Information Sciences 559 (2021) 153–170

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods

Jie Sun [a], Hamido Fujita Ph.D [b,c,d,*], Yujiao Zheng [a], Wenguo Ai [e]

[a] School of Accountancy, Tianjin University of Finance and Economics, Tianjin, China
[b] Faculty of Information Technology, Ho Chi Minh City University of Technology (HUTECH), Ho Chi Minh City, Viet Nam
[c] Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, Spain
[d] Faculty of Software and Information Science, Iwate Prefectural University, Iwate Japan
[e] School of Management, Harbin Institute of Technology, Harbin, China

A R T I C L E I N F O

Article history:
Received 30 August 2020
Received in revised form 6 November 2020
Accepted 17 January 2021
Available online 5 February 2021

Keywords:
Financial distress prediction
Multi-class classification
Decomposition and fusion method
Support vector machine

A B S T R A C T

Binary financial distress prediction (FDP), which categorizes corporate financial status into the two classes of distress and nondistress, cannot provide enough support for effective financial risk management. This paper focuses on research on multiclass FDP based on the support vector machine (SVM) integrated with the decomposition and fusion methods. Corporate financial status is subdivided into four states: financial soundness, financial pseudosoundness, moderate financial distress and serious financial distress. Three multi-class FDP models are built by integrating the SVM with three decomposition and fusion methods, i.e., one-versus-one (OVO), one-versus-rest (OVR), and error-correcting output coding (ECOC), and they are, respectively called OVO-SVM, OVR-SVM and ECOC-SVM. Empirical research based on data from Chinese listed companies shows that OVO-SVM overall outperforms OVR-SVM and ECOC-SVM and is preferred for multiclass FDP. In addition, all three models trained on the original highly class-imbalanced training dataset cannot obtain satisfying performance, and the data level preprocessing mechanisms that make class distributions balanced in the training dataset can greatly improve their multiclass FDP performance. Compared with multivariate discriminant analysis (MDA) and multinomial logit (MNLogit), OVO-SVM has significantly higher accuracy for financial pseudosoundness and moderate financial distress and lower accuracy for financial soundness and serious financial distress, resulting in no significant difference among their overall multiclass FDP performance. However, OVO-SVM is still more competitive than MDA and MNLogit in that financial pseudosoundness and moderate financial distress are much more difficult to predict by human expertise than the other two financial states.

1. Introduction

At present, the world economy is characterized by an overall downward trend. In such an uncertain economic environment, enterprises are under tremendous pressure, and many companies are confronted with financial distress due to unsci-

* Corresponding author at Faculty of Information Technology, Ho Chi Minh City University of Technology (HUTECH), Ho Chi Minh City, Viet Nam.
E-mail addresses: h.fujita@hutech.edu.vn, hfujita-799@acm.org (H. Fujita).

https://doi.org/10.1016/j.ins.2021.01.059
0020-0255/

entific financial management. When an enterprise falls into financial distress, not only is its own going-concern capability directly and negatively affected but also the interests of its external stakeholders are threatened. For example, enterprises in financial distress may be unable to repay bank loans due to deficiency of cash and cash equivalents, which may lead to an increase in bad debts of banks and admit massive hidden dangers to the entire financial system. One of the reasons that enterprises fall into financial distress is that financial managers lack effective supportive tools to help them forecast different types of financial distress. Consequently, an enterprise in mild financial distress may fail to perceive its increasing financial risk and miss the most opportune time to take measures. As a result, serious financial distress in a variety of forms, such as vast or continuous losses, debt default, bankruptcy, and so on, may break out when financial risk accumulates to a certain level. Therefore, it is important to predict financial distress more scientifically based on the financial information disclosed by enterprises.

Exploring effective financial distress prediction (FDP) models has always been an essential research topic that attracts both academics and practitioners. In the 20th century, some classical statistical FDP models were developed, such as the Z-score model [3], the Zeta-score model [4], the logistic regression analysis (Logit) model [32] and the probit model [65], which set a foundation for further research on FDP modeling. With the development of artificial intelligence, an increasing number of researchers have devoted themselves to developing more effective FDP models based on artificial intelligence technologies, such as decision trees (DTs) [14], neural networks (NNs) [52], support vector machines (SVMs) [39], case-based reasoning (CBR) [41], and genetic algorithms (GAs) [38]. Either statistical or artificial intelligence FDP models have their own advantages and disadvantages. For example, statistical FDP models usually have fewer parameters and show more stable performance on different datasets, but they lack satisfactory learning capability from the complex data distribution. In contrast, artificial intelligence FDP models present strong learning capability, which may require many more parameters and modeling time and sacrifice some generalization ability.

In the 21st century, more attention has been given to hybrid and ensemble FDP models. The hybrid FDP models integrate two or more algorithms to improve FDP performance, e.g., applying GA, rough set theory, and the sequential floating forward selection algorithm. to optimize the FDP features for the above statistical or artificial intelligence classifiers [6,42,60], integrating fuzzy theory and the SVM algorithm to propose the fuzzy-SVM model [8]. The ensemble FDP models output the prediction results by combining the outputs of multiple base classifiers, e.g., using different classification algorithms to train multiple base classifiers on the same dataset [44], or utilizing the same classification algorithm to train multiple base classifiers on different versions of subsets extracted from the original dataset by random subspace [23], bagging [19], boosting [20], etc. In recent years, FDP models have been further developed from stationary modeling to dynamic modeling, which is based on instance selection mechanisms [46], adaptive and dynamic ensemble mechanisms based on data batch combinations [47], incremental bagging [27], and time-weighting Adaboost [40,48]. In addition, class-imbalanced FDP modeling also caught researchers' attention, and the problem of a biased distribution between the distressed and the nondistressed samples is solved by random undersampling of the majority class or upsampling of the minority class [63], synthetic minority oversampling technique (SMOTE) [56], oneclass classification methods [66], and some imbalance-oriented classifier ensemble approaches [43,50].

Although the above research on FDP modeling has obtained great achievement in the literature, all approaches focused on binary-class FDP, i.e., distressed and nondistressed, bankrupt and nonbankrupt. In fact, the problem of multiclass FDP is also very important and worth studying because stakeholders usually need more refined predictions of corporate financial states to make more scientific decisions. Some researchers have attempted to build multiclass FDP models based on approaches such as logit [16,22,33], MDA [5], NNs [1,59,67], and DT [64]. However, determining how to build an effective multiclass FDP model based on SVMs remains to be further explored. Hence, this paper attempts to explore multiclass FDP models based on SVMs combined with the decomposition and fusion methods.

This study has the following three contributions. First, it enriches the theoretical system of multiclass FDP modeling. Second, it broadens the application of machine learning approaches to corporate financial risk management. Third, it innovatively explores four-state multiclass FDP models based on SVMs combined with decomposition and fusion methods, which are tested by empirical experiments on the real-world data of Chinese listed companies.

The rest of the paper is organized as follows. Section 2 reviews the current literature on FDP. Section 3 defines the multiclass financial states of Chinese listed companies. Section 4 describes the approaches of multiclass FDP modeling based on SVMs combined with the decomposition and fusion methods. Section 5 shows the results and analysis of empirical research, and Section 6 concludes.

2. Literature review

Research on FDP modeling has been widely studied by both academics and practitioners and has achieved fruitful research results. FDP models have evolved from traditional statistical models to state-of-the-art artificial intelligence models.

Altman [3] proposed a Z-score model based on multivariate discriminant analysis (MDA) to analyze corporate bankruptcy and default risk. The Z-score model has become a classic model in the field of corporate credit evaluation. Altman [4] further revised the Z-score model and established a more comprehensive "Zeta" discriminant analysis model, which increased the discriminant variables from five to seven. Serrano-Cinca and Gutiérrez-Nieto [36] applied a partial least squares discriminant analysis method to establish a bankruptcy prediction model, which has the advantage of being unaffected by multicollinearity. All of the above models strictly require the sample data to obey the assumption of normal distribution and equivalent covariance. The logit model does not require the above two assumptions. Ohlson [32] first established the logit model to represent the relationship between corporate bankruptcy probability and financial ratios and found that the probabilistic prediction of bankruptcy showed more rationality. Similar to the logit model, the probit model is also based on maximum likelihood estimation, and Zmijewski [65] used the probit model to predict the probability of corporate financial distress. Tseng and Lin [54] used the quadratic programming method in logistic regression to construct the second interval logit model for forecasting bankruptcy, and Jabeur [15] used a partial least squares logistic regression method to construct a model for bankruptcy prediction, and both models improved the prediction performance.

Currently, DT, NNs, SVM, CBR, etc., have become very mature artificial intelligence algorithms for classification problems. FDP models based on artificial intelligence approaches have gradually caught increasing attention. McKee and Greenstein [29] used the recursive partitioning method to construct a DT model for bankruptcy prediction. Sun and Li [44,45] proposed a data mining method based on attribute induction and information gain to build a DT FDP model for Chinese listed companies. After Tam [52] attempted to use NNs for FDP, an increasing number of studies showed that NN FDP models have better performance than traditional statistical models [17,24,34]. The SVM is famous for good performance even when the sample size is relatively small, and empirical comparison indicated that FDP models based on SVMs showed higher accuracy than logit, MDA and NNs [7,25,39,62]. Mselmi et al. [31] found that compared with logit models, NNs, partial least squares models, etc., the SVM model has the highest FDP accuracy for one year prior to financial distress. Sun and Hui [41] carried out FDP by similarity-weighted voting CBR, which is based on k-nearest neighbors. Cheng et al. [9] proposed a novel purity-based k-nearest neighbors imputation method to build FDP models, which are robust to missing values and noise of the dataset.

The above FDP models are mainly based on a single artificial intelligence algorithm, but each algorithm has its own limitations. Therefore, some studies have integrated a variety of artificial intelligence algorithms to build hybrid or ensemble FDP models. For hybrid FDP models, there are three types of hybridization: one algorithm selects features for another algorithm, one algorithm optimizes parameters for another algorithm, or two or more algorithms hybridize to generate a new classification algorithm [49]. Yeh et al. [60] used rough set theory and the SVM algorithm to establish a two-stage hybrid model RS-SVM, in which RS is used to reduce redundant attributes and then combined with the SVM algorithm to predict business failure. Lin, Yeh and Lee [26] incorporated an isometric feature mapping algorithm into SVM training after feature dimension reduction, and their empirical evidence showed that the prediction performance is better than the model based on principal component analysis and SVMs. Fallahpour et al. [12] used the sequential floating forward selection algorithm to select the best eigenvalues and then combined it with the SVM classifier to construct the FDP model and experimentally verified its effectiveness. Zhang and Hu [61] used the nonlinear subspace multicore learning method to solve the optimal weight of the base kernel in the linear combination and then optimized the SVM to construct an optimized FDP model. Huang and Yen [18] found that the hybrid model, by integrating the deep belief network (DBN) and SVM, was able to generate more accurate FDP than the use of either the DBN or the SVM in isolation.

