is defined as follows:

$$P_i = \frac{class_i}{k} \tag{9}$$

Here, P_i represents the probability that the new instance may be category i , $class_i$ indicates the number of instances of category i among the k nearest neighbors.

Probability ESN and Probability ELM When dealing with classification problems, neural networks such as ESN and ELM will usually perform binary coding on the category based on One-Hot Encoding, and fit the training data, then conversed regression result will be the class label. Generally, when the artificial neural network needs to output the probability of the category, the Softmax function will be added to complete the classification model training through the feedback gradient. However, ELM is a feedforward neural network, and ESN uses the linear fitting of random neurons to simulate the nonlinear prediction ability. If the Softmax function is added, the feedback gradient may greatly reduce the performance of the two classifiers. Therefore, a distance-based classification probability mapping function is proposed in this paper.

Here, we obtain an intuitive mapping function by analyzing the distance between the prediction result and the expected value. Fig. 2a shows the statistical information of the forecasting value of ELM on Iris dataset. Here, the forecasting $n \times 3$ (n is the number of instances, '3' is the number of categories) matrix is converted into a vector, then sort the vector in ascending order, finally use equation $\bar{y} = 1 - |1 - y|$ (y indicates each forecasting value) to convert the predicted value to the distance in $(-\infty, 1]$.

In Fig. 2a, the horizontal axis represents the index of the sort result, and the vertical axis represents the converted distance. The closer the points in Fig. 2a are to the labeling threshold, the denser the distribution is, which is very similar to the normal distribution in Fig. 2b. Therefore, we consider using the normal distribution function to estimate the probability when the corresponding category is true, i.e., the posterior probability of the corresponding category. Since the label True is 1 and the label False is 0, the mapping method based on normal function ($X \sim N(\mu, \sigma^2)$) can be obtained by setting the parameters of normal function. Here, we set $\mu = 1, \sigma = 1/3$, and the mapping equation is as follows:

$$\hat{P}_i = \frac{3}{\sqrt{2\pi}} e^{-\frac{9(y_i-1)^2}{2}} \tag{10}$$

Here, y_i represents the forecast value of the i -th category by ESN or ELM. Normalize \hat{P}_i to get P_i , which is the estimated value of the posterior probability of the i -th category.

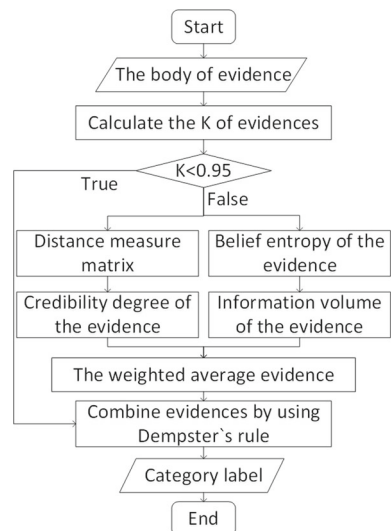
$$P_i = \frac{\hat{P}_i}{\sum_i \hat{P}_i} \tag{11}$$

3.2 Fusion of Evidence

After obtaining the BPAs, how to fuse these redundant and complementary evidences to generate more valuable classification results is another difficulty in this paper. Here, D-S evidence theory is used to fuse the output of multiple classifiers. However, when there is highly conflicting evidence, it will inevitably make mistakes and draw counter-intuitive conclusions. Therefore, it is very necessary to pre-process the conflicting evidence before fusion. The degree of conflict between the body of evidence can be measured by the normalization constant K . The result of preprocessing conflict evidence and the original evidence without conflict will be fused through Dempster’s rule, and the fusion category probability will be used to decide the label of a new instance.

In order to have an intuitive understanding of the process of fusing pieces of evidence, we have drawn the flow chart of fusing pieces of evidence as shown in Fig. 3. Firstly, the output value BPA_i ($i = 1, 2, 3, 4.$) of the probabilistic classifiers SVM, ESN, KNN, and ELM is used as a piece of evidence to form the mutually independent and complete body of evidence. Next, the algorithm will judge the conflict degree in the body of evidence according to the normalized constant K . Here, K is the judgment condition for the acceptability of highly conflicting evidence [27, 41]. When $K < 0.95$, it indicates that the conflict is within the acceptable range, the body of evidence can be directly fused through Dempster’s rule. Otherwise, it is very necessary to preprocess the conflicting evidence before fusion. In the preprocessing of conflict evidence, the weight of each evidence is generated based on two aspects: credibility degree and information volume. The credibility degree is obtained from the distance measure matrix of evidence, and information volume is another form of information entropy. Then

Fig. 3 Flow chart of fusion evidence



Algorithm 1 Combining the body of evidence by the Dempster's rule

Require: The body of evidence including BPA1, BPA2, BPA3, and BPA4 generated by SVM, ESN, KNN, and ELM respectively

