

Research Article

Creating Ensemble Classifiers with Information Entropy Diversity Measure

Jiangbo Zou,^{1,2} Xiaokang Fu,^{3,4} Lingling Guo,⁵ Chunhua Ju,³ and Jingjing Chen ⁶

¹Chinese Academy of International Trade and Economic Cooperation, Beijing 10071, China

²School of Business Administration, Zhejiang Gongshang University, Hangzhou 310018, China

³School of E-Commerce & Management Science, Zhejiang Gongshang University, Hangzhou 310018, China

⁴Sunyard System Engineering Co., Ltd., Hangzhou 310053, China

⁵College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, China

⁶Zhijiang College, Zhejiang University of Technology, Shaoxing 312030, China

Correspondence should be addressed to Jingjing Chen; joyjchan@gmail.com

Received 25 March 2021; Revised 20 April 2021; Accepted 9 May 2021; Published 24 May 2021

Academic Editor: Xiaokang Zhou

Copyright © 2021 Jiangbo Zou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ensemble classifiers improve the classification accuracy by incorporating the decisions made by its component classifiers. Basically, there are two steps to create an ensemble classifier: one is to generate base classifiers and the other is to align the base classifiers to achieve maximum accuracy integrally. One of the major problems in creating ensemble classifiers is the classification accuracy and diversity of the component classifiers. In this paper, we propose an ensemble classifier generating algorithm to improve the accuracy of an ensemble classification and to maximize the diversity of its component classifiers. In this algorithm, information entropy is introduced to measure the diversity of component classifiers, and a cyclic iterative optimization selection tactic is applied to select component classifiers from base classifiers, in which the number of component classifiers is dynamically adjusted to minimize system cost. It is demonstrated that our method has an obvious lower memory cost with higher classification accuracy compared with existing classifier methods.

1. Introduction

The ensemble method was firstly proposed by Hansen and Salamon to optimize neural networks [1]. It is well known that an ensemble learning model is usually more accurate than a single learning model [2–8]. According to Singh's work, classifier combination is now widely applied in the area of machine learning and pattern recognition, such as text classification, speech recognition, seismic wave analysis, communication network, and online transaction log analysis [9]. Instead of constructing a monolithic system, ensemble learning is used to construct a pool of learners and combine them in a smart way into an overall system. In the research area of dynamic data stream classification, ensemble learning has become one of the hot spots [10].

Recently, a great number of researches propose various kinds of classifiers especially in the field of data stream

mining [11, 12]. To cope with concept drift, some published papers focus on dynamic weight mechanism. Previously, Wang and Pineau proposed online cost-sensitive boosting algorithms for online ensemble algorithms, which can achieve the similar accuracy of traditional boosting with simpler base models [13]. Tennant et al. presented a real-time data stream classifier to address the overlap of the velocity and volume aspects of big data analytics, which is adaptive to concept drift [14].

These ensemble classification approaches have good stability and can overcome the general concept drift phenomenon in data stream classification. However, there is no evidence that an ensemble classifier system with more single base classifiers is better than the ensemble classifier system with fewer single base classifiers. Sometimes, the classifier fusion method creates large-scale classifiers that require a great deal of memory and computing resources, leading to

low efficiency. To solve this problem, Zhou offered valuable sights about the diversity metric, which can be leveraged to select a subset of learners to comprise the final ensemble [15]. It was proved that ensemble classifiers with greater diversity have stronger generalization ability. However, Bi demonstrated that the accuracy of classifiers was not strongly correlated with the diversity; in some contexts, the relationship was negative [16]. Luo put forward a self-adapted classifier ensemble method with particle classification information, considering both ensemble classifier accuracy and diversity [17]. In this model, particle classification information was used to mark the learning effect, and the product of weighted accuracy and diversity was the selection criteria in a based classifier filter named C-Lib. Thus, whether diversity correlated with ensemble performance is still unclear.

In this paper, we propose a method for generating ensemble classifiers by measuring the diversity between base classifiers, coming up with an incremental classification algorithm to maximize the diversity of component classifiers as well as minimizing the system cost of an ensemble classifier. We verify the approach in data stream mining along with other traditional algorithms, suggesting that the proposed model is efficient and promising.

2. Materials and Methods

2.1. Ensemble Classifier Diversity. The diversity of an ensemble classifier is the difference of base classifiers, and an ensemble classifier with high diversity means complementarity. According to the previous work, data misclassified by one classifier can be probably correctly classified by others, leading to higher overall performance and better stability of an ensemble classifier with several different base classifiers than that of a single classifier [18, 19].

Figure 1(a) presents the diversity between two linear classifiers in an ensemble classifier with different data distributions. Suppose Dataset A and Dataset B are datasets from two different classes, and the data distributions are denoted by two anomalous curves. Linear classifier p and linear classifier q are two selected base classifiers of an ensemble classifier, which are trained by test dataset. The correctly classified data by linear classifier p is denoted by the regions marked with horizontal bars: areas S1 and S2. The correctly classified data by linear classifier q is denoted by the regions marked with vertical bars: area H1. It is clear that there are still many blank regions left without correction and the diversity between the two linear classifiers is not high, resulting in ineffective ensemble classifiers.

Suppose we select another two base classifiers, linear classifier i and linear classifier j , which are combined to an ensemble classifier with the same data distribution in Figure 1(a). Figure 1(b) shows a better classification result. More data is correctly classified as shown in areas of S1', S2', H1', and H2'. Comparing with the two ensemble classifiers, it is noted that the ensemble linear classifier with base classifier set (i, j) is significantly superior to the ensemble linear classifier with the base classifier set (p, q) , which may

attribute to the base classifier selection and optimization [20, 21].

