IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 7 March 2023, accepted 26 March 2023, date of publication 29 March 2023, date of current version 3 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3262836

APPLIED RESEARCH

An Enhanced Dynamic Ensemble Selection Classifier for Imbalance Classification With Application to China Corporation Bond Default Prediction

YU WANG[1], JUNBIN ZHANG[1], AND WEI YAN[2,3]

[1]Faculty of Finance, City University of Macau, Macau, China
[2]Guangzhou Institute of International Finance, Guangzhou University, Guangzhou 510006, China
[3]Zhuhai DeltaFit Technology Company Ltd., Zhuhai 519031, China

Corresponding author: Wei Yan (yanwei.macau@connect.um.edu.mo)

ABSTRACT China corporation bond default prediction is important and can be formulated as an imbalance classification problem solved by static ensemble classifiers. However, dynamic ensemble selection (DES) classifiers have not been applied to this typical problem in the context of business research. DES classifiers are capable of selecting an ensemble classifier for each test sample, leading to better classification performance than static ensemble classifiers. Most existing DES classifiers can not address imbalance classification optimally and only use single criteria of competent for classifier selection, resulting in sub-optimal classification performance. In this paper, we propose an enhanced DES classifier, named META-DES-Diversity, that inherits strengths of data sampling, meta-learning, criteria of diversity, and dynamic weighting fusion scheme to alleviate such limitations. Specifically, the synthetic minority over-sampling technique is initially used to balance the training set before generating a candidate classifier pool. To select an ensemble classifier with a highly competent, the meta-learning framework META-DES is utilized to consider multiple criteria of competence. In complement with the meta-leaning framework, a two-phase selection strategy is utilized to perform competence and diverse ensemble classifier selection. Note that a competence-driven decision fusion scheme is employed to effectively fuse classification results from selected ensemble classifiers. Experiments on 14 two-class imbalanced data sets from the KEEL repository and one self-collected China corporation bond data set show the effectiveness and superiority of the proposed enhanced DES classifier.

INDEX TERMS Default prediction, dynamic ensemble selection (DES), imbalance classification, meta learning.

I. INTRODUCTION

China corporate bond market is the largest bond market, with over 306 billion USD.[1] According to the Fitch Ratings, one of the big three credit rating agencies, China corporate bond default rate touches a record high in 2022.[2] In the first half of 2021, China corporate bond defaults hit 116 billion RMB. The rising defaults are caused by two factors: intense regulation and a deleveraging policy. Both factors narrow Chinese enterprises' financing channels. That means it is a great challenge for highly leveraged firms to repay their principal and interest, resulting in an increasing number of bond defaults. Corporate bond default prediction can be formulated as a two-class classification problem and has received considerable attention [1].

In the classification context, ensemble classifiers (ECs) have shown their theoretic and empirical superiority over statistical classifiers, such as logistic regression [2], [3], decision

The associate editor coordinating the review of this manuscript and approving it for publication was Szidonia Lefkovits.

[1]https://www.bloomberg.com/news/articles/2022-09-07/china-overtakes-us-with-306-billion-corporate-credit-boom
[2]https://www.fitchratings.com/

tree [4], and neural networks [5]. Technically, the ECs aggregate multiple base classifiers to obtain a fused one, enabling the ECs to exploit all the strengths of base classifiers [6]. The ECs are success under the expert assumption, assuming that data are with complex feature space and each base classifier in ensemble is an expert of a specific region of feature space.

Existing ECs can be arranged into two categories: static and dynamic based on the way of classifier selection [7]. Specifically, the ECs perform prediction in three stages: generation, selection, and integration. In generation, a pool of competent base classifiers are generated. In selection, a single or an ensemble competent base classifiers from this pool is (are) selected. In integration, the prediction results of these selected classifiers are fused with a predefined rule, leading to final prediction. Among these stages, the selection is vital and attracts considerable attention. Existing strategies of selection can be classified into two categories: static selection (SS) and dynamic selection (DS). SS aims to select ensemble classifiers based on average performance of base classifiers on a validation set, which is constructed from training data set. Typical SS methods are BoostForest [8], AdaBoost [9] and XGBoost [10]. SS is all-sample-oriented, that all test samples share the same ensemble classifier. SS obtains good prediction results when it well fit the distribution of all test samples. Despite acceptable results, it has been shown that the SS cannot guarantee to obtain optimal result for every test sample. In contrast, the DS is single-sample-specific [11]. Specifically, a single competent classifier (this is known as dynamic classifier selection, DCS) or an ensemble classifier (this is known as dynamic ensemble selection, DES), is selected for predicting each test example. Generally, he competence of a classifier is evaluated by its classification performance in the local region of the test sample based on the competency criterion. For your understanding, the frameworks of static selection, dynamic classifier selection, and dynamic ensemble selection are illustrated in Fig. 1.

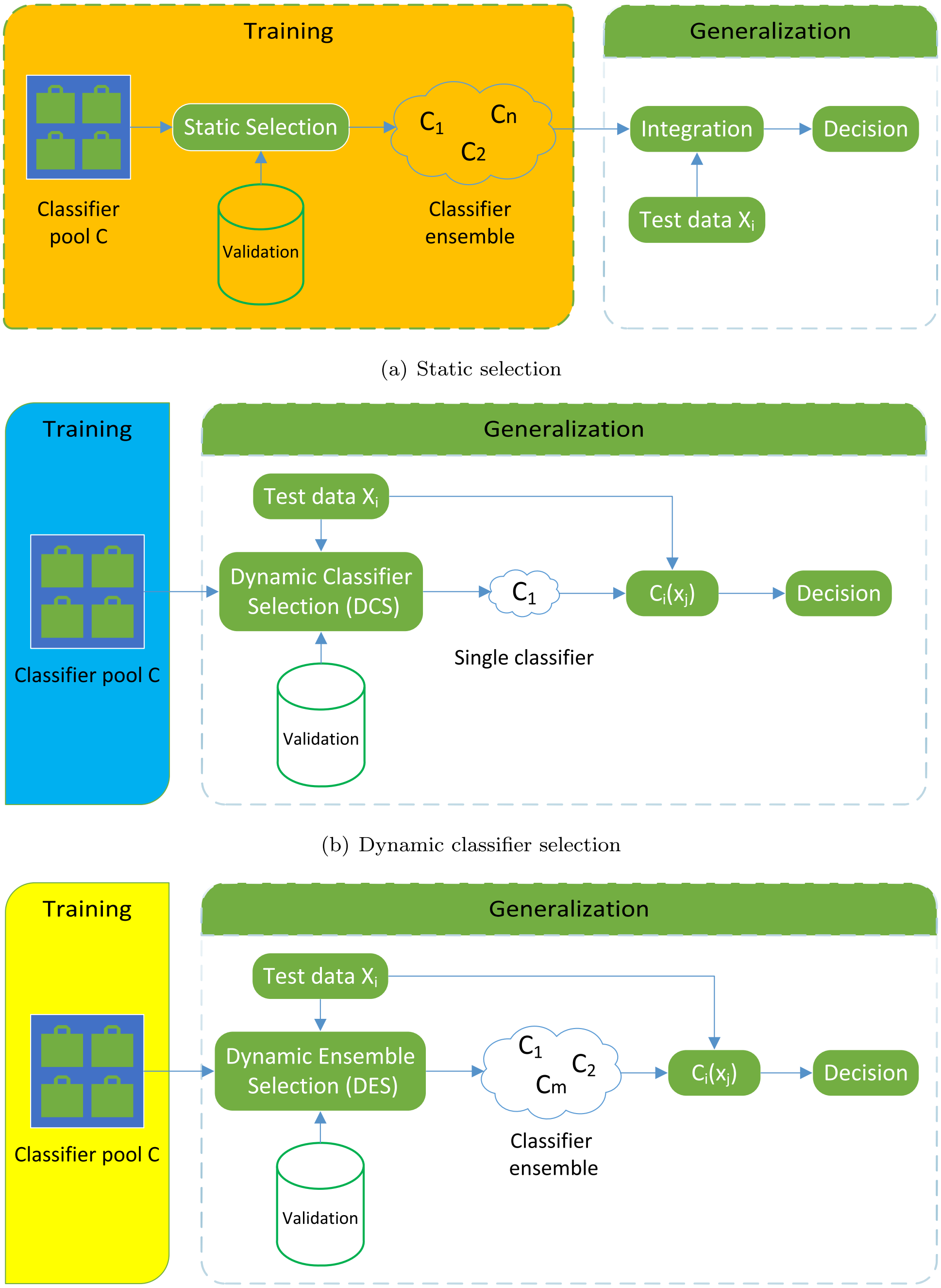
Duo to effectiveness and efficiency, the DES classifiers have drawn widely consideration [12], [13], [14], [15]. In these methods, the selection stage is critical, and the criterion used to evaluate the competence of base classifier have be investigated. Zhang et al. [16] proposed the DES-MI which used the criterion of weighted accuracy rate as to evaluate the competence level of the base classifiers for classification. García et al. [17] introduced the DES-KNN which utilized both diversity and accuracy, ensuring it can be able to select base classifiers complement each other and corporate well the ensemble. Cruz et al. [18] constructed META-DES that used multiple individual-based criteria, including probability, accuracy, behaviour, etc. Despite acceptable results, all the above-mentioned methods are sub-optimal, since they cannot exploit the diversity, accuracy and other multiple criteria, simultaneously.

