Journal of Business Research 120 (2020) 137–146

Contents lists available at ScienceDirect

Journal of Business Research

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

Combining corporate governance indicators with stacking ensembles for financial distress prediction

Deron Liang[a], Chih-Fong Tsai[b,*], Hung-Yuan (Richard) Lu[c], Li-Shin Chang[a]

[a] Department of Computer Science and Information Engineering, National Central University, Taiwan
[b] Department of Information Management, National Central University, Taiwan
[c] Mihaylo College of Business and Economics, California State University-Fullerton, USA

ARTICLE INFO

Keywords:
Data mining
Financial distress prediction
Bankruptcy prediction
Corporate governance indicators
Stacking ensembles

ABSTRACT

In this paper, we use a stacking ensemble to construct a bankruptcy prediction model. We collect a comprehensive list of 40 financial ratios (FRs) and 21 corporate governance indicators (CGIs) for US companies, and conduct two experiments. In the first, we utilize all FRs and CGIs to build our model. Our results show that this model does not perform significantly better than the baseline models. In the second experiment, we use 6 specific FRs and 6 specific CGIs selected by a stepwise discriminant analysis to construct another model. We find that this model performs better than the baseline models, and exhibits strong performance when the costs of misclassifying bankruptcy companies are high.

1. Introduction

Predicting corporate bankruptcy has long been an important research topic because bankruptcies impose tremendous costs on market participants, as well as the economy as a whole. Consequently, shareholders, corporations, and financial institutions are all interested in models of financial distress prediction (FDP) or bankruptcy prediction to help identify troubled firms in early stages of distress (Kumar & Ravi, 2007; Lin, Hu, & Tsai, 2012; Olson, Delen, & Meng, 2012; Sun, Li, Huang, & He, 2014).

In the past, FDP methods have largely utilized statistical models. For instance, Fitzpatrick (1932) compared 13 financial ratios (FRs) of bankrupt and normal companies. Subsequently, other FDP methods such as univariate analysis (Beaver, 1966), discriminant analysis (Altman, 1968), and logit analysis (Ohlson, 1980) have been used. Among them, the model developed by Altman (1968), which uses a linear combination of five FRs, provides good prediction accuracy; it has been widely used in the literature to proxy for bankruptcy risk.

In the 1990s, studies reported that machine learning techniques, such as artificial neural networks, can provide better results than statistical ones, such as discriminant analysis and logit analysis, in terms of prediction accuracy (Lacher, Coats, Sharma, & Fant, 1995; Lee, Han, & Kwon, 1996; Serrano-Cinca, 1997). Since then, many studies started using various machine learning techniques to predict financial distress, such as support vector machines, k-nearest neighbor, decision trees, and neural networks (Barboza, Kimura, & Altman, 2017; Lin et al., 2012; Olson et al., 2012; Tsai, Hsu, & Yen, 2014). Among them, ensemble learning methods, which combine multiple classification techniques, have shown superior prediction performance over single classification techniques (du Jardin, 2016; Fallahpour, Lakvan, & Zadeh, 2017; Liang, Tsai, Dai, & Eberle, 2018; Tsai, 2014; Tsai et al., 2014; Volkov, Benoit, & Van den Poel, 2017; Zieba, Tomczak, & Tomczak, 2016). Specifically, stacking ensemble performs better than various classifier ensemble methods, such as bagging and boosting in different applications (Graczyk, Lasota, Trawinski, & Trawinski, 2010; Kim, 2018; Liang et al., 2018; Opitz & Maclin, 1999; Patil, Aghav, & Sareen, 2016; Zenko, Todorovski, & Dzeroski, 2001; Xia, Liu, Da, & Xie, 2018).

While a stacking ensemble is a widely used machine-learning technique, its application in FDP has not been not fully explored. First, it is unclear whether a stacking ensemble would perform better or worse than other ensemble methods in predicting bankruptcy. For example, Kim (2018) and Liang et al. (2018) show that stacking ensembles outperform bagging and boosting ensembles. Kim (2018) uses a sample of US hospitality firms, and Liang et al. (2018) uses Australian and German credit datasets, as well as bankruptcy datasets from China and Taiwan. In contrast, Pisula (2020) shows that boosting ensembles outperform bagging and stacking ensembles using a Polish bankruptcy dataset.

Second, the bankruptcy prediction models developed in the literature mostly use only financial ratios as predictors. Other variables, such

* Corresponding author at: Department of Information Management, National Central University, Taiwan.
E-mail address: cftsai@mgt.ncu.edu.tw (C.-F. Tsai).

https://doi.org/10.1016/j.jbusres.2020.07.052
Received 21 December 2019; Received in revised form 22 July 2020; Accepted 24 July 2020

as corporate governance indicators (CGIs), may provide additional power in bankruptcy prediction (Chaganti, Mahajan, & Sharma, 1985; Daily, 1994, 1996; Daily & Dalton, 1994). Corporate governance includes the mechanisms, processes and relations by which corporations are controlled and directed (Shailer, 2004). Effective corporate governance allows shareholders to exercise appropriate oversight of a company to maximize firm value and ensure that it generates a return on their holdings (Chen, 2014). The lack of proper corporate governance mechanisms may allow managers to take on excess risks and increase the risks of bankruptcy.

In recent literature, several papers have used CGIs to build bankruptcy prediction models, but the capabilities of CGIs in predicting bankruptcy have not been fully explored. Bredart (2014) and Platt and Platt (2012) use both FRs and CGIs to build their bankruptcy prediction models, but the number of CGIs used in their studies are limited. Liang, Lu, Tsai, and Shih (2016) include a comprehensive set of CGIs for their analysis, such as board structure, ownership structure, key person retained, deviation, etc., but their studies focus on Taiwanese and Chinese firms. In this study, we use a comprehensive list of FRs and CGIs retrieved from U.S companies, and examine the effect of combining financial ratios and CGIs on the prediction performance of stacking ensembles. Our study aims to provide unbiased insights on the role of CGIs in bankruptcy prediction as we use a relative recent sample (1996–2014) that includes both large and small U.S. firms.

We contribute to the corporate governance literature in two ways. First, by manually reading annual reports and proxy statements, we collect some novel corporate governance indicators. For example, C_{18} is a variable equal to one when the candidates for the board of directors recommended by management or the nominating committee are not elected to the board, and zero otherwise. This variable is collected by comparing the candidates from the proxy statements with the elected board members from the annual reports. Under normal circumstances, candidates recommended by the nominating committees would be elected to the board. In rare cases, when the nominated candidates are not elected, we consider the firm as having tension within the board or conflicts between shareholders and management.

Second, we use recent corporate governance data to conduct our experiments. Our sample starts in 1996 and ends in 2014, during which time the United States went through significant changes in corporate governance regime. For example, Section 302 of the Sarbanes-Oxley (SOX, hereafter) Act of 2002 requires public-traded companies to assess the effectiveness of the firm's internal controls, and makes directors and officers personally liable for the accuracy of financial statements. Section 404 of SOX further requires auditors to attest to management's assessment of the company's internal controls and Section 407 requires disclosure of financial experts in the audit committee. These regulations made significant changes to companies' corporate governance practices and our results from using the recent data may provide a fresh perspective to academics and policy makers.

This study also contributes to the machine learning literature in two ways. First, we use a comprehensive list of FRs and CGIs to build a bankruptcy prediction model and provide evidence on the usefulness of stacking ensemble techniques. Second, we use feature selection to identify specific FRs and CGIs with high relevance. We test whether including only selected FRs and CGIs in our model performs better than when we use a comprehensive set of predictors.