Ensure: Categories forecasted by multi-classifier ensemble algorithm

```

1:  $[m_1, m_2, m_3, m_4] \leftarrow [\text{BPA1}, \text{BPA2}, \text{BPA3}, \text{BPA4}]$  //Construct a matrix of the body of evidence
2:  $K \leftarrow$  Calculate the conflict degree in the body of evidence by Equation 4 or Equation 5
3: if  $K < 0.95$  then
4:   for  $i=1 \rightarrow 4$  do
5:     The final fusion probability of each category  $\leftarrow (m_i \oplus m_{i+1})$  // fusion by Dempster's rule as
     Equation 3 and Equation 4
6:   end for
7:   Convert to category label by using One-Hot Encoding
8: else
9:    $DMM[4, 4]$  //Initialize  $4 \times 4$  distance measure matrix
10:  for  $i=1 \rightarrow 4$  do
11:    for  $j=1 \rightarrow 4$  do
12:       $DMM[i, j] \leftarrow$  the distance between  $m_i$  and  $m_j$  by Equation
      
$$\sqrt{\frac{[\sum_s m_i(B_s) \log \sqrt{\frac{2m_i(B_s)^2}{m_i(B_s)^2 + m_j(B_s)^2}}]^2 + [\sum_s m_j(B_s) \log \sqrt{\frac{2m_j(B_s)^2}{m_i(B_s)^2 + m_j(B_s)^2}}]^2}{2}}$$

13:    end for
14:  end for
15:   $Sup[4], Crd[4], IV[4]$  //Initialize  $1 \times 4$  vectors of support degree, credibility degree and information
    volume
16:  for  $i=1 \rightarrow 4$  do
17:     $Sup[i] = \frac{3}{\sum_{j=1}^4 DMM[i, j]}$ 
18:     $IV[i] \leftarrow e^{-\sum_{B \subseteq \Omega} m(B) \log(m(B))}$ ,  $i \in [1, 4]$ 
19:  end for
20:   $SUM \leftarrow IV \times IV$ 
21:  for  $i=1 \rightarrow 4$  do
22:     $Crd[i] = \frac{Sup[i]}{\sum_{s=1}^4 Sup[s]}$ 
23:     $IV[i] = \frac{IV[i]}{SUM}$ 
24:  end for
25:   $Crd\_IV[4]$  //Initialize  $1 \times 4$  vector of credibility degree jointed information volume by multiplication
26:  for  $i=1 \rightarrow 4$  do
27:     $Crd\_IV[i] = Crd[i] \times IV[i]$ 
28:  end for
29:   $SUM \leftarrow Crd\_IV \times Crd\_IV$ 
30:  for  $i=1 \rightarrow 4$  do
31:     $Crd\_IV[i] = \frac{Crd\_IV[i]}{SUM}$ 
32:  end for
33:   $WAE = \sum_{i=0}^4 Crd\_IV[i] \times m_i$  //the weighted average evidence vector
34:  for  $i=1 \rightarrow 4$  do
35:    The final fusion probability of each category  $\leftarrow (WAE \oplus WAE)$  // fusion by Dempster's rule as
    Equation 3 and Equation 4
36:  end for
37:  Convert to category label by using One-Hot Encoding
38: end if
39: return Category label

```

according to the weight of the evidence, the weighted average evidence will be generated as the preprocessing result of the conflict evidence. Finally, the fusion result is converted into category labels to be used as the forecast result of the multi-classifier ensemble algorithm. The pseudo-code of the evidence fusion process is shown in Algorithm 1.

4 Experiment and Analysis

In order to verify the multi-classifier ensemble algorithm proposed in this paper, comparative experiments have been completed on eight public data sets. The data sets, experimental setup, results, and analysis will all be described in this section.

4.1 Data Set and Experimental Settings

Data Set The attributes of the eight public data sets in UCI database are as follows:

1. Post is an abbreviated form of Postoperative Patient Data, and its classification task is to determine where patients in the postoperative recovery area should be sent;
2. Nursery is an abbreviated form of Nursery Database, and it is derived from the hierarchical decision model originally developed for the ranking of nurseries' applications;
3. Iris is an abbreviated form of Iris Data Set, and it is the most famous database in the pattern recognition literature that contains the attribute information of three iris plants;
4. Wine is an abbreviated form of Wine Recognition Data, and it is the result of the chemical analysis of wines originated in the same region in Italy, which draw from three different varieties;
5. Breast is an abbreviated form of Breast Tissue Data Set, and it is a data set with electrical impedance measurements of fresh tissue samples excised from the breast;
6. Banance is an abbreviated form of Balance Scale Data Set, and it is a database of balance weight and distance.
7. Hayes is an abbreviated form of Hayes-Roth & Hayes-Roth (1977) Database, and it is a theme library of human subjects study;
8. Page is an abbreviated form of Page Blocks Classification Data Set, and it is the problem including classifying all the blocks of the document page layout detected by the segmentation process.