For ensemble FDP models, Sun and Li [44] built an ensemble FDP model by weighted majority voting for diverse classifiers, such as MDA, logit, DT, SVM and CBR, and found that an ensemble model outperforms a single classifier model in terms of accuracy and stability. Kim and Kang [19] experimentally compared the Bagging-NN ensemble model and the Boosting-NN ensemble model with the traditional NN model for FDP of Korean firms and indicated that the bagged and boosted NN ensemble FDP models showed improved performance over the traditional NN FDP model. Kim and Upneja [20] proposed an AdaBoosted DT ensemble model for FDP of publicly traded US restaurants, which demonstrated better FDP performance with smaller error in overall and type I error rates compared with single DT models. Li et al. [23] proposed the random subspace logit model by using the random subspace approach to generate a group of diverse logit decision agents and then combined their outputs to produce more accurate FDP results. Wang et al. [57] used the random subspace method improved by fusing the lasso regularized sparse method to build an ensemble FDP model, which solved the high-dimensionality and class-imbalance problems in FDP as well. Wang et al. [58] proposed the ensemble FDP model named the regularized sparse-based random subspace with evidential reasoning rule (RS^2_ER), which incorporates the feature regularizing module for identifying the discriminatory predictive power of multiple features and the probabilistic fusion module for enhancing the aggregation over base classifiers.

Since Sun and Li [46] proposed the concept of financial distress concept drift, an increasing number of dynamic ensemble models have been proposed for dynamic FDP. Sun et al. [47] proposed an ensemble mechanism based on combination of different data batches for base classifiers' dynamic construction and adaptive selection. Liu and Wu [27] carried out dynamic FDP modeling based on the hybrid use of incremental bagging and a genetic algorithm for treating financial distress concept drift. Sun et al. [40] integrated the time-weighting mechanism into the AdaBoost-SVM ensemble model for dynamic FDP with concept drift. In addition, class imbalance is another problem that should be considered in FDP modeling. Zhou [63] solved the problem of the biased distribution between distressed and nondistressed samples by random undersampling of the majority class or upsampling of the minority class. Veganzones and Séverin [56] indicated that the SMOTE approach that generates artificial data interpolated between existing minority samples is a good choice to serve as an efficient solution for a real bankruptcy prediction problem. In addition, Zoričák et al. [66] adopted one-class classification methods for class-imbalanced bankruptcy. Sun et al. [50] and Sun et al. [43] integrated SMOTE with bagging to build imbalance-oriented classifier ensembles for FDP, respectively, using SVM and DT as base classifiers. Su et al. [48] built an ensemble FDP model based on an AdaBoost-SVM model combined with SMOTE and time weighting, which treats both problems of concept drift and class imbalance in FDP. Shen et al. [37] proposed a dynamic FDP model, the adaptive neighbor SMOTE recursive ensemble approach (ANS-REA), which considers financial distress concept drift and allows for multiple forecast results from unbalanced data streams.

In addition to the above studies on two-class FDP, in which the sample is divided into two groups, i.e., the distressed and the nondistressed, or the bankrupt and the nonbankrupt, some literature also focused on multistate FDP. Lau [22] constructed a five-state FDP model by using the multinomial logistic analysis approach (MNLogit). Peel and Peel [33] also applied the MNLogit model to three-state FDP and found that it outperformed the MDA model in predicting private company failure. Altman et al. [5] adopted MDA to estimate two submodels working in sequence. The first model is used to discriminate between sound and unsound companies, and the second model is implemented after the first model has diagnosed a business as unsound and further discriminates between unsound and vulnerable companies. Jones and Hensher [16] applied the mixed logistic analysis approach to three-state FDP, which was superior to the MNLogit model according to FDP performance. Wilson et al. [59], Zurada et al. [67] and Agarwal et al. [1] performed NN simulations for three-state and four-state FDP and found that it was superior to the MNLogit model. Zhou et al. [64] used DT as the base classifier and explored a one-versus-one multiclass classification fusion using an optimizing decision-directed acyclic graph for predicting public companies' four-category listing status. Ultimately, research on SVM-based multiclass FDP has been neglected according to our knowledge. Given the impressive performance of SVMs in binary FDP, research on SVM-based multiclass FDP needs to be further explored.

3. Definition of multiclass financial states for Chinese listed companies

Most literature related to FDP modeling classified companies into two categories, i.e., distressed and nondistressed. They usually gave a definition of financial distress according to the institutional background of a certain country, e.g., legal bankruptcy, default of debt, inability to pay obligations, negative net assets, special treatment by stock exchanges, and so on [3,11,35,44]. Companies meeting predefined conditions are regarded as distressed, and the rest are considered nondistressed. Lau [22] categorized the financial states of American firms into financial stability and four states of increasing severity of financial distress, i.e., omitting or reducing dividend payments, technical default and default on loan payments, protection under Chapter X or XI of the Bankruptcy Act, and bankruptcy and liquidation. Peel and Peel [33] categorized company status into three types, i.e., failed companies, nonfailed profit-making companies and nonfailed loss-making companies. Wilson et al. [59] carried out FDP for three corporate outcomes, i.e., failed firms, distressed acquired firms and nonfailed firms. Zurada et al. [67] and Agarwal et al. [1] attempted to make FDP of four financial states, namely, healthy firms, firms that experienced dividend cuts/reduction, firms that experienced loan/debt default, and firms that filed for bankruptcy. Alam et al. [2] tried to identify potentially failing banks and used a sample composed of failed banks, extreme performance banks and control banks. Jones and Hensher [16] studied the multiclass FDP problem by categorizing Australian companies into three financial states: nonfailed firms, insolvent firms and firms filed for bankruptcy. Zhou et al. [64] tried to build multiclass classification models for the four listing states of Chinese listed companies in the Shanghai and Shenzhen Stock Exchange, i.e., normal status without any risk warning, abnormal status with other risk warning, abnormal status with delisting risk warning, and delisted status, and considered the latter three states as indications of financial distress. However, they found that it is very difficult to distinguish between the two abnormal statuses with other risk warnings and with delisting risk warnings.

Since corporate bankruptcy-related data are not public in China, studies on the FDP of Chinese listed companies almost define financial distress based on the risk warning mechanism defined in the Listing Rules of Shanghai or Shenzhen Stock Exchange. They usually apply the label of financial distress to firms specially treated as delisting risk warnings or other risk warnings and apply the label of financial nondistress to normal firms that have not been specially treated for a certain period. However, some companies labeled as financial nondistress in one year may fall in financial distress in the next year, which means its financial state is not sound enough and prone to easily deteriorate, although it is currently labeled financial nondistress. In contrast, companies labeled financial distress in one year may not necessarily suffer very serious financial difficulty. Some of them may only have temporary and mild financial distress and can quickly recover, and others may be truly in serious financial distress and even further deteriorate in future years, which may finally lead to stock suspension or delisting. Therefore, FDP models based on binary categorization are not refined enough to provide scientific decision support for corporate inside and outside stakeholders.

This study tries to build SVM-based multiclass FDP models by extending Chinese listed companies' financial status into four categories: financial soundness, financial pseudosoundness, moderate financial distress and serious financial distress. Suppose t_0 represents the standard year in which a listed company is labeled with a certain financial status. The four financial states are defined as follows:

(1) Class A: financial soundness, which means that the listed companies should have not been specially treated in both t_0 year and $t_0 + 1$ year.
(2) Class B: financial pseudosoundness, which means that the listed companies have not been specially treated in t_0 year but are specially treated in $t_0 + 1$ year.

(3) Class C: moderate financial distress, which means that the listed companies are specially treated in t_0 year but recover in $t_0 + 1$ year.
(4) Class D: serious financial distress, which means that the listed companies are specially treated in t_0 year and do not recover in $t_0 + 1$ year.

Among these four states, financial soundness represents a state of financial safety, and financial pseudosoundness, moderate financial distress and serious financial distress represent increasing degrees of financial distress, which need corporate managers and outside stakeholders to keep increasing levels of vigilance.

4. Multiclass FDP modeling based on SVMs combined with the decomposition and fusion methods

4.1. Overall framework

Fig. 1 shows the overall framework of multiclass FDP modeling based on SVMs combined with the decomposition and fusion methods.

The first step for FDP modeling is the collection of the initial training dataset, which means collecting sample companies and their feature data of the four financial states. We call it the initial training dataset, which is characterized by a class-imbalanced data distribution. To build FDP models capable of predicting financial distress two years ahead, we should collect the financial data of t_0-2 year for each sample company labeled with a financial status of t_0 year and use them as feature data for multiclass FDP modeling. Suppose the initial training dataset is denoted as Eq. (1):

$$\begin{aligned} IT_d &= \{(x_i, y_i)\}(i = 1, 2, \cdots, N) \\ &= \left\{\left(x_a^A, y_a^A\right)\right\} \cup \left\{\left(x_b^B, y_b^B\right)\right\} \cup \left\{\left(x_c^C, y_c^C\right)\right\} \cup \left\{\left(x_d^D, y_d^D\right)\right\} \\ &(a = 1, 2, \cdots, m; b = 1, 2, \cdots, n; c = 1, 2, \cdots, h; d = 1, 2, \cdots, z) \end{aligned} \tag{1}$$

where $y_i \in \{1, 2, 3, 4\}$ is the class label of financial status for sample company i, $y_a^A = 1, y_b^B = 2, y_c^C = 3, y_d^D = 4$, m, n, h and z are all positive integers and represent the sample numbers of class A (financial soundness), class B (financial pseudosoundness), class C (moderate financial distress) and class D (serious financial distress), respectively, and $N = m + n + h + z$ represents the total sample number of the four classes. As mentioned above, the sample distribution among the four financial states has the characteristic of class imbalance; hence, $m \neq n \neq h \neq z$, and they may even have large differences.

Traditionally, the SVM classification algorithm was developed for the class-balanced binary classification problem. On the one hand, the SVM model trained on a class-imbalanced training dataset is usually not able to effectively identify the minority class [51]. Therefore, an imbalanced data preprocessing mechanism needs to be designed to effectively reduce the imbalance degree among different financial states to prevent FDP models from biasing toward the majority class and lacking the capability of effectively predicting the minority class. After imbalanced data preprocessing, a modeling training dataset MT_d for multiclass FDP modeling is obtained.