The rest of this paper is organized as follows. Section 2 reviews the literature and describes classifier ensembles and feature selection techniques. Section 3 introduces the stacking ensemble approach and experimental setup. Section 4 presents the experimental results and Section 5 concludes the paper.

2. Literature review

2.1. Variables

2.1.1. Financial ratios

The objective of financial distress prediction (FDP) is to predict whether a debtor (i.e. individuals or companies) will become bankrupt or not. In the past, the commonly used methods to predict financial distress or bankruptcy were based on statistical analysis, particularly univariate analysis (Beaver, 1966).

Multivariate analysis was first used for the FDP problem by Altman (1968), who created the Z-score. The Z-score is a linear combination of five FRs, weighted by coefficients:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5 \quad (1)$$

where X_1 (working capital/total assets) measures liquidity, X_2 (retained earnings/total asset) measures cumulative profitability, X_3 (earnings before interest and taxes/total assets) measures operating efficiency, X_4 (market value of equity/book value of total liabilities) measures market-based leverage, and X_5 (sales/total assets) measures asset turnover.

2.1.2. Corporate governance indicators

Corporate governance refers to mechanisms instituted by capital providers or the government to ensure that the suppliers of finance receive returns on their investment in firms run by professional managers or entrepreneurs (Shleifer & Vishny, 1997). In legal or business environments that lack sound corporate governance mechanisms, capital providers can be hesitant to invest in firms and managers are more likely to make decisions that are detrimental to the firm for private gain. In some cases, this may lead to financial distress.

Prior studies have shown that corporate governance is positively associated with firm value (Gompers, Ishii, & Metrick, 2003). More specifically, corporate governance has been shown to be positively associated with cash holdings of a company (Harford et al., 2008). This idea posited in Harford et al. is closely related to our hypothesized relation between corporate governance and bankruptcy. They argue that when the economy is booming, managers are tempted to over-invest cash reserves to expand the firm at the expense of flexibility during economic recessions. In firms with sound corporate governance mechanisms, managers are likely to be restricted in their ability to invest the free cash flow (Jensen, 1986; Stulz, 1990), leading to higher cash reserves and a lower probability of bankruptcy during recessions.

In the prior literature, corporate governance indicators (CGIs) have been used to analyze financial distress of US (Bredart, 2014; Platt & Platt, 2012) and Taiwanese firms (Lee & Yeh, 2004; Liang et al., 2016; Lin, Liang, & Chu, 2010). In particular, Lee and Yeh (2004) and Bredart (2014) show that using some specific CGIs can make a logistic regression model perform significantly better, in terms of prediction accuracy, than models using some specific FRs. Platt and Platt (2012) analyze the differences between bankrupt and non-bankrupt firms based on mean comparisons of related CGIs and FRs. They find that some CGIs are significantly related to corporate success, which can effectively distinguish healthy firms from those that fail. Other studies also demonstrate that a combination of FRs and CGIs allows different machine learning-based prediction models, such as support vector machines, k-nearest neighbor, naïve Bayes, decision trees, and neural networks to perform better than using FRs or CGIs alone (Lin et al., 2010; Liang et al., 2016).

Following prior literature, we classify corporate governance into three categories: board structures (Collier & Esteban, 1999; Yeh & Woidtke, 2005; Liang, Xu, & Jiraporn, 2013), ownership structures (Berkman, Cole, & Fu, 2009; Cheung, Chung, Tan, & Wang, 2013; Jian & Wong, 2010; La Porta, Lopez-de-Silanes, & Shleifer, 1999), and others (Liang et al., 2016). Board structures include variables such as the percentage of inside/independent/gray directors, compensation committee members, and audit committee members. Ownership structures include variables such as the percentage of shares held by the board of inside/outside directors, supervisors, and ultimate controller (through individual, unlisted company, and a juridical person). Other CGIs may include seats controlled by the ultimate controller, the number of times financial forecasts are released in a year, and number of financial restatements in a year.

2.2. Ensemble learning techniques

Classifier ensembles are based on combining multiple classifiers, and aim to produce better classification results than single classifiers. In particular, a number of classifiers are trained and their outputs are combined for a final decision (Rokach, 2010). Two widely applied methods to construct classifier ensembles are bagging and boosting. In bagging, data sampling is performed over the original training set to train a fixed number of classifiers. Boosting focuses on adjusting the weights of misclassified training samples in an iterative manner to minimize expected error over different input data distributions, and the final classifier aggregates the learned classifiers by voting (Bauer & Kohavi, 1999; Wozniak, Grana, & Corchado, 2014).

In the bankruptcy prediction literature, a number of studies have demonstrated that combining multiple classifiers can provide better prediction accuracy than single classifiers (du Jardin, 2016; Fallahpour et al., 2017; Tsai, 2014; Tsai et al., 2014; Volkov et al., 2017; Zieba et al., 2016).

Besides bagging and boosting, stacking is another type of ensemble learning technique, which was proposed by Wolpert (1992). In general, stacking ensembles are based on a two-level architecture. The first level constructs a number of different classifiers (i.e. base learners) whose outputs are used to train the second level classifier (i.e. meta learner) for the final prediction result. Machine learning studies have shown that stacking ensembles outperform bagging and boosting methods (Graczyk et al., 2010; Opitz & Maclin, 1999; Zenko et al., 2001).

Recently, stacking ensemble techniques have been widely applied in FDP problems, including credit scoring (Armaki, Fallah, Alborzi, & Mohammadzadeh, 2017; Patil et al., 2016; Wei, Yang, Zhang, & Zhang, 2019; Xia et al., 2018; Zhang, He, & Zhang, 2019) and bankruptcy prediction (Kim, 2018; Liang et al., 2018; Pisula, 2020). However, they only consider financial ratios to construct stacking ensembles. In other words, it is unknown whether combining financial ratios with corporate governance indicators can improve the prediction performance of stacking ensembles.

2.3. Research hypotheses

The main purpose of this study is to examine whether CGIs can provide incremental power in predicting bankruptcy for a comprehensive sample of U.S firms. Prior literature has shed some light on this issue and found that specific CGIs or combining CGIs and FRs can allow some well-known prediction models to provide better prediction accuracies than those based on FRs alone (Lee & Yeh, 2004; Liang et al., 2016; Lin et al., 2010). Most prior studies use data from other countries or have a limited number of CGIs. Our focus is on U.S. firms which recently have gone through tremendous regulatory reforms in corporate governance. In addition, prior studies examining the effectiveness of different specific CGIs use single learning-based prediction models. The stacking ensemble technique, an ensemble learning-based model which has shown its superiority over single learning-based techniques, has not been fully explored in this research topic.

In sum, we state our first hypothesis in the alternative form as follows.

H1. A bankruptcy prediction model that incorporates both CGIs and FRs more accurately predicts bankruptcy than models that include only FRs.

Specifically, our first hypothesis tests whether combining the five well-known FRs from Altman (1968) Z-score (which has been recognized as a standard) with the 21 CGIs can enhance the prediction accuracy of stacking ensembles. To test this hypothesis, the baseline prediction model for performance comparison is based on the single learning technique using the FRs alone.

In this study, we collect 61 indicators, consisting of 40 FRs and 21 CGIs. Including such a large number of input variables in our model may not be optimal as some of the variables may increase noise, rather than accuracy. Therefore, to improve the performance of our model, we use a feature selection filter to remove non-representative features (i.e. input variables) (Liang, Tsai, & Wu, 2015).