Apart from classifier selection, data characteristic is also important in designing a classifier system [19]. Imbalance is a key characteristic of data from many real applications, such as extreme events prediction [20], heart disease classification [21] and plant identification [22]. In context of business research, China corporate bond data is also highly imbalanced. Specifically, the number of undefault bond (i.e., the majority class) is much higher than that of default ones (the minority class). In imbalanced classification context, traditional algorithms usually favor the majority class and fail to correctly classify the minority class, resulting in performance loss. In addition, few DES algorithms can optimally handle imbalanced data. This observation motivates us to propose a novel DES algorithm to make it suitable for imbalance classification.

To perform imbalance classification, some strategies have been proposed and can be divided into three categories: (1) algorithm-level methods directly perform imbalanced classification with sample-weighting techniques [15]. Specifically, the minority class are assigned additional weights to increase their impact. Among DES methods, DES-MI is capable of dealing with multi-class imbalanced classification. DES-MI involves a weighting mechanism to highlight the competence of classifiers that are powerful in classifying examples in the region of underrepresented competence. The minority class in the competent region of a test sample will be given higher weights, and thus helping select the competent base classifiers which correctly classify the minority class; (2) Data-level methods rely on data resampling, such as under-sampling or over-sampling techniques, for training data balancing [23]; (3) Cost-sensitive learning methods lie in between the data-level and algorithm-level approaches. It assigns different costs to samples and modifies the learning algorithm to accept the costs. Typical example is the cost sensitive SVM [24].

In this paper, we propose an enhanced DES method for imbalance classification and apply it to China corporate bond default prediction. The proposed technology is inspired by technique of data sampling [25], meta-learning framework [26], [27], technique of diversity [28] and weighted fusion strategy [29]. In particular, it directly address the imbalanced data sets and can comprehensively consider multiple criteria for selecting competent and diverse classifiers. The main contributions of this paper are summarized as follows:

- Guided by the META-DES framework, we propose an enhanced DES method, named META-DES-Diversity for binary imbalance classification. META-DES-Diversity improves META-DES through exploiting the strengths of the DES algorithms, data-sampling algorithm and decision-weighting fusion. Specifically, a over-sampling mechanism is used on the minority class for balancing training data set. Moreover, the criteria of diversity is incorporated with the META-DES framework under a two-step selection scheme, helping select competent and diverse classifiers for optimal classification. In addition, a competence driven weighted majority vote scheme is introduced for decision integration.



(a) Static selection

(b) Dynamic classifier selection

(c) Dynamic ensemble selection

FIGURE 1. Flow charts of three strategies of classifier selection: (a) Static selection; (b) Dynamic classifier selection; (3) Dynamic ensemble selection.

- We apply our META-DES-Diversity to the China corporate bond default prediction. To our best knowledge, it is the first work of applying the DES algorithm to China corporate bond default. That means we provide a future research direction for addressing task of imbalanced China corporate bond default prediction.
- We construct a China corporate bond data set and verify the effectiveness of our META-DES-Diversity on this data set. In addition, we conduct experiments on 14 data sets from the KEEL repository to demonstrate its superiority over six classifiers including five typical DES-based classifiers and a stacked classifier.

The paper is organized as follows. Section II reviews the processes of DES. Section III describes our META-DES-Diversity, including processes and algorithms. Section IV performs experiments on imbalanced data sets and compares

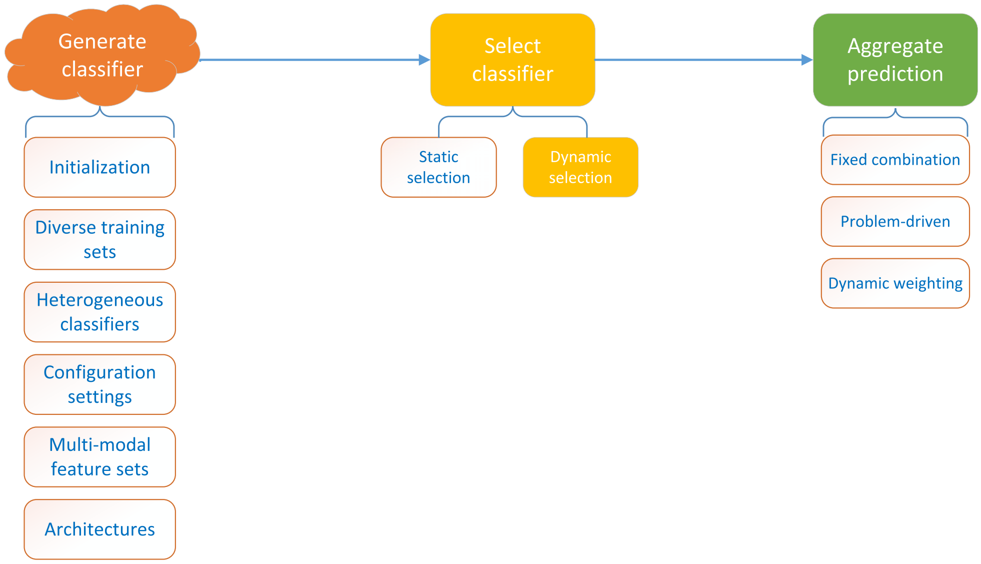


FIGURE 2. Taxonomy of three stages involves in an ensemble classifier system. We focus on the dynamic selection in this work.

it with related ensemble algorithms. Section V applies proposed method to the China corporate bond default prediction. We conclude this work in section VI.