On the other hand, since the traditional SVM model cannot be directly used for multiclass classification, it should be combined with certain kinds of decomposition and fusion methods. In the training stage, the multiclass FDP modeling problem is transformed into multiple two-class FDP modeling problems based on a certain decomposition method, and then the multiple binary SVM models are built by supervised training. In the testing stage, the multiple binary SVM models are combined to output multiclass FDP results based on a certain fusion method. The effectiveness of multiclass FDP models can be evaluated by comparing the predicted labels of testing samples with their true labels.

4.2. SMOTE data preprocessing mechanism

The traditional binary SVM model requires a relatively balanced number of training samples for the two classes. An SVM model trained directly on an imbalanced dataset usually has a classification hyperplane obviously shifting to the majority class, which increases the risk of misclassifying the minority class. For the multiclass FDP modeling problem, the sample distribution among the four financial states is seriously imbalanced. The number of companies in financial soundness far exceeds the numbers of financial pseudosoundness, moderate financial distress and serious financial distress. Namely, the class of financial soundness is the majority class, and the other three classes of financial pseudosoundness, moderate financial distress and serious financial distress are the minority classes. Therefore, to improve the prediction and recognition ability to classify the minority classes, an imbalanced data preprocessing mechanism is introduced before model training to weaken the class imbalance among different financial states. In this paper, the SMOTE approach is adopted as the class-imbalanced data preprocessing mechanism, which is a technique for oversampling the minority class by linear interpolation.

Take the minority class of moderate financial distress as an example. It is assumed that the number of samples to be interpolated is p times the number of original samples, and $x_c^C (c = 1, 2, \cdots, h)$ represents a sample of moderate financial distress. The SMOTE approach generates new synthetic samples of moderate financial distress by the following steps. First, it uses the Euclidean distance to find the k-nearest neighbors, $k \geq p$, of the minority samplex_c^C and then randomly selects p from the k-

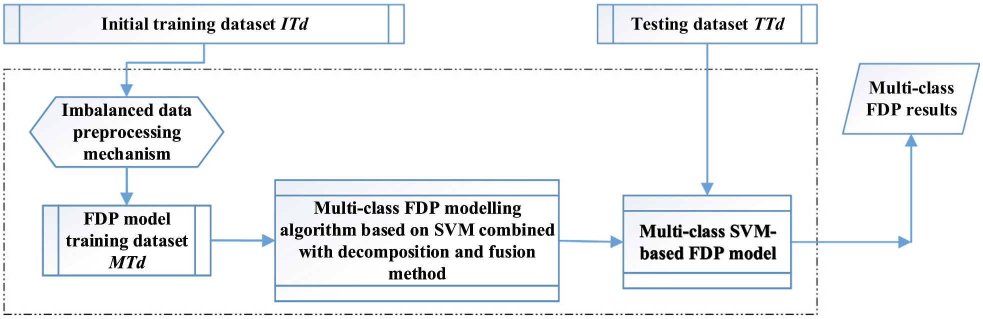

Fig. 1. Overall framework of multiclass FDP modeling.

nearest neighbors as the parent samples, i.e., $x'^{C}_{j}(j = 1, 2, \cdots, p)$. Then, according to Eq. (2), interpolation is performed between the original sample x^C_c and its p neighbors x'^{C}_{j} to generate p new synthetic samples $x^{newC}_{j}(j = 1, 2, \cdots, p)$.

$$x^{newC}_{j} = x^C_c + rand(0, 1) \times \left(x^C_c - x'^{C}_{j}\right) \tag{2}$$

4.3. Basic model of the SVM

The SVM is a supervised machine learning method for binary classification based on statistical learning theory proposed by Vapnik, and it seeks to improve the generalization ability of classifiers by structural risk minimization [55]. There are two cases according to whether the sample data are linearly separable or not. In the case where the sample data are linearly separable, training an SVM model proceeds by finding the optimal classification hyperplane: $w \cdot x + b = 0$, which can accurately distinguish the two classes and maximize the classification interval. When the sample data are linearly inseparable, a nonlinear function $\phi(x)$ is used to map the dataset of the input space into a feature space G, which is more easily linearly separable. We call $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$ the kernel function. In the feature space G, a linear maximum model is used to find the nonlinear maximum interval hyperplane: $w \cdot \phi(x) + b$. The support vectors are the training samples closest to the separating hyperplanes with maximum intervals. It is obvious that the samples for FDP modeling are linearly inseparable.

$$\begin{cases} w \cdot \phi(x_i) + b \geq +1 if y_i = +1 \\ w \cdot \phi(x_i) + b \leq -1 if y_i = -1 \end{cases} \tag{3}$$

$$\begin{cases} \quad min\left(\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i\right) \\ s.t. \begin{cases} y_i[w \cdot \phi(x_i) + b] \geq 1 - \xi_i \\ \xi_i \geq 0 (i = 1,\ 2,\ \cdots,\ N) \end{cases} \end{cases} \tag{4}$$

The binary SVM classifier should satisfy the conditions in Eq. (3), in which w is the weight vector and b is the offset vector. The interval between the separating hyperplanes of the two classes is $2/\|w\|$, so the problem of finding the optimal hyperplane can be transformed into the minimization of $\|w\|$ under the fixed empirical risk, which is represented as Eq. (4). ξ_i is a slack variable that is used to allow a small number of training samples to be misclassified, thus avoiding the overfitting problem. $C \in R^+$ indicates the penalty parameter, which is the degree of the penalty imposed when the training sample is misclassified. The SVM classifier obtained by solving the optimization problem is expressed as formula (5).

$$f(x) = \mathrm{sgn}(w \cdot \phi(x) + b) \tag{5}$$

4.4. Decomposition and fusion method for multiclass FDP modeling

Similar to the two-class FDP, the goal of multiclass FDP modeling is to train a classification model that can predict the financial status of a new sample. However, with the increase in classes, the training of the multiclass FDP model is more complicated. The traditional SVM algorithm applicable to two-class FDP modeling cannot be directly used for multiclass FDP modeling. Certain kinds of decomposition and fusion methods should be adopted to build multiple base SVM models for multiclass FDP. That is, the multiclass FDP problem is decomposed into multiple two-class FDP problems based on SVMs by a certain decomposition method, and the prediction results output by the base SVM models will be combined by a certain fusion method to output multiclass FDP results. It not only reduces the complexity of multiclass FDP modeling problems but also overcomes the limitations of the traditional SVM algorithm that cannot directly solve multiclass classification [13,53].

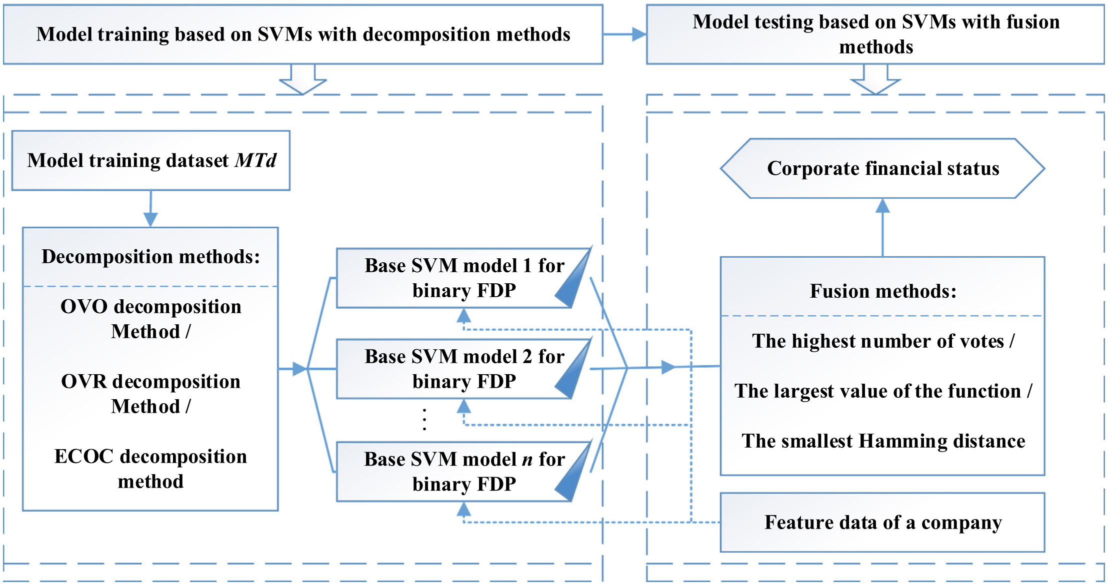

Fig. 2. Multiclass FDP modeling based on SVMs combined with decomposition and fusion methods.

The multiclass FDP modeling based on SVMs combined with decomposition and fusion methods is shown in Fig. 2, which is divided into two steps: model training based on SVMs with decomposition methods and model testing based on SVMs with fusion methods. SVM model training for multiclass FDP is performed on the model training dataset MT_d. Base SVM models for binary classification are trained based on decomposition methods such as one-versus-one (OVO), one-versus-rest (OVR), and error-correcting output coding (ECOC). For the three decomposition methods, the three fusion methods based on the highest number of votes, the largest value of the function, and the smallest Hamming distance are adopted. In the stage of model testing based on SVMs with fusion methods, the feature data of a company is input into each base SVM model for binary classification, and the outputs of the base SVM models are fused to obtain the final prediction result for multiclass FDP.

