IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 22, 2019, accepted June 19, 2019, date of publication June 26, 2019, date of current version July 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924923

A MCDM-Based Evaluation Approach for Imbalanced Classification Methods in Financial Risk Prediction

YONGMING SONG[1] **AND YI PENG**[2]
[1]School of Business Administration, Shandong Technology and Business University, Yantai 264005, China
[2]School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Yi Peng (pengyi@uestc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1811462 and Grant 71771037.

ABSTRACT Various classifiers have been proposed for financial risk prediction. The traditional practice of using a singular performance metric for classifier evaluation is not sufficient for imbalanced classification. This paper proposes a multi-criteria decision making (MCDM)-based approach to evaluate imbalanced classifiers in credit and bankruptcy risk prediction by considering multiple performance metrics simultaneously. An experimental study is designed to provide a comprehensive evaluation of imbalanced classifiers using the proposed evaluation approach over seven financial imbalanced data sets from the UCI Machine Learning Repository. The TOPSIS, a well-known MCDM method, was applied to rank three categories of imbalanced classifiers using six popular evaluation criteria. The rankings results indicate that: 1) the rankings generated by the TOPSIS, which combine the results of six evaluation criteria, provide a more reasonable evaluation of imbalanced classifiers over any single performance criterion; and 2) Synthetic Minority Oversampling Technique (SMOTE)-based ensemble techniques outperform other groups of imbalanced learning approaches. Specifically, SMOTEBoost-C4.5, SMOTE-C4.5, and SMOTE-MLP were ranked as the top three classifiers based on their performances on the six criteria.

INDEX TERMS Financial risk prediction, imbalanced classification, multiple criteria decision making (MCDM), algorithm evaluation.

I. INTRODUCTION

Financial risk prediction has been a hot topic for years due to its great importance [1]–[4]. Bankruptcy or default prediction is one of the most important tasks in financial risk management. Since the number of default or bankruptcy is significantly outnumbered by non-default or non-bankruptcy [5]–[7], bankruptcy classification is a typical imbalanced classification problem.

Many methods have been developed to learn from imbalanced data sets over the decades. They can be categorized into three major groups: resampling, cost-sensitive learning, and ensemble techniques. Previous researches have proved that class imbalance is likely to result in a degradation for the final prediction [8]–[10]. The class imbalance problem has always been regarded as a challenging task in a broad scope of financial problems. In last years, some works have studied the performance of imbalanced models on financial risk prediction. He et al. [11] introduced a model based on resampling the credit scoring data sets according to their imbalance ratio and a threshold. Sun et al. [12] proposed an ensemble for imbalanced credit evaluation based on the SMOTE algorithm and the BAGGING technique with different sampling rates. Veganzones and Séverina [13] investigated the performance of bankruptcy prediction models in imbalanced datasets by analyzing three key notions: degree of imbalance, loss of performance, and sampling techniques. García et al. [14] investigated whether or not there exists any potential difference in their performance due to the distribution of sample types in a database. As can be seen from the above analysis, few studies comprehensively investigate the performance of various types of financial risk prediction models in imbalanced data sets. Thus, it is interesting to investigate the effects of imbalanced classification techniques on financial risk classification and compare their performances. The objective of this paper is to propose a multi-criteria decision making (MCDM) based

The associate editor coordinating the review of this manuscript and approving it for publication was Fatih Emre Boran.

approach for a comprehensive assessment of imbalanced classifiers in credit and bankruptcy risk prediction. The basic idea of the proposed approach is to rank imbalanced classifiers in credit and bankruptcy risk prediction according to their performances on a selection of metrics, rather than singular metric, using MCDM methods [15], [16]. Although there have been some studies evaluating the performance of imbalanced classification methods, few, if any, have analyzed this problem using a combination of multiple criteria.

An experiment is designed to assess four base classifiers (SVM, MLP, LR and C4.5) and their combinations with resampling, cost-sensitive learning, and ensemble techniques using six evaluation metrics (i.e., G-mean, F-measure, AUC, FP rate, FN rate, and time) over seven public imbalanced credit and bankruptcy risk data sets. The results show that the SMOTE-based ensemble techniques outperform other group of techniques.

The contributions of the proposed MCDM-based evaluation approach for imbalanced classification methods in financial risk prediction with respect to previous studies are summarized as follows.

- A MCDM approach based on six key criteria (G-mean, F-measure, AUC, FP rate, FN rate, and time) is proposed to evaluate imbalanced financial risk classification methods, integrated using TOPSIS method.
- This article makes a systematic analysis about resampling, cost-sensitive learning, and ensemble techniques in financial risk prediction.
- An objective determining weights of assessment criteria based on Entropy method is put forward.
- Some instructive results are obtained for imbalanced classification methods in financial risk prediction.

The rest of this paper is organized as follows. Section 2 reviews the background and related works including existing algorithms in financial risk classification, imbalanced learning techniques, and performance metrics for imbalanced classification. Section 3 describes TOPSIS, the MCDM method used in this study. Section 4 presents the experiment design and results. Section 5 concludes the paper.

II. BACKGROUND AND RELATED WORKS

A. FINANCIAL RISK PREDICTION MODELS

Numerous classification algorithms have been proposed for financial risk prediction, such as logistic regression (LR), neural networks (NN), support vector machines (SVM), decision trees (DT), and partial least squares [17], [18]. Ensemble learning techniques, which have demonstrated notable improvement over a single classification algorithm, have been applied to financial risk classification. Ravikumar and Ravi [19] presented ensemble classifiers by simple majority voting scheme based on seven algorithms. Sun and Li [20] investigated weighted majority voting combination of multiple diversified classifiers and obtained higher average accuracy than any base classifier. Furthermore, ensembles of classifiers [21] attempt to increase the accuracy of individual classifiers by their combination. Ensemble learning refers to the combination of several classifiers to produce a strong classifier. The key to the integrated algorithm lies in the diversity of the base classifiers. One of the most common approaches to construct ensembles by data variation are Boosting [22] and Bagging [23]. A strong classifier is obtained from multiple classifiers by resampling, which is the basic principle for Bagging. Boosting integrations base classifiers based on the weights. Bagging and Boosting-based ensemble methods have been received increasing attention [24]–[27]. Bagging and Boosting ensembles based on NN were applied [24], [25]. Kim and Upneja [26] compared the predictive and discriminatory performances of AdaBoosted DT models with single DT models and AdaBoosted DT model based on C4.5 demonstrated the best prediction performance. Sun et al. [27] established AdaBoost ensemble respectively with single attribute test (SAT) and DT and found that AdaBoost-SAT outperformed AdaBoost-DT.