II. LITERATURE REVIEW

A. DYNAMIC ENSEMBLE SELECTION

To perform classification, dynamic ensemble selection based classifiers share three same stages: generation, selection, and aggregation.

1) GENERATE A POOL OF BASE CLASSIFIERS

This pool should contain base classifiers with accuracy and diversity [18]. Here, diversity means that no classifiers in this pool share same classification results. To pursue diversity, there are six practical strategies: (1) try different initialization settings; (2) equip various configurations; (3) construct various layers of deep architectures for neural network classifiers; (4) combine heterogeneous classifiers; (5) construct training sets with different distributions; (6) represent data with multi-modal feature sets. In practical, one can obtain accurate and diverse base classifiers through optimal combination of these strategies.

2) SELECT COMPETENT CLASSIFIER

After generation, it comes selection. Selection can be conducted either in static or dynamic manner. With static selection, one selects an ensemble classifier with a specific criterion (e.g., diversity or accuracy) [28]. Note that this ensemble classifier is then used to predict all test samples. Despite acceptable results, these static classifiers share a same shortcoming that they cannot work well on small-size datasets. To address this shortcoming, dynamic selection was introduced with local-region assumption. This assumption refers to each base classifier is an expert in a specific region of feature space. In contrast to the static selection, dynamic selection selects a competent single base classifier or a competent ensemble for every test sample.

Technically, dynamic selection consists of three aspects:

- Define the region of competence; one need to construct a local region, which contains a subset of training samples that are around the test sample. In addition, this local region is used to estimate the competence level of the base classifiers.
- Determinate the selection criteria; with the above-defined local region, one can estimate competence levels of base classifiers for classifier selection, (e.g., accuracy, probabilistic, and ranking).
- Determinate the selection mechanism; once obtaining base classifiers, it comes selection. One can choose to select a single classifier (i.e., dynamic classifier system) or an ensemble classifiers (i.e., dynamic ensemble system). See Fig. 3 for the taxonomy of the above-mentioned three aspects.

3) AGGREGATE PREDICTION

In this stage, one needs to fuse those prediction results from selected classifiers. This fusion is performed under a specific scheme. Existing schemes can be classified into three groups: fixed combination, problem-driven and dynamic weighting.

1) Fixed combination rules [30]:
   Typical rule in this category is the Majority Voting and other common used fixed combination rules are the Sum, Product, Maximum, Minimum, Median and Majority voting. These rules share a problem that they require certain assumptions about the base classifiers in order to obtain a good performance. For example,

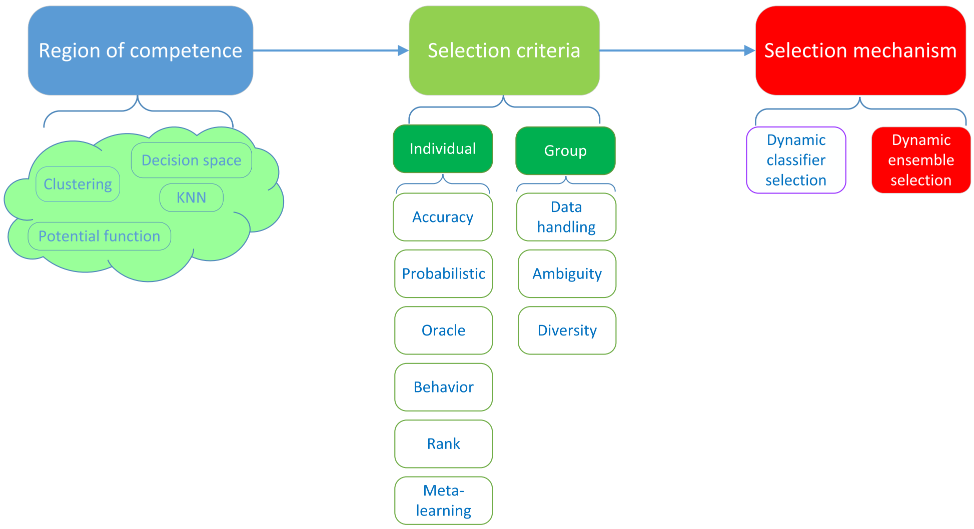


FIGURE 3. Taxonomy of three steps involving in dynamic selection. Both individual and group criteria are considered in this work.

the Majority Voting and Product rule are effective if the base classifiers are independent, while the Sum rule produces good results when the base classifiers have independent noise behavior.

2) Problem-driven rules [31]:
These rules assume that the fusion process is adapted to specific classification problem. In these rules, predictions of the base classifier are taken as features for another learning algorithm, which learns the aggregation function. Several works have shown the superiority of problem-driven rules over fixed combination rules. For example, an MLP neural network used to combine the outputs of the base experts trained using distinct feature sets outperformed all fixed combination rules for recognition of handwritten digit and character.

3) Dynamic weighting:
Essentially, dynamic weighting is the same with dynamic selection methods [7]. They are all built on the local competence of the base classifiers. Recall that local region contains samples that are around the test sample. Dynamic weighting fuses the results of all classifiers and imposes a high weight value on the competent classifier. Note that a hybrid dynamic selection and weighting scheme is optimal. In this scheme, the base classifiers that presented a certain competence level are first selected. Then, one can fuse their prediction results with imposing are weighted based on their estimated competence levels. Experimental results conducted in demonstrate that the hybrid dynamic selection and weighting approaches usually present the best classification performances when compared to performing only dynamic weighting [29].

III. PROPOSED METHOD

Recall that our motivation is to propose an enhanced DES classifier for imbalance classification. According to a recent review work [7] and an investigating work [28] on DES classifiers, one can improve a DES classifier from two aspects: optimize candidate classifier pool and hybrid ensemble selection criteria, and thus obtaining optimal imbalance classification performance. Specifically, data balancing strategy has shown good results in optimizing candidate classifier pool [7]. As for ensemble selection criteria, the DES methods that used the Double Fault diversity measures achieved more accurate ensemble systems [28]. Motivated by these observations, we proposed an enhanced DES classifier, termed META-DES-Diversity. META-DES-Diversity is built on the framework of META-DES, which is capable of considering multiple criteria on classifier competence evaluation. META-DES-Diversity improves META-DES through using data sampling technique for optimizing candidate classifier pool; in addition, it borrows the idea of two-step selection strategy from DES-KNN [17] to complement with classifier diversity. Note that a competence driven weighted majority vote scheme is introduced for precise decision integration. The flowchart of the proposed method is given in Figure 4. The details of construction are illustrated as follows.

A. CLASSIFIER POOL GENERATION

Before generation, that the oversampling technique SMOTE [32] is utilized to balance training data set. After data balancing, a pool of base classifiers are generated through utilizing the Bagging [33]. The Bagging aims to build a diverse ensemble of classifiers through randomly selecting different subsets of training data, on which each classifier is trained.