4.4.1. Multiclass FDP modeling based on OVO and SVMs

Suppose K is the number of classes; then, multiclass FDP modeling based on OVO and SVMs (OVO-SVMs) is implemented by training $K(K-1)/2$ base SVM models of binary FDP, each of which is trained on a training data subset composed of any two classes [21], and the binary SVM FDP model for discriminating class i and class j is expressed as Eq. (6).

$$f_{ij}(x) = w_{ij} \cdot \phi(x) + b_{ij}(1 \le i < j \le K) \tag{6}$$

For the four financial states represented as A (financial soundness), B (financial pseudosoundness), C (moderate financial distress) and D (serious financial distress), the OVO-SVM modeling framework for multiclass FDP is shown in Fig. 3. In the stage of OVO-SVM model training, six binary SVM models are built based on the decomposition method of OVO. That is, each pair of the four financial states, i.e., (A, B), (A, C), (A, D), (B, C), (B, D) and (C, D), are used as the positive and negative classes, respectively, and the corresponding samples' data subset is extracted from the model training dataset MT_d to construct a two-class training subset. In the stage of OVO-SVM model testing, the fusion method based on the highest number of votes is adopted to output the final FDP results. Let M_1, M_2, M_3 and M_4 denote the number of votes for financial status A, B, C and D, respectively, and the initial votes are all set as zero, namely, $M_1 = M_2 = M_3 = M_4 = 0$. Each binary SVM FDP model predicts the financial status for the enterprise with feature data of $\times$ and votes for the corresponding financial state. When a financial state obtains one vote from a binary SVM FDP model, its corresponding number of votes is incremented by one, as Eq. (7) shows. Finally, according to Eq. (8), the financial state with the highest number of votes is output as the final FDP result [30].

$$\begin{cases} M_1 = M_1 + 1 if one\ \ of\ \ the\ \ binary\ \ SVM\ \ models\ \ outputs\ \ A \\ M_2 = M_2 + 1 if one\ \ of\ \ the\ \ binary\ \ SVM\ \ models\ \ outputs\ \ B \\ M_3 = M_3 + 1 if one\ \ of\ \ the\ \ binary\ \ SVM\ \ models\ \ outputs\ \ C \\ M_4 = M_4 + 1 if one\ \ of\ \ the\ \ binary\ \ SVM\ \ models\ \ outputs\ \ D \end{cases} \tag{7}$$

$$F(x) = \arg\max(M_j)(j = 1,\ 2,\ 3, 4) \tag{8}$$

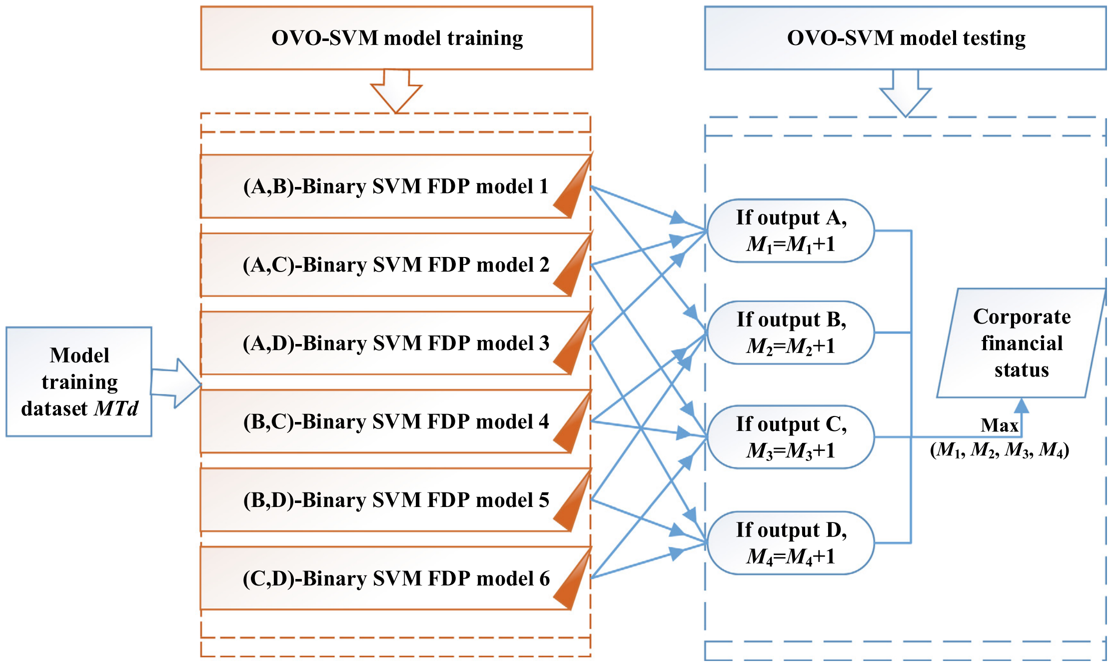

Fig. 3. OVO-SVM modeling framework for multiclass FDP.

4.4.2. Multiclass FDP modeling based on OVR and SVMs

Suppose K is the number of classes; then, multiclass FDP modeling based on OVR and SVMs (OVR-SVMs) is implemented by training K base SVM models of binary FDP, each of which is trained on the whole model training dataset MT_d by relabeling one of the K classes as the positive class and the remaining classes together as the negative class. Then, the K base SVM models of binary FDP are expressed in Eq. (9).

$$f_i(x) = w_i \cdot \phi(x) + b_i (i = 1, 2, \cdots, K) \tag{9}$$

For the four financial states represented as A, B, C and D, the OVR-SVM modeling framework for multiclass FDP is shown in Fig. 4. In the stage of OVR-SVM model training, four binary SVM models are built based on the decomposition method of OVR, i.e., A versus B + C + D, B versus A + C + D, C versus A + B + D and D versus A + B + C. In the stage of OVR-SVM model testing, the fusion method based on the largest value of the function is adopted to output the final FDP results. That is, each binary SVM model predicts the financial status for the enterprise with feature data of x according to the value of function $f_i(x)$, and a larger value of function $f_i(x)$ indicates a higher probability of being classified as class i, which corresponds to statuses A, B, C and D when i is equal to 1, 2, 3 and 4, respectively. According to the fusion method shown in Eq. (10), only the binary SVM FDP model with the highest value of function $f_i(x)$ is effective to output the final FDP result [28].

$$F(x) = \arg\max(f_j(x)) (j = 1,\ 2,\ \cdots,\ 4) \tag{10}$$

4.4.3. Multiclass FDP modeling based on ECOC and SVMs

The ECOC method uses the error-correcting coding idea to decompose the multiclass classification problem into multiple binary classification problems in supervised learning. In the encoding process, if each class can only be designated as positive or negative, it is called a binary ECOC code. If each class can be designated as positive, negative or deactivated, it is called a ternary ECOC code [10]. The framework of ECOC-SVM model for multiclass FDP is shown in Fig. 5, which is divided into two processes: encoding and decoding.

In the encoding process, each financial state is mapped to a label code, and the whole mapping scheme is recorded in a code matrix. That is, each financial state is designated as positive, negative or deactivated, respectively. Suppose there are r types of coding in total, in terms of which we can extract r training data subsets for binary FDP from the model training dataset MT_d and train r binary SVM FDP models. The encoding process constructs a code matrix with error correction capability and is represented as $V = [v_{k,j}]$ $(k = 1,\ 2,\ 3,\ 4;\ \ j = 1,\ 2,\ \cdots,\ r)$. The row number k represents an index of financial status, and the column number j represents an index of the coding scheme, which also corresponds to a binary SVM FDP model.

The code matrix V is also used to guide the decoding process of ECOC-SVM for multiclass FDP. First, each binary SVM FDP model predicts the financial status for the enterprise with feature data of x according to the value of function $f_i(x)$, and the

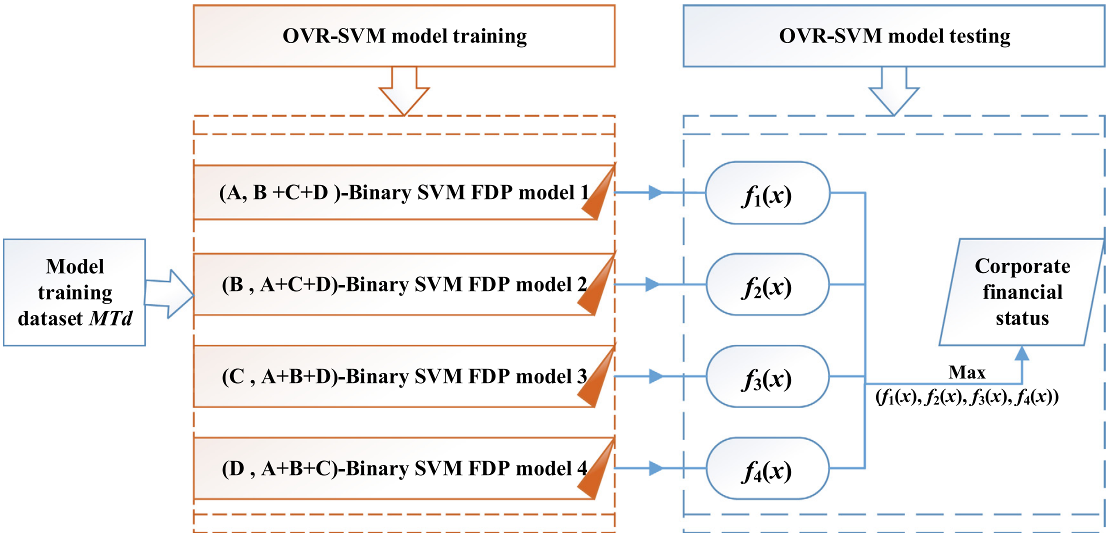

Fig. 4. OVR-SVM modeling framework for multiclass FDP.

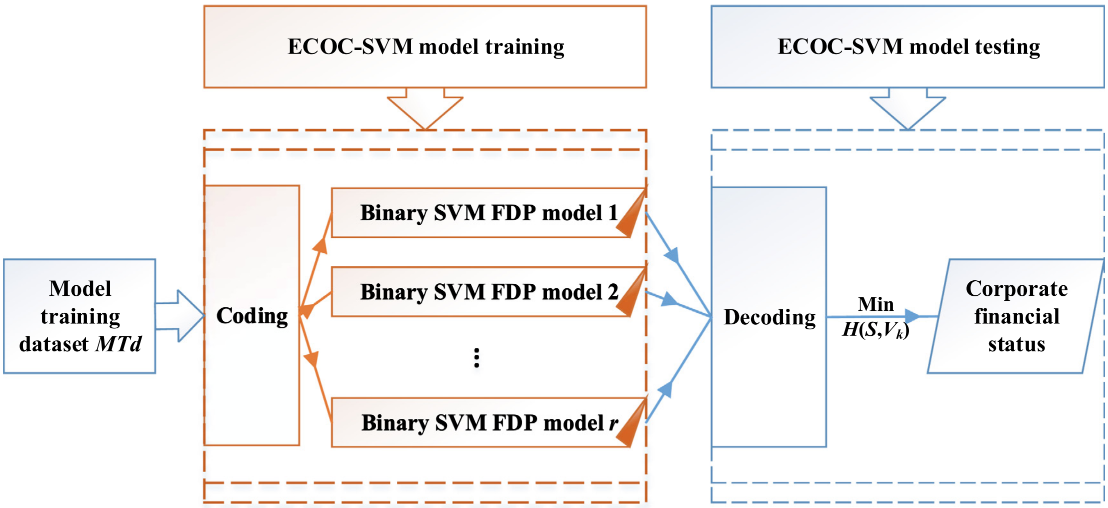

Fig. 5. ECOC-SVM modeling framework for multiclass FDP.