In data stream mining, the distribution of the dataset changes rapidly over time. In Figure 1(c), the dataset is changed to dataset M and dataset N that are two typical data stream datasets, and the base classifiers of the ensemble classifier remain the same. It is suggested that the proportion of blank regions in Figure 1(b) increases and the accuracy of the ensemble classifier (i, j) for the dataset (A, B) is lower than that for the dataset (M, N). Such classification performance is far from the requirements for dynamical stream data mining. Therefore, the diversity of the same ensemble classifiers can be different when the dataset changed. The diversity measure method is particularly important in classifier combination optimization for better selection decision support, as well as the low computing resource consumption, especially in data stream classifiers.

Further, suppose another base linear classifier E is added to the ensemble classifier in Figure 1(b). If all the data in dataset A and dataset B can be correctly classified by the new ensemble classifier with E without blank area, such ensemble classifier is considered as the best ensemble classifiers and the diversity is positively correlated with ensemble classification accuracy. Instead, if the blank area is bigger than that in Figure 1(b), the performance of the new ensemble classifier with another base classifier F is lower and the diversity among base classifiers is negatively correlated with ensemble classification accuracy.

2.2. Diversity Measure Method in Ensemble Classifiers. Diversity among the members of a team of classifiers is deemed to be a key issue in the classifier ensemble problem. Unfortunately, diversity measurement is not straightforward because there is no generally accepted definition [14, 22–24]. According to Zhukov et al. [11], the diversity measure methods for ensemble classifiers can be divided into two categories: pairwise measure and nonpairwise measure. Pairwise diversity measures emphasis on local optimum calculates the average (dis) similarity metric between all possible pairs of individual classifiers in an ensemble, such as Q-statistic and correlation coefficient. Nonpairwise measure emphasizes on global optimum, which often calculates a statistic using the notion of entropy or using (dis) similarity metrics between individual classifiers and the averaged classifier [25–27]. Both methods combine accuracy and diversity together.

Relevant concepts are defined to describe the two types of diversity measures as follows: let $Z = \{z_1, \dots, z_N\}$ be a training dataset with labels with M different classes in total, $z_j \in \mathcal{R}^n$ coming from the classification problem in question. Let $D = \{D_1, D_2, \dots, D_M\}$ be a set of base classifiers, D_i an N-dimensional binary vector, and vector $C = \{1, 2, \dots, M\}$ the class label set. Assume z_j is a sample of training data from dataset Z, $z_j = \{A_1, A_2, \dots, A_s, C_j\}$, described by s features value A and one class label value C_j belongs to C. The output of a base classifier D_i for z_j is denoted y_{ij} , $i = 1, \dots, L$ if D_i classified z_j to a class correctly, and 0; otherwise, $i = 1, \dots, L$ by an N-dimensional binary vector $y_i = [y_{i1}, \dots, y_{iN}]^T$. So, come

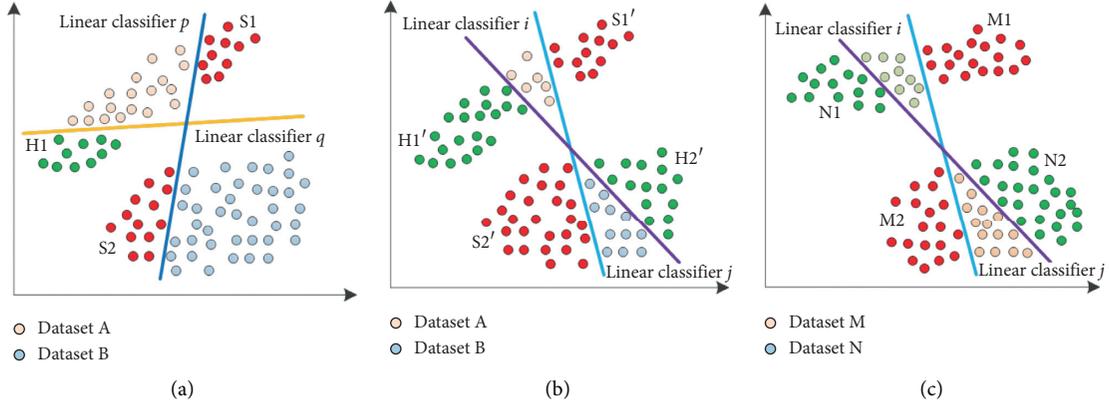


FIGURE 1: (a) Linear ensemble classifier; (b) linear ensemble classifier with higher diversity; and (c) linear ensemble classifier in data stream classification.

up the output matrix $Y = [y_j, i]^{L \times M}$, including all of the classifying results from the training dataset Z and base classifiers set D . Let D_i and D_k be a pair of base classifiers from D ; the relationship between them can be described as Table 1.

N^{ab} means the amount of training data that can be correctly classified by base classifier D_i , D_k or not. For example, N^{10} represents the amount of training data samples which are correctly classified by base classifier D_i , which are incorrectly classified by D_k . The table is from the conception of the confusion matrix. The size of training dataset Z is N , obviously, $N = N^{11} + N^{10} + N^{01} + N^{00}$, and two commonly used measures of diversity will be given as follows.

According to Yule's Q -statistic, the diversity between two base classifiers D_i and D_k can be calculated by the equation:

$$Q_{ik} = \frac{N^{11}N^{00} - N^{10}N^{01}}{N^{11}N^{00} + N^{10}N^{01}}, \quad (1)$$

where N^{ab} is the number of elements z_j of Z for which $y_i, j = a$ and $y_k, j = b$ (see Table 1). Q_{ik} ranges between -1 and 1 ; classifiers that classify more common objects correctly have a positive Q value. In contrast, those that classify more objects to different classes will result in a negative Q value. If two base classifiers are statistically independent, the expectation of Q_{ik} is 0 [6, 28].

The correlation coefficient between two base classifiers can be calculated as follows:

$$\rho_{ik} = \frac{N^{11}N^{00} - N^{10}N^{01}}{\sqrt{(N^{11} + N^{10})(N^{01} + N^{00})(N^{11} + N^{01})(N^{10} + N^{00})}}, \quad (2)$$

ρ_{ik} has the range as Q_{ik} , and they have the same changing trend. It can be proved that $|\rho_{ik}| < |Q_{ik}|$ [10]. For this comparison, diversity that measures by Q -statistic is more accurate and sensitive than that by the correlation coefficient.