B. IMBALANCED LEARNING TECHNIQUES

The class imbalance problem refers to a situation in which the class distribution is highly skewed. Many techniques have been developed to address the class imbalance problem over the years. López et al. [28] categorize them into three major groups: resampling, cost-sensitive learning, and ensemble techniques.

As the number of imbalanced learning approaches increases, how to select an effective one for a given task becomes an important yet difficult issue. The traditional practice of choosing a single measure to evaluate imbalanced classification algorithms is not sufficient and several studies [29], [30] have proved that the choice of evaluation measures can have a substantial effect on the conclusions. For instance, the experiments conducted by Raeder et al. [30] showed that Naive Bayes was ranked as the best classifier by the area under the ROC curve (AUC) and the worst classifier by Brier score on the same data sets.

This section introduces the three groups of imbalanced classification techniques and the major evaluation measures that have been used in imbalanced classification. For comprehensive and up-to-date reviews of classification approaches for imbalanced data, please refer to [28], [29].

Most existing classification approaches for the imbalance problem can be categorized into three groups: preprocessing, ensemble, and cost-sensitive learning [28]. The following subsections provide brief descriptions of each group.

1) PREPROCESSING IMBALANCED DATASETS: RESAMPLING TECHNIQUES

Resampling approaches target the imbalanced classification problem by reducing skewed distributions of imbalanced data sets using preprocessing techniques. According to the underlying principles, resampling approaches can be classified as undersampling, oversampling and hybrid methods. While undersampling changes class distribution by removing data records from the majority class, oversampling creates new minority data records by replicating or utilizing sophisticated techniques. Hybrid methods combine both undersampling and oversampling techniques to handle the class imbalance. Because resampling concerns only about preprocessing imbalanced data, it can be used with any standard classifier or specially designed imbalance learning algorithms.

This paper chooses random undersampling (RUS) and Synthetic Minority Oversampling Technique (SMOTE) to represent resampling techniques in the experiment. RUS randomly removes majority examples from the original data to reduce the imbalance [31]. Despite its simplicity, RUS performed better than some more sophisticated techniques [32]. Synthetic Minority Oversampling Technique (SMOTE) [33] is one of the most well-known approaches in the area of preprocessing imbalanced data. It creates synthetic minority class data by generating neighbors from real minority examples [33]. SMOTE improves the classification performance for a minority class because it creates a larger and more general decision region [33].

2) ENSEMBLE METHODS

Ensemble methods have also been combined with preprocessing algorithms [34]–[36] to address imbalanced classification problem. This study chooses UnderBagging and SMOTEBoost to represent this category of techniques. UnderBagging [34] randomly undersamples the majority data in each Bagging iteration and keep all minority class instances in every iteration. SMOTEBoost [35] introduces synthetic minority class instances using SMOTE algorithm. Since new instances are created, new weights must be assigned, which are proportional to the total number of instances in the new dataset. The weights of the instances from the original data-set are normalized to form a distribution with the new instances.

3) COST-SENSITIVE LEARNING

In real-life imbalanced classification problems, misclassifying data instances from different classes have different costs. Most likely, misclassification cost of the minority class is higher than the majority class. For example, in medical diagnosis, the cost of having a disease undetected is much higher than the cost of having a false alarm. Based on this observation, cost-sensitive methods deal with the class imbalance problem by assigning different costs to different types of misclassifications [37], [38].

4) SUMMARY AND COMMENTS

Many studies have been conducted to compare imbalanced learning techniques. VanHulse et al. [32] introduced a comprehensive experiment with eight sampling methods. It showed that random sampling approach performs better than intelligent sampling approach like SMOTE. García et al. [39] surveyed the influence of imbalance ratio for classifier results on several resampling methods. Experiments showed that oversampling consistently outperforms undersampling when data sets are strongly imbalanced. Khoshgoftaar et al. [40] compared bagging with boosting based on imbalanced data and noisy. The experiments showed that bagging generally outperform boosting in noisy data environments. Galar et al. [41] established an empirical comparison with a wide range of ensembles. Their main conclusion is that SMOTEbagging, RUSBoost, and UnderBagging have the best AUC results. López et al. [28] carried out an experimental analysis to contrast sampling, cost-sensitive learning and ensemble techniques. The results show the dominance of ensemble approaches UnderBagging and SMOTEBagging as weak classifiers are C4.5 and K-NN while the best results are acquired by SMOTE and cost-sensitive learning when SVM is used.

In financial risk prediction, some studies have considered the effect of the imbalanced data on classification results [5], [6], [42]–[45]. Li and Sun [6] used an oversampling method to balance the training dataset, and showed that the constructed model based on the corrected balanced training data set significantly outperformed the model trained on the original imbalanced data set. Crone and Finlay [42] applied both over-sampling and under-sampling methods to balance the original imbalanced credit datasets. It showed that over-sampling significantly increases the accuracy relative to under-sampling across all algorithms. Besides, Brown and Mues [5] implemented experimental comparisons with several techniques based on imbalanced credit scoring data sets. The results have shown that random forest and gradient boosting classifiers have good performance in a credit scoring context with noticeable class imbalances.

C. EVALUATION MEASURES

The evaluation criteria is a key factor in assessing a classifiers' performance. The performance of a binary classification algorithm can be evaluated using the information provided by a confusion matrix shown in *Table 1*, which summarizes correctly and incorrectly recognized examples of each class.

Traditionally, frequently used performance metrics in evaluating classifiers are accuracy, recall, F-measure, G-mean, and Area under the ROC Curve (AUC). The following paragraphs describe these performance metrics and their components.

(1)*Overall accuracy (ACC):* Accuracy is the percentage of correctly classified instances. $ACC = \frac{TP+TN}{TP+FN+FP+TN}$.

ACC is not effective in evaluating imbalanced classifiers because it is sensitive to data distributions [46].