prediction results of all binary SVM FDP models are integrated to obtain a prediction code represented as $S = (s_1, s_2, \cdots, s_r)$. Second, according to Eq. (11), the Hamming distance between the prediction code S and each row in the code matrix V is calculated. Then, according to the fusion strategy based on the smallest Hamming distance, as shown in Eq. (12), we can find the row closest to the prediction code S in the code matrix V and the row index corresponds to the predicted financial status of multiclass FDP. The multiclass FDP model based on ECOC-SVM has some tolerance and correction capability for errors because the calculation of the Hamming distance is comprehensive. Therefore, in the decoding process, even if a prediction error occurs in an individual binary SVM FDP model, it does not have an excessive impact on the final prediction result of multiclass FDP.

$$H(S, V_k) = \sum_{j=1}^{r} \left\{ \frac{1}{2} (1 - \mathrm{sgn}(s_j \cdot v_{k,j})) \right\} \ (\mathrm{k} = 1,\ 2,\ 3,\ 4; j=1,\ \cdots,\ r) \tag{11}$$

$$F(x) = \mathrm{argmin}(H(S,\ V_k))(\mathrm{k} = 1,\ 2,\ 3,\ 4) \tag{12}$$

5. Empirical research

Empirical research is carried out to test the effectiveness of the above proposed multiclass FDP approaches based on the real-world data of Chinese listed companies. Four training datasets with different levels of class-imbalance preprocessing are used to train the multiclass FDP models, which are tested on the original class-imbalanced testing datasets. The performances of the above three multiclass FDP models based on SVMs integrated with decomposition and fusion methods are compared with each other as well as with the MDA and MNLogit models.

5.1. Sample and data

From the public companies listed in the Shanghai and Shenzhen Stock Exchange during 2006–2017, the samples of the four financial states are selected according to the definition stated in Section 3. To prevent a sample from having overlapping features of different financial states, each selected sample of the t_0 year, regardless of whether it belongs to classes A, B, C or D, should be a normal company without being specially treated before the t_0 year. The information about whether listed companies were specially treated or not comes from the China Wind Database, and the listed companies' financial indicator data are selected from the China Stock Market and Accounting Research (CSMAR) Database. In addition, we delete the sample companies that have missing data or noise data deviating from the mean by more than triple the standard deviation. The total number of ultimate samples is 2,893. Among them, the samples belonging to classes A, B, C and D are numbered as 2,235, 352, 143 and 163, respectively, as shown in Table 1.

5.2. Variable selection

We use a set of financial ratios as the input variables for FDP modeling because former literature has shown the usefulness of financial ratios in predicting financial distress or bankruptcy. As shown in Table 1, 49 financial ratios are selected as the candidate input variables, which cover proportion ratio, solvency, development capability, risk level, operational capability, cash flow capability and profitability. To establish multiclass FDP models for predicting financial status two years ahead, sample companies' t_0-2 year financial ratio data are used as the input variable values to predict the t_0 year financial status[1]. For example, when a company is labeled a certain class of financial status in 2017, the financial ratio data derived from the company's 2015 annual financial report are used as the feature values.

Using an independent samples t-test, we test the mean difference of each candidate financial ratio for each pair of two financial states, i.e., A&B, A&C, A&D, B&C, B&D and C&D, to select the financial ratios good at discriminating the two financial states. In this way, we first select six financial ratio subsets for A&B, A&C, A&D, B&C, B&D and C&D. Then, we take the intersection of the six financial ratio subsets to obtain the ultimate input variables for multiclass FDP, as shown in Table 2.

5.3. Experimental design

In the empirical research, experiments based on the multiclass FDP models of OVO-SVM, OVR-SVM, and ECOC-SVM are conducted 30 times. In each experiment, 70% of the sample is randomly selected from each class to compose a training dataset, and the testing dataset consists of the remaining 30% of the sample from each class. It is worth mentioning that we use the original class-imbalanced testing datasets to test the performance of the multiclass FDP models, which are built on the original class-imbalanced training datasets as well as the less class-imbalanced and nearly class-balanced training datasets preprocessed by certain mechanisms. More specifically, multiclass FDP experiments are carried out on the following four types of training datasets, and their sample numbers and proportions are listed in Table 3.

Original: This means that the original training datasets with a high degree of class imbalance and without preprocessing are used for model training.

Smote(Half): This means that no preprocessing is carried out for the class A sample, and the SMOTE approach is used to magnify the class B sample three times and to magnify the class C sample and class D sample seven times.

Smote(All): This means that no preprocessing is carried out for the class A sample, and the SMOTE approach is used to magnify the class B sample six times and to magnify the class C sample and class D sample fourteen times.

US&Smote: This means that the random undersampling method is adopted to extract 50% of the class A sample, and the SMOTE approach is used to magnify the class B sample three times and to magnify the class C sample and class D sample seven times.

For the *original* training dataset, the sample number of class A is far more than those of the other three classes, and the proportions of classes A, B, C and D are 77.17%, 12.18%, 4.98% and 5.67%, respectively, which is characterized by high class imbalance. After data preprocessing, the degree of class imbalance is alleviated. For *Smote(Half)* training datasets, the ratio between class A and each of the other three classes equals approximately 2:1. For a *Smote(All)* or *US&Smote* training dataset,

[1] For a Chinese listed company, whether it shall be specially treated in the t_0 year depends on its annual financial report of the t_0-1 year. Therefore, to provide multiclass FDP models with forecasting capability, we use sample companies' financial ratio data of the t_0-2 year as the input variable values to predict the financial status of the t_0 year.

Table 1
The sample numbers of training data and testing data.

Data type	Class A	Class B	Class C	Class D	Total
Training data	1565	247	101	115	2028
Testing data	670	105	42	48	865
Total	2235	352	143	163	2893

Table 2
The ultimate input variables for multi-class FDP.

Variable name	Financial ratio name	Calculation formula	Type
V1	Working capital ratio	(Current assets − Current liabilities)/Total assets	Proportion ratio
V2	Fixed assets ratio	Net fixed assets/Total assets	
V3	Operating liabilities ratio	Operating liabilities/Total liabilities	
V4	Interest coverage ratio	(Net profit + Financial expense)/Financial expense	Solvency
V5	Operating cash flow ratio	Net cash flow from operating activities/Current liabilities	
V6	Operating revenue growth rate	(Operating revenue of current year − Operating revenue of last year)/Operating revenue of last year	Development capability
V7	Financial leverage	(Net profit + Income tax expense + Financial expense)/(Net profit + Income tax expense)	Risk level
V8	Inventory turnover	Operating cost/Average balance of inventory	Operational capability
V9	Net cash flow to operating profit	Net cash flow from operating activities/Operating profit	Cash flow capability
V10	Net profit to total assets	Net profit/Average balance of total assets	Profitability
V11	Net profit to current assets	Net profit/Average balance of current assets	
V12	Return on invested capital	(Net profit + Financial expenses)/(Total assets − Current liabilities + Notes payable + Short-term borrowing + long-term liabilities due within one year)	
V13	Operating profit ratio	Operating profit/Operating revenue	

Table 3
The sample numbers and proportions of training datasets.

Training datasets		Class A	Class B	Class C	Class D	Total number
Original	Number	1565	247	101	115	2028
	Proportion	77.17%	12.18%	4.98%	5.67%	100%
Smote(Half)	Number	1565	741	707	805	3818
	Proportion	40.99%	19.41%	18.52%	21.08%	100%
Smote(All)	Number	1565	1482	1414	1610	6071
	Proportion	25.78%	24.41%	23.29%	26.52%	100%
US&Smote	Number	783	741	707	805	3036
	Proportion	25.79%	24.41%	23.29%	26.52%	100%

the sample number of each class is nearly equal, and the proportion of each class is approximately 25%, but the total sample number of a *Smote(All)* training dataset is nearly twice as many as that of a *US&Smote* training dataset.

We use the Matlab software to conduct the experiments. The radial basis kernel function is used for training the SVM models and the parameters are set as the default values in the SVM tool box. There are several coding schemes for the ECOC method in the Matlab software and the *Ordinal*[2] coding design is applied because of better performance than the others.

5.4. Empirical results and analysis

5.4.1. Effect of class imbalance on the proposed models' FDP performance

To demonstrate the effect of class imbalance on the multiclass FDP performance of the above three models, this study compares the average results of FDP performance measures for models built on training datasets with different levels of class imbalance, which are listed in Table 4.

[2] For the first binary learner, one class is negative and the rest are positive. For the second binary learner, the first two classes are negative and the rest are positive. For the third binary learner, the first three classes are negative and the rest are positive, and so on.

Table 4
The average results of FDP performance measures for models built on training datasets with different levels of class imbalance.

Models	Measures	*Original*	*Smote(Half)*	*Smote(All)*	*US&Smote*
OVO-SVM	Overall accuracy	**79.19%**	76.79%	<u>63.35%</u>	63.44%
	Class A accuracy	**99.74%**	89.82%	<u>61.53%</u>	62.57%
	Class B accuracy	<u>0.00%</u>	0.29%	**65.59%**	61.46%
	Class C accuracy	<u>0.00%</u>	66.75%	**75.16%**	70.87%
	Class D accuracy	<u>34.86%</u>	71.04%	**73.61%**	73.40%
	MacroR	<u>33.65%</u>	56.97%	**68.97%**	67.08%
OVR-SVM	Overall accuracy	**77.77%**	73.62%	69.04%	<u>66.34%</u>
	Class A accuracy	**97.58%**	82.24%	72.34%	<u>67.33%</u>
	Class B accuracy	<u>3.94%</u>	41.52%	52.98%	**59.02%**
	Class C accuracy	<u>2.86%</u>	36.27%	44.05%	**48.10%**
	Class D accuracy	<u>28.33%</u>	56.18%	79.93%	**84.51%**
	MacroR	<u>33.18%</u>	54.05%	62.33%	**64.74%**
ECOC-SVM	Overall accuracy	**77.53%**	71.30%	69.09%	<u>65.95%</u>
	Class A accuracy	**94.97%**	78.09%	73.24%	<u>67.23%</u>
	Class B accuracy	<u>9.56%</u>	50.98%	57.37%	**64.98%**
	Class C accuracy	<u>7.94%</u>	34.21%	44.21%	**58.17%**
	Class D accuracy	<u>43.75%</u>	53.40%	**58.61%**	56.88%
	MacroR	<u>39.05%</u>	54.17%	58.36%	**61.82%**

In Table 4, the overall accuracy is calculated by dividing the correctly classified testing sample number by the total testing sample number. The accuracy of a certain class is determined as the correctly classified testing sample number of the class divided by the total testing sample number of the class. MacroR is the average of single class accuracies. It is worth mentioning that each number in Table 4 is the mean value of the corresponding performance measures obtained in 30 experiments, a number in bold font represents the maximum value of a row, and a number underlined denotes the minimum value of a row.