B. META LEARNING

Inspired by [18] and [27], we transfer dynamic ensemble selection problem into meta-learning problem. In fact, it is insufficient to accurately estimate the level of competence of a base classifier using only one criterion. Meta-learning

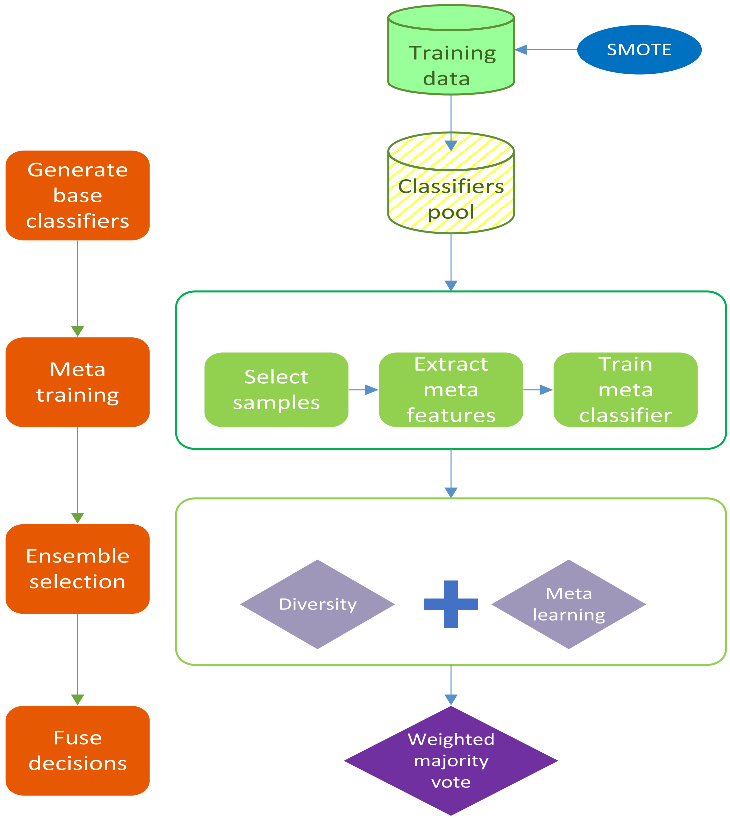


FIGURE 4. The framework of proposed META-DES-Diversity.

helps to make optimal selection, through using different kinds of criterion to decide whether a base classifier is competent enough to classify a given test sample. Technically, meta-learning refers to a two-class classification problem. Both classes are either competent or incompetent. A meta classifier is trained based on meta features and aims to classify a base classifier is whether competent or incompetent. Here, each meta-feature corresponds to a specific criterion to measure the level of competence of a base classifier. In detail, this meta-learning consists of three steps:

- Select sample: in this stage we construct a subset from training data set for meta-feature extraction. Note that a consensus threshold is used to deal with a practical case in which the extent of consensus of the classifier pool is low. Specifically, it refers to that the number of votes from the winning class is close to or even equal to the number of votes from the second class. Technically, for each training sample, the degree of consensus of the pool is computed. If this degree falls below the threshold, this sample is selected to extract meta-features.
- Extract meta-features: five distinct sets of meta-features are extracted, and then they are encoded into a unify meta-feature vector for each base classifier. (See Table 1 for details about these five meta-features.) Each feature corresponds to a specific criterion, measuring the level of competence of a base classifier, such as the classification performance estimated in a local region of the feature space and the classifier confidence for the classification of the input sample. In detail, meta-feature extraction is conducted in three steps. First, the region of competence is defined by using the k-Nearest Neighbor algorithm. Second, we get output profiles, that are prediction results of base classifiers, for this region. Third, the output profile of the query sample is calculated, following by obtaining the set with similar output profiles of the query sample through the Euclidean distance.
- Train the meta-classifier: with extracted meta-features, we then train a meta-classifier which predicts whether or not a base classifier is competent enough to classify test sample. Follow [27], we consider the Naive Bayes as the meta-classifier for optimal performance. We conclude the above meta-leaning in algorithm 1.

C. ENSEMBLE SELECTION

The task at this phase is to select competent and diverse classifiers for each test sample. We resort to a two-stages strategy that includes two steps:

- select competent classifiers from the candidate classifiers pool for each test sample. First, we estimate two competent regions: one is the feature space-based, and

Algorithm 1 Meta Learning

Require: Meta training data D_{mtrain}, Classifier pool $\mathcal{C}$, the nearest neighbors k, the number of similar output files K_p, the scaling coefficient α.
Ensure: meta-classifier γ

$\mathcal{S}_\gamma \leftarrow \emptyset$
for all $x \in D_{mtrain}$ **do**
  Computer the consensus of the pool $C(x, \mathcal{C})$
  if $C(x, \mathcal{P}) < h_c$ **then**
    Find k nearest neighbors as competence region $\mathcal{N}$ of x using meta training data.
    Compute the output profile $\bar{x}$ of x
    Find K_p most similar output files $\mathcal{O}$ of $\bar{x}$.
    for all $c_i \in C$ **do**
      f = Meta Feature Extraction $(\mathcal{O}, \mathcal{N}, c_i, x)$
      if c_i correctly classifiers x **then**
        $\alpha_{ij} = 1$, c_i is competent.
      else
        $\alpha_{ij} = 0$, c_i is not competent.
      end if
      $\mathcal{S}_\gamma = \mathcal{S}_\gamma \cup \{f\}$
    end for
  end if
end for
Train γ using the Multinomial naive Bayes.
Return the meta-classifier γ.

another is decision space-based competent region. Subsequently, the performance of each classifier is extracted as the meta-feature vectors, which are input into the meta-classifier. Finally, we estimate the competence of each base classifier; if the competence is large than threshold 0.5, the classifier is competent and is selected into ensemble.

- select the J most diverse classifiers from the above competent classifiers for each test sample. To evaluate diversity, we resort to the Double Fault (DF) [28]. The double fault diversity measure is a pairwise measure. Given classifiers c_i and c_j, the DF is obtained by can be calculated by the following equation [28].

$$DF_{ij} = \frac{N^{00}}{N^{11} + N^{01} + N^{10} + N^{00}} \tag{1}$$

where N^{ab} are listed in Table 2. The classifiers can be ranked according to its overall diversity; the predefined J classifiers are selected.

D. WEIGHTED INTEGRATION

To optimally fuse the prediction results, we decide to exploit the strengths of competence and dynamic weighting. Specifically, we use the majority vote scheme and resort to a hybrid combination approach. First, the competence base classifiers are selected to compose the ensemble. Next, the decision of each of these classifier is weighted by its level of competence. Thus, the decisions obtained by the classifiers with a high level of competence have a high influence in the final decision.