(2) *True positive rate (recall):* $TP_{rate} = \frac{TP}{TP+FN}$ is the percentage of positive instances correctly classified.

(3) *True negative rate:*$TN_{rate} = \frac{TN}{FP+TN}$ is the percentage of negative instances correctly classified.

(4) *False positive rate:* $FP_{rate} = \frac{FP}{FP+TN}$ is the percentage of negative instances misclassified.

(5) *False negative rate:* $FN_{rate} = \frac{FN}{TP+FN}$ is the percentage of positive instances misclassified.

(6) *F-measure:* It is the harmonic mean of precision and recall, $F - measure = \frac{2precision \times recall}{precision + recall}$.

(7) *G-mean:* is the geometric mean of the true rates, which can be defined as: $G - mean = \sqrt{\frac{TP}{TP+FN} \times \frac{TN}{TN+FP}}$.

(8) *AUC:* The Area under the ROC (Receiver Operating Characteristic) Curve (AUC) shows the tradeoff between TP rate and FP rate and measures the ability of a classifier to correctly predict positive instances [46]. Although AUC is a powerful metric and provides an informative evaluation, it is overly optimistic when the data is highly imbalanced [47]. The reason is that even a large change of the number of false positives in highly skewed data sets will not greatly affect the FP rate used in AUC.

As the number of imbalanced learning approaches available increases, how to select an effective one for a given task becomes an important yet difficult issue. The traditional practice of choosing a single measure to evaluate imbalanced classification algorithms is not sufficient and several studies [29], [30] have proved that the choice of evaluation measures can have a substantial effect on the conclusions. For instance, the experiments conducted by Raeder et al. [30] showed that Naive Bayes was ranked as the best classifier by the area under the ROC curve (AUC) and the worst classifier by Brier score on the same data sets.

III. MCDM METHOD

Multiple criteria decision making (MCDM), which evaluates alternatives by considering two or more criteria, has made remarkable progress during the past 40 years and many approaches have been developed to solve MCDM problems, such as goal programming [48], AHP [49], TOPSIS [50], VIKOR [51], DEA [52], PROMETHEE [53] and ELECTRE-TRI [54]. Since MCDM is used to rank discrete alternative problems in this study, any approach developed for multiple criteria discrete alternative problems can be used. We choose Technique for order preference by similarity to ideal solution (TOPSIS), which is a simple and widely used multiple criteria decision method, for the experimental study.

A. TECHNIQUE FOR ORDER PREFERENCE BY SIMILARITY TO IDEAL SOLUTION (TOPSIS)

TOPSIS finds the best alternatives by minimizing the distance to the idea solution and maximizing the distance to the negative-ideal solution [55]. The TOPSIS procedure used in this paper is summarized as follows [56]:

Step 1: Calculate the normalized decision matrix. The normalized value r_{ij} is calculated as:

$$r_{ij} = x_{ij} \Big/ \sqrt{\sum_{i=1}^{n} x_{ij}^2}, \quad i = 1, \cdots, n; \quad j = 1, \cdots, m. \tag{1}$$

where n and m denote the number of alternatives and the number of criteria, respectively. The performance value of alternative A_i on the criterion C_j is represented by x_{ij}.

Step 2: Calculate the weighted normalized decision matrix according to obtaining the criterion weights using entropy method. The weighted normalized value v_{ij} is calculated as:

$$v_{ij} = \omega_j r_{ij}, \quad i = 1, \cdots, n; \ j = 1, \cdots, m. \tag{2}$$

where ω_j is the weight of the *jth* criterion, and $\sum_{j=1}^{m} \omega_j = 1$.

Step 3: Find the ideal alternative solution A^+, which is calculated as follows:

$$A^+ = \{v_1^+, \cdots, v_m^+\} = \left\{ \left(\max_i v_{ij} \middle| j \in I' \right), \left(\min_i v_{ij} \middle| j \in I'' \right) \right\}. \tag{3}$$

where I' indicates benefit criteria and I'' indicates cost criteria. For the evaluation of classification algorithms, G-mean, F-measure, AUC are benefit criteria to be maximized, while FP rate, FN rate, and time are cost criterion to be minimized.

Step 4: Find the anti-ideal alternative solution A^-, which is calculated as follows:

$$A^- = \{v_1^-, \cdots, v_m^-\} = \left\{ \left(\min_i v_{ij} \middle| j \in I' \right), \left(\max_i v_{ij} \middle| j \in I'' \right) \right\}. \tag{4}$$

Step 5: Calculate the degree of separation using the n dimensional Euclidean distance. The distance of each alternative from the ideal solution is calculated as follows:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} \left(v_{ij} - v_j^+ \right)^2}, \quad i = 1, \cdots, n. \tag{5}$$

The distance of each alternative from the anti-ideal solution is calculated as follows:

$$D_i^- = \sqrt{\sum_{j=1}^{m} \left(v_{ij} - v_j^- \right)^2}, \quad i = 1, \cdots, n. \tag{6}$$

Step 6: Calculate relative approach degree as follows:

$$R_i^+ = D_i^- \big/ (D_i^- + D_i^+), \quad i = 1, \cdots, n. \tag{7}$$

Step 7: Rank alternatives by maximizing the relative approach degree R_i^+.

B. ENTROPY METHOD –DETERMINING CRITERIA WEIGHTS

The weights of criteria play an important role in MCDM models and have crucial impact on the final ranking of alternatives. Various approaches have been developed to determine criteria weights [57]–[60]. The information entropy [61] is a measure of the average unpredictability of a random variable. The advantage of the entropy-based weights computing method is that it calculates the criteria weights from the given evaluating matrix and requires no input from the decision maker. This method has been used to assign criterion weights in some literature [62], [63].

In the experimental study, the criteria weights are estimated using the following procedure. Let X be the set of evaluating objects, Y be the set of evaluating index. The standardization of evaluating matrix is presented as:

$$D = \begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_i \\ \vdots \\ A_m \end{array} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1i} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2i} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ii} & \cdots & x_{in} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mi} & \cdots & x_{mn} \end{bmatrix}. \quad (8)$$

where A_i is the *ith* alternative and x_{ij} is the representing value of the *ith* alternative in relation to the *jth* criterion.