In addition to these two pairwise measures of diversity, there are many other methods. The disagreement measure and the double-fault measure are two popular measures. In

TABLE 1: A 2×2 table of the relationship between a pair of classifiers.

	D_k correct (1)	D_k wrong (0)
D_i correct (1)	N^{11}	N^{10}
D_i wrong (2)	N^{01}	N^{00}
Total, $N = N^{00} + N^{01} + N^{10} + N^{11}$		

processing data stream by ensemble classifiers, pairwise diversity measure is an effective way to incrementally adjust the number of base classifiers. However, this paper applies nonpairwise diversity measures to classifier ensemble in the processing data stream, because nonpairwise measures can ensure the global optimal among classifiers when learning ensemble classifier. In this paper, information entropy is incorporated into the diversity measure. Entropy is defined as a measure of uncertainty in information theory; the greater the entropy value, the smaller the information uncertain degrees, and vice versa. Information entropy can be applied to the diversity measures of nonpairwise classifiers through the transformation of entropy.

For a data sample $Z_j, Z_j \in Z$, the output of base classifier D_i for the training data Z_j is denoted by y_{ji} . If Z_j is successfully classified by D_i , $y_{ji} = 1$, and otherwise $0, i = 1, \dots, L$. If the outputs of $\lfloor L/2 \rfloor$ of the L base classifiers for Z_j are the same (0 or 1), the outputs of the left $L - \lfloor L/2 \rfloor$ of the L base classifiers are the alternative value, coming up to the highest diversity among classifiers for Z_j . If all the y_{ji} values of the L base classifiers are the same, 0s or all 1s, there is no disagreement among base classifiers, coming up to the lowest diversity among classifiers for Z_j . For N training data, the measure of diversity based on information entropy is as the following equation:

$$E = \frac{1}{N} \sum_{j=1}^N \frac{1}{(L - L/2)} \min\{R(Z_j), L - R(Z_j)\}. \quad (3)$$

In equation (3), $R(Z_j)$ denotes the number of classifiers from D with the same output value y_{ji} , and entropy E varies between 0 and 1, where 0 indicates no difference and 1 indicates the highest possible diversity among the base

classifiers in D . In the context of data stream mining, E equals 0 means the lowest diversity among the base classifiers, and the number of base classifiers in the ensemble classifier can be reduced due to the reasonable classifier effectiveness. In contrast, the E value close to 1 means the diversity of the classifiers is high; several new base classifiers can be added to the ensemble classifier for better classification effectiveness. Based on the above concepts, we design an incremental classification algorithm based on information entropy diversity measures to optimize the effectiveness of ensemble classifiers data stream processing.

2.3. An Incremental Classification Algorithm Based on Information Entropy Diversity Measure. A typical data stream processing flow chart is shown in Figure 2. A data stream is inputted in an incremental ensemble classifier continuously chronologically. The data stream is processed according to the time period and the time granularity, which is set based on different requirements. For example, the weblog data stream frequently changes so a fine time granularity is required. However, for the credit-rating data stream, a wide time granular can be accepted.

In the time period from $[t-f]$ to $[t]$, ensemble classifier L_{t-f} deals with coming data which arrive during the f times period, while at the time $[t]$, the model will be incrementally updated coming with a new ensemble classifier L_t to process data during $[t]$ to $[t+f]$. In order to make an ensemble model to prevent concept drift when processing data stream, an incremental process is necessary which can be achieved by iterating the process of updating the model in each time period.

Taking time $[t]$ for example, the training dataset of L_t is mainly composed of labeled data, which has already been classified by ensemble classifier L_{t-f} in the period of $[t-f]$ to $[t]$. First, base classifiers are generated from the labeled training dataset by selected classification algorithms. Second, a certain number of base classifiers are selected to combine an incremental ensemble classifier L_t at time of $[t]$. The base classifiers in the new ensemble classifier are selected from ensemble new learning classifiers and old classifiers. The selection is based on two criteria, accuracy and diversity, which are measured by transformed information entropy. On one hand, we use accuracy as a criterion to remove base classifiers which have poor classification performance. On the other hand, the diversity criterion is used to adjust the number of base classifiers to achieve the global optimization of incremental ensemble classifier [29–31].

2.4. Incremental_SEM Algorithm. The most important process in generating an incremental ensemble classifier is selecting the most suitable classifiers with great accuracy and a proper number of classifiers. In this paper, the basic tactic for base classifier selection is integrating information entropy measure to the cyclic iterative selection algorithm, along with the accuracy performance data. The pseudocode of the base classifier selection algorithm for the proposed incremental classification model is given in Algorithm 1 Incremental_SEM.

Incremental_SEM uses cyclic iterative optimization selection method to maximize the information entropy difference and dynamically adjust the number of ensemble classifiers. The key part of the algorithm lies in the setting of the interval threshold of classification diversity, which should be set according to different applications. Since the initialization and preprocessing part is the same as the traditional method of the processing data stream, it is skipped in the paper. Starting from computing the diversity of ensemble classifier L_{t-f} , we compare its value to the interval threshold and take different actions according to the comparison (line 3). If the value is higher than the upper limit of the interval threshold, keep generating a new base classifier and add it into the ensemble classifier. Recompute the diversity of a new ensemble classifier until the diversity is located in the interval threshold (lines 4–13). If the value is lower than the lower limit of the interval threshold, compute the accuracy of each base classifier and kick out the base classifier with the lowest accuracy (lines 14–19). Otherwise, if the value is located in the interval threshold, it is no need to update the ensemble classifier for the next time stage (lines 21–23).

3. Results

This section lists the experiments conducted to evaluate the performance of the proposed algorithm on data stream classification. Trace based simulation approach has been used to evaluate and compare the performance of the proposed algorithm with other baseline algorithms.

3.1. Experimental Data. The proposed algorithm was evaluated on steam data generated by a massive online analysis (MOA) system. MOA is a software environment for implementing algorithms and running experiments for online learning from evolving data streams. We select the following stream generators to generate data.