Algorithm 2 Ensemble Selection

Require: Dynamic selection dataset D_{sel}, test dataset D_{test}, Classifiers pool $\mathcal{C}$, the meta-classifier γ, the number of nearest neighbors k, the number of similar output files K_p, the scaling coefficient α, the number of diverse classifiers J.
Ensure: Classifier ensemble C_e for each test sample

for all $x_i \in D_{test}$ **do**
  ◇ Step1: select the competent classifiers
  $C_t = \emptyset$
  Find k nearest neighbors as competence region θ_i of x_i using D_{sel}
  Compute the output profile $\bar{x}_i$ of x_i
  Find K_p similar output files $\mathcal{O}$ of $x_{t,test}$ using D_{sel}
  for all $c_i \in C$ **do**
    f_t = Meta Feature Extraction $\left(\theta_t, \mathcal{O}, c_i, x_{t,\text{test}}\right)$
    input f_t to the meta classifier γ
    $\alpha_t = \gamma\,(v_t)$
    if $\alpha_t > 0.5$ **then**
      $C'_t = C'_t \cup \{c_i\}$
    end if
  end for
  ◇ Step 2: Select J most diverse classifiers
  $C''_t = \varnothing$
  for every $c_i \in C$ **do**
    for every $c_j \in C$ **do**
      if $i \neq j$ **then**
        Measure diversity between c_i and c_j : DF^t_{ij}
        Measure diversity of c_i : $DF^t_i = \sum_{j=1}^{M} DF^t_{ij}$
      end if
    end for
    if $\left|C'_t\right| > J$ **then**
      Rank the classifiers in C'_t based on DF^t_i
      The J diverse classifiers in C'_t are added into C_e
    else
      Rank the classifiers in C according to α_t
      The J most competent classifiers in C are added into C_e
    end if
  end for
end for
Return classifier ensemble C_e.

We compute a weighted majority vote by associating a weight w_j with classifier C_j:

$$\hat{y} = \arg\max_i \sum_{j=1}^{m} w_j \chi_A \left(C_j(\mathbf{x}) = i\right), \tag{2}$$

where w_j is the weight and equals to the value of competence. χ_A is the characteristic function $\left[C_j(\mathbf{x}) = i \in A\right]$, and A is the set of unique class labels.

TABLE 1. Descriptions of five distinct sets of meta-features. They are adapted from the META-DES [18].

Meta feature	Criterion	Paradigm	Number of features
f_{Hard}	Classification accuracy of the K-Nearest Neighbors	Classifier accuracy over a local region	K
f_{Prob}	Posterior probability of the K-Nearest Neighbors	Classifier consensus	K
$f_{Overall}$	Overall accuracy in the region of competence	Accuracy over a local region	1
f_{Conf}	Degree of confidence for the input sample	Classifier confidence	1
f_{OP}	Output profiles classification	Decision templates	K_p

[1] K is the size of the region of competence.
[2] K_p is the size of the output profiles set, containing the K_p most similar output profiles of the query sample.

TABLE 2. Pairwise contingency.

	c_j correct	c_j wrong
c_i correct	N^{11}	N^{10}
c_i wrong	N^{01}	N^{00}

[1] N^{11} represents the number of samples that are correctly classified by both classifiers c_i and c_j.

TABLE 3. Characteristics of 14 imbalanced data sets from the KEEL repository.

Name	Dimension	Samples	Imbalance rate
glass1	9	241	1.82
glass-0-1-2-3vs4-5-6	9	214	3.2
glass-0-1-6vs2	9	192	10.29
glass4	9	214	15.47
vehicle1	18	846	2.9
vehicle2	18	846	2.88
vehicle3	18	846	2.99
vowel0	13	988	9.98
new-thyroid2	5	215	5.14
yeast-1vs7	7	459	14.3
yeast-2vs8	8	482	23.1
yeast3	8	1484	8.1
yeast-2vs4	8	514	9.08
yeast-0-5-6-7-9vs4	8	528	9.35

TABLE 4. Statistical comparison of all classifiers using the Wilcoxon signed-rank test.

Comparison	Hypothesis	p-value
Our classifier vs. StackedClassifier	5%	0.000915
Our classifier vs. DES-KL	5%	0.000968
Our classifier vs. DES-Clustering	5%	0.000974
Our classifier vs. DES-KNN	5%	0.000981
Our classifier vs. DES-MI	5%	0.000979
Our classifier vs. META-DES	5%	0.000980

IV. EXPERIMENTS

To validate the effectiveness of our proposed new DES method, we conducted experiments on 14 two-class imbalanced data sets from the Knowledge Extraction based on Evolutionary Learning (KEEL) repository [34]. The characteristics of these data sets are concluded in Table 3. To verify its superiority, we compared it with five typical DES techniques, including DES-KL [35], DES-Clustering [36], DES-KNN [36], DES-MI [16], and META-DES [18]. Among them, DES-MI is a DES method for imbalance classification. Both DES-Clustering and DES-KNN use accuracy and diversity as selection criteria. META-DES uses different criteria regarding the behavior of a base classifier. For comprehensive comparison, we also compared it with a stacking classifier, termed StackedClassifier [37].

TABLE 5. Time cost of proposed META-DES-Diversity on 14 imbalanced data sets from the KEEL repository.

Dataset	Time cost (in seconds)
glass1	185
glass-0-1-2-3vs4-5-6	176
glass-0-1-6vs2	170
glass4	180
vehicle1	353
vehicle2	362
vehicle3	395
vowel0	413
new-thyroid2	192
yeast-1vs7	364
yeast-2vs8	415
yeast3	408
yeast-2vs4	416
yeast-0-5-6-7-9vs4	423

A. EXPERIMENTAL SETUP

For fair comparison, each compared method is the same experimental setup as that in their original paper. Initially, the Bagging uses the decision tree as the base classifier. The number of base classifier is set to 100. The same candidate classifier pool is used for all classifiers. As for the metric of KNN is the minkowski and the number of neighbors is set to 7. The number of output profiles used to estimate the competence of the base classifiers is set to 5. The percentage of input data to fit the dynamic selection data set is set at 50%. The number of base classifiers selected with accuracy and diversity is set to 50% and 30% of pool of base classifiers, respectively; In META-DES, the instance selection threshold is set to 50%. All compared classifiers are implemented in Python under framework of the library of dynamic classifier system [38]. A 10-fold cross validation procedure is conducted to evaluate classification performance.

B. RESULTS AND DISCUSSION

Imbalance classification aims to improve prediction of minority samples while maintaining performance of the majority. The receiver operating characteristics (ROC) [39] curve is a

TABLE 6. Classification results in terms of AUC on 14 two-class imbalanced data sets from the KEEL repository.

Dataset	StackedClassifier	DES-KL	DES-Clustering	DES-KNN	DES-MI	META-DES	META-DES-Diversity
glass1	0.8317	0.8069	0.8260	0.8545	0.8785	0.8524	**0.8928**
glass-0-1-2-3vs4-5-6	0.9835	0.9858	0.9864	0.9849	0.9920	0.9845	**0.9968**
glass-0-1-6vs2	0.8546	0.8705	0.8245	0.8365	0.8768	0.8607	**0.9814**
glass4	0.9488	0.9585	0.9528	0.9774	0.9725	0.9665	**0.9808**
vehicle1	0.8270	0.8276	0.8115	0.8506	0.8552	0.8408	**0.8655**
vehicle2	0.9808	0.9905	0.9886	0.9804	0.9855	0.9960	**0.9995**
vehicle3	0.7846	0.7868	0.7586	0.8008	0.8198	0.8045	**0.8364**
vowel0	0.9490	0.9564	0.9485	0.9665	0.9778	0.9656	**0.9997**
new-thyroid2	0.9548	0.9665	0.9648	0.9702	0.9785	0.9778	**0.9991**
yeast-1vs7	0.8168	0.8106	0.8327	0.8249	0.8486	0.8247	**0.8825**
yeast-2vs8	0.9042	0.8716	0.8758	0.9360	0.9448	0.9125	**0.9850**
yeast3	0.9256	0.9074	0.8750	0.9370	0.9545	0.9406	**0.9770**
yeast-2vs4	0.9505	0.9148	0.9560	0.9564	0.9780	0.9745	**0.9955**
yeast-0-5-6-7-9vs4	0.5486	0.5546	0.5000	0.6045	0.8246	0.8007	**0.8840**

widely used criterion of imbalance classification. ROC provides visualization of the trade-off between the false positive rate and the true positive rate. In this paper, area under the ROC curve (AUC) was utilized to measure the performance of all classifiers [40]. The results in terms of AUC are provided in Table 6. The higher AUC values means the better classifier.