Step 1: Calculate the normalized decision matrix R

$$R = \left[r_{ij}\right]_{n\times m}, \quad i = 1, \cdots, n; j = 1, \cdots, m. \quad (9)$$

The normalized value r_{ij} is calculated for the benefit criteria as follows

$$r_{ij} = \left(\frac{\max x_{ij} - x_{ij}}{\max x_{ij} - \min x_{ij}}\right). \quad (10)$$

Step 2: Calculate information entropy value. The entropy of each index j is defined as follows

$$E_j = -k \sum_{i=1}^{n} f_{ij} \ln f_{ij}, \quad j = 1, \cdots, m. \quad (11)$$

Where value of$f_{ij}$is defined as $f_{ij} = {r_{ij}} \Big/ \sum_{i=1}^{n} r_{ij}, k = {1}\big/{\ln(n)}$, which guarantee $0 \leq E_j \leq 1$ and suppose when $f_{ij} = 0, f_{ij} \ln f_{ij} = 0$.

Step 3: Calculate difference degree. The difference degree of each index j can be calculated as follows:

$$G_j = 1 - E_j, \quad j = 1, 2, \cdots, m. \quad (12)$$

Step 4: Calculate index weigh $\omega = (\omega_1, \omega_2, \cdots, \omega_m)^T$

$$\omega_j = {G_j} \Big/ \sum_{j=1}^{n} G_j, \quad j = 1, 2, \cdots, m. \quad (13)$$

Since the lower value of entropy indicates the higher diversification and more information of the criterion, the weight of the criterion would be higher.

IV. EXPERIMENTAL STUDY

An experimental study is designed to evaluate the effectiveness of the proposed approach. Utilizing 7 imbalanced binary data sets representing credit approval risk and bankruptcy risk from the UCI Machine Learning repository [64], the experiment compares the performances of three groups of imbalanced classification approaches. Four base classifiers have been selected from commonly used classification techniques in financial risk prediction [52]: a decision tree C4.5 [65], SVM [66], LR [67] and multilayer perceptron (MLP) [68]. The three groups of imbalanced techniques (resampling techniques (RUS and SMOTE), cost sensitive, and ensembles (bagging and boosting)) are combined with the four base classifiers.

The experiment was carried out according to the following process:

Input: 7 binary financial imbalanced classification data sets

Output: Rankings of 20 classifiers

Step 1: Prepare target imbalanced data sets.

Step 2: Setting cost-matrix $C(+,-) = IR$, $C(-,+) = 1$ and then carrying out cost-sensitive classification algorithms on 10-fold cross-validation using WEKA 3.7 [69].

Step 3: Executing SMOTE ($k = 5$) [33] and RUS by means of WEKA 3.7 to obtain balanced data set.

Step 4: Carrying out SMOTE and RUS-based ensemble algorithms on 10-fold cross-validation of the obtained balanced data set by means of WEKA 3.7.

Step 5: Determine the weights of six evaluation criteria (G-mean, F-measure, AUC, FP rate, FN rate, and time) by means of entropy method following the procedure described in Section 3.2 using MATLAB- R2012b.

Step 6: Evaluate classification approaches based on six evaluation criteria (G-mean, F-measure, AUC, FP rate, FN rate, and time) using TOPSIS, which is implemented using MATLAB R2012b to generate a ranking of all the classification approaches.

END

A. IMBALANCED DATA SETS

This study chose 7 highly imbalanced financial risk-related binary data sets from the UCI Machine Learning repository. *Table 2* summarizes the data name, number of features, number of instances, percentage of positive (bankrupt or default) and negative (normal) instances, and class imbalance ratios (IR), which is the ratio of the number of instances of the majority class and the minority class.

B. EXPERIMENTAL SETUP

Four classification algorithms: LR, SVM, MLP, and C4.5 are selected as the base classifiers. All these four classifiers have been implemented in the Weka learning environment [69] using the default parameters. The Cost-Sensitive Classifier from the Weka environment [69] was utilized to provide cost-sensitive versions of the four basic classifiers. SMOTE-Boost and Under-Bagging are representatives of ensemble techniques.

The experimental study was conducted using the 10-fold cross validation strategy. Each data set was divided into ten folds and each fold has similar number of instances. Then for each fold, a learning algorithm was trained on the remaining nine folds and then tested on the current fold. To obtain stable and reliable results, the 10-fold cross-validation strategy was repeated 10 times and each time the ordering of instances was shuffled.

C. RESULTS AND DISCUSSION

1) RESULTS

Table 3 summarizes the average results of all 20 algorithms on the six criteria. The mean value across all data sets generated by each algorithm on each metric is used to represent the performance of that algorithm. We can observe that no algorithm achieves the best performances across all criteria.

The weights of the six criteria used in TOPSIS is summarized in *Table 4* based on the entropy approach. Shown in *Table 4*, the most important performance measures are AUC, F-measure, and FN rate. Finally, *Table 5* reports the rankings of classification algorithms generated by TOPSIS method using the average classification results on the 7 imbalanced data sets.

Furthermore, some observations are summarized as follows:

1) SMOTE, as a single or hybrid imbalanced learning approach, outperforms any other groups of algorithms, including cost sensitive classifiers, resampling techniques (RUS), and hybrid approaches (Under-Bagging). SMOTE-BOOST-classifier and SMOTE-classifier are the top two ranked groups of algorithms for financial imbalanced classification.

2) SMOTEBoost-C4.5, SMOTE- C4.5, and SMOTE- MLP are ranked as the top three classifiers based on their performances on the six criteria. The results are in concordance with the studies done in [28], [70].

3) The resampling technique RUS is outperformed by SMOTE, which is in concordance with the study done in [28], [39]. All the four SMOTE-classifiers rank higher than the RUS-classifiers, which may due to the removal of significant samples during the learning process.

4) As a group, the CS-classifiers ranked lower than the RUS-classifiers, SMOTE-classifiers and hybrid groups of algorithms.

2) COMPARATIVE ANALYSIS AND DISCUSSION

In this section, we draw a comparison with previous study [71], which proposed an accurate multi-criteria decision making methodology based on four evaluation criterion (Wgt.Avg.F-score, CPUTimeTesting, CPUTimeTraining, and Consistency measures) to empirically evaluate and rank classifiers.