In sum, we state our second hypothesis in the alternative form as follows.

H2. A bankruptcy prediction model that incorporates a parsimonious set of CGIs and FRs, identified from feature selections, more accurately predicts bankruptcy than models that include a comprehensive set of CGIs and FRs.

Specifically, our second hypothesis tests whether using a parsimonious set of CGIs and FRs can enhance the prediction performance of stacking ensembles. To test this hypothesis, we first perform three well-known feature selection algorithms since different algorithms can produce different selection results over the same dataset. Next, we use the features that produce the best prediction performance to run our model and the results are compared with the baseline model, which is based on stacking ensembles trained by the combined 61 indicators.

3. The experimental procedure

3.1. Stacking ensembles

Fig. 1 shows the proposed stacking ensembles for bankruptcy prediction. For the first level of classifiers, i.e. the base learners, two classifiers are constructed based on using FRs alone and a combination of FRs and CGIs (i.e. FRs + CGIs). A common strategy for diversifying classifier ensembles is to use different types of input features to individually construct different classifiers, which can be subsequently combined in a later stage (Opitz & Maclin, 1999; Rokach, 2010). However, this strategy for constructing stacking ensembles has not been considered in bankruptcy prediction. Third base learners have not been constructed using CGIs alone because they will not perform well unless they are combined with FRs.

For the second level of classifier (i.e. the meta learner), the input

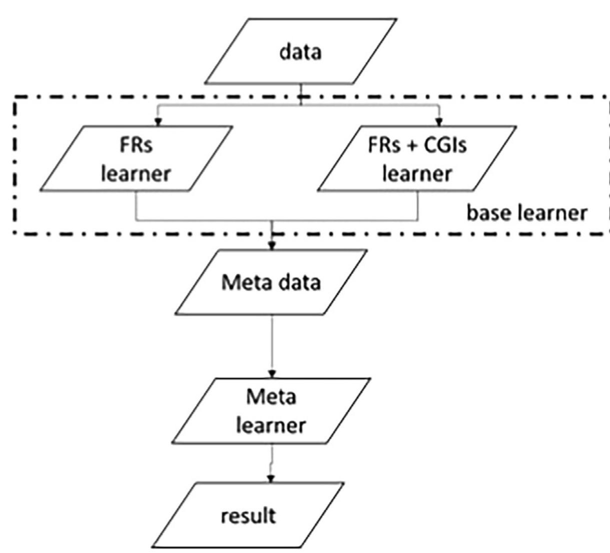

Fig. 1. The proposed stacking ensembles.

features to train the meta learner are based on the outputs produced by the base learners. More specifically, the probability of being bankrupt is used as the input feature. For example, given a training case, which is inputted into the two base learners, suppose that the FRs and FRs + CGIs based classifiers predict that the input case has a 75% and a 51% probability of going bankrupt, respectively. Then, 0.75 and 0.51 are the input features, whereas the original class label (i.e. either bankrupt or not) of the case is the final prediction output for training the meta learner.

Notably, we use support vector machines (SVM) by the linear kernel function to construct the base and meta learners. This is because it has been widely used as one of the representative baseline classifiers in many related works, regardless of whether the studies aimed to propose novel approaches (Fallahpour et al., 2017; Liang et al., 2018; Sun, Fujita, Chen, & Li, 2017; Sun, Li, Fujita, Fu, & Ai, 2020; Wei et al., 2019; Wu, Xiao, Dang, Yang, & Yang, 2014; Xia et al., 2018; Zhang et al., 2019; Zieba et al., 2016) or compare existing techniques (Armaki et al., 2017; Barboza et al., 2017; Graczyk et al., 2010; Kim, 2018; Liang et al., 2016; Lin et al., 2010; Olson et al., 2012; Patil et al., 2016; Sun et al., 2014; Tsai et al., 2014; Volkov et al., 2017; Zhou, Lu, & Fujita, 2015).

In addition, the 10-fold cross validation method is used to divide the dataset into 90% and 10% training and testing sets, respectively (Kohavi, 1995). In the related literature mentioned above, there are many studies that use the 10-fold cross validation method to train and test their prediction models, including Armaki et al. (2017), Fallahpour et al. (2017), Graczyk et al. (2010), Kim (2018), Liang et al. (2016), Liang et al. (2018), Lin et al. (2010), Olson et al. (2012), Tsai et al. (2014), Wu et al. (2014), Zhang et al. (2019), Zieba et al. (2016).

3.2. The dataset and input variables

We collected a list of bankrupt companies from the UCLA-LoPucki Bankruptcy Research Dataset, which can be purchased from the website: https://lopucki.law.ucla.edu/. The sample period spans from 1996 to 2014. Bankruptcies are relatively rare events and a comprehensive list of bankrupt firms improves the reliability of our study. Even though the UCLA-LoPucki database is not open access, it has more bankrupt firms than other databases, such as Compustat. It also has more complete data on the characteristics of the bankruptcies. For the period of 1996–2014, there are 764 bankrupt firms available from the database. This database has been used in a large number of published works and is well-accepted in bankruptcy prediction research.[1]

Using Central Index Keys (CIK) or company names, we obtain financial data from the Compustat database. We construct our control (non-bankrupt) sample by matching each bankrupt firm with a non-bankrupt control firm within the same four-digit SIC industry group. We also require the difference in total assets between the pair to be less than 50% of the larger firm. If we are not able to find a match within the four-digit SIC code, we use three digit-SIC and then two digit-SIC to conduct our search.

We used the Institutional Shareholder Services (ISS) database to begin collecting the CGI variables. However, since ISS only covers firms included in the S&P 1500 index, we are only able to obtain CGIs for 69 sample firms from this database. As a set of 69 firms is too small for us to conduct a meaningful experiment, we hand-collected the CGI variables from companies' proxy statements (DEF-14a) and annual reports (10-k) through EDGAR (Electronic Data Gathering and Retrieval) from the United States Securities and Exchange Commission's website (www.sec.gov). We identify each company by searching partial company names identified from the UCLA-LoPucki and Compustat databases. We collected most of the CGI variables from proxy statements and annual reports. Our sample consists of 286 US firms, of which 143 firms are bankrupt and the other 143 firms are not bankrupt.

The list of 40 financial variables and 21 CGIs collected for our sample firms are shown in Tables 1 and 2, respectively. We provide the descriptive statistics of these variables in Appendix A.

As shown in Column 2 of Table 1, we classify the financial ratios into three categories: Liquidity, Leverage (or Solvency), and Profitability. Liquidity is defined as how easily an asset can be converted into cash. Since cash is the only medium acceptable for satisfying liabilities, liquidity ratios measure a company's ability to pay off its short-term obligations. The second set of ratios is leverage ratios. A company is generally financed by either equity or debt. Leverage refers to the amount of capital financed by debt in relation to the amount financed by equity. As capital structures are considered long-term decisions, leverage ratios measure a company's ability to satisfy its long-term obligations. The last set of ratios is profitability ratios, which measures a company's ability to generate profits from its operations to satisfy the cost of debt.

In our study, we first include the 5 FRs used in Altman (1968). We then survey the literature to obtain a comprehensive list of 34 widely used FRs in other FDP studies (Bauweraerts, 2016; Charitou, Neophytou, & Charalambous, 2004; Kim, Jo, & Shin, 2016; Liang et al., 2016; Tian & Yu, 2017; Wu et al., 2014). We then use feature selection to identify the representative FRs that are of higher relevance in distinguishing bankrupt from non-bankrupt firms.