Table 1 shows some information about the eight public data sets during the analysis process. In the Table 1, the first column is the name of the data set; the second column is the distribution of the number of records in different categories for the corresponding data set (the symbol “/” distinguishes different categories, and the number indicates the number of instances in different categories); the third column is the number of feature dimensions of the corresponding dataset; the last column is the total number of instances in the data set. In the comparative experiment in this paper, the 30% instances are randomly selected as test set for each dataset, and the rest as the training set.

Table 1 Data set overview

Dataset	Class distribution	Attributes	Total size
Post	2/24/64	8	90
Nursery	4320/2/328/4266/4044	8	12960
Iris	50/50/50	4	150
Wine	59/71/48	13	178
Breast	21/15/18/16/14/22	9	106
Banance	49/288/288	4	625
Hayes	51/51/30	5	132
Page	4913/329/28/88/115	10	5473

Experimental Set In the comparative experiment, we set the same parameters for each dataset and run the program in python 2.7. In SVM, the error value *tol* for stopping training is $1e-3$, and the penalty is set to $l2$. In ESN, the number of reservoir neurons is 200, the spectral radius of the recurrent weight matrix is 0.25, the proportion of recurrent weights is set to 0.95, and the output activation function is *tanh*. In KNN, the number of neighbors is set to 5, and the power parameter for the Minkowski metric is 2. In ELM, the number of hidden layer nodes is 200, and the activation function is *tanh*. In addition, the three indicators used as classifier evaluation: Accuracy(ACC), F1 score(F1), and Area Under the Receiver Operating Characteristic curve(AUROC) are implemented by using “sklearn.metrics” interface.

4.2 Experimental Results and Discussion

On the eight public data sets, the training set and the test set are randomly divided according to the test set ratio 30%, and the classification accuracy of 100 repeated experiments is plotted on Fig. 4. Its subgraphs (a), (b), (c), (d), (e), (f), (g), and (h) correspond to the classification accuracy of 100 random repeated experiments of Post, Nursery, Iris, Wine, Breast, Banance, Hayes, and Page data sets respectively. Here, the horizontal axis marks the index of the repeated trials, the vertical axis is the classification accuracy(ACC) of SVM, ESN, KNN, ELM, Voting-based method, and the proposed ensemble algorithm which distinguishes by the legend in the upper right corner of each subgraph.

We analyze the results shown in Fig. 4 and found that our proposed ensemble algorithm achieve the best classification results in most cases, even if a base classifier is almost invalid in the scenario shown in Fig. 4b, e. Compared with the Voting-based ensemble method, the proposed algorithm also achieves significantly better performance in multiple scenarios, as shown in six subgraph (a), (b), (e), (f), (g) and (h) of Fig. 4. Therefore, the comprehensive analysis can draw the conclusion that the multi-classifier ensemble classification algorithm proposed in this paper has quite good performance in the classification accuracy of different applications.

In order to further explore the mechanism why the proposed multi-classifier ensemble algorithm can have better classification performance, we have listed three typical examples of conflicting evidence fusion in data set Post in Table 2. “No.1”, “No.2” and “No.3” mark three classification instances in Post. “BPAs” and “Target” respectively correspond to the basic probability assignment of each category with the symbol “/” to distinguish and the recognized category label. The last line “Real” is the real label of the instance. In Table 2, when a classifier generates BPAs that conflict with other evidence in the classification task of three instances, the fusion weighted average evidence can often get the correct classification result. Based on this idea, this paper proposes a multi-classifier fusion classification algorithm.

In order to further analyze the performance of the proposed algorithm, the accuracy(ACC), F1 score(F1), and area under the receiver operating characteristic curve(AUROC) of 100 randomized trials are used to evaluate the classification performance of the proposed ensemble algorithm. In Table 3, the evaluation values of 100 trials in eight public data sets are listed in the form of “mean \pm standard deviation”. In addition, the statistics of Table 3, additionally including the significance of the proposed algorithm relative to other ensemble methods, are visualized in Fig. 5. The horizontal lines on each subplot show the significance (p-value) of the proposed algorithm relative to other ensemble methods on the three statistical values of ACC, F1, and AUROC. The red horizontal line indicates that both are significant, and the value on the horizontal line is the p-value, “*” indicates moderately significant, and “***” indicates very significant. By reasonably analyzing the statistical results of repeated experiments, it is found