The basic ideas of the two articles are the similar. While, Ref [71] cannot evaluate imbalanced classifiers in financial risk prediction very well because the evaluation criteria are not comprehensive. There is only one evaluation index: F-score for the accuracy of characterization in [71]. For example, consider a credit data set where only 10 companies are bankruptcy and 100 companies are non-bankrupt; suppose the confusion matrix for two classifiers are shown in *Tables 6* and *7* , respectively. Based on *Tables 6* and *7*, F-score and FN rate of two classifiers in given data set are shown in *Table 8*. According to evaluation criterion F-score [71], we conclude that classifier 2 is superior to classifier 1. However, in assessing the performance of the models, we considered FN rate as more important, because the economic cost of classifying a bankruptcy company as non-bankrupt is higher than that of the reverse classification. Whereas the classifier 1 had a 10% FN rate, the classifier 2 indicated a 30% FN rate as inferred from *Table 8*. This is strong evidence that the only one evaluation index: F-score for the accuracy of characterization in [71] are insufficient to evaluate the imbalanced classification problem about bankruptcy classification. The proposed MCDM method based on six key criteria (G-mean, F-measure, AUC, FP rate, FN rate, and time) can well evaluate the problem of unbalanced classification in financial risk prediction.

TABLE 1. Confusion matrix for a two-class problem.

	Positive prediction	*Negative prediction*
Positive class	True Positive(TP)	False Negative(FN)
Negative class	False Positive(FP)	True Negative(TN)

D. STATISTICAL SIGNIFICANCE TESTS

In general, the non-parametric tests should be preferred over the parametric ones because they do not assume normal distributions and are independent for any evaluation measure.

To verify the significance of the experimental results obtained by this study and based on the recommendations of previous research [72]–[74], Wilcoxon test [75] is employed in this paper. To save the space of this paper, we only take the process of Wilcoxon test for the top five algorithms in terms of AUC, F-measure, and FN rate, respectively. For simplicity, the top five algorithms (SMOTEBoost-C4.5, SMOTE-C4.5, SMOT-MLP, SMOT-LR, SMOTEBoost-LR) derived by TOPSIS method are denoted by 1-5, respectively. The Wilcoxon test results are shown in *Table 9*. It can be seen from *Table 9* that, in terms of AUC, significant differences are found in cases of 1 vs. 3, 1 vs. 4, 1 vs. 5, and 2 vs. 5 (with $\alpha = 0.01$). Significant difference can also be found in the case of 2 vs. 4, 3 vs. 4, and 3 vs. 5 (with $\alpha = 0.05$). But, significant difference cannot be found in the case of 1 vs. 2, 2 vs. 3, 4 vs. 5. In terms of F-measure, significant differences are found in cases of 1 vs. 3, 1 vs. 4, 1 vs. 5, 2 vs. 3, 2 vs. 4, and 2 vs. 5 (with $\alpha = 0.01$), whereas significant difference cannot be found in the cases of 1vs. 2, 3 vs. 4, 3 vs. 5, and 4 vs. 5. In terms of FN rate, significant differences are found in cases of 1 vs. 3, 1 vs. 4, 1 vs. 5, 2 vs. 3, 2 vs. 4, and 2 vs. 5 (with $\alpha = 0.01$), significant difference can also be found in the case of 1 vs. 2 (with $\alpha = 0.05$). Whereas significant difference cannot be found in the cases of 3 vs. 4, 3 vs. 5, and 4 vs. 5.

Through the above analysis, we can draw the conclusions that significant difference cannot be found based on a single evaluation criteria for ranking algorithms (1 vs. 2, 2 vs. 3, 3 vs. 4, and 4 vs. 5). In this case, the proposed MCDM-based

TABLE 2. Description of imbalanced data sets.

data name	Features	Instances	% Positive instances	% Negative instances	IR
Polish companies bankruptcy -1stYear	64	7027	3.86	96.14	25
Polish companies bankruptcy -2ndYear	64	10173	3.93	96.07	24.4
Polish companies bankruptcy -3rdYear	64	10503	4.71	95.29	20.2
Polish companies bankruptcy -4thYear	64	9792	5.26	94.74	18
Polish companies bankruptcy -5thYear	64	5910	6.94	93.06	13.4
default of credit card clients	23	30000	22	78	3.52
German Credit	24	1000	30	70	2.33

TABLE 3. The average classification results based on 7 financial imbalanced data sets.

Algorithms		G-mean	F-measure	AUC	FP rate	FN rate	Time (s)
RUS	SVM	0.6114	0.581	0.6327	0.2643	0.4706	3.65
	MLP	0.7269	0.7129	0.7904	0.2123	0.3279	28.64
	LR	0.706	0.7024	0.7581	0.2799	0.3074	**0.327**
	C4.5	0.68	0.6823	0.6863	0.322	0.3156	0.66
SMOTE	SVM	0.6289	0.6097	0.6541	0.2864	0.4053	65.67
	MLP	0.7936	0.7921	0.8739	0.194	0.2164	416.69
	LR	0.754	0.747	0.83	0.2107	0.2786	5.77
	C4.5	0.8841	0.8824	0.8974	0.1011	0.1296	9.63
CS	SVM	0.5883	0.2484	0.6053	0.2883	0.501	24.29
	MLP	0.4183	0.218	0.7189	0.6503	0.184	212.31
	LR	0.6813	0.2867	0.749	0.3199	0.3081	2.84
	C4.5	0.6063	0.3537	0.6646	0.1189	0.5636	5.12
UnderBagging	SVM	0.6426	0.6114	0.7064	0.2341	0.449	38.97
	MLP	0.7494	0.7421	0.825	0.217	0.279	344.43
	LR	0.71846	0.7159	0.7846	0.2709	0.2916	4.6
	C4.5	0.7157	0.7213	0.7836	0.3001	0.267	5.29
SMOTEBoost	SVM	0.6537	0.645	0.698	0.3211	0.36541	311.27
	MLP	0.7979	0.7897	0.8586	0.1429	0.2544	2594.37
	LR	0.7559	0.7499	0.7976	0.2147	0.2713	23.47
	C4.5	**0.9193**	**0.9189**	**0.9619**	**0.07586**	**0.08543**	98.89

evaluation approach, which integrates performance values from multiple evaluation criteria, provides a new perspective.