- (i) Hyperplane generator generates a problem of predicting the class of a rotating hyperplane. HP1 and HP2 are the data stream generated by hyperplane generator with 5% noise data in the experiment.
- (ii) Random tree generator generates a random radial basis function stream. It constructs a decision tree by choosing attributes at random to split and assigning a random class label to each leaf. RT1 and RT2 are data stream generated by random tree generator with both label attribute and number attribute.
- (iii) SEA generator generates SEA concept functions. This dataset contains abrupt concept drift. SEA1 is the data stream generated by SEA generator with 5% noise data and concept drift.
- (iv) STAGGER generator generates STAGGER Concept functions which were introduced by Schlimmer. SG1 is generated by STAGGER generator.

A detailed description of the experimental data stream is shown in Algorithm 1. Due to the infinite nature of data

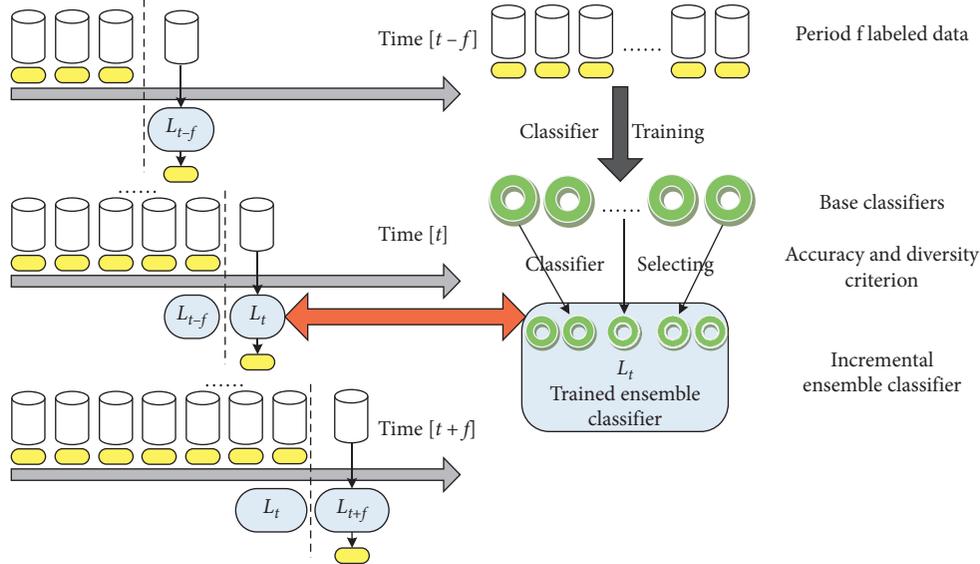


FIGURE 2: Incremental model of processing data stream.

stream in a real environment, it is not easy to do simulations in experiments. Massive data is used to simulate infinite data stream, the experiment data size of each dataset is shown in column 2 of Table 2.

Figure 3 shows the scatter diagrams of each dataset that helps understand data more intuitively. Since the volume of each dataset is large, partial data is selected to be shown in the diagrams. Usually, a dimension reduction operation is needed for the preprocessing dataset. As shown in Figure 3, it can be found that each attribute is nonlinear relativity.

3.2. Experiment Setup. An open-source mining software, WEKA, has been used to realize the ensemble classifier algorithms. The baseline algorithms in Weka are Naïve Bayes, Sequential Minimal Optimization (SMO), J48 that is the implementation of C4.5 for building a decision tree, IBk that is the implementation of the K -nearest neighbor algorithm (KNN), Kstar that is an instance-based classifier, NNge, PART that builds a “partial” C4.5 decision tree in each iteration and makes the “best” leaf into a rule and AOD [32, 33]. The algorithms are shown in Table 3.

A computer with 1.73 GHz CUP and 2 G memory is used as the experiment computer, installed with the operating system Windows XP. In order to study the effectiveness of the proposed approach, experiments were setup to compare Incremental_SEM with Bagging and AdaBoost on different datasets. In all ensemble methods, decision trees were used as the base classifier. Based on WEKA 3.6, the decision tree construction method was J48 from the Weka library, which are selected to generate base classifiers with the default parameter sets [10]. The performance of each ensemble classifier was evaluated using a stratified 10-fold cross-validation procedure, in which the original dataset was partitioned randomly into 10 equal size subsamples and each fold contains roughly the same proportions of class labels. The experiment settings were as follows: the parameters of

Bagging and AdaBoost were kept at their default values in Weka. The ensemble size can be regarded as a hyper-parameter of the ensemble method. It can be tuned through cross-validation or using a separate validation set. It can also be thought of as an indicator of the operating complexity of the ensemble. For Incremental_SEM, different information entropy intervals were set for the six generated datasets, interval [0.21, 0.43] for HP1, HP2 and SEA1, [0.63, 0.85] for SG1, and [0.46, 0.69] for RT1 and RT2.

3.3. Results and Analysis. As shown in Figure 2, f is set as a time interval in the incremental model of the processing data stream, and the classifier is adjusted every time period. For each algorithm, the accuracy of the current ensemble classifier was calculated in every time period. We verified algorithms from two aspects: classification accuracy and system memory cost. Suppose at time t , ensemble classifier has m base classifiers; each base classifiers classification accuracy is $a_i (i = 1, 2, \dots, m)$. Take A_t as ensemble classification accuracy: $A_t = (a_1 + a_2 + \dots + a_m) / m$. The ensemble classification results of the dataset in Algorithm 1 are shown in Figure 4.