We largely follow the prior literature to identify CGIs to be included our study. As shown in Table 2, we classified the CGIs into three categories: board structure, ownership structure, and board conflict. The first category, board structure, includes CGIs that are related to the composition and the independence of the board and its committees. We follow prior literature and collect 14 CGIs in this category (Collier & Esteban, 1999; Yeh & Woidtke, 2005; Liang et al., 2013). The second category, ownership structure, include CGIs that are related to ownership stakes of inside/outside directors and majority shareholders. We follow prior literature and collect 8 CGIs in this category (Berkman et al., 2009; Cheung et al., 2013; Jian & Wong, 2010; La Porta et al., 1999). The last category, board conflict, includes one novel variable that has not been examined in prior literature. C_{18} is a variable equal to one when the candidates for the board of directors recommended by management or the nominating committee are not elected to the board, and zero otherwise. This variable is collected by comparing the candidates from the proxy statements with the elected board members from the annual reports. Normally, candidates recommended by the nominating committees are elected to the board if the board, management, and major stakeholders are congruent. In rare cases when the nominated candidates are not elected, we consider the firm as having tensions or conflicts within the board, or between shareholders and management.

For each of our variables, the feature values are normalized to range between 0 and 1 using the following equation:

$$\forall\, x \in F, \quad normalize(x) = \frac{\mathrm{x} - \min(\mathrm{F})}{\max(F) - \min(F)} \tag{2}$$

where F means the set of all feature values in a specific variable, x for the feature value to be normalized, and max(F) and min(F) are the maximum and minimum values of F, respectively.

3.3. The evaluation metrics

To evaluate prediction performance, the detection error tradeoff (DET) curve and misclassification cost are used. The DET curve is used for detection tasks that involve a tradeoff of two error types, which are missed detections and false alarms (Martin, Doddington, Kamm, Ordowski, & Przybocki, 1997). For the FDP problem, a type I error means that the model misclassifies a bankrupt company as a normal company. In contrast, a type II error means that the model misclassifies a normal company as a bankrupt company. Since type I and II errors are

[1] The list of published works is available at: https://lopucki.law.ucla.edu/published_research.htm.

Table 1
The 40 FRs.

Variables	Categories	Descriptions
Z_1	Altman	Working capital/total assets
Z_2	Altman	Retained earnings/total assets
Z_3	Altman	Earnings before interest and taxes/total assets
Z_4	Altman	Market value of equity/book value of total liabilities
Z_5	Altman	sales/total assets
Z_6	Altman	Z_SCORE
FR_1	Liquidity	(Current Assets)/(Current Liabilities)
FR_2	Liquidity	(Current Assets)/(Total Assets)
FR_3	Liquidity	(Current Assets-Inventory)/(Total Assets)
FR_4	Liquidity	Quick Ratio
FR_5	Leverage	(Current Liabilities)/(Total Assets)
FR_6	Leverage	(Financial Debt)/(CashFlow)
FR_7	Liquidity	(Cash + Mark.Sec.)/(Total Sales)
FR_8	Liquidity	(Cash + Mark.Sec.)/(Total Assets)
FR_9	Profitability	EBITDA/(Total Sales)
FR_{10}	Liquidity	Cash/(Current Liabilities)
FR_{11}	Liquidity	Cash/(Total Assets)
FR_{12}	Liquidity	Cash/(Total Debt)
FR_{13}	Leverage	(Shareholder Funds)/(Total Assets)
FR_{14}	Leverage	(Long Term Debt)/(Shareholder Funds)
FR_{15}	Leverage	(Long Term Debt)/(Total Assets)
FR_{16}	Leverage	(Total Debt)/(Shareholder Funds)
FR_{17}	Leverage	(Total Debt)/(Total Assets)
FR_{18}	Profitability	EBITDA/(Total Assets)
FR_{19}	Profitability	(Profit before Tax)/(Shareholder Funds)
FR_{20}	Profitability	(Net Income)/(Shareholder Funds)
FR_{21}	Profitability	(Net Income)/(Total Assets)
FR_{22}	Profitability	(Total Sales)/(Shareholder Funds)
FR_{23}	Profitability	(Operating CashFlow)/(Total Assets)
FR_{24}	Profitability	(Operating CashFlow)/(Total Sales)
FR_{25}	Profitability	EBIT/(Total Sales)
FR_{26}	Liquidity	(Current Assets)/(Total Sales)
FR_{27}	Liquidity	(Net Op.Work.Capital)/(Total Sales)
FR_{28}	Profitability	(Accounts Receivable)/(Total Sales)
FR_{29}	Profitability	(Accounts Payable)/(Total Sales)
FR_{30}	Profitability	Inventory/(Total Sales)
FR_{31}	Liquidity	Cash/(Total Sales)
FR_{32}	Leverage	Change in Other Debts
FR_{33}	Leverage	Change in Equity Position
FR_{34}	Leverage	(Financial Expenses)/(Total Sales)

Table 2
The 21 CGIs.

Variables	Descriptions
Board structure	
C_1	Number of directors
C_2	Number of inside directors
C_3	Number of IND directors
C_4	Number of gray directors
C_5	C_3/C_1
C_6	C_2/C_1
C_7	C_4/C_1
C_8	Number of Compensation members
C_9	The IND director number in compensation members/C_8
C_{10}	Number of Audit members
C_{11}	The IND director number in audit members/C_{10}
C_{14}	=1 if the board is staggered
Ownership structure	
C_{12}	inside director total ownership%
C_{13}	outside director total ownership%
C_{15}	=1 if ultimate controller > 50%
C_{16}	=1 if ultimate controller > 30%
C_{17}	=1 if the company issues special shares
C_{19}	Total shareholding of audit members%/total shareholding%
C_{20}	Total shareholding of compensation members%/total shareholding%
C_{21}	Total shareholding of external directors that is larger than 5%
Board Conflict	
C_{18}	=1 if the board of directors reported on proxy statement and annual report do not match, and 0 otherwise.

mutually exclusive, the DET curve for type I and II errors can be produced based on adjusting the threshold values (Liang et al., 2016, 2018).

On the other hand, misclassification costs can be used to solve the limitation of the DET curve when two models exhibit similar performance. The misclassification cost can be measured by fixing the cost ratios to understand model performance, as well as quantify the DET curve result (Liang et al., 2016, 2018). Specifically, misclassification costs can be obtained by the following equation:

$$(Type\ \ I\ \ error \times \text{bankrupt firms} \times \text{cost ratio}) + (Type\ \ II\ \ error \times \text{non bankrupt firms}) \tag{3}$$

In this study, we define cost ratios as the cost of a type I error (misclassifying bankrupt companies as non-bankrupt companies) relative to the cost of a type II errors (misclassifying non-bankrupt companies as bankrupt companies). Since different users may use the prediction model in different ways, evaluating model performance over a range of cost ratios may provide additional insights.

For example, an ordinary equity investor who is using the model to make investment decisions may have lower cost ratios since the maximum he/she can lose is the initial investment if the firm goes bankrupt. But for banks that make loans or auditors who provide attestation services, the cost ratios may be much higher because these parties may suffer tremendous losses in the event of a default. In this study, we fix the cost ratios at 1, 1.5, 2, 3, 5, 7.5, 10, 12.5, and 15.[2]

After obtaining the misclassification costs across different cost ratios, we use the Wilcoxon test (Demsar, 2006) to compare the performance of two models, where $p < 0.05$ indicates the difference is statistically significant.