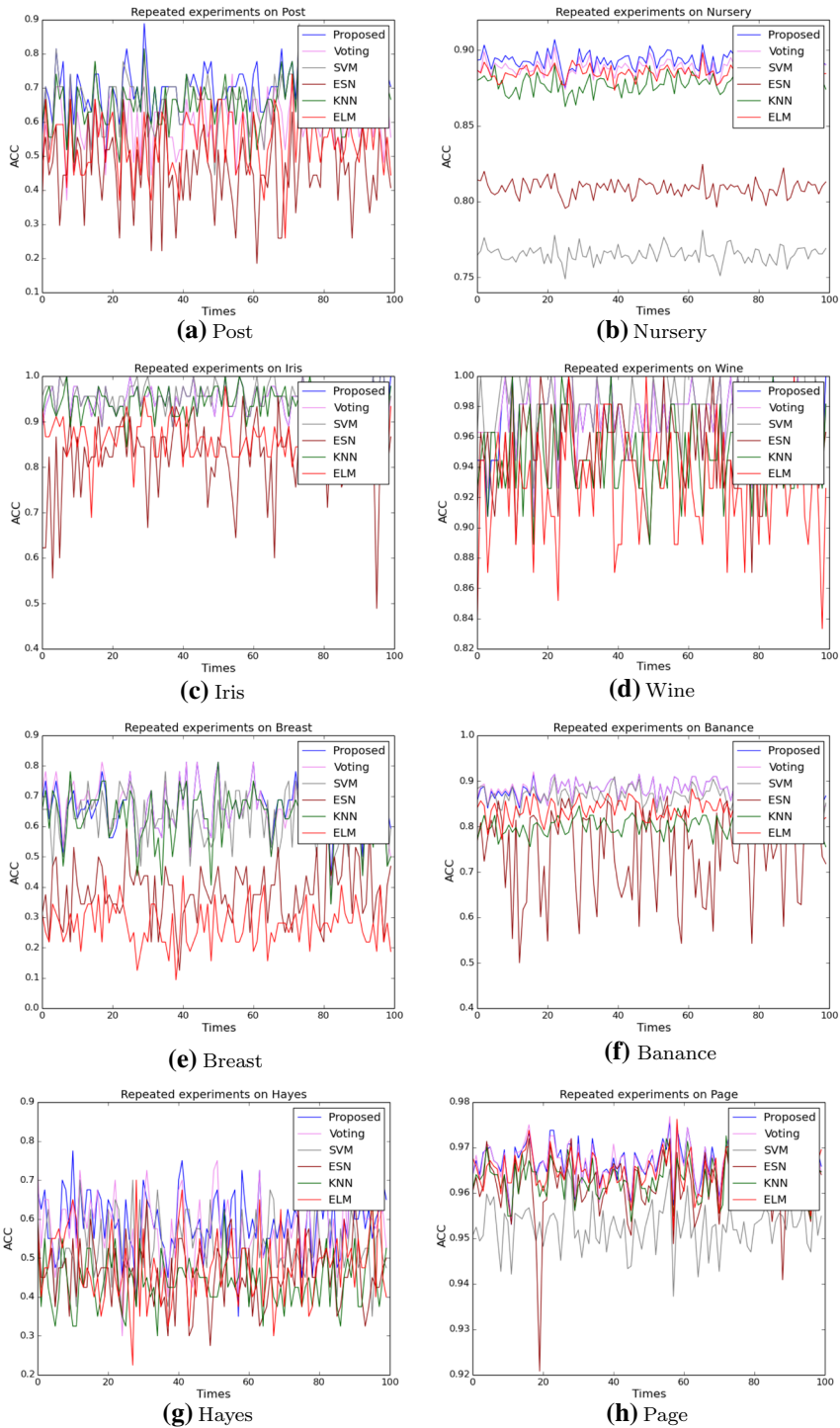


Fig. 4 Repeated experiments on eight public datasets

Table 2 Conflict evidence fusion for Post

Method		No.1	No.2	No.3
SVM	BPAs	0.7358/0.0484/0.2156	0.5858/0.1789/0.2352	0.7667/0.1163/0.1170
	Target	A	A	A
ESN	BPAs	0.3654/0.4719/0.1625	0.0287/0.2914/0.6799	0.0802/0.5764/0.3434
	Target	I	S	I
KNN	BPAs	0.7500/0.0000/0.2500	1.0000/0.0000/0.0000	1.0000/0.0000/0.0000
	Target	A	A	A
ELM	BPAs	0.7594/0.0074/0.2331	0.0256/0.0738/0.9005	0.4064/0.5854/0.0081
	Target	A	S	I
Proposed	BPAs	0.9868/0.0000/0.0132	0.2338/0.0170/0.7492	1.0000/0.0000/0.0000
	Target	A	S	A
Real		A	S	A

that although the proposed algorithm has few advantages over other methods in the accuracy and F1 score of wine data set, and AUROC of page data set, it shows the best performance in the evaluation indexes of most data sets. In general, the proposed multi-classifier fusion idea based on D-S evidence theory achieves higher performance than all single classifiers, the Voting-based method, and the Multi-modal method. This further verifies that the previous conclusion is correct, i.e., the proposed algorithm has better classification performance than others.

Through the above analysis and discussion, reasonable conclusions can be drawn as follows:

1. A single classifier is limited by its analysis perspective, which is determined by its basic principles, and often leads to uncertain and fuzzy decisions in real and complex classification tasks;
2. Fusion of conflicting evidence based on D-S evidence theory can largely avoid the bias of a single classifier, and infer more credible artificial intelligence decisions;
3. The proposed multi-classifier ensemble algorithm performs better than all single classifiers, the Voting-based method, and the Multi-modal method on eight public data sets with different attributes and achieves better classification performance.