In Figure 4(a), it is clear that the accuracy of the Incremental_SEM algorithm is slightly higher than the Bagging algorithm and both of them are obviously higher than a single algorithm when comparing the experiment results of Incremental_SEM algorithm with Bagging and Single classifier in datasets HP1 and RT1 at the time interval value of 10 seconds. However, the execution time of Incremental_SEM algorithm and bagging is longer than the single classifier, mainly because diversity computing in Incremental_SEM is time-consuming. Moreover, in order to test the memory cost while adding entropy diversity in ensemble classifiers, Incremental_SEM with traditional incremental algorithms, bagging, and without diversity measure are investigated. The experiment results are shown

```

Input:
Training dataset with labels by ensemble classifier  $L_{t-f}$ ;
The interval threshold of classification diversity:  $[a, b]$ ;
Iterate number (each iterate creates a new base classifier:  $k$ )
Ensemble classifier at period of  $[t-f, t]$ :  $L_{t-f}$ 
Output: incremental ensemble classifier  $L_t$  at time  $t$ .
(1)   Begin
(2)   Loop
(3)   Compute diversity value  $\lambda_0$  of ensemble classifier  $L_{t-f}$ ;
(4)   If  $\lambda_0 \in [b, 1]$ 
(5)     For  $i = 1$  to  $k$ 
(6)       Sampling training data from labeled dataset at period of  $[t-f, t]$  by  $L_{t-f}$ ;
(7)       Generate a new base classifier  $L_i$ ;
(8)       Add  $L_i$  to  $L_{t-f}$ ;
(9)       Compute the diversity value  $\lambda_1$ ;
(10)    If  $\lambda_1 \in [a, b]$ 
(11)       $L_t = L_{t-f}$  ;
(12)      Return  $L_t$ 
(13)    End for
(14)  else if  $\lambda_0 \in [0, a]$ 
(15)    Compute the accuracy of each base classifier at  $L_{t-f}$ ;
(16)    Sort base classifiers in decreasing order of accuracy as baselist;
(17)    Delete some member base classifiers with the lowest accuracy at  $L_{t-f}$ ;
(18)    Update the  $L_{t-f}$ ;
(19)     $L_t = L_{t-f}$ ;
(20)    return  $L_t$ ;
(21)  else
(22)     $L_t = L_{t-f}$ ;
(23)    return  $L_t$ 
(24)  End if
(25)  Break;
(26)  End loop

```

ALGORITHM 1: Incremental_SEM algorithm.

in Table 4 with the average classification accuracy (ACA) and average system memory cost (ASM).

In Figure 4(b), it is noted that Incremental_SEM classification accuracy is almost the same as AdaBoost algorithm and both of them are higher than a single classifier when comparing Incremental_SEM with AdaBoost classification algorithm in datasets HP2 and RT2 at the time interval value of 20 seconds. Due to the higher dimension of the two datasets, the algorithm executing time average is longer than that in Figure 4(a), illustrating that learning a new base classifier is time-consuming. The results support the conclusion that adding diversity can increase classification time without improvement of entire effectiveness through comparing Incremental_SEM with a single algorithm.

In Figure 4(c), it is clearly noted that sharp accuracy drops (such as at times 30, 55, 60, 75) since the concept drift phenomenon existed in both two datasets when comparing Incremental_SEM with the AdaBoost algorithm at the time interval value of 5 seconds. Comparing with a single algorithm, Incremental_SEM and AdaBoost are more stable,

suggesting that ensemble classifier has an advantage when concept drift exists in the dataset. It can be concluded that adding diversity into the ensemble classifier can improve algorithm performance, which is consistent with the view by Nan and Zhou [34–37].

From Tables 4–6, it can be found that the accuracy of Incremental_SEM classification is not significantly higher than AdaBoost and Bagging algorithm and all of them are nearly the same in our experiment. However, the average system memory cost of Incremental_SEM is much lower than AdaBoost and Bagging. It can be demonstrated that, for system memory cost, Incremental_SEM classification is better than traditional ensemble classification algorithms.

In order to testify the advantages of adopting entropy as a diversity measure when processing data stream, an experiment with the dataset in Table 3 was conducted to compare Q-statistic with correlation coefficient diversity measure. Table 7 shows that Incremental_SEM average accuracy is higher with Q-statistic than that with correlation coefficient ρ .