4. Experimental results

4.1. Study one for hypothesis one

The first experimental study aims to examine the first hypothesis (c.f. Section 2.3) that combining the five FRs from Altman's Z-score and 21 CGIs can improve the prediction performance of the stacking ensembles. Fig. 2 shows the DET curves of stacking ensembles by combining FRs and CGIs (i.e. M1) and the baseline by FRs alone (i.e. M0). Since we are only concerned about Type I errors, we only show results ranging from 0 to 0.25. Table 3 shows the results of testing the difference in misclassification costs between stacking ensembles and the baseline model, using a Wilcoxon test.

According to the results of the DET curves and misclassification costs, using FRs based on the Z-score alone make the prediction model perform better than the model using the full combination of FRs and CGIs. Particularly, no matter what the cost ratio is, the baseline model always significantly outperforms the stacking ensembles ($p < 0.05$). In other words, adding the features of CGIs does not improve the performance of the prediction model. As a result, our first hypothesis, which suggests that the 21 CGIs are useful for bankruptcy prediction, is rejected.

However, in Liang et al. (2016), stepwise discriminant analysis (SDA) is used in the feature selection step to examine the discrimination power of combining FRs with different types of CGIs for bankruptcy

[2] We choose the list of cost ratios by considering the use of a bankruptcy prediction model by two groups: equity investors and debt investors. For equity investors, the cost ratio of one would be reasonable because the cost of Type I error (buying the wrong firm) would be roughly the same as the cost of Type II error (shorting the wrong firm) in a well-diversified portfolio. For debt investors, the cost ratios would be significantly large than one because the cost of type I error (the loan amount lost when a firm defaults) is generally many times larger than the cost of type II error (the interest forfeited by not lending to a non-default firm).

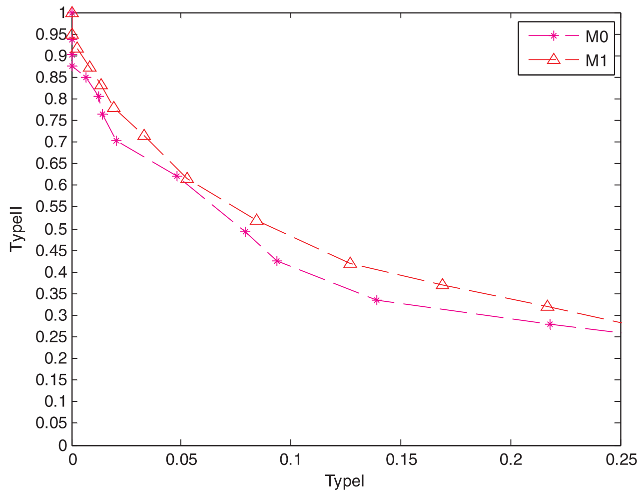

Fig. 2. The DET curves of the baseline (M0) and stacking ensembles (M1).

Table 3
The level of significant difference between the baseline and stacking ensembles in terms of misclassification costs.

Cost	Baseline	Stacking ensembles
1	1	0.044
1.5	1	0.006
2	1	0.022
3	1	0.043
5	1	0.001
7	1	0.001
10	1	0.000
12.5	1	0.000
15	1	0.000

prediction. They find that using specific CGIs, i.e. board structure and ownership structure, can significantly impact model performance. Therefore, it is possible that combining FRs with select CGIs may improve model performance.

4.2. Study two for hypothesis two

To investigate hypothesis two, feature selection is performed over the combined FRs and CGIs to select representative features. In this paper, three feature selection methods are compared, which are *t*-test, SDA, and stepwise logistic regression (SLR). In addition to the baseline model constructed in the previous study, the OR ensembles proposed by Liang et al. (2018) are also compared to see whether stacking ensembles can perform better than the others.

First, we examine which feature selection method performs better than the others. Tables 4 and 5 show the misclassification costs under different feature selection methods for the features of FRs and combined FRs and CGIs, respectively. Note that 'All' means the result without performing feature selection.

Table 4
The level of significant difference between *t*-test, SDA, and SLR for FRs in terms of misclassification costs.

Cost	All	*t*-Test	SDA	SLR
1	0.735	1	0.256	0.217
1.5	0.832	1	0.177	0.339
2	0.872	1	0.336	0.089
3	0.442	1	0.401	0.015
5	0.292	1	0.479	0.157
7	1	0.502	0.002	0.594
10	1	0.285	0.003	0.594
12.5	1	0.001	0.003	0.594
15	0.001	1	0.002	0.000

Table 5
The level of significant difference between *t*-test, SDA, and SLR for FRs + CGIs in terms of misclassification costs.

Cost	All	*t*-Test	SDA	SLR
1	0.023	0.001	1	0.005
1.5	0.048	0.238	1	0.133
2	0.657	1	0.731	0.132
3	0.219	1	0.116	0.006
5	0.717	1	0.301	0.020
7	0.830	0.634	1	0.006
10	0.000	0.665	1	0.000
12.5	0.000	0.003	1	0.000
15	0.000	0.003	1	0.000

These results indicate that in most cases (i.e. different cost ratios), a *t*-test performs the best for the features of FRs (c.f. Table 4), whereas SDA performs the best for the features of FRs + CGIs (c.f. Table 5). This finding is consistent with Liang et al. (2016).

Next, we compare the prediction performances of the baseline by performing *t*-test over FRs (denoted as M0), stacking ensembles by performing SDA over FRs + CGIs (denoted as M1), and OR ensembles by performing SDA over FRs + CGIs (denoted as M2). Fig. 3 shows the DET curves of the baseline, stacking ensembles, and OR ensembles. In addition, Table 6 shows the results of testing the differences in misclassification costs of different classifiers, using a Wilcoxon test.

These results show that when the cost ratios are larger than 7 (i.e. 10, 12.5, and 15), stacking ensembles perform significantly better than the other two classifiers, and OR ensembles outperform the baseline. On the other hand, there is no clear winner between baseline and stacking ensemble when the cost ratio is smaller than 7 (i.e. 1, 1.5, 2, 3, and 5). However, when the cost ratios are larger than 7, stacking ensembles outperform the baseline (c.f. Table 7).

To compare stacking and OR ensembles, Table 8 shows the misclassification costs of these two ensembles. It can be seen that OR ensembles perform better than stacking ensembles in some cases, i.e. the cost ratios are 2 and 3. However, on average, stacking ensembles are the better choice than OR ensembles.

These results suggest that combining specific FRs and CGIs can effectively improve the prediction performance of stacking ensembles. Table 9 lists the selected features of FRs and CGIs by using SDA. Note that the negative feature weight value means that it is positively related to normal companies, whereas the positive feature weight value is

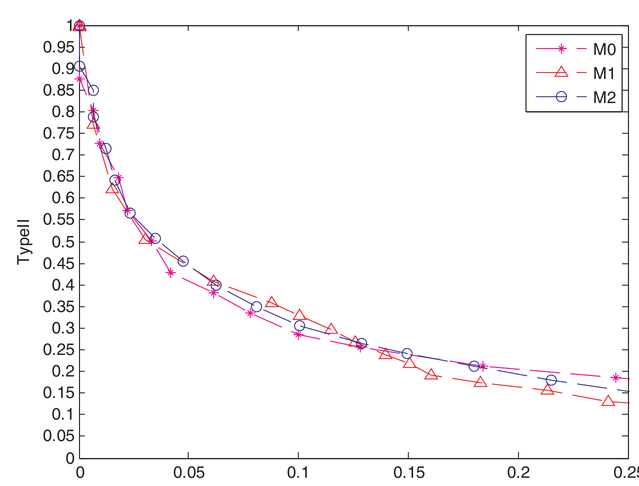

Fig. 3. The DET curves of the baseline (M0), stacking ensembles (M1), and OR ensembles (M2).