In summary, whether the line graph of the accuracy of 100 repeated experiments, the conflict evidence fusion process of the classification examples, and the statistical table of the experimental data all believe that the multi-classifier ensemble algorithm proposed in this paper achieves the best classification performance than others.

5 Conclusion and Outlook

In this paper, a multi-classifier ensemble algorithm based on D-S evidence theory is proposed. The main contributions are as follows: the mapping method between the base classifier and the posterior probability of their prediction results is developed to form four probability classifiers to obtain multi-source complementary evidence; On the basis of measuring the conflict evidence by using the conflict coefficient, the body of evidence will be distinguished and fused to obtain the final classification result. In the comparative experiment, eight dif-

Table 3 Comparison of experimental results

Data	Index	Unit Model		ESN	KNN	ELM	Voting	Multi-modal	Proposed
		SVM	ELM						
Post	ACC	0.6742 ± 0.0677	0.4206 ± 0.1828	0.6613 ± 0.0501	0.6164 ± 0.0455	0.6746 ± 0.0613	0.5952 ± 0.1033	0.6905 ± 0.0587	
	F1	0.4609 ± 0.0828	0.3033 ± 0.1318	0.4094 ± 0.0342	0.4345 ± 0.1088	0.4607 ± 0.0695	0.4529 ± 0.1228	0.5285 ± 0.0509	
	AUROC	0.5022 ± 0.0508	0.4828 ± 0.0893	0.4656 ± 0.0379	0.5158 ± 0.0615	0.5017 ± 0.0499	0.5239 ± 0.1089	0.5431 ± 0.0375	
Nursery	ACC	0.76575 ± 0.0058	0.7924 ± 0.0307	0.8856 ± 0.0052	0.8879 ± 0.0064	0.8844 ± 0.0056	0.8537 ± 0.0062	0.8881 ± 0.0055	
	F1	0.5931 ± 0.0080	0.5988 ± 0.0161	0.7189 ± 0.0107	0.6758 ± 0.0034	0.6729 ± 0.0061	0.6481 ± 0.0067	0.7196 ± 0.0129	
	AUROC	0.9191 ± 0.0025	0.8705 ± 0.0215	0.9747 ± 0.0034	0.9715 ± 0.0017	0.9789 ± 0.0016	0.9499 ± 0.0042	0.9748 ± 0.0033	
Iris	ACC	0.9542 ± 0.0305	0.8287 ± 0.0848	0.9464 ± 0.0320	0.8696 ± 0.0458	0.9515 ± 0.0275	0.9042 ± 0.0432	0.9548 ± 0.0266	
	F1	0.9534 ± 0.0313	0.8265 ± 0.0843	0.9454 ± 0.0326	0.8686 ± 0.0457	0.9502 ± 0.0284	0.9025 ± 0.0416	0.9536 ± 0.0278	
	AUROC	0.9972 ± 0.0037	0.9268 ± 0.0544	0.9919 ± 0.0087	0.8909 ± 0.0552	0.9938 ± 0.0065	0.9823 ± 0.0124	0.9949 ± 0.0083	
Wine	ACC	0.9775 ± 0.0174	0.9529 ± 0.0254	0.9631 ± 0.0242	0.9512 ± 0.0270	0.9790 ± 0.0160	0.9731 ± 0.0180	0.9787 ± 0.0168	
	F1	0.9774 ± 0.0170	0.9528 ± 0.0252	0.9630 ± 0.0242	0.9516 ± 0.0272	0.9789 ± 0.0162	0.9731 ± 0.0181	0.9785 ± 0.0168	
	AUROC	0.9995 ± 0.0009	0.9927 ± 0.0072	0.9956 ± 0.0058	0.9937 ± 0.0064	0.9994 ± 0.0008	0.9997 ± 0.0005	0.9965 ± 0.0056	
Breast	ACC	0.6497 ± 0.0809	0.4040 ± 0.0924	0.6206 ± 0.0901	0.2758 ± 0.0747	0.6325 ± 0.0741	0.6023 ± 0.0848	0.6638 ± 0.0694	
	F1	0.6291 ± 0.0694	0.3974 ± 0.0925	0.5970 ± 0.0778	0.2350 ± 0.0676	0.6049 ± 0.0516	0.5797 ± 0.0726	0.6370 ± 0.0535	
	AUROC	0.8790 ± 0.0321	0.7290 ± 0.0653	0.8725 ± 0.0456	0.7715 ± 0.0535	0.8777 ± 0.0369	0.8453 ± 0.0412	0.8800 ± 0.0361	
Banance	ACC	0.8697 ± 0.0197	0.7506 ± 0.0902	0.8335 ± 0.0205	0.8442 ± 0.0197	0.8875 ± 0.0182	0.8678 ± 0.0211	0.8895 ± 0.0181	
	F1	0.6053 ± 0.0129	0.6057 ± 0.0524	0.6094 ± 0.0226	0.6273 ± 0.0301	0.6338 ± 0.0280	0.6477 ± 0.0352	0.6552 ± 0.0272	
	AUROC	0.9067 ± 0.0084	0.8096 ± 0.0194	0.8098 ± 0.0264	0.8511 ± 0.0084	0.8964 ± 0.0264	0.9029 ± 0.0194	0.9074 ± 0.0182	
Hayes	ACC	0.5560 ± 0.0665	0.4758 ± 0.0720	0.4491 ± 0.0780	0.5267 ± 0.0890	0.6327 ± 0.0909	0.5801 ± 0.0755	0.6353 ± 0.0941	
	F1	0.5866 ± 0.0666	0.4749 ± 0.0810	0.4561 ± 0.0906	0.4578 ± 0.1087	0.6589 ± 0.0916	0.6036 ± 0.0754	0.6583 ± 0.0983	
	AUROC	0.7821 ± 0.0481	0.6864 ± 0.0588	0.7096 ± 0.0514	0.7620 ± 0.0606	0.7947 ± 0.0549	0.7596 ± 0.0529	0.7954 ± 0.0578	
Page	ACC	0.9577 ± 0.0045	0.9611 ± 0.0066	0.9627 ± 0.0041	0.9634 ± 0.0043	0.9672 ± 0.0041	0.9646 ± 0.0043	0.9676 ± 0.0040	
	F1	0.7733 ± 0.0353	0.7775 ± 0.0433	0.7962 ± 0.0333	0.7455 ± 0.0426	0.8083 ± 0.0329	0.7991 ± 0.0338	0.8155 ± 0.0309	
	AUROC	0.9823 ± 0.0052	0.9738 ± 0.0070	0.9442 ± 0.0135	0.9676 ± 0.0152	0.9854 ± 0.0042	0.9883 ± 0.0039	0.9864 ± 0.0098	