As Table 4 shows, all three multiclass FDP models of OVO-SVM, OVR-SVM and ECOC-SVM demonstrate the highest overall accuracy and class A accuracy and the lowest class B/C/D accuracies and MacroR in column *Original*. This indicates that the multiclass FDP models based on SVMs integrated with decomposition and fusion methods tend to have very high accuracy for class A, very low accuracy for classes B and C, and relatively low accuracy for class D when the training dataset is highly class-imbalanced because the models are severely biased toward the majority class. Such an effect is particularly serious for the OVO-SVM model because it completely loses the capability of recognizing class B and class C, which are the two financial states between financial soundness and serious financial distress, and the ECOC-SVM model is relatively less affected by class imbalance than the OVO-SVM model and the OVR-SVM model.

The overall accuracy and class A accuracy in column *Smote(Half)* are much lower than those in column *Original*, and class B/C/D accuracies and MacroR in column *Smote(Half)* are much higher than those in column *Original*, except that the class B accuracy of OVO-SVM is still only 0.29%, at a much lower level than that of OVA-SVM or ECOC-SVM. Such results indicate that the SMOTE approach, which generates some synthetic samples of minority classes in the training dataset, can help the multiclass FDP models based on SVMs integrated with decomposition and fusion methods to improve FDP performance for minority classes by decreasing the level of model bias toward the majority class to some extent. Furthermore, the OVO-SVM model is more sensitive to class imbalance than the OVA-SVM model and the ECOC-SVM model.

The results in columns *Smote(All)* and *US&Smote* show evidently better multiclass FDP performance than those in columns *Original* and *Smote(Half)* for each of the three models, which provides explicit evidence that balancing the sample numbers of different classes in the training dataset by a certain data preprocessing mechanism is of paramount importance to guarantee the multiclass FDP performance for all three models based on SVMs integrated with decomposition and fusion methods. It is worth noting that the OVO-SVM model demonstrates better FDP performance in column *Smote(All)* than in column *US&Smote*, and the OVR-SVM model and the ECOC-SVM model both demonstrate better FDP performance in column *US&Smote* than in column *Smote(All)*. Therefore, the OVO-SVM model prefers the training data preprocessing mechanism of SMOTE for the minority classes, and the OVR-SVM model and the ECOC-SVM model prefer the integration of random undersampling for the majority class and SMOTE for the minority classes, in the condition that the degree of class imbalance does not exceed that of our dataset.

The above results also provide evidence that overall accuracy alone is not a good performance measure for class-imbalanced multiclass FDP because very high majority class accuracy accompanied by very low minority class accuracies may also result in a relatively high overall accuracy. Instead, MacroR is a better measure for evaluating multiclass FDP performance since a high MacroR is usually the result of the relatively high accuracy of each single class. Therefore, we only analyze each class's accuracy together with MacroR in the following text.

5.4.2. Comparison of multiclass FDP performance among the proposed models

Table 4 shows that regardless of column *Smote(All)* or *US&Smote*, the OVO-SVM model shows better multiclass FDP performance than the OVR-SVM model and the ECOC-SVM model. To make the comparison among the three models clearer, we further calculate the mean and standard deviation of each performance measure according to each model's best experimental results and list them in Table 5. Namely, we compare the multiclass FDP performance of the OVO-SVM models based on the *Smote(All)* training datasets and the OVR-SVM and ECOC-SVM models based on the *US&Smote* training datasets in this section. The values of each performance measure for each model obtained in 30 experiments are graphed as a curve in Fig. 6.

For class A (financial soundness), the OVR-SVM model has the highest average prediction accuracy of 67.33% and a relatively small standard deviation of 1.84%. The ECOC-SVM model's average class A accuracy is very similar to that of OVR-SVM, but its standard deviation is the largest. The OVO-SVM model has the lowest average prediction accuracy of 61.53% with the smallest standard deviation of 1.76%. For class B (financial pseudosoundness), the OVO-SVM model has the highest average prediction accuracy of 65.59% with the smallest standard deviation of 3.27%. The OVR-SVM model has the lowest average prediction accuracy of 59.02% and an intermediate standard deviation, and the ECOC-SVM model has an intermediate average prediction accuracy with the largest standard deviation of 4.98%. For class C (moderate financial distress), the OVO-SVM model also has the highest average prediction accuracy of 75.16% and the lowest standard deviation of 5.85%. The OVR-SVM model has the lowest average prediction accuracy of 48.10% and an intermediate standard deviation, and the ECOC-SVM model has an intermediate average prediction accuracy with the largest standard deviation of 12.04%. For class D (serious financial distress), the OVR-SVM model has the highest average prediction accuracy of 84.51% with the smallest standard deviation of 4.26%. The ECOC-SVM model has the lowest average prediction accuracy of 56.88% with the largest standard deviation of 6.14%. The OVO-SVM model has intermediate average prediction accuracy and standard deviation, which are much better than those of the ECOC-SVM model. As Fig. 6(a–d) show, the OVO-SVM curves for classes A, B, C and D are all located at similar heights with relatively small waves, showing relatively stable and satisfying accuracies for each class. However, the OVR-SVM model and the ECOC-SVM model both have some curves at relatively higher places with smaller waves and have some curves at very low places with large waves, e.g., the OVR-SVM curves for class B and class C and the ECOC-SVM curves for classes C and class D, showing uneven forecasting performance for the four classes.

According to the results of MacroR, the OVO-SVM model achieves the highest MacroR of 68.97%, showing the best comprehensive multiclass FDP performance among the three models. However, the OVR-SVM model and the ECOC-SVM model show much lower values of MacroR. On the whole, the OVO-SVM model shows the best overall capability for multiclass FDP because of the highest MacroR and the stable performance for each class. Although the OVO-SVM model shows the lowest class A accuracy among the three models, it is still at an acceptable level. In addition, a certain sacrifice of class A (financial soundness) accuracy is necessary and worthwhile for obtaining higher performance of forecasting the other three classes (certain degree of financial distress). In contrast, the OVR-SVM model and the ECOC-SVM model show much worse capability for multiclass FDP because of the relatively bad performance of MacroR and very poor performance in recognizing some kinds of financially distressed firms. Because failure in recognizing financially distressed firms usually makes stakeholders incur much higher economic losses than failure in recognizing financially sound firms, OVO-SVM is the best choice among the three models for multiclass FDP.

Since 30 experiments are conducted and the training sample and testing sample are randomly selected from each class according to the proportion of 70% and 30% in each experiment, Fig. 6 also shows the sensitivity of model performance to different training data and testing data. It is clear that the above conclusion about the performance comparison among the three multiclass FDP models almost holds for each experiment, indicating that the performance of the three multiclass FDP models are not sensitive to the change of training data and testing data.

5.4.3. Statistical test of performance comparison among the proposed models

To test whether the differences among the three models' multiclass FDP performance have statistical significance, this study further conducts a normal distribution test and mean comparison test for the best experimental results of each model.

Table 5
The mean and standard deviation of model performance measures.

Measures	Statistics	OVO-SVM	OVR-SVM	ECOC-SVM
Class A accuracy	Mean	<u>61.53%</u>	**67.33%**	67.23%
	Standard deviation	**1.76%**	1.84%	<u>2.08%</u>
Class B accuracy	Mean	**65.59%**	<u>59.02%</u>	64.98%
	Standard deviation	**3.27%**	4.76%	<u>4.98%</u>
Class C accuracy	Mean	**75.16%**	<u>48.10%</u>	58.17%
	Standard deviation	**5.85%**	8.33%	<u>12.04%</u>
Class D accuracy	Mean	73.61%	**84.51%**	<u>56.88%</u>
	Standard deviation	5.72%	**4.26%**	<u>6.14%</u>
MacroR	Mean	**68.97%**	64.74%	<u>61.82%</u>
	Standard deviation	**2.24%**	2.61%	<u>2.75%</u>