Table 6
The level of significant difference between the baseline, stacking ensembles, and OR ensembles in terms of misclassification costs.

Cost	Baseline	Stacking ensembles	OR ensembles
1	0.058	1	0.003
1.5	0.726	1	0.127
2	1	0.509	0.237
3	1	0.100	0.104
5	1	0.124	0.001
7	0.141	1	0.080
10	0.452	1	0.220
12.5	0.001	1	0.007
15	0.001	1	0.007

Table 7
The level of significant difference between the baseline and stacking ensembles.

Cost	Baseline	Stacking ensembles
1	0.058	1
1.5	0.726	1
2	1	0.509
3	1	0.100
5	1	0.124
7	0.141	1
10	0.452	1
12.5	0.001	1
15	0.001	1

Table 8
The level of significant difference between stacking and OR ensembles.

Cost	Stacking ensembles	OR ensembles
1	1	0.003
1.5	1	0.127
2	0.826	1
3	0.433	1
5	1	0.269
7	1	0.080
10	1	0.220
12.5	1	0.007
15	1	0.007

positively related to bankrupt companies.

In summary, the results demonstrate that combining the 5 related FRs from Altman's Z-score and all of the 21 CGIs cannot provide better prediction performance than models based on the FRs alone. Moreover, using specific CGIs alone cannot make provide good prediction performance. On the contrary, better discriminative power can be obtained by combining specific FRs and CGIs to allow the prediction models to effectively distinguish between bankrupt and normal companies.

The proposed stacking ensembles based on combining specific FRs and CGIs perform significantly better than the baseline model by using FRs alone and OR ensembles when the cost ratios are larger than 7 (c.f. Table 7). Therefore, it may be sufficient for general investors to use the traditional FRs for reasonable FDP performance. However, for other investors, such as institutional investors, who have very low tolerance to prediction error of bankrupt companies (i.e. high cost ratios are regarded as very critical), stacking ensembles trained by using the combined specific FRs and CGIs are likely more optimal.

5. Conclusion

In this study, we examine the usefulness of CGIs in a bankruptcy prediction model that uses the stacking ensemble technique. Our first experiment tests the effectiveness of stacking ensembles in a model that includes the five FRs from Altman's Z-score and 21 CGIs as input variables. We compare our model's performance with baseline model of using FRs alone. Our second experiment tests the effectiveness of stacking ensembles in a model that uses selected features from a comprehensive list of 40 FRs and 21 CGIs. We test the model against the OR ensemble model and baseline model of using FRs alone.

In the first experiment, we find that combining the five FRs from Altman's Z-score and 21 CGIs does not improve the prediction performance of stacking ensembles. We find insignificant results consistent across all cost ratios when compared with the baseline model. In our second experiment, we first perform feature selection to identify representative features from a comprehensive list of 40 FRs and 21 CGIs. We include only the selected features in our experiment and find that the selected FRs and CGIs can make stacking ensembles outperform the OR ensembles (Liang et al., 2018) and the baseline model, particularly when the cost ratios are high.

We contribute to the corporate governance literature in two ways. First, using hand-collected data, we find that one of the novel CGIs we collected, which signals board conflict, is useful in predicting bankruptcy. Second, we show that using recent CGIs retrieved from U.S companies, which went through significant changes in corporate governance regime, are useful in FDP when combine with FRs. We also contribute to the machine learning literature by showing that stacking ensemble outperforms other ensemble methods in FDP, in particular when we use a parsimonious set of CGIs and FRs identified from feature selection.

Our study has several limitations. First, we only use data from one year prior to bankruptcy. Thus, we do not consider the role of CGIs and FRs in predicting bankruptcy in the more distant future. Second, in the last twenty years the U.S. Securities and Exchange Commission and the U.S. stock exchanges continuously change corporate governance regulations and requirements (e.g. the independence and financial expert

Table 9
The selected features of FRs and CGIs by SDA.

	Variable descriptions	Feature weight	Normal mean	Bankruptcy mean
FRs				
Z1	Working capital/total assets	−1.974	0.14	−0.15
FR_4	Quick Ratio	−2.579	1.5	0.86
FR_{17}	(Total Debt)/(Total Assets)	2.481	0.37	0.69
FR_{18}	EBITDA/(Total Assets)	−2.084	0.09	−0.01
FR_{23}	(Operating CashFlow)/(Total Assets)	−3.502	0.06	−0.05
FR_{29}	(Accounts Payable)/(Total Sales)	1.531	0.1	0.18
CGIs				
C_5	Number of IND director/number of Directors	1.101	0.64	0.65
C_{11}	The IND director number in audit members/number of audit members	−1.200	0.96	0.93
C_{15}	=1 if ultimate controller > 50%	−0.574	25/143	28/143
C_{16}	=1 if ultimate controller > 30%	0.902	5/143	14/143
C_{17}	=1 if the company issues special shares	0.676	14/143	24/143
C_{18}	=1 if proxy statement and annual report are not consistent	0.463	0.22	0.29

requirement of the board and the committees). Consequently, we are not able to satisfy the ceteris paribus assumption. Last, despite our best efforts, our sample size is small with 286 observations, which calls into question the power of our tests.

There are several interesting and important issues that are left unexplored. First, since most related studies constructing FDP models are only based on static data, the time weight of input variables is ignored. That is, input variables collected immediately before the outcome variable should be weighted more heavily than those collected in the more distant past. Sun et al. (2017) show that time weighting combined with Adaboost SVM ensembles outperform many related prediction models. Therefore, it is worth investigating the effect of using the time weight of data on the combination of FRs and CGIs. Second, in practice, FDP related domain datasets are usually class imbalanced, where the number of bankrupt cases is much lesser than the number of normal cases. Over-sampling techniques, such as SMOTE (Synthetic Minority Over-sampling Technique), are one type of solution to balance the original datasets, which focus on creating synthetic samples for the bankrupt class (Sun et al., 2020). It would be useful to examine prediction performance after combining the data over-sampling and feature selection processes for the imbalanced FDP datasets. Third, to perform feature selection, there are two relevant research streams. One stream is based on domain knowledge from financial and accounting theory and the other is based on data mining techniques. Zhou et al. (2015) show that the combination of domain knowledge and data mining techniques in feature selection methods can outperform unique domain knowledge and unique data mining methods in FDP domain problem datasets. This combined feature selection method can be employed for selecting more representative FRs and CGIs. Last, but not the least, for constructing stacking ensembles, there are many other classification techniques available for the FDP related problems. It would be useful to compare the prediction accuracy of different classification techniques for the base and meta learners, such as k-nearest neighbor, neural networks, random forests, etc., based on FRs, CGIs, and the combined FRs and CGIs, respectively.