V. CONCLUSIONS

Default and bankruptcy are rare events compared to normal accounts and companies functioning well, which indicate that financial risk data are imbalanced by nature. Many techniques have been developed to deal with the problem of learning from imbalanced data sets How to select an effective and appropriate algorithm for financial risk classification is an importance task. The goal of this paper is to evaluate imbalanced classifiers in financial risk prediction by considering multiple performance measures simultaneously using a multi-criteria decision making (MCDM) method.

TABLE 4. The criteria weights by means of entropy method.

G-mean	F-measure	AUC	FP rate	FN rate	Time (s)
0.1319	0.2287	0.2749	0.0942	0.1956	0.0747

To ensure the objectiveness of the final ranking of classifiers, the entropy-based method was used to calculate the criteria weights from the given evaluating matrix and requires no input from the decision maker.

An experiment was designed to evaluate the proposed approach using 7 financial imbalanced binary data sets from the UCI Machine Learning repository. The experiment makes use of four standard classifiers (i.e., LR, SVM,

TABLE 5. Ranking result for algorithms using TOPSIS.

Algorithms		R_i^+	Ranking
RUS	SVM	0.5628	16
	MLP	0.70258	9
	LR	0.6948	10
	C4.5	0.6622	11
SMOTE	SVM	0.5948	13
	MLP	0.7898	**3**
	LR	0.752	**4**
	C4.5	0.9316	**2**
CS	SVM	0.4834	20
	MLP	0.5222	17
	LR	0.5756	15
	C4.5	0.5092	18
UnderBagging	SVM	0.5916	14
	MLP	0.727	6
	LR	0.7144	8
	C4.5	0.7231	7
SMOTEBoost	SVM	0.6064	12
	MLP	0.4952	19
	LR	0.7499	**5**
	C4.5	0.9779	**1**

TABLE 6. Confusion matrix for classifier 1.

	Positive prediction	*Negative prediction*
Positive class	9	1
Negative class	10	90

TABLE 7. Confusion matrix for classifier 2.

	Positive prediction	*Negative prediction*
Positive class	7	3
Negative class	5	95

MLP and C4.5) combined with three groups of imbalanced techniques, namely cost-sensitive learning, resampling (RUS and SMOTE), and hybrid approaches. Six frequently used performance metrics for imbalanced learning: G-mean, F-measure, AUC, FP rate, FN rate, and time were used in the experiment. TOPSIS, a well-known MCDM method, was applied to rank the imbalanced learning approaches. The final

TABLE 8. F-score and FN rate for classifiers 1 and 2.

	F-score	*FN rate*
Classifier 1	0.621	0.10
Classifier 2	0.636	0.30

TABLE 9. Wilcoxon tests of the top five algorithms in terms of AUC, F-measure, and FN rate.

Comparison	Measure	R^+	R^-	Hypothesis	P-value
1 vs. 2	AUC	53	38	Not rejected	0.138
	F-measure	52	39	Not rejected	0.181
	FN rate	26	65	**Rejected at 5%**	0.022
1 vs. 3	AUC	61	30	**Rejected at 1%**	0.005
	F-measure	61	30	**Rejected at 1%**	0.005
	FN rate	23	68	**Rejected at 1%**	0.005
1 vs. 4	AUC	62.5	28.5	**Rejected at 1%**	0.001
	F-measure	62	29	**Rejected at 1%**	0.002
	FN rate	22	69	**Rejected at 1%**	0.002
1 vs. 5	AUC	63	28	**Rejected at 1%**	0.001
	F-measure	62	29	**Rejected at 1%**	0.002
	FN rate	22	69	**Rejected at 1%**	0.002
2 vs. 3	AUC	51	40	Not rejected	0.234
	F-measure	61	30	**Rejected at 1%**	0.005
	FN rate	23	68	**Rejected at 1%**	0.005
2 vs. 4	AUC	58	33	**Rejected at 5%**	0.022
	F-measure	62	29	**Rejected at 1%**	0.002
	FN rate	22	69	**Rejected at 1%**	0.002
2 vs. 5	AUC	61	30	**Rejected at 1%**	0.005
	F-measure	62	29	**Rejected at 1%**	0.002
	FN rate	22	69	**Rejected at 1%**	0.002
3 vs. 4	AUC	57	34	**Rejected at 5%**	0.035
	F-measure	56	35	Not rejected	0.051
	FN rate	33	58	Not rejected	0.234
3 vs. 5	AUC	57	34	**Rejected at 5%**	0.035
	F-measure	55	36	Not rejected	0.073
	FN rate	36	55	Not rejected	0.445
4 vs. 5	AUC	50	41	Not rejected	0.295
	F-measure	41	50	Not rejected	0.945
	FN rate	45.5	45.5	Not rejected	0.628

1 indicates SMOTEBoost-C4.5; 2 indicates SMOTE-C4.5; 3 indicates SMOT-MLP; 4 indicates SMOT-LR; 5 indicates SMOTEBoost-LR

ranking results indicate that SMOTE-based ensemble classifiers outperform other groups of imbalanced learning algorithms, SMOTEBoost-C4.5, SMOTE-C4.5, and SMOT-MLP were ranked as the top three classifiers based on their performances on the six criteria.

From the above discussion, the proposed MCDM-based evaluation approach for imbalanced learning approaches can make up the shortfall of single criteria evaluation. Hence, it is interesting topic to establish an assembled algorithm based on MCDM method to classify the financial imbalanced data sets in the future. Besides, as future work, the performance of the classifiers for imbalanced data sets in regard to class imbalance ratios (IR) is another research, which may provide a useful guide to select a suitable classification method in financial risk prediction.

REFERENCES

[1] S. Birla, K. Kohli, and A. Dutta, "Machine Learning on imbalanced data in Credit Risk," in *Proc. IEEE 7th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Oct. 216, pp. 1–6.

[2] M. L. Challa, V. Malepati, and S. N. R. Kolusu, "Forecasting risk using auto regressive integrated moving average approach: An evidence from S&P BSE Sensex," *Financial Innov.*, vol. 4, no. 1, pp. 1–17, Oct. 2018.

[3] I. Yousaf, S. Ali, and S. Z. A. Shah, "Herding behavior in ramadan and financial crises: The case of the Pakistani stock market," *Financial Innov.*, vol. 4, pp. 1–14, Jul. 2018.