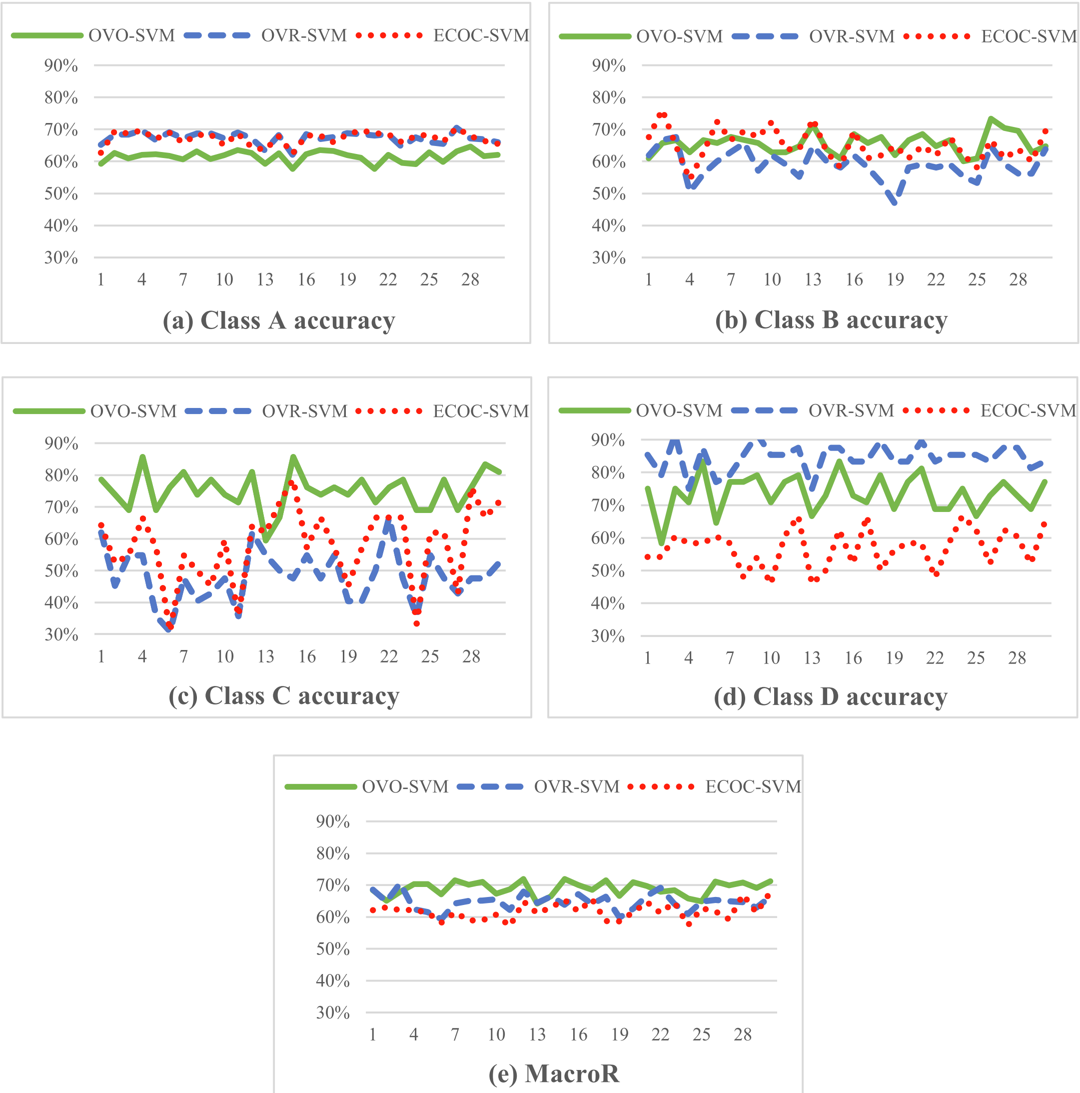

Fig. 6. The curves of multiclass FDP performance measures obtained in the 30 experiments based on the proposed models.

First, the Kolmogorov-Smirnov test is conducted to test whether the values of each performance measure for each model follow the normal distribution, and the results in Table 6 show that all the model performance measures follow the normal distribution.

Then, a paired-samples *t*-test of means comparison was carried out to compare the multiclass FDP performance for each pair of models, and the results are shown in Table 7. Except that the OVR-SVM model and the ECOC-SVM model do not have significant differences in class A accuracy, and the OVO-SVM model and the ECOC-SVM model do not have significant differences in class B accuracy, the other performance measures all show significantly different levels between each pair of models at the 1% significance level. Such results indicate that the above comparison of the multiclass FDP performance among the three models is further supported by the evidence of statistical tests.

As Table 8 and Table 9 show, there is no significant difference among the three models in terms of MacroR. According to the single class accuracy, MDA is overall outperformed by MNLogit, which is consistent with the findings of earlier literature [33]. OVO-SVM has significantly worse performance than MDA and MNLogit in forecasting class A (financial soundness) and class D (serious financial distress), but it demonstrates significantly better performance than MDA and MNLogit in predicting class B (financial pseudosoundness) and class C (moderate financial distress), which are more difficult and more important to be recognized in practice. Therefore, we argue that all three models have acceptable multiclass FDP performance, but MNLogit and OVO-SVM are preferred over MDA for multiclass FDP. We also consider the OVO-SVM model to be more competitive than MNLogit, given that financial pseudosoundness and moderate financial distress are very difficult to predict by human expertise, while financial soundness and serious financial distress may be much easier to recognize by human expertise.

Table 6
The results of Kolmogorov-Smirnov test for model performance measures.

Performance measure	Model	Significance level	Test result
Class A accuracy	OVO-SVM	0.430	Retain the null hypothesis
	OVR-SVM	0.636	Retain the null hypothesis
	ECOC-SVM	0.183	Retain the null hypothesis
Class B accuracy	OVO-SVM	0.905	Retain the null hypothesis
	OVR-SVM	0.939	Retain the null hypothesis
	ECOC-SVM	0.798	Retain the null hypothesis
Class C accuracy	OVO-SVM	0.869	Retain the null hypothesis
	OVR-SVM	0.563	Retain the null hypothesis
	ECOC-SVM	0.669	Retain the null hypothesis
Class D accuracy	OVO-SVM	0.710	Retain the null hypothesis
	OVR-SVM	0.226	Retain the null hypothesis
	ECOC-SVM	0.424	Retain the null hypothesis
MacroR	OVO-SVM	0.650	Retain the null hypothesis
	OVR-SVM	0.941	Retain the null hypothesis
	ECOC-SVM	0.891	Retain the null hypothesis

The null hypothesis: The values of an FDP performance measure for a certain model in 30 experiments follow the normal distribution.

Table 7
The *t*-test results for mean comparison of performance measures between each pair of models.

Performance measure	Model pairs	Difference	Significance level
Class A accuracy	OVO-SVM VS OVR-SVM	−5.80%	0.000***
	OVO-SVM VS ECOC-SVM	−5.70%	0.000***
	OVR-SVM VS ECOC-SVM	0.10%	0.635
Class B accuracy	OVO-SVM VS OVR-SVM	6.57%	0.000***
	OVO-SVM VS ECOC-SVM	0.61%	0.515
	OVR-SVM VS ECOC-SVM	−5.96%	0.000***
Class C accuracy	OVO-SVM VS OVR-SVM	27.06%	0.000***
	OVO-SVM VS ECOC-SVM	16.99%	0.000***
	OVR-SVM VS ECOC-SVM	−10.07%	0.000***
Class D accuracy	OVO-SVM VS OVR-SVM	−10.90%	0.000***
	OVO-SVM VS ECOC-SVM	16.73%	0.000***
	OVR-SVM VS ECOC-SVM	27.63%	0.000***
MacroR	OVO-SVM VS OVR-SVM	4.23%	0.000***
	OVO-SVM VS ECOC-SVM	7.15%	0.000***
	OVR-SVM VS ECOC-SVM	2.92%	0.000***

*** indicate the significant level of 1% and * indicate the significant level of 10%.

Table 8
The average values of multi-class FDP performance measures for MDA, MNLogit and OVO-SVM.

Measure	MDA	MNLogit	OVO-SVM
Class A accuracy	68.87%	**69.08%**	<u>61.53%</u>
Class B accuracy	<u>61.90%</u>	63.65%	**65.59%**
Class C accuracy	68.73%	66.90%	**75.16%**
Class D accuracy	75.42%	**76.46%**	73.61%
MacroR	68.73%	**69.02%**	68.97%

6. Conclusion

Binary FDP models can only recognize a firm to be distressed or nondistressed, which is not refined enough. Multiclass FDP models can further help financial managers improve the precision of financial risk management. Consequently, financial managers can be expected to find it much easier to take more targeted measures to prevent serious financial distress because multiclass FDP models can predict more precise labels of multiple financial states. In addition, an effective multiclass FDP model is important not only for corporate financial managers but also for other stakeholders, such as creditors, shareholders, potential investors and regulators. This paper categorizes corporate financial status into four states: A (financial soundness), B (financial pseudosoundness), C (moderate financial distress) and D (serious financial distress). Financial soundness is a relatively safe financial status, and the latter three represent increasing levels of financial distress, about which management should monitor and take precautions. We build three multiclass FDP models by integrating SVMs with the three decomposition and fusion methods of OVO, OVR and ECOC, which transfer the problem of multiclass FDP modeling into multiple

Table 9
The mean comparison results of FDP performance measure among MDA, MNLogit and OVO-SVM.

Measure	Statistics	MDA VS MNLogit	MDA VS OVO-SVM	MNLogit VS OVO-SVM
Overall accuracy	Mean difference	−0.35%	5.03%	5.37%
	Significance level	0.058*	0.000***	0.000***
Class A accuracy	Mean difference	0.21%	7.34%	7.55%
	Significance level	0.244	0.000***	0.000***
Class B accuracy	Mean difference	−1.75%	−3.68%	−1.94%
	Significance level	0.000***	0.000***	0.009***
Class C accuracy	Mean difference	1.83%	−6.43%	−8.25%
	Significance level	0.053*	0.000***	0.000***
Class D accuracy	Mean difference	−1.04%	1.81%	2.85%
	Significance level	0.030**	0.003***	0.000***
MacroR	Mean difference	−0.29%	−0.24%	0.05%
	Significance level	0.304	0.491	0.906

problems of binary FDP modeling. We carry out empirical research to test the performance of multiclass FDP models based on the data of Chinese listed companies during 2006–2017, in which the proportions of the four financial states are 77.17%, 12.18%, 4.98% and 5.67%, showing a high degree of class imbalance. Empirical results show that the three multiclass FDP models of OVO-SVM, OVR-SVM and ECOC-SVM are very sensitive to the class imbalance of training datasets, and the data-level preprocessing mechanisms that make training datasets nearly class-balanced can greatly improve their multiclass FDP performance. Comparative analysis indicates that the OVO-SVM model outperforms the OVR-SVM model and the ECOC-SVM model for multiclass FDP, which is statistically significant by the *t*-test of means comparison. Furthermore, the OVO-SVM model is preferred over MDA and MNLogit for multiclass FDP because of higher accuracies for financial pseudosoundness and moderate financial distress, which are very difficult to predict by human expertise.

The results indicating that OVO-SVM outperforms OVR-SVM and ECOC-SVM may be partially attributed to the feature selection procedure, which is based on the candidate features' discriminant power with respect to pairwise classification (A&B, A&C, A&D, B&C, B&D and C&D). In the future, different feature selection procedures may be designed for different decomposition and fusion methods. In addition, the decomposition and fusion methods can also be integrated with other classification algorithms to propose more effective multiclass FDP models.

CRediT authorship contribution statement

Jie Sun: Conceptualization, Methodology, Software, Writing - original draft. **Hamido Fujita:** Supervision, Methodology, Formal analysis. **Yujiao Zheng:** Resources, Investigation. **Wenguo Ai:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research is supported by the National Natural Science Foundation of China [Grant numbers 71771162].