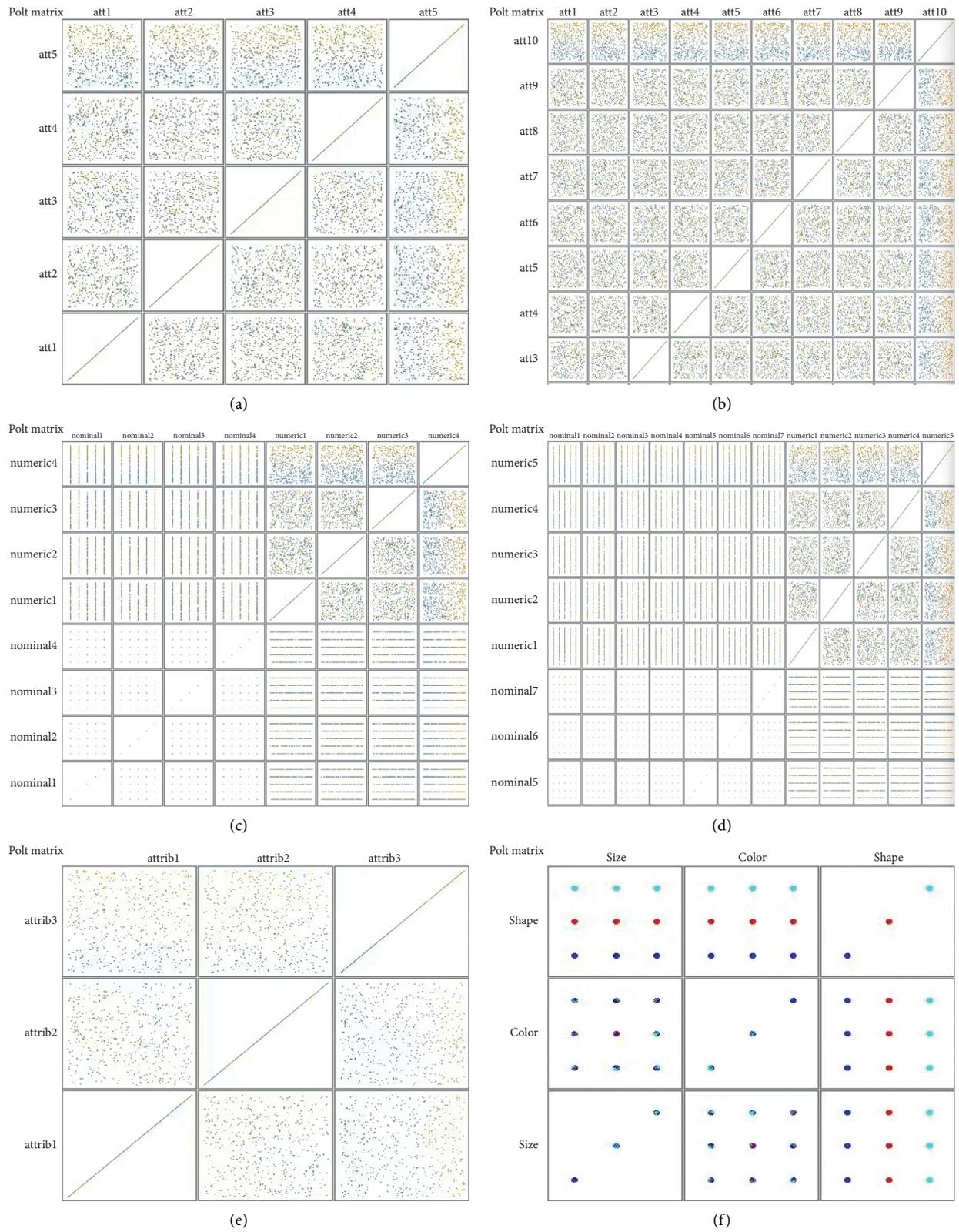
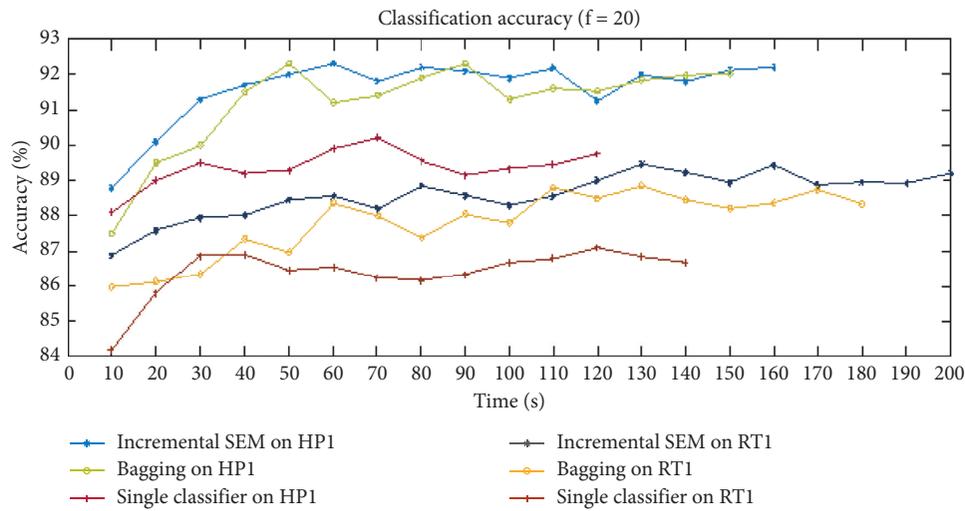


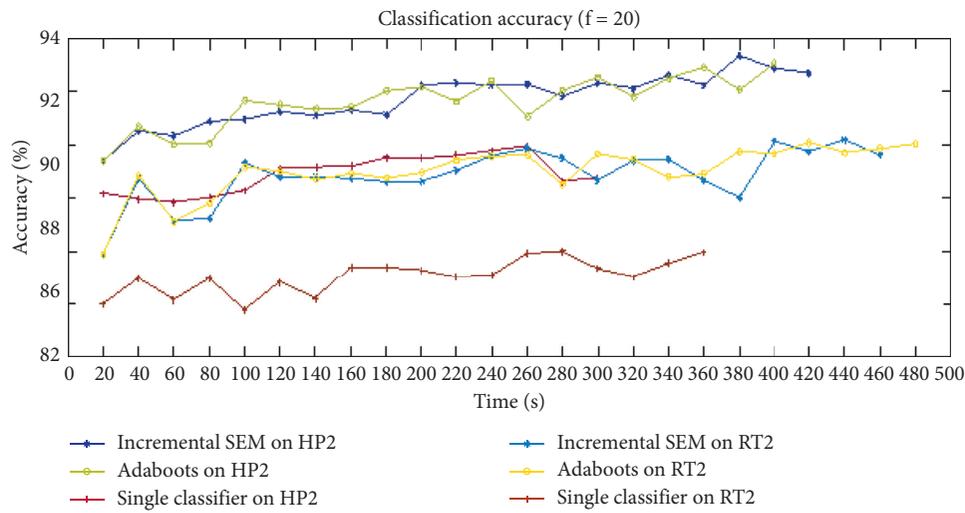
FIGURE 3: The scatter diagrams of the dataset. (a) HP1. (b) HP2. (c) RT1. (d) RT2. (e) SEA1. (f) SG1.

TABLE 2: Dataset description.

Dataset name	Data size	Classification number	Attribute number (label attribute/number attribute)
HP1	200,000	10	5 (0/5)
HP2	400,000	5	10 (0/10)
RT1	200,000	8	8 (4/4)
RT2	400,000	4	12 (7/5)
SEA1	500,000	2	3 (0/3)
SG1	500,000	2	3 (3/0)



(a)



(b)

FIGURE 4: Continued.