[4] S. Moradi and F. M. Rafiei, "A dynamic credit risk assessment model with data mining techniques: Evidence from iranian banks," *Financial Innov.*, vol. 5, pp. 1–27, Mar. 2019.

[5] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, Feb. 2012.

[6] H. Li and J. Sun, "Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples—Evidence from the Chinese hotel industry," *Tourism Manage*, vol. 33, no. 3, pp. 622–634, Jan. 2012.

[7] J. Sun, H. Li, Q.-H. Huang, and K.-Y. He, "Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches," *Knowl.-Based Syst.*, vol. 57, pp. 41–56, Feb. 2014.

[8] P. R. Kumar and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques—A review," *Eur. J. Oper. Res.*, vol. 180, no. 1, pp. 1–28, Jul. 2007.

[9] D. Liang, C.-C. Lu, C.-F. Tsai, and G.-A. Shih, "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study," *Eur. J. Oper. Res.*, vol. 252, no. 2, pp. 561–572, Jul. 2016.

[10] J. Sun, H. Li, and H. Adeli, "Concept drift-oriented adaptive and dynamic support vector machine ensemble with time window in corporate financial risk prediction," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 4, pp. 801–813, Jul. 2013.

[11] H. He, W. Zhang, and S. Zhang, "A novel ensemble method for credit scoring: Adaption ofdifferent imbalance ratios," *Expert Syst. Appl.*, vol. 98, pp. 105–117, May 2018.

[12] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Inf. Sci.*, vol. 425, pp. 76–91, Jan. 2018.

[13] D. Veganzones and E. Séverin, "An investigation of bankruptcy prediction in imbalanced datasets," *Decis. Support Syst.*, vol. 112, pp. 111–124, Aug. 2018.

[14] V. García, A. I. Marqués, and J. S. Sánchez, "Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction," *Inf. Fusion*, vol. 47, pp. 88–101, May 2019.

[15] G. Kou, Y. Peng, and G. X. Wang, "Evaluation of clustering algorithms for financial risk analysis using MCDM methods," *Inf. Sci.*, vol. 275, pp. 1–12, Aug. 2014.

[16] G. Kou, Y. Peng, and C. Lu, "MCDM approach to evaluating bank loan default models," *Technol. Econ. Develop. Economy*, vol. 20, pp. 292–311, Jan. 2014.

[17] N. Chen, B. Ribeiro, and A. Chen, "Financial credit risk assessment: A recent review," *Artif. Intell. Rev.*, vol. 45, no. 1, pp. 1–23, Jan. 2016.

[18] C. Serrano-Cinca and B. Gutiñrrez-Nieto, "Partial least square discriminant analysis for bankruptcy prediction," *Decis. Support Syst.*, vol. 54, no. 3, pp. 1245–1255, Feb. 2013.

[19] P. Ravikumar and V. Ravi, "Bankruptcy prediction in banks by an ensemble classifier," in *Proc. IEEE Int. Conf. Ind. Technol.*, vol. 12, Jun. 2007, pp. 2032–2036.

[20] J. Sun and H. Li, "Listed companies,' financial distress Predict. based weighted majority voting combination multiple classifiers," *Expert Syst. Appl.*, vol. 35, pp. 818–827, Oct. 2008.

[21] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, nos. 1–2, pp. 1–39, 2010.

[22] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

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

[24] E. Alfaro, N. García, M. Gámez, and D. Elizondo, "Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks," *Decis. Support Syst.*, vol. 45, no. 1, pp. 110–122, Apr. 2008.

[25] M.-J. Kim and D.-K. Kang, "Ensemble with neural networks for bankruptcy prediction," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3373–3379, Apr. 2010.

[26] S. Y. Kim and A. Upneja, "Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models," *Econ. Model.*, vol. 36, pp. 354–362, Jan. 2014.

[27] J. Sun, M.-Y. Jia, and H. Li, "AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9305–9312, Aug. 2011.

[28] 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.

[29] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[30] T. Raeder, G. Forman, and N. V. Chawla, "Learning from imbalanced data: Evaluation matters," in *Data Mining: Foundations and Intelligent Paradigms*, New York, NY, USA: Springer, vol. 23, pp. 315–331, 2012.

[31] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 40, no. 1, pp. 185–197, Oct. 2009.

[32] J. Van Hulse, T. M. Khoshgoftaar, and A. Napolitano, "Experimental perspectives on learning from imbalanced data," *Proc. 24th Int. Conf. Mach. Learn.*, Corvallis, OR, USA, 2007, pp. 935–942.

[33] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[34] R. Barandela, R. M. Valdovinos, and J. S. Sanchez, "New applications of ensembles of classifiers," *Pattern Anal. Appl.*, vol. 6, no. 3, pp. 245–256, Dec. 2003.

[35] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," in *Proc. Knowl. Discovery Databases*, 2003, pp. 107–119.

[36] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 2, pp. 539–550, Apr. 2009.

[37] P. Domingos, "MetaCost: A general method for making classifiers cost-sensitive," *Proc. 5th Int. Conf. Knowl. Discovery Data Mining*, 1999, pp. 155–164.

[38] C. Elkan, "The foundations of cost-sensitive learning," in *Proc. Int. Joint Conf. Artif. Intell.*, Seattle, WA, USA, Aug. 2001, pp. 973–978.

[39] V.García, J. S. Sánchez, and R. A. Mollineda, "On the effectiveness of preprocessing methods when dealing with different levels of class imbalance," *Knowl.-Based Syst.*, vol. 25, no. 1, pp. 13–21, Feb. 2012.

[40] T. M. Khoshgoftaar, J. V. Hulse, and A. Napolitano, "Comparing boosting and bagging techniques with noisy and imbalanced data," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 3, pp. 552–568, May 2011.