We observed that our proposed classifier META-DES-Diversity perform better than that of other classifiers on all data sets. That verifies its effectiveness on two-class imbalance classification. Note that our method shows big advantage on the data set yeast-2vs8 which has the highest imbalance rate. This is because our classifier combines strengths of data sampling for data balancing of minority class, the meta learning framework for competent classifier evaluation, the characteristic of diversity, and competence-driven decision fusion.

From Table 4, we observed that all hypotheses are rejected at 5% (95% confidence). From these results, we conclude that the proposed META-DES-Diversity is efficient in dealing with two-class imbalance classification.

V. CHINA CORPORATE BOND DEFAULT PREDICTION

In this section we applied our DES classifier on China corporate bond default prediction. This task aims to predict weather one bond is default or not. Technically, the problem of default prediction can be formulated as a two-class imbalance classification problem [41].

A. DATA ACQUISITION

The China corporate bond data set is collected from the Wind Economic Database.[3] This economic database provides the Chinese financial market data and information to analysts, fund managers and traders, with a full coverage of equities, bonds, funds, indexes, warrants, commodity futures, foreign exchanges, and economy. This bond data set contains 90749 records issued between 2014 and 2021, with 1061 default bonds. This data set contains two types of features: statistic and finance. Specifically, the number of statistic and finance features is 16 and 31, respectively. The imbalance ratio of the default prediction data set is 21.2. Our China corporate bond data set is public available.[4] Feature description is provided in Table 7.

B. PREDICTION RESULTS

To verify the superiority of the proposed method, we make comparisons with five related methods, including static classifiers and dynamic classifiers. Among them, one typical static ensemble method XGBoost [42], two dynamic classifier selection methods (OLA [43] and MCB [44]), and two dynamic ensemble selection methods (DES-MI [16] and META-DES [18]). Decision tree is used as the base classifier for all methods.

The confusion matrix in Table 8 is used to express the prediction results. Based on this confusion matrix,

[3] https://www.wind.com.cn/en/edb.html

[4] https://github.com/williamyan24/ChinaBondDefaultPrediction

TABLE 7. Features of China corporate bond data set.

Index	Name	Category	Value
1	Inventory turnover	finance	numeric
2	Liability with interest	finance	numeric
3	Quarterly earnings per share (year-over-year growth rate)	finance	numeric
4	Fixed asset turnover	finance	numeric
5	Money funds vs short-term debt	finance	numeric
6	Trade financial assets	finance	numeric
7	Return on equity ROE average	finance	numeric
8	Total profit (year-on-year growth rate)	finance	numeric
9	Interest expense	finance	numeric
10	Interest cost	finance	numeric
11	Retained earnings	finance	numeric
12	Current ratio	finance	numeric
13	Total current assets	finance	numeric
14	Current asset turnover	finance	numeric
15	Quick ratio	finance	numeric
16	Total owner's equity	finance	numeric
17	Undistributed profits (general business)	finance	numeric
18	Cash-to-debt ratio	finance	numeric
19	Cash flow interest coverage ratio	finance	numeric
20	Sales profit margin (1 Year)	finance	numeric
21	Cash from the sale of goods to provide labor income	finance	numeric
22	Earned interest multiple	finance	numeric
23	Bill receivable	finance	numeric
24	Accounts receivable turnover	finance	numeric
25	Year-on-year growth rate of operating income	finance	numeric
26	Working capital to total assets	finance	numeric
27	Assets and liabilities	finance	numeric
28	Total assets	finance	numeric
29	Return on total assets ROA	finance	numeric
30	Total asset turnover	finance	numeric
31	Total current liabilities	finance	numeric
32	Entity rating at issue	statistics	numeric
33	Debt rating at issue	statistics	numeric
34	Most recent entity rating	statistics	numeric
35	Latest debt ratings	statistics	numeric
36	Penalties in one year	statistics	numeric
37	Latest debt rating adjustments	statistics	numeric
38	Debt rating adjustment within one year	statistics	numeric
39	Number of downgrades of related debt ratings within one year	statistics	numeric
40	Number of related debt rating upgrades within one year	statistics	numeric
41	Number of times the bond has been deferred in a year	statistics	numeric
42	Number of times the issuer has been deferred rating within one year	statistics	numeric
43	Entity rating at issue 2	statistics	numeric
44	Issuers rating score at Issue	statistics	numeric
45	Debt rating score at Issue	statistics	numeric
46	Latest issuer rating scores	statistics	numeric
47	Latest debt rating scores	statistics	numeric

many measures can be constructed and used to evaluate the performance of a classifier. When it comes to bond default prediction, the precision is not used since the bond data set is imbalanced and the defaults belong to the minority class. In evaluation, the recall [45], F1-score and the aforementioned AUC are used as the evaluation criteria. The higher the value, the better the classification on default prediction. These measures are defined as follows.

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{3}$$

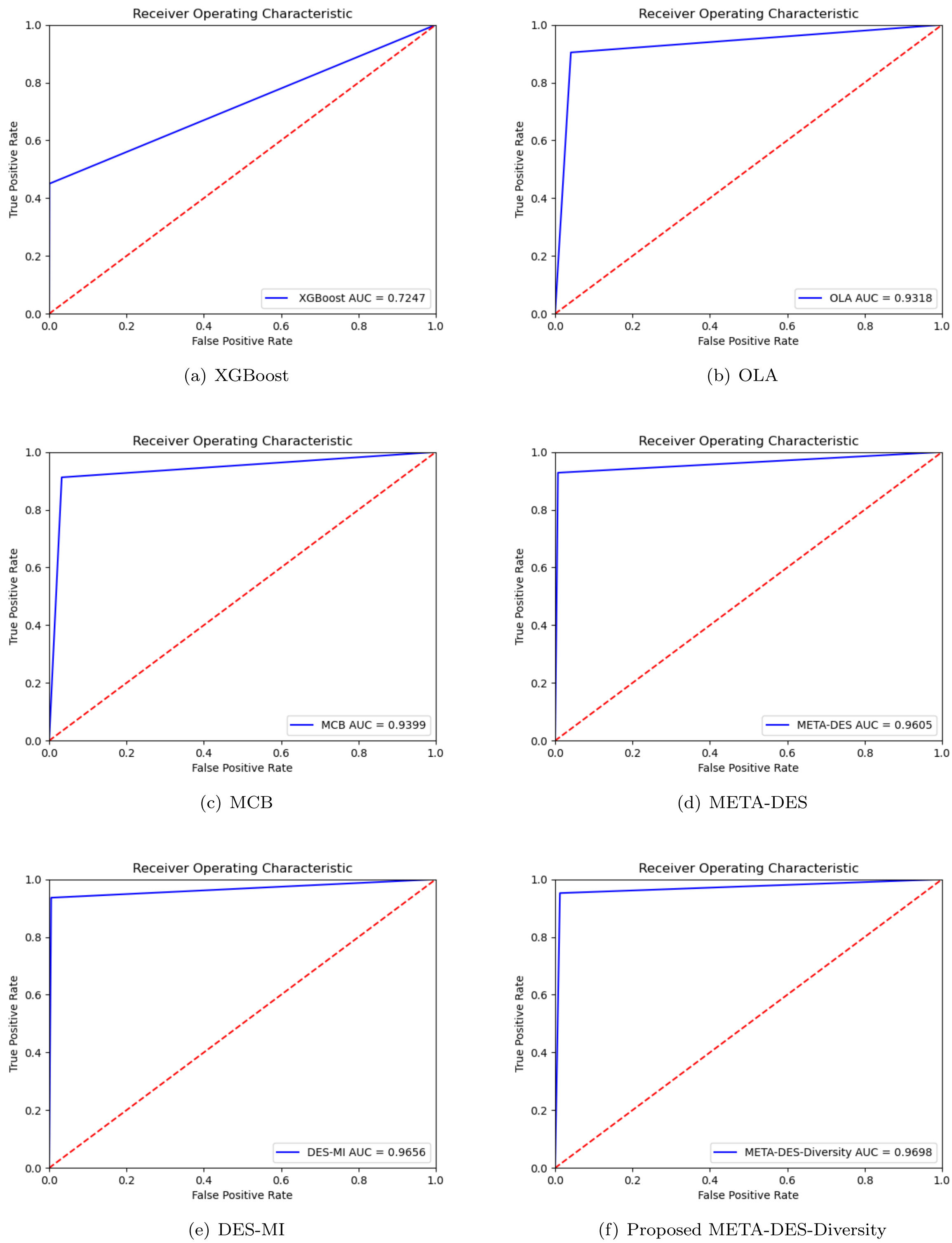
$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{4}$$