Bold indicates that the method performs best on the specified dataset for the corresponding evaluation metric

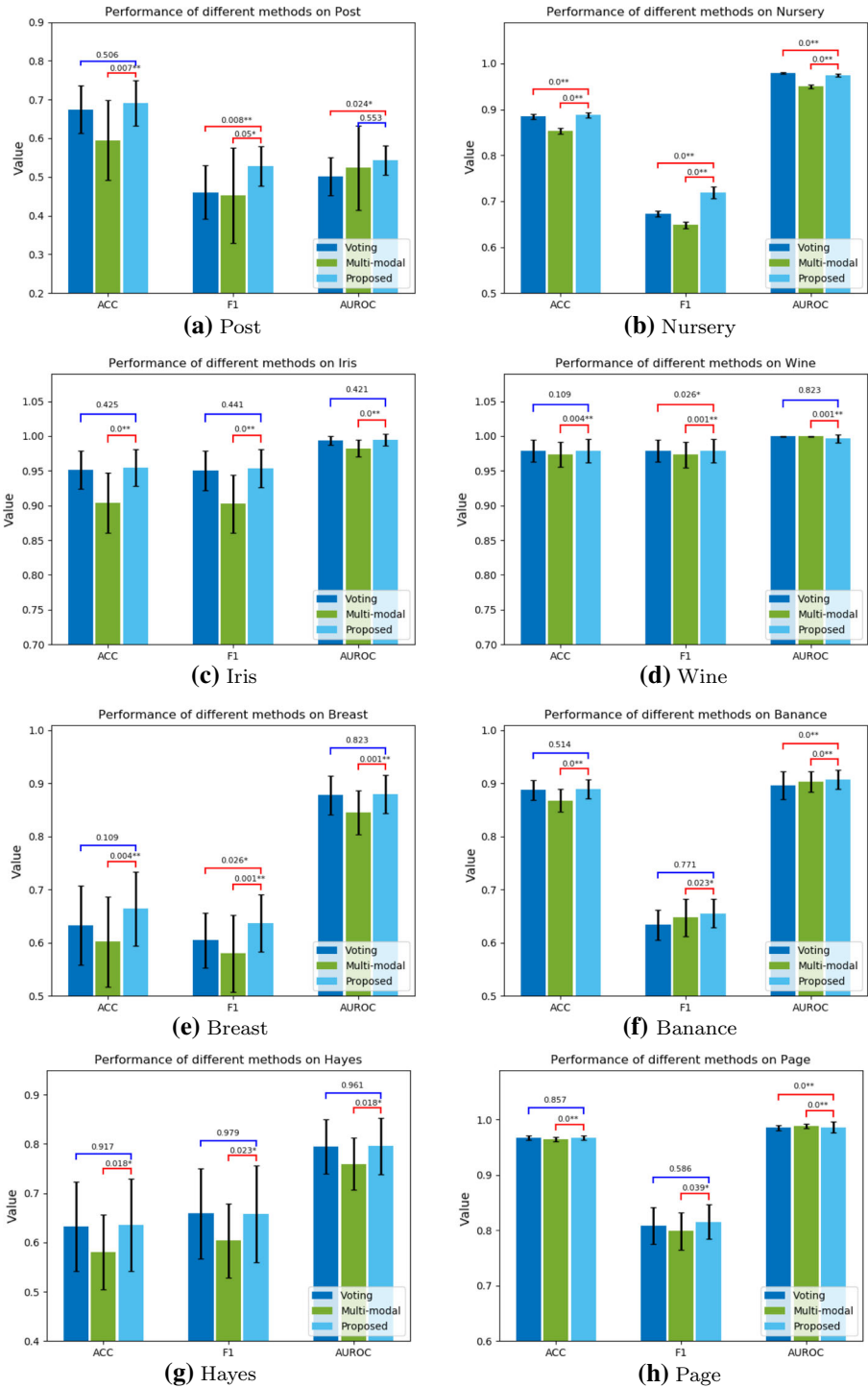


Fig. 5 The performance of different methods in different data sets

ferent public data sets are introduced to test the performance of the proposed algorithm in different applications by Accuracy, F1 and AUROC. The experimental results confirm that the proposed multi-classifier ensemble algorithm performs better than all single classifiers, the Voting-based method, and the Multi-modal method in classification performance of different application scenarios.

In future research, we will further explore two aspects: First, how does the number of classifiers affect the performance of the ensemble method? Second, when the neural network with reverse recursion mechanism is used as the basic classifier, the parameter optimization method of the basic classifier will be studied.

Acknowledgements This research was funded by Application of collaborative precision positioning service for mass users (2016YFB0501805-1), National Development and Reform Commission integrated data service system infrastructure platform construction project (JZNYYY001), Guangxi Key Lab of Multi-source Information Mining & Security (MIMS21-M-04).

References

1. Altman N (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat* 46(3):175–185. <https://doi.org/10.1080/00031305.1992.10475879>
2. Campos GO, Zimek A, Sander J, Campello R, Micenkova B, Schubert E, Assent I, Houle ME (2015) On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data Min Knowl Disc* 30:891–927
3. Chen Q, Shi L, Na J, Ren X, Nan Y (2018) Adaptive echo state network control for a class of pure-feedback systems with input and output constraints. *Neurocomputing* 275:1370–1382. <https://doi.org/10.1016/j.neucom.2017.09.083>
4. Chen W, Li Y, Tsangaratos P, Shahabi H, Ilia I, Xue W, Bian H (2020) Groundwater spring potential mapping using artificial intelligence approach based on kernel logistic regression, random forest, and alternating decision tree models. *Appl Sci*. <https://doi.org/10.3390/app10020425>
5. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1023/A:1022627411411>
6. Dempster AP (1967) Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* 54(3–4):515–528. <https://doi.org/10.1093/biomet/54.3-4.515>
7. Deng W, Yao R, Zhao H, Yang X, Li G (2019) A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. *Soft Comput* 23:2445–2462
8. Denoeux T (1995) A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Trans Syst* 25(5):804–813. <https://doi.org/10.1109/21.376493>
9. Dietterich TG (2000) Ensemble methods in machine learning. In: *Multiple classifier systems*. Springer, Berlin, pp 1–15
10. Duijn RPW, Tax DMJ (2000) Experiments with classifier combining rules. In: *Multiple classifier systems*. Springer, Berlin, pp 16–29
11. Erfani SM, Rajasegarar S, Karunasekera S, Leckie C (2016) High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognit* 121–134
12. Farooq A, Anwar S, Awais M, Rehman S (2017) A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *2017 IEEE international conference on imaging systems and techniques (IST)*, pp 1–6
13. Gerhardt N, Schwolow S, Rohn S, Pérez-Cacho PR, Galán-Soldevilla H, Arce L, Weller P (2019) Quality assessment of olive oils based on temperature-ramped HS-GC-IMS and sensory evaluation: comparison of different processing approaches by LDA, kNN, and SVM. *Food Chem* 278:720–728. <https://doi.org/10.1016/j.foodchem.2018.11.095>
14. Hansen L, Salamon P (1990) Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell* 12(10):993–1001. <https://doi.org/10.1109/34.58871>
15. Hasan Sonet KMM, Rahman MM, Mazumder P, Reza A, Rahman RM (2017) Analyzing patterns of numerously occurring heart diseases using association rule mining. In: *2017 twelfth international conference on digital information management (ICDIM)*, pp 38–45
16. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1):489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>