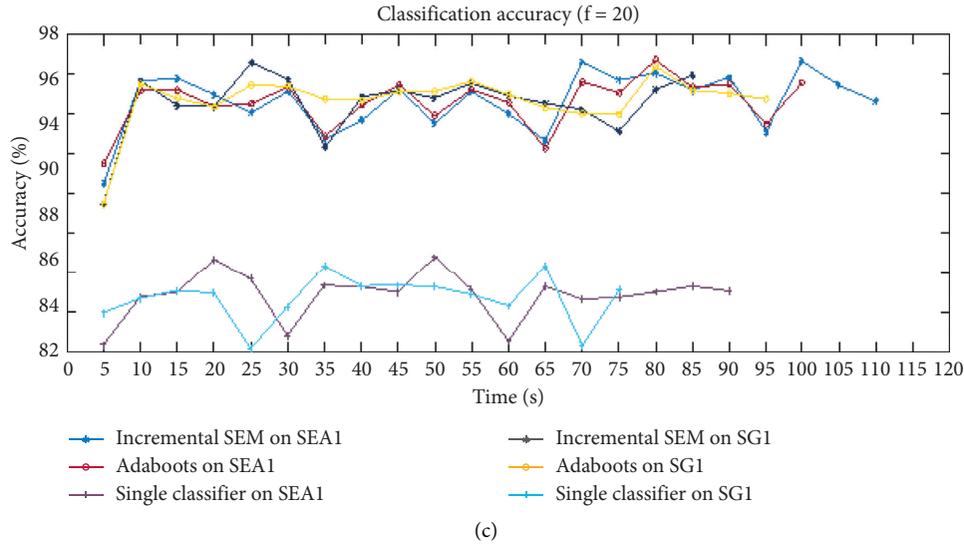


FIGURE 4: The comparison experiments on (a) dataset HP1 and RT1, (b) dataset HP2 and RT2, and (c) dataset SEA1 and SG1.

TABLE 3: A list of classification algorithms available by Weka.

Number	Classifier name	Simple description of classifiers
1	Naïve Bayes	The Naïve Bayes classifier using kernel density estimation over multiple values for continuous attributes, instead of assuming a simple normal distribution
2	SMO	Sequential minimal optimization algorithm for training a support vector classifier using polynomial kernels
3	J48	Decision tree, the implementation of C4.5
4	IBk	An instance-based learning algorithm, the implementation of k-nearest neighbor algorithm (kNN)
5	KStar	The K instance-based learner using all nearest neighbors and an entropy-based distance
6	NNge	Nearest neighbor-like algorithm using nonnested generalized exemplars
7	PART	Generating a PART decision list for classification
8	AOD	Perform classification by averaging over all of a small space of alternative Naive Bayes-like models that have weaker independence

TABLE 4: The comparison of HP1 and RT1.

		HP1	RT1 (%)
Incremental_SEM	ACA	91.608%	88.599
	ASM	217MB	344
Bagging	ACA	91.19%	87.805
	ASM	250MB	485
Single	ACA	89.372	86.398
	ASM	157MB	237

TABLE 5: The comparison of HP2 and RT2.

		HP2 (%)	RT2 (%)
Incremental_SEM	ACA	91.732	88.861
	ASM	271	292
AdaBoost	ACA	91.677	88.943
	ASM	315	358
Single	ACA	88.859	84.976
	ASM	186	198

TABLE 6: The comparison of SEA1 and SG1.

		SEA1 (%)	SG1 (%)
Incremental_SEM	ACA	94.697	94.798
	ASM	106	97
AdaBoost	ACA	94.606	94.671
	ASM	138	115
Single	ACA	85.042	84.946
	ASM	78	82

TABLE 7: Comparison experiment between Q-statistic and correlation coefficient ρ .

	HP1 (%)	HP2 (%)	RT1 (%)	RT2 (%)	SEA1 (%)	SG1 (%)
Q-statistic	1.939	1.516	1.943	2.445	0.915	1.263
ρ	1.629	1.478	2.089	2.038	0.917	0.986

4. Conclusions

Ensemble classifier, as a common algorithm at processing data stream, is famous for its high classification accuracy and stability. We proposed an ensemble algorithm incorporating entropy as the diversity measure. It is proved that our Incremental_SEM algorithm has a higher classification accuracy rate than a single classifier and lower system memory cost than the Bagging and AdaBoost algorithm. It is also suggested that the Q-statistic diversity measure outperforms the correlation coefficient diversity measure. Future research will focus on how to verify the relativeness between accuracy and diversity in theory.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request as the size of the experimental data is too large to upload via this submission interface.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

J.Z. and J.C. conceptualized the study; X.F. was responsible for methodology; J.Z., L.G., and J.C. validated the study; X.F. contributed to formal analysis; L.G. investigated the study; X.F. provided resources; J.Z. was responsible for data curation; J.Z. prepared the original draft; J.C. reviewed and edited the manuscript; C.J. was responsible for project administration. All the authors have read and agreed to the published version of the manuscript.

Acknowledgments

This research was funded by Zhejiang Gongshang University Youth Program Project (grant no. 3090JYN9920001G-332). The funders had no role in the design of the study; the collection, analyses, or interpretation of data; the writing of the manuscript, or the decision to publish the results.