$$\text{F1}-\text{score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{5}$$

TABLE 8. Confusion matrix.

	Prediction	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

These prediction results are recorded in Table 9, and the AUC results are provided in Figure 5. One can observe that the dynamic selection classifiers perform better than that of static ensemble classifier. That verify that effectiveness of the dynamic selection strategy on the China corporate bond data



(a) XGBoost (b) OLA

(c) MCB (d) META-DES

(e) DES-MI (f) Proposed META-DES-Diversity

FIGURE 5. **The classification results of all classifiers on the China corporate bond data set in terms of AUC: (a) XGBoost; (b) OLA; (c) MCB; (d) META-DES; (e) DES-MI; (f) Proposed META-DES-Diversity.**

set. In addition, our proposed method obtain better results than that of DES-MI and META-DES, which demonstrates that our method can well handle imbalance classification. This advantage owns to we resort to data sampling method and successfully select classifiers with various criteria and diversity.

TABLE 9. Classification results in terms of Recall, F1-score and AUC on the China corporate bond data set.

	Recall	F1-score	AUC
XGBoost	0.4501	0.5809	0.7247
OLA	0.9123	0.2105	0.9318
MCB	0.9163	0.2904	0.9399
DES-MI	0.9282	0.6862	0.9656
META-DES	0.9362	0.7099	0.9605
Proposed META-DES-Diversity	**0.9403**	**0.7292**	**0.9698**

VI. CONCLUSION

In this paper, we have proposed an enhanced DES classifier based on the framework of META-DES, named META-DES-Diversity, with application to the China corporate bond default prediction. META-DES-Diversity improves META-DES through incorporating it with data sampling, the technique of diversity and hybrid dynamic weighting. Compared with existing DES classifiers, our META-DES-Diversity can directly address imbalanced data and select competent classifiers with various criterion, leading to optimal classification performance. Note that we effectively fuse predictions from ensemble classifiers with a weighted Majority Vote scheme in which the competence of classifiers is used as weights for classifiers. Experiments on 14 two-class imbalanced data sets from KEEL repository demonstrate the superiority of the META-DES-Diversity. In addition, it has been successfully applied on the China corporate bond default prediction with considerable results in terms of three measurements (e.g., Recall, F1-score and AUC), showing that the META-DES-Diversity can address default prediction better than other five typical classifiers.

In the future, we would like to continue our research on the following two directions. First, we will propose another version based on formal definition of the Oracle, which is an abstract method that represents an ideal classifier selection scheme. Second, we will consider the case of classification with noisy label.

ACKNOWLEDGMENT

The authors would like to thank Prof. Rafael M. O. Cruz for sharing the codes of DESLib. They would also like to thank Yiye Liu and Liqun Zhang for their suggestions on China bond data set collection.

REFERENCES

[1] M. Chi, S. Hongyan, W. Shaofan, L. Shengliang, and L. Jingyan, "Bond default prediction based on deep learning and knowledge graph technology," *IEEE Access*, vol. 9, pp. 12750–12761, 2021.

[2] Y.-X. Wu, Z.-N. Hu, Y.-Y. Wang, and F. Min, "Rare potential poor household identification with a focus embedded logistic regression," *IEEE Access*, vol. 10, pp. 32954–32972, 2022.

[3] R. Wang, N. Xiu, and C. Zhang, "Greedy projected gradient-Newton method for sparse logistic regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 2, pp. 527–538, Feb. 2020.

[4] C.-Y. Guo and Y.-J. Lin, "Random interaction forest (RIF)—A novel machine learning strategy accounting for feature interaction," *IEEE Access*, vol. 11, pp. 1806–1813, 2023.

[5] D. Song, Y. Tang, B. Wang, J. Zhang, and C. Yang, "Two-branch generative adversarial network with multiscale connections for hyperspectral image classification," *IEEE Access*, vol. 11, pp. 7336–7347, 2023.

[6] L. Jia, Z. Wang, S. Lv, and Z. Xu, "PE_DIM: An efficient probabilistic ensemble classification algorithm for diabetes handling class imbalance missing values," *IEEE Access*, vol. 10, pp. 107459–107476, 2022.

[7] R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, "Dynamic classifier selection: Recent advances and perspectives," *Inf. Fusion* vol. 41, pp. 195–216, May 2018.

[8] C. Zhao, D. Wu, J. Huang, Y. Yuan, H.-T. Zhang, R. Peng, and Z. Shi, "BoostTree and BoostForest for ensemble learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Dec. 7, 2022, doi: 10.1109/TPAMI.2022.3227370.

[9] N. Zhang, Y. Xu, Q.-X. Zhu, and Y.-L. He, "Improved locality preserving projections based on heat-kernel and cosine weights for fault classification in complex industrial processes," *IEEE Trans. Rel.*, vol. 72, no. 1, pp. 204–213, Mar. 2023.

[10] K. Guo, C. Shen, X. Zhou, S. Ren, M. Hu, M. Shen, X. Chen, and H. Guo, "Traffic data-empowered XGBoost-LSTM framework for infectious disease prediction," *IEEE Trans. Intell. Transp. Syst.*, early access, May 13, 2022, doi: 10.1109/TITS.2022.3172206.

[11] R. M. O. Cruz, D. V. R. Oliveira, G. D. C. Cavalcanti, and R. Sabourin, "FIRE-DES++: Enhanced online pruning of base classifiers for dynamic ensemble selection," *Pattern Recognit.*, vol. 85, pp. 149–160, Jan. 2019.