17. Jaeger H (2007) Echo state network. *Scholarpedia* 2(9):2330. <https://doi.org/10.4249/scholarpedia.2330>
18. Jaeger H, Haas H (2004) Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* 304(5667):78–80. <https://doi.org/10.1126/science.1091277>
19. Johnson JM, Khoshgoftaar TM (2019) Survey on deep learning with class imbalance. *J Big Data* 6(1):27. <https://doi.org/10.1186/s40537-019-0192-5>
20. Kuncheva L (2002) Switching between selection and fusion in combining classifiers: an experiment. *IEEE Trans Syst Man Cybernet Part B Cybernet Publ IEEE Syst Man Cybernet Soc* 32(2):146
21. Ma Q, Shen L, Chen W, Wang J, Wei J, Yu Z (2016) Functional echo state network for time series classification. *Inf Sci* 373:1–20. <https://doi.org/10.1016/j.ins.2016.08.081>
22. Maldonado S, López J (2018) Dealing with high-dimensional class-imbalanced datasets: embedded feature selection for SVM classification. *Appl Soft Comput* 67:94–105. <https://doi.org/10.1016/j.asoc.2018.02.051>
23. Martins JG, Oliveira LES, Sabourin R, Britto AS (2018) Forest species recognition based on ensembles of classifiers. In: 2018 IEEE 30th international conference on tools with artificial intelligence (ICTAI), pp 371–378. <https://doi.org/10.1109/ICTAI.2018.00065>
24. Mirza B, Lin Z (2016) Meta-cognitive online sequential extreme learning machine for imbalanced and concept-drifting data classification. *Neural Netw* 80:79–94. <https://doi.org/10.1016/j.neunet.2016.04.008>
25. Murugavel ASM, Ramakrishnan S (2016) Hierarchical multi-class SVM with elm kernel for epileptic EEG signal classification. *Med Biol Eng Comput* 54(1):149–161
26. Alaa MB, Samy AN, Basseem A-M, Ahmed K, Musleh M, Eman A (2019) Predicting Liver patients using artificial neural network, pp 1–11
27. Peng Y, Lin JR, Zhang JP, Hu ZZ (2017) A hybrid data mining approach on bim-based building operation and maintenance. *Build Environ* 126:483–495. <https://doi.org/10.1016/j.buildenv.2017.09.030>
28. Platt J (1999) Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In: *Advances in large margin classifiers*. MIT Press, pp 61–74
29. Plawiak P (2017) Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals. *Swarm Evolut Comput* 39C(2018):192–208
30. Sagi O, Rokach L (2018) Ensemble learning: a survey. *WIREs Data Mining Knowl Discov* 8(4):e1249. <https://doi.org/10.1002/widm.1249>
31. Saritas MM, Yasar A (2019) Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *Int J Intell Syst Appl Eng* 7:88–91
32. Shafer G (1978) A mathematical theory of evidence. *Technometrics* 20(1):106
33. Sumaiya Thaseen I, Aswani Kumar C (2017) Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *J King Saud Univ Comput Inf Sci* 29(4):462–472. <https://doi.org/10.1016/j.jksuci.2015.12.004>
34. Tan CJ, Lim CP, Cheah Y (2014) A multi-objective evolutionary algorithm-based ensemble optimizer for feature selection and classification with neural network models. *Neurocomputing* 125:217–228. <https://doi.org/10.1016/j.neucom.2012.12.057>
35. Uriz M, Paternain D, Bustince H, Galar M (2018) A first approach towards the usage of classifiers' performance to create fuzzy measures for ensembles of classifiers: a case study on highly imbalanced datasets. In: 2018 IEEE international conference on fuzzy systems (FUZZ-IEEE), pp 1–8. <https://doi.org/10.1109/FUZZ-IEEE.2018.8491440>
36. Wang F, Zhang B, Chai S, Xia Y (2018) An extreme learning machine-based community detection algorithm in complex networks. *Complexity* 2018:1–10
37. Wang L, Wang Z, Liu S (2016) An effective multivariate time series classification approach using echo state network and adaptive differential evolution algorithm. *Expert Syst Appl* 43(C):237–249. <https://doi.org/10.1016/j.eswa.2015.08.055>
38. Wei H, Kehtarnavaz N (2020) Simultaneous utilization of inertial and video sensing for action detection and recognition in continuous action streams. *IEEE Sens J* 20(11):6055–6063. <https://doi.org/10.1109/JSEN.2020.2973361>
39. Weinberger KQ, Saul LK (2009) Distance metric learning for large margin nearest neighbor classification. *J Mach Learn Res* 10:207–244
40. Xiao W, Zhang J, Li Y, Zhang S, Yang W (2017) Class-specific cost regulation extreme learning machine for imbalanced classification. *Neurocomputing* 261:70–82. <https://doi.org/10.1016/j.neucom.2016.09.120>
41. Zhang L, Ding L, Wu X, Skibniewski MJ (2017) An improved dempster-shafer approach to construction safety risk perception. *Knowl-Based Syst* 132:30–46. <https://doi.org/10.1016/j.knosys.2017.06.014>
42. Zhao K, Sun R, Li L, Hou M, Yuan G, Sun R (2021) An improved evidence fusion algorithm in multi-sensor systems. *Appl Intell*. <https://doi.org/10.1007/s10489-021-02279-5>

43. Zhao K, Sun R, Li L, Hou M, Yuan G, Sun R (2021) An optimal evidential data fusion algorithm based on the new divergence measure of basic probability assignment. *Soft Comput.* <https://doi.org/10.1007/s00500-021-06040-5>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.