[41] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.
[42] S. Crone and S. Finlay, "Instance sampling in credit scoring: An empirical study of sample size and balancing," *Int. J. Forecasting*, vol. 28, pp. 224–238, Jan./Mar. 2012.
[43] J.-J. Liao, C.-H. Shih, M.-F. Hsu, and T.-F. Chen, "An ensemble-based model for two-class imbalanced financial problem," *Econ. Model.*, vol. 37, pp. 175–183, Feb. 2014.
[44] V. García, A. I. Marqués, and J. S. Sánchez, "Improving risk predictions by preprocessing imbalanced credit data," in *Proc. Int. Conf. Neural Inf. Process.*, 2012, pp. 68–75.
[45] J. Sun, Y. C. Lee, H. Li, and Q. H. Huang, "Combining B&B-based hybrid feature selection and the imbalance-oriented multiple-classifier ensemble for imbalanced credit risk assessment," *Technol. Econ. Develop. Economy*, vol. 21, no. 3, pp. 351–378, May 2015.
[46] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[47] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2006, pp. 233–240.
[48] A. Charnes and W. W. Cooper, *Management Models and Industrial Applications of Linear Programming*. Hoboken, NJ, USA: Wiley, 1961.
[49] D. Ergu, G. Kou, Y. Peng, and Y. Shi. "A simple method to improve the consistency ratio of the pair-wise comparison matrix in ANP," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 246–259, Aug. 2011.
[50] Y.-K. Lin and C.-T. Yeh, "Multi-objective optimization for stochastic computer networks using NSGA-II and TOPSIS," *Eur. J. Oper. Res.*, vol. 218, no. 3, pp. 735–746, 2012.
[51] M. T. Lua, S.-W. Lin, and G. H. Tzeng, "Improving RFID adoption in Taiwan's healthcare industry based on a DEMATEL technique with a hybrid MCDM model," *Decis. Support Syst.*, vol. 56, pp. 259–269, Dec. 2013.
[52] G. L. Yang, J. B. Yang, W. B. Liu, and X. X. Li, "Cross-efficiency aggregation in DEA models using the evidential-reasoning approach," *Eur. J. Oper. Res.*, vol. 231, pp. 393–404, Dec. 2013.
[53] H. S. Kilic, S. Zaim, and D. Delen, "Selecting "The Best,' ERP system for SMEs using a combination ANP PROMETHEE methods," *Expert Syst. Appl.*, vol. 42, pp. 2343–2352, Apr. 2015.
[54] D. Bouyssou and T. Marchant, "On the relations between ELECTRE TRI-B and ELECTRE TRI-C and on a new variant of ELECTRE TRI-B," *Eur. J. Oper. Res.*, vol. 242, pp. 201–211, Apr. 2015.
[55] D. L. Olson, "Comparison of weights in TOPSIS models," *Math. Comput. Model.*, vol. 40, pp. 721–727, Oct. 2004.
[56] Y. Peng, G. Wang, and H. Wang, "User preferences based software defect detection algorithms selection using MCDM," *Inf. Sci.*, vol. 191, pp. 3–13, May 2012.
[57] E. U. Choo and W. C. Wedley, "Optimal criterion weights in multicriteria decision making," *J. Oper. Res. Soc.*, vol. 36, pp. 983–1075, Dec. 2017.
[58] J. S. Dyer and R. K. Sarin, "Measurable multiattribute value functions," *Oper. Res.*, vol. 27, pp. 810–832, Aug. 1979.
[59] D. Pekelman and S. Sen, "Mathematical programming models for determination of attribute weights," *Manage. Sci.*, vol. 20, pp. 1217–1246, Apr. 1974.
[60] D. Zhang, P. L. Yu, and P. Z. Wang, "State-dependent weights in multicriteria value functions," *J. Optim. Theory Appl.*, vol. 74, no. 1, pp. 1–22, Jul. 1992.
[61] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 79–423, Jul. 1948.
[62] S. Chen and C. Hwang, *Multiple Attribute Decision Making: Method and Applications*. New York, NY, USA: Springer, 1992.
[63] Z. Zhi-Hong, Y. Yi, and S. Jing-Nan, "Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment," *J. Environ. Sci.*, vol. 18, no. 5, pp. 1020–1023, 2006.
[64] A. Frank and A. Asuncion. (2015). *UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science*. [Online]. Available: http://archive.ics.uci.edu/ml
[65] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.
[66] J. Platt, "Machines using sequential minimal optimization," in *Advances in Kernel Methods-Support Vector Learning*, Cambridge, MA, USA: MIT Press, 1998.
[67] D. W. Hosmer and L. Stanley, *Applied Logistic Regression*, New York, NY, USA: Wiley, 2013.
[68] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
[69] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.
[70] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, and Y. Zhou, "A novel ensemble method for classifying imbalanced data," *Pattern Recognit.*, vol. 48, no. 5, pp. 1623–1637, 2015.
[71] R. Ali, S. Y. Lee, and T. C. Chung, "Accurate multi-criteria decision making methodology for recommending machine learning algorithm," *Expert Syst. Appl.*, vol. 71, pp. 257–278, Apr. 2017.
[72] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[73] Y. Liu, J. W. Bi, and Z. P. Fan, "Multi-class sentiment classification: The experimental comparisons offeature selection and machine learning algorithms," *Expert Syst. Appl.*, vol. 80, pp. 323–339, Sep. 2017.
[74] L. Zhou, "Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods," *Knowl.-Based Syst.*, vol. 41, no. 1, pp. 16–25, Mar. 2013.
[75] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945.

YONGMING SONG received the M.S. degree from the School of Mathematical Science, Chongqing Normal University, Chongqing, China, and the Ph.D. degree from the School of Management and Economics, University of Electronic Science and Technology of China, Chengdu, China, in 2009 and 2018, respectively.

He is currently a Lecturer with the School of Business Administration, Shandong Technology and Business University. His research has been published or accepted in the *Applied Soft Computing*, the *Journal of the Operational Research Society*, *Soft Computing*, *PLOS ONE*, and the *Mathematical Problems in Engineering*. His current research interests include group decision making, multicriteria decision making, data mining, E-commerce, and logistics management.

YI PENG received the B.S. degree in management information systems from Sichuan University, China, in 1997, and the M.S. degree in management information systems and the Ph.D. degree in information technology from the University of Nebraska at Omaha, Omaha, NE, USA, in 2007. She was an Assistant Professor, from 2007 to 2011, and has been a Professor with the School of Management and Economics, University of Electronic Science and Technology of China, since 2011. Her current research interests include data mining, multiple criteria decision making, and data mining applications. He has been selected to the list of Highly Cited Researchers 2016 in the field of Computer Science published by Clarivate Analytics (formerly Thomson Reuters).