[12] B. Jiao, Y. Guo, D. Gong, and Q. Chen, "Dynamic ensemble selection for imbalanced data streams with concept drift," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 22, 2022, doi: 10.1109/TNNLS.2022.3183120.

[13] Y. Feng, Y. Yin, D. Wang, and L. Dhamotharan, "A dynamic ensemble selection method for bank telemarketing sales prediction," *J. Bus. Res.*, vol. 139, pp. 368–382, Feb. 2022.

[14] Y.-R. Choi and D.-J. Lim, "DDES: A distribution-based dynamic ensemble selection framework," *IEEE Access*, vol. 9, pp. 40743–40754, 2021.

[15] W.-H. Hou, X.-K. Wang, H.-Y. Zhang, J.-Q. Wang, and L. Li, "A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment," *Knowl.-Based Syst.*, vol. 208, Nov. 2020, Art. no. 106462.

[16] S. García, Z.-L. Zhang, A. Altalhi, S. Alshomrani, and F. Herrera, "Dynamic ensemble selection for multi-class imbalanced datasets," *Inf. Sci.*, vols. 445–446, pp. 22–37, Jun. 2018.

[17] R. G. F. Soares, A. Santana, A. M. P. Canuto, and M. C. P. de Souto, "Using accuracy and diversity to select classifiers to build ensembles," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2006, pp. 1310–1316.

[18] R. M. O. Cruz, R. Sabourin, G. D. C. Cavalcanti, and T. Ing Ren, "META-DES: A dynamic ensemble selection framework using meta-learning," *Pattern Recognit.*, vol. 48, no. 5, pp. 1925–1935, May 2015.

[19] A. Roy, R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, "A study on combining dynamic selection and data preprocessing for imbalance learning," *Neurocomputing*, vol. 286, pp. 179–192, Apr. 2018.

[20] X. Chen, L. Gupta, and S. Tragoudas, "Improving the forecasting and classification of extreme events in imbalanced time series through block resampling in the joint predictor-forecast space," *IEEE Access*, vol. 10, pp. 121048–121079, 2022.

[21] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, and H. M. Gheni, "Improving the heart disease detection and patients' survival using supervised infinite feature selection and improved weighted random forest," *IEEE Access*, vol. 10, pp. 67363–67372, 2022.

[22] Y.-R. Chen, J.-S. Leu, S.-A. Huang, J.-T. Wang, and J.-I. Takada, "Predicting default risk on peer-to-peer lending imbalanced datasets," *IEEE Access*, vol. 9, pp. 73103–73109, 2021.

[23] C. Liu, S. Jin, D. Wang, Z. Luo, J. Yu, B. Zhou, and C. Yang, "Constrained oversampling: An oversampling approach to reduce noise generation in imbalanced datasets with class overlapping," *IEEE Access*, vol. 10, pp. 91452–91465, 2022.

[24] A. Rosales-Perez, S. Garcia, and F. Herrera, "Handling imbalanced classification problems with support vector machines via evolutionary bilevel optimization," *IEEE Trans. Cybern.*, early access, Apr. 13, 2022, doi: 10.1109/TCYB.2022.3163974.

[25] R. M. O. Cruz, M. A. Souza, R. Sabourin, and G. D. C. Cavalcanti, "Dynamic ensemble selection and data preprocessing for multi-class imbalance learning," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 11, Oct. 2019, Art. no. 1940009.

[26] I. Khan, X. Zhang, M. Rehman, and R. Ali, "A literature survey and empirical study of meta-learning for classifier selection," *IEEE Access*, vol. 8, pp. 10262–10281, 2020.

[27] R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, "META-Des.Oracle: Meta-learning and feature selection for dynamic ensemble selection," *Inf. Fusion*, vol. 38, pp. 84–103, Nov. 2017.

[28] J. A. S. Lustosa Filho, A. M. P. Canuto, and R. H. N. Santiago, "Investigating the impact of selection criteria in dynamic ensemble selection methods," *Expert Syst. Appl.*, vol. 106, pp. 141–153, Sep. 2018.

[29] R. M. O. Cruz, R. Sabourin, and G. D. C. Cavalcanti, "META-DES.H: A dynamic ensemble selection technique using meta-learning and a dynamic weighting approach," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–8.

[30] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[31] S. Masoudnia and R. Ebrahimpour, "Mixture of experts: A literature survey," *Artif. Intell. Rev.*, vol. 42, no. 2, pp. 275–293, 2014.

[32] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, Apr. 2018.

[33] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[34] J. Derrac, S. Garcia, L. Sanchez, and F. Herrera, "Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *J. Mult. Valued Log. Soft Comput.*, vol. 17, pp. 1–36, Jun. 2015.

[35] T. Woloszynski, M. Kurzynski, P. Podsiadlo, and G. W. Stachowiak, "A measure of competence based on random classification for dynamic ensemble selection," *Inf. Fusion*, vol. 13, no. 3, pp. 207–213, Jul. 2012.

[36] M. C. P. de Souto, R. G. F. Soares, A. Santana, and A. M. P. Canuto, "Empirical comparison of dynamic classifier selection methods based on diversity and accuracy for building ensembles," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1480–1487.

[37] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.

[38] R. M. O. Cruz, L. G. Hafemann, R. Sabourin, and G. D. C. Cavalcanti, "DESlib: A dynamic ensemble selection library in Python," *J. OF Mach. Learn. Res.*, vol. 21, pp. 1–5, Jan. 2020.

[39] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.

[40] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Inf. Sci.*, vol. 250, pp. 113–141, Nov. 2013.

[41] Z. Lu and Z. Zhuo, "Modelling of Chinese corporate bond default—A machine learning approach," *Accounting Finance*, vol. 61, no. 5, pp. 6147–6191, Dec. 2021.

[42] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho, "Xgboost: Extreme gradient boosting," *R Package Version* vol. 1, no. 4, pp. 1–4, Aug. 2015.

[43] K. Woods, W. P. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 405–410, Apr. 1997.

[44] G. Giacinto and F. Roli, "Dynamic classifier selection based on multiple classifier behaviour," *Pattern Recognit.*, vol. 34, no. 9, pp. 1879–1882, 2001.

[45] M. Buckland and F. Gey, "The relationship between recall and precision," *J. Amer. Soc. Inf. Sci.*, vol. 45, no. 1, pp. 12–19 2014.

YU WANG was born in Hubei, China. He received the master's degree from the City University of Macau, where he is currently pursuing the Ph.D. degree with the Faculty of Finance. His research interests include machine learning, block-chain finance, applied economics, bond default prediction, and credit risk assessment.

JUNBIN ZHANG received the master's degree from the Macau University of Science and Technology. He is currently pursuing the Ph.D. degree with the Faculty of Finance, City University of Macau. His research interests include machine learning techniques and their applications on finance like digital inclusive finance, block-chain finance, and household finance.

WEI YAN received the Ph.D. degree from the Faculty of Science and Technology, University of Macau. He is currently a Postdoctoral Fellow with the Guangzhou Institute of International Finance, Guangzhou University, Guangzhou, China. He has published more than eight articles, including top journals, such as IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (TSMC), *Knowledge-Based Systems* (KBS), and IEEE TRANSACTIONS ON CYBERNETICS (TCYB). His research interests include machine learning, pattern recognition, and signal processing.