References

[1] A. Agarwal, J.T. Davis, T. Ward, Supporting ordinal four-state classification decisions using neural networks, Inf. Technol. Manage. 2 (1) (2001) 5–26.
[2] P. Alam, D. Booth, K. Lee, T. Thordarson, The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: An experimental study, Expert Syst. Appl. 18 (3) (2000) 185–199.
[3] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, J. Finance 23 (4) (1968) 589–609.
[4] E.I. Altman, R.G. Haldeman, P. Narayanan, Zeta analysis: a new model to identify bankruptcy risk of corporations, J. Bank. Finance 1 (1) (1977) 29–54.
[5] E.I. Altman, G. Marco, F. Varetto, Corporate distress diagnosis: comparisons using linear discriminant analysis and neural network – The Italian experience, J. Bank. Finance 18 (3) (1994) 505–529.
[6] B. Back, T. Laitinen, K. Sere, Neural networks and genetic algorithms for bankruptcy predictions, Expert Syst. Appl. 11 (1996) 407–413.
[7] J.K. Bae, Predicting financial distress of the South Korean manufacturing industries, Expert Syst. Appl. 39 (10) (2012) 9159–9165.
[8] A. Chaudhuri, K. De, Fuzzy support vector machine for bankruptcy prediction, Appl. Soft Comput. 11 (2) (2011) 2472–2486.
[9] C.H. Cheng, C.P. Chan, Y.J. Sheu, A novel purity-based k nearest neighbors imputation method and its application in financial distress prediction, Eng. Appl. Artif. Intell. 81 (2019) 283–299.
[10] T. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, J. Artif. Intell. Res. 2 (1995) 263–286.
[11] M. Doumpos, C. Zopounidis, A multinational discrimination method for the prediction of financial distress: the case of Greece, Multinational Finance J. 3 (1999) 71–101.
[12] S. Fallahpour, E.N. Lakvan, M.H. Zadeh, Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem, J. Retail. Consum. Serv. 34 (1) (2017) 159–167.
[13] B. Fei, J. Liu, Binary tree of SVM: A new fast multiclass training and classification algorithm, IEEE Trans. Neural Networks 17 (3) (2006) 696–704.

[14] H. Frydman, E.I. Altman, D. Kao, Introducing recursive partitioning for financial classification: The case of financial distress, J. Finance 40 (1985) 269–291.
[15] S.B. Jabeur, Bankruptcy prediction using partial least squares logistic regression, J. Retail. Consum. Serv. 36 (2017) 197–202.
[16] S. Jones, D.A. Hensher, Predicting firm financial distress: A mixed logit model, Account. Rev. 79 (4) (2004) 1011–1038.
[17] R. Geng, I. Bose, X. Chen, Prediction of financial distress: An empirical study of listed Chinese companies using data mining, Eur. J. Oper. Res. 241 (2) (2015) 236–247.
[18] Y.P. Huang, M.F. Yen, A new perspective of performance comparison among machine learning algorithms for financial distress prediction, Appl. Soft Comput. 83 (2019) 105663.
[19] M. Kim, D. Kang, Ensemble with neural networks for bankruptcy prediction, Expert Syst. Appl. 37 (2010) 3373–3379.
[20] S.Y. Kim, A. Upneja, Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models, Econ. Model. 36 (2014) 354–362.
[21] Kressel, U. (1999). Pairwise classification and support vector machines. In: Scholkopf, B., Burges, C. and Smola, A. (Eds.), Advances in kernel methods: Support vector learning. MIT Press, Cambridge, 255–268.
[22] A.H.L. Lau, A five-state financial distress prediction model, J. Account. Res. 25 (1) (1987) 127–138.
[23] H. Li, Y. Lee, Y. Zhou, J. Sun, The random subspace binary logit (RSBL) model for bankruptcy prediction, Knowl.-Based Syst. 24 (8) (2011) 1380–1388.
[24] T.-H. Lin, A cross model study of corporate financial distress prediction in Taiwan: multiple discriminant analysis, logit, probit and neural networks models, Neurocomputing. 72 (16–18) (2009) 3507–3516.
[25] F. Lin, D. Liang, C.C. Yeh, J.C. Huang, Novel feature selection methods to financial distress prediction, Expert Syst. Appl. 41 (5) (2014) 2472–2483.
[26] F. Lin, C.C. Yeh, M.Y. Lee, The use of hybrid manifold learning and support vector machines in the prediction of business failure, Knowl.-Based Syst. 24 (1) (2011) 95–101.
[27] J. Liu, C. Wu, Dynamic forecasting of financial distress: the hybrid use of incremental bagging and genetic algorithm—empirical study of Chinese listed corporations, Risk Manag. 19 (1) (2017) 32–52.
[28] W. Maass, On the computational power of winner-take-all, Neural Comput. 12 (2000) 2519–2535.
[29] T.E. McKee, M. Greenstein, Predicting bankruptcy using recursive partitioning and a realistically proportioned data se, J. Forecast. 19 (2000) 219–230.
[30] E. Montanes, J. Barranquero, J. Diez, J.J. Coz, Enhancing directed binary trees for multi-class classification, Inf. Sci. 223 (20) (2013) 42–55.
[31] N. Mselmi, A. Lahiani, T. Hamza, Financial distress prediction: The case of French small and medium-sized firms, Int. Rev. Finan. Anal. 50 (2) (2017) 67–80.
[32] J.A. Ohlson, Financial ratios and the probabilistic prediction of bankruptcy, J. Account. Res. 18 (1) (1980) 109–131.
[33] M.J. Peel, D.A. Peel, Some further empirical evidence on predicting private company failure, Account. Business Res. 18 (69) (1987) 57–66.
[34] F.M. Rafiei, S.M. Manzari, S. Bostanian, Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence, Expert Syst. Appl. 38 (8) (2011) 10210–10217.
[35] S.A. Ross, R.W. Westerfield, J.F. Jaffe, Corporate Finance, 6th Edition., McGraw-Hill, Irwin, 2002, pp. 859–860.
[36] C. Serrano-Cinca, B. Gutiérrez-Nieto, Partial least square discriminant analysis for bankruptcy prediction, Decis. Support Syst. 54 (3) (2013) 1245–1255.
[37] F. Shen, Y. Liu, R. Wang, W. Zhou, A dynamic financial distress forecast model with multiple forecast results under unbalanced data environment, Knowl.-Based Syst. 192 (2020) 1–16.
[38] K.-S. Shin, Y.-J. Lee, A genetic algorithm application in bankruptcy prediction modeling, Expert Syst. Appl. 23 (2002) 321–328.
[39] K.-S. Shin, T.S. Lee, H.-J. Kim, An application of support vector machines in bankruptcy prediction model, Expert Syst. Appl. 28 (2005) 127–135.
[40] J. Sun, H. Fujita, P. Chen, H. Li, Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble, Knowl.-Based Syst. 120 (2017) 4–14.
[41] J. Sun, X.-F. Hui, Financial distress prediction based on similarity weighted voting CBR, Lecture Notes Artif. Intell. 4093 (2006) 947–958.
[42] J. Sun, K. He, H. Li, SFFS-PC-NN optimized by genetic algorithm for dynamic prediction of financial distress with longitudinal data streams, Knowl.-Based Syst. 24 (2011) 1013–1023.
[43] J. Sun, J. Lang, H. Fujita, H. Li, Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates, Inf. Sci. 425 (2018) 76–91.
[44] J. Sun, H. Li, Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers, Expert Syst. Appl. 35 (3) (2008) 818–827.
[45] J. Sun, H. Li, Data mining method for listed companies' financial distress prediction, Knowl.-Based Syst. 21 (1) (2008) 1–5.
[46] J. Sun, H. Li, Dynamic financial distress prediction using instance selection for the disposal of concept drift, Expert Syst. Appl. 38 (2011) 2566–2576.
[47] J. Sun, H. Li, H. Adeli, Concept drift-oriented adaptive and dynamic support vector machine ensemble with time window in corporate financial risk prediction, IEEE Trans. Systems Man Cybernet. Systems 43 (4) (2013) 801–813.
[48] J. Sun, H. Li, H. Fujita, B. Fu, W. Ai, Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting, Inf. Fusion 54 (2020) 128–144.
[49] J. Sun, H. Li, Q.H. Huang, 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. 57 (2014) 41–56.
[50] J. Sun, Z. Shang, H. Li, Imbalance-oriented SVM methods for financial distress prediction: a comparative study among the new SB-SVM-ensemble method and traditional methods, J. Operat. Res. Soc. 65 (2014) 1905–1919.
[51] Y. Sun, A.K.C. Wong, M.S. Kamel, Classification of imbalanced data: a review, Int. J. Pattern Recognit Artif Intell. 23 (4) (2009) 687–719.
[52] K. Tam, Neural network models and the prediction of bank bankruptcy, Omega 19 (5) (1991) 429–445.
[53] D. Tomar, S. Agarwal, A comparison on multi-class classification methods based on least squares twin support vector machine, Knowl.-Based Syst. 81 (2015) 131–147.
[54] F. Tseng, L. Lin, A quadratic interval logit model for forecasting bankruptcy, Omega 33 (1) (2005) 85–91.
[55] V. Vapnik, Statistical Learning Theory, John Wiley, New York, 1998.
[56] D. Veganzones, E. Séverin, An investigation of bankruptcy prediction in imbalanced datasets, Decis. Support Syst. 112 (2018) 111–124.
[57] G. Wang, G. Chen, Y. Chu, A new random subspace method incorporating sentiment and textual information for financial distress prediction, Electron. Commer. Res. Appl. 29 (2018) 30–49.
[58] G. Wang, J. Ma, G. Chen, Y. Yang, Financial distress prediction: Regularized sparse-based Random Subspace with ER aggregation rule incorporating textual disclosures, Appl. Soft Comput. J. 90 (2020) 1–18.
[59] N. Wilson, K.S. Chong, M.J. Peel, Neural network simulation and the prediction of corporate outcomes: Some empirical findings, Int. J. Econ. Business 2 (1) (1995) 31–50.
[60] C.-C. Yeh, D.-J. Chi, M.-F. Hsu, A hybrid approach of DEA, rough set and support vector machines for business failure prediction, Expert Syst. Appl. 37 (2) (2010) 1535–1541.
[61] X. Zhang, L. Hu, A nonlinear subspace multiple kernel learning for financial distress prediction of Chinese listed companies, Neurocomputing 177 (12) (2016) 636–642.
[62] H. Zhang, F. Yang, Y. Li, H. Li, Predicting profitability of listed construction companies based on principal component analysis and support vector machine—Evidence from China, Autom. Constr. 53 (2015) 22–28.
[63] L. Zhou, Performance of corporate bankruptcy prediction models on imbalanced dataset: the effect of sampling method, Knowl.-Based Syst. 41 (2013) 16–25.
[64] L. Zhou, Q. Wang, H. Fujita, One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies, Information Fusion 36 (2017) 80–89.
[65] M.E. Zmijewski, Methodological issues related to the estimation of financial distress prediction models, J. Account. Res. 22 (1984) 59–82.

[66] M. Zoričák, P. Gnip, P. Drotár, V. Gazda, Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets, Econ. Model. 84 (2020) 165–176.
[67] J.M. Zurada, B.P. Foster, T.J. Ward, R.M. Barker, Neural networks versus Logit regression models for predicting financial distress response variables, J. Appl. Bus. Res. 15 (1) (1998) 21–29.