References

- [1] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.
- [2] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine Learning*, vol. 51, no. 2, pp. 181–207, 2003.
- [3] L. I. Kuncheva, "Diversity in classifier ensembles," in *Combining Pattern Classifiers: Methods and Algorithms* John Wiley and Sons, Hoboken, NJ, USA, 2004.
- [4] F. Schwenker, F. Roli, and J. Kittler, "Multiple classifier systems," in *Proceedings of the 12th International Workshop, MCS 2015, Güzburg, Germany, July 2015*.
- [5] H. M. Gomes, J. P. Barddal, F. Enembreck, and A. Bifet, "A survey on ensemble learning for data stream classification," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, pp. 23–32, 2017.
- [6] D. V. R. Oliveira, G. D. C. Cavalcanti, and R. Sabourin, "Online pruning of base classifiers for dynamic ensemble selection," *Pattern Recognition*, vol. 72, pp. 44–58, 2017.
- [7] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Information Fusion*, vol. 53, pp. 155–164, 2020.
- [8] X. Zhou, W. Liang, S. Shimizu, J. Ma, and Q. Jin, "Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5790–5798, 2021.
- [9] P. K. Singh, R. Sarkar, and M. Nasipuri, "Correlation-based classifier combination in the field of pattern recognition," *Computational Intelligence*, vol. 34, no. 3, pp. 839–874, 2018.
- [10] M. Tennant, F. Stahl, O. Rana, and J. B. Gomes, "Scalable real-time classification of data streams with concept drift," *Future Generation Computer Systems*, vol. 75, pp. 187–199, 2017.
- [11] A. V. Zhukov, D. N. Sidorov, and A. M. Foley, "Random forest based approach for concept drift handling," in *Proceedings of the 2017 International Conference on Analysis of Images, Social Networks and Texts*, Moscow, Russia, July 2017.
- [12] C. M. Salgado, S. M. Vieira, L. F. Mendonça, S. Finkelstein, and J. M. C. Sousa, "Ensemble fuzzy models in personalized medicine: application to vasopressors administration," *Engineering Applications of Artificial Intelligence*, vol. 49, pp. 141–148, 2016.
- [13] B. Wang and J. Pineau, "Online bagging and boosting for imbalanced data streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3353–3366, 2016.

- [14] Z. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," *Artificial Intelligence*, vol. 137, no. 12, pp. 239–263, 2002.
- [15] X. Zhou, W. Liang, Z. Luo, and Y. Pan, "Periodic-aware intelligent prediction model for information diffusion in social networks," *IEEE Transactions on Network Science and Engineering*, 2021.
- [16] J. Luo and D. Chen, "Application of adaptive ensemble algorithm based on correctness and diversity," *Journal of Zhejiang University (Engineering Science)*, vol. 45, no. 3, pp. 558–562, 2011.
- [17] K. Yan, A. Chong, and Y. Mo, "Generative adversarial network for fault detection diagnosis of chillers," *Building and Environment*, vol. 172, Article ID 106698, 2020.
- [18] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, and F. Herrera, "A survey on data preprocessing for data stream mining: current status and future directions," *Neurocomputing*, vol. 239, pp. 39–57, 2017.
- [19] M. Mojirsheibani and C. Shaw, "Classification with incomplete functional covariates," *Statistics & Probability Letters*, vol. 139, pp. 40–46, 2018.
- [20] K. Yan, J. Su, J. Huang, and Y. Mo, "Chiller fault diagnosis based on VAE-enabled generative adversarial networks," *IEEE Transactions on Automation Science and Engineering*, 2020.
- [21] X. Zhou, X. Xu, W. Liang et al., "Intelligent small object detection based on digital twinning for smart manufacturing in industrial CPS," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.
- [22] J. Obregon, A. Kim, and J.-Y. Jung, "RuleCOSI: combination and simplification of production rules from boosted decision trees for imbalanced classification," *Expert Systems with Applications*, vol. 126, pp. 64–82, 2019.
- [23] L. Guo, Q. Liu, K. Shi, Y. Gao, J. Luo, and J. Chen, "A blockchain-driven electronic contract management system for commodity procurement in electronic power industry," *IEEE Access*, vol. 9, pp. 9473–9480, 2021.
- [24] J. Lu, D. Chen, G. Wang, D. Kiritsis, and M. Törngren, "Model-based systems engineering tool-chain for automated parameter value selection," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–15, 2021.
- [25] P. Porwik, R. Doroz, and K. Wrobel, "An ensemble learning approach to lip-based biometric verification, with a dynamic selection of classifiers," *Expert Systems with Applications*, vol. 115, pp. 673–683, 2019.
- [26] X. Zhou, Y. Hu, W. Liang, J. Ma, and Q. Jin, "Variational LSTM enhanced anomaly detection for industrial big data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3469–3477, 2021.
- [27] X. Zhou, Y. Li, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, p. 1, 2020.
- [28] J. Xia, M. Dalla Mura, J. Chanussot, P. Du, and X. He, "Random subspace ensembles for hyperspectral image classification with extended morphological attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 4768–4786, 2015.
- [29] J. Y. Choi, D. H. Kim, K. N. Plataniotis, and Y. M. Ro, "Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography," *Expert Systems with Applications*, vol. 46, pp. 106–121, 2016.
- [30] K. Yan, L. Liu, Y. Xiang, and Q. Jin, "Guest editorial: AI and machine learning solution cyber intelligence technologies: new methodologies and applications," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6626–6631, 2020.
- [31] Z. Zhang, Y. Zeng, and K. Yan, "A hybrid deep learning technology for PM 2.5 air quality forecasting," *Environmental Science and Pollution Research*, pp. 1–14, 2021.
- [32] X. Zhou, W. Liang, K. I.-K. Wang, R. Huang, and Q. Jin, "Academic influence aware and multidimensional network analysis for research collaboration navigation based on scholarly big data," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 1, pp. 246–257, 2021.
- [33] S. B. Kotsiantis, "Supervised machine learning: a review of classification techniques," in *Proceedings of the 2007 conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, Amsterdam, Netherlands, June 2007.
- [34] L. Nan and Z. Zhou, "Selective ensemble of classifier chains," in *Proceedings of the 2013 11th International Workshop on Multiple Classifier Systems, MCS 2013*, Nanjing, China, May 2013.
- [35] G. D. C. Cavalcanti, L. S. Oliveira, T. J. M. Moura, and G. V. Carvalho, "Combining diversity measures for ensemble pruning," *Pattern Recognition Letters*, vol. 74, pp. 38–45, 2016.
- [36] Y. Bi, "The impact of diversity on the accuracy of evidential classifier ensembles," *International Journal of Approximate Reasoning*, vol. 53, no. 4, pp. 584–607, 2012.
- [37] N. Jin, Y. Zeng, K. Yan, and Z. Ji, "Multivariate air quality forecasting with nested LSTM neural network," *IEEE Transactions on Industrial Informatics*, 2021.