Appendix A. Descriptive statistics of the input variables

	Bankrupt firms					Normal firms				
	Mean	*Min*	*Max*	*Std*	*Median*	*Mean*	*Min*	*Max*	*Std*	*Median*
Z_1	0.689	0	0.926	0.157	0.736	0.784	0.403	1	0.070	0.776
Z_2	0.813	0	0.945	0.134	0.857	0.883	0.513	1	0.070	0.896
Z_3	0.694	0.10985	0.852	0.102	0.730	0.743	0	1	0.100	0.761
Z_4	0.023	0	0.292	0.041	0.007	0.121	0	1	0.173	0.063
Z_5	0.175	0.003802	1	0.162	0.127	0.193	0	0.967	0.155	0.154
Z_6	0.420	0.009934	0.595	0.103	0.449	0.527	0	1	0.113	0.517
FR_1	0.133	0	0.696	0.120	0.103	0.226	0.006	1	0.171	0.178
FR_2	0.306	0	0.951	0.213	0.268	0.364	0.022	1	0.242	0.322
FR_3	0.231	0	0.879	0.163	0.193	0.281	0.016	1	0.202	0.216
FR_4	0.117	0	0.851	0.132	0.084	0.206	0.008	1	0.181	0.154
FR_5	0.181	0.005719	1	0.181	0.117	0.085	0	0.461	0.070	0.064
FR_6	0.046	0	0.050	0.004	0.047	0.056	0.029	1	0.080	0.048
FR_7	0.031	1.98E−06	1	0.111	0.005	0.021	0	0.597	0.063	0.005
FR_8	0.108	6.96E−05	0.620	0.122	0.060	0.137	0	1	0.174	0.071
FR_9	0.898	0	0.994	0.129	0.931	0.929	0.257	1	0.073	0.939
FR_{10}	0.068	0	0.567	0.110	0.022	0.120	0	1	0.170	0.052
FR_{11}	0.141	0122	0.879	0.161	0.078	0.177	0	1	0.200	0.109
FR_{12}	0	0	0.004	0	0	0.009	0	1	0.084	0
FR_{13}	0.838	0	1	0.113	0.869	0.911	0.551	0.999	0.055	0.917
FR_{14}	0.125	0	1	0.085	0.113	0.114	0.090	0.168	0.006	0.114
FR_{15}	0.242	0	1	0.215	0.219	0.173	0	0.870	0.151	0.146
FR_{16}	0.109	0	1	0.085	0.098	0.098	0.077	0.148	0.006	0.098
FR_{17}	0.360	0.003967	1	0.216	0.326	0.191	0	0.908	0.144	0.156
FR_{18}	0.544	0	0.971	0.136	0.573	0.628	0.075	1	0.110	0.644
FR_{19}	0.414	0	1	0.087	0.408	0.410	0.340	0.463	0.014	0.410
FR_{20}	0.600	0.032561	0.910	0.119	0.635	0.670	0	1	0.114	0.694
FR_{21}	0.918	0	1	0.091	0.930	0.930	0.902	0.940	0.004	0.930
FR_{22}	0.697	0.217162	0.796	0.097	0.729	0.751	0	1	0.095	0.770
FR_{23}	0.067	0	1	0.080	0.059	0.059	0.055	0.072	0.002	0.059
FR_{24}	0.501	0	0.801	0.139	0.534	0.606	0.101	1	0.106	0.614
FR_{25}	0.786	0	1	0.094	0.805	0.806	0.469	0.845	0.036	0.810
FR_{26}	0.926	0.173875	0.991	0.125	0.962	0.950	0	1	0.108	0.971
FR_{27}	0.046	0	1	0.114	0.019	0.035	0.002	0.634	0.067	0.019
FR_{28}	0.266	0	1	0.102	0.275	0.291	0.128	0.691	0.049	0.282
FR_{29}	0.076	0	0.434	0.073	0.060	0.078	0	1	0.088	0.068
FR_{30}	0.107	0.001654	1	0.144	0.061	0.060	0	0.477	0.065	0.044
FR_{31}	0.125	0	1	0.170	0.057	0.123	0	0.999	0.172	0.047
FR_{32}	0.027	3.56E−06	0.636	0.067	0.008	0.027	0	1	0.091	0.007
FR_{33}	0.540	0.435526	1	0.078	0.515	0.517	0	0.938	0.078	0.510
FR_{34}	0.449	0.358915	1	0.090	0.421	0.423	0	0.815	0.065	0.421
C_1	0.378	0.071429	0.857	0.163	0.357	0.363	0	1	0.173	0.357
C_2	0.246	0	0.857	0.147	0.143	0.254	0	1	0.151	0.143
C_3	0.460	0.090909	1	0.204	0.455	0.436	0	1	0.208	0.455
C_4	0.101	0	1	0.169	0	0.101	0	0.600	0.140	0
C_5	0.623	0.030303	1	0.255	0.702	0.605	0	0.909	0.258	0.664
C_6	0.319	0	1	0.238	0.225	0.301	0	0.938	0.187	0.250
C_7	0.109	0	1	0.183	0	0.102	0	0.583	0.149	0
C_8	0.343	0	0.667	0.106	0.333	0.353	0	1	0.138	0.333
C_9	0.939	0	1	0.165	1	0.948	0	1	0.157	1
C_{10}	0.365	0	0.778	0.101	0.333	0.361	0.111	1	0.117	0.333

C_{11}	0.934	0	1	0.180	1	0.964	0.333	1	0.113	1
C_{12}	0.099	0	1	0.195	0.027	0.102	0	0.919	0.174	0.032
C_{13}	0.147	0	1	0.242	0.030	0.079	0	0.759	0.140	0.016
C_{14}	0.566	0	1	0.497	1	0.545	0	1	0.500	1
C_{15}	0.091	0	1	0.288	0	0.035	0	1	0.184	0
C_{16}	0.168	0	1	0.375	0	0.098	0	1	0.298	0
C_{17}	0.196	0	1	0.398	0	0.175	0	1	0.381	0
C_{18}	0.294	0	1	0.457	0	0.224	0	1	0.418	0
C_{19}	0.251	0	0.997	0.300	0.116	0.222	0	1	0.303	0.080
C_{20}	0.176	0	0.964	0.208	0.076	0.165	0	1	0.235	0.071
C_{21}	0.130	0	1	0.243	0	0.062	0	0.708	0.138	0

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*, 589–609.

Armaki, A. G., Fallah, M. F., Alborzi, M., & Mohammadzadeh, A. (2017). A hybrid meta-learner technique for credit scoring of banks' customers. *Engineering, Technology & Applied Science Research, 7*(5), 2073–2082.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications, 83*, 405–417.

Bauweraerts, J. (2016). Predicting bankruptcy in private firms: Towards a stepwise regression procedure. *International Journal of Financial Studies, 7*, 147–153.

Bauer, E., & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning, 36*(1–2), 105–139.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71–111.

Berkman, H., Cole, R. A., & Fu, L. J. (2009). Expropriation through loan guarantees to related parties: Evidence from China. *Journal of Banking & Finance, 33*, 141–156.

Bredart, X. (2014). Financial distress and corporate governance: The impact of board configuration. *International Business Research, 7*(3), 72–80.

Chaganti, R. S., Mahajan, V., & Sharma, S. (1985). Corporate board size, composition and corporate failures in retailing industry. *Journal of Management Studies, 22*(4), 400–417.

Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: Empirical evidence for the UK. *European Accounting Review, 13*(3), 465–497.

Chen, I.-J. (2014). Financial crisis and the dynamics of corporate governance: Evidence from Taiwan's listed firms. *International Review of Economics and Finance, 32*, 3–28.

Cheung, Y.-L., Chung, C.-W., Tan, W., & Wang, W. (2013). Connected board of directors: A blessing or a curse? *Journal of Banking & Finance, 37*, 3227–3242.

Collier, J., & Esteban, R. (1999). Governance in the participative organisation: Freedom, creativity and ethics. *Journal of Business Ethics, 21*, 173–188.

Daily, C. M. (1994). Bankruptcy and corporate governance: The impact of board composition and structure. *The Academy of Management Journal, 37*(6), 1603–1617.

Daily, C. M. (1996). Governance patterns in bankruptcy reorganizations. *Strategic Management Journal, 17*(5), 355–375.

Daily, C. M., & Dalton, D. R. (1994). Corporate governance and the bankrupt firm: An empirical assessment. *Strategic Management Journal, 15*(8), 643–654.

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*, 1–30.

du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research, 254*, 236–252.

Fallahpour, S., Lakvan, E. N., & Zadeh, M. H. (2017). Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem. *Journal of Retailing and Consumer Services, 34*, 159–167.

Gompers, P. A., Ishii, J. L., & Metrick, A. (2003). Corporate governance and equity prices. *Quarterly Journal of Economics, 118*(1), 107–155.

Graczyk, M., Lasota, T., Trawinski, B., & Trawinski, K. (2010). Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. *International Conference on Intelligent Information and Database Systems: Part II,* 340–350.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review, 76*(2), 323–329.

Jian, M., & Wong, T. J. (2010). Propping through related party transactions. *Review of Accounting Studies, 15*, 70–105.

Kim, H. J., Jo, N. O., & Shin, K. S. (2016). Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications, 59*, 226–234.

Kim, S. Y. (2018). Predicting hospitality financial distress with ensemble models: The case of US hotels, restaurants, and amusement and recreation. *Service Business, 12*, 483–503.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence, 2*, 137–1143.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review. *European Journal of Operational Research, 180*, 1–28.

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (1999). Corporate ownership around the world. *The Journal of Finance, 54*, 471–517.

Lacher, R. C., Coats, P. K., Sharma, S. C., & Fant, L. F. (1995). A neural network for classifying the financial health of a firm. *European Journal of Operational Research, 85*, 53–65.

Lee, K. C., Han, I., & Kwon, Y. (1996). Hybrid neural network models for bankruptcy predictions. *Decision Support Systems, 18*, 63–72.

Lee, T.-S., & Yeh, Y.-H. (2004). Corporate governance and financial distress: Evidence from Taiwan. *Corporate Governance: An International Review, 12*(3), 378–388.

Liang, D., Lu, C.-C., Tsai, C.-F., & Shih, G.-A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research, 252*, 561–572.

Liang, D., Tsai, C.-F., Dai, A.-J., & Eberle, W. (2018). A novel classifier ensemble approach for financial distress prediction. *Knowledge and Information Systems, 54*, 437–462.

Liang, D., Tsai, C.-F., & Wu, H.-T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems, 73*, 289–297.

Liang, Q., Xu, P., & Jiraporn, P. (2013). Board characteristics and Chinese bank performance. *Journal of Banking & Finance, 37*, 2953–2968.

Lin, F.-Y., Liang, D., & Chu, W.-S. (2010). The role of non-financial features related to corporate governance in business crisis prediction. *Journal of Marine Science and Technology, 18*(4), 504–513.

Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2012). Machine learning in financial crisis prediction: A survey. *IEEE Transactions on Systems, Man and Cybernetics – Part C: Applications and Reviews, 42*(4), 421–436.

Martin, A., Doddington, G., Kamm, T., Ordowski, M., & Przybocki, M. (1997). The DET curve in assessment of detection task performance. *European Conference on Speech Communication and Technology, 4*, 1895–1898.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*, 109–131.

Opitz, D. W., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research, 11*, 169–198.

Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems, 52*(2), 464–473.

Patil, P. S., Aghav, J. V., & Sareen, V. (2016). An overview of classification algorithms and ensemble methods in personal credit scoring. *International Journal of Computer Science and Technology, 7*(2), 183–188.

Pisula, T. (2020). An ensemble classifier-based scoring model for predicting bankruptcy of Polish companies in the Podkarpackie Voivodeship. *Journal of Risk and Financial Management, 13*(2), 1–35.

Platt, H., & Platt, M. (2012). Corporate board attributes and bankruptcy. *Journal of Business Research, 65*(8), 1139–1143.

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review, 33*, 1–39.

Serrano-Cinca, C. (1997). Feedforward neural networks in the classification of financial information. *The European Journal of Finance, 3*, 183–202.

Shailer, G. (2004). *An introduction to corporate governance in Australia.* Pearson Education Australia.

Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance, 52*(2), 737–783.

Stulz, R. M. (1990). Managerial discretion and optimal financing policies. *Journal of Financial Economics, 26*(1), 3–27.

Sun, J., Fujita, H., Chen, P., & Li, H. (2017). Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. *Knowledge-Based Systems, 120*, 4–14.

Sun, J., Li, H., Fujita, H., Fu, B., & Ai, W. (2020). Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. *Information Fusion, 54*, 128–144.

Sun, J., Li, H., Huang, Q.-H., & He, K.-Y. (2014). Prediction financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems, 57*, 41–56.

Tian, S., & Yu, Y. (2017). Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance, 51*, 510–526.

Tsai, C.-F. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. *Information Fusion, 16*, 46–58.

Tsai, C.-F., Hsu, Y.-F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing, 24*, 977–984.

Volkov, A., Benoit, D. F., & Van den Poel, D. (2017). Incorporating sequential information in bankruptcy prediction with predictors based on Markov for discrimination. *Decision Support Systems, 98*, 59–68.

Wei, S., Yang, D., Zhang, W., & Zhang, S. (2019). A novel noise-adapted two-layer ensemble model for credit scoring based on backflow learning. *IEEE Access, 7*, 99217–99230.

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259.

Wozniak, M., Grana, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion, 16*, 3–17.

Wu, W., Xiao, Z., Dang, X., Yang, D., & Yang, X. (2014). Financial ratio selection for business failure prediction using soft set theory. *Knowledge-Based Systems, 63*, 59–67.

Xia, Y., Liu, C., Da, B., & Xie, F. (2018). A novel heterogeneous ensemble credit scoring model based on bstacking approach. *Expert Systems with Applications, 93*, 182–199.

Yeh, Y.-H., & Woidtke, T. (2005). Commitment or entrenchment?: Controlling shareholders and board composition. *Journal of Banking & Finance, 29*, 1857–1885.
Zenko, B., Todorovski, L., & Dzeroski, S. (2001). A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods. *IEEE International Conference on Data Mining,* 669–670.
Zhang, W., He, H., & Zhang, S. (2019). A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. *Expert Systems with Applications, 121*, 221–232.
Zhou, L., Lu, D., & Fujita, H. (2015). The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches. *Knowledge-Based Systems, 85*, 52–61.
Zieba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic feature generation in application to bankruptcy prediction. *Expert Systems with Applications, 58*, 93–101.

Dr. Deron Liang is a professor at the Department of Computer Science and Information Engineering, National Central University. His research interests cover business intelligence and distributed systems.

Dr. Chih-Fong Tsai is a professor at the Department of Information Management, National Central University, Taiwan. His current research focuses on multimedia information retrieval and data mining.

Dr. Hung-Yuan Lu is an associate professor at the Department of Accounting, California State University-Fullerton. His research interests include financial reporting, voluntary disclosure and financing constraints.

Mr. Li-Shin Chang received a Master's degree from the Department of Computer Science and Information Engineering, National Central University. His research interests cover machine learning and data mining applications.