Received: 18 September 2019 | Accepted: 12 January 2020
DOI: 10.1002/for.2661

RESEARCH ARTICLE

WILEY

Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods

Xiaobo Tang | **Shixuan Li** | **Mingliang Tan** | **Wenxuan Shi**

School of Information Management, Wuhan University, Wuhan, China

Correspondence
Shixuan Li, School of Information Management, Wuhan University, No. 299 Bayi Road. Wuhan, Hubei, China.
Email: shixuan.li@hotmail.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 71673209

Abstract

Financial distress prediction (FDP) has been widely considered as a promising approach to reducing financial losses. While financial information comprises the traditional factors involved in FDP, nonfinancial factors have also been examined in recent studies. In light of this, the purpose of this study is to explore the integrated factors and multiple models that can improve the predictive performance of FDP models. This study proposes an FDP framework to reveal the financial distress features of listed Chinese companies, incorporating financial, management, and textual factors, and evaluating the prediction performance of multiple models in different time spans. To develop this framework, this study employs the wrapper-based feature selection method to extract valuable features, and then constructs multiple single classifiers, ensemble classifiers, and deep learning models in order to predict financial distress. The experiment results indicate that management and textual factors can supplement traditional financial factors in FDP, especially textual ones. This study also discovers that integrated factors collected 4 years prior to the predicted benchmark year enable a more accurate prediction, and the ensemble classifiers and deep learning models developed can achieve satisfactory FDP performance. This study makes a novel contribution as it expands the predictive factors of financial distress and provides new findings that can have important implications for providing early warning signals of financial risk.

KEYWORDS

deep learning, financial distress prediction, machine learning, management factors, textual factors

1 | INTRODUCTION

Financial distress prediction (FDP) has attracted a high degree of attention in both academic and industrial fields over the last few decades as it is an effective tool of risk management (Geng, Bose, & Chen, 2015; Wanke, Barros, & Faria, 2015). Such predictions can provide early warning signals of financial risks for relevant agents in the economy, which is important for stakeholders, listed companies, and even the development of the economy itself (Farooq & Qamar, 2019). Based on such signals, and prior to a crisis, the relevant investors can realize investment strategy adjustments, and firms and governments are enabled to develop remedial measures, thereby avoiding financial losses to a certain extent (Wang, Chen, & Chu, 2018). For these reasons, numerous studies have focused on the prediction of financial distress; specifically, exploring predictive factors and prediction methods has been a key undertaking in this body of research.

The factors that are involved in FDP include financial and nonfinancial factors. Early studies in this field mainly examined financial factors, including profitability, solvency, operational capabilities, and finance structure (Hua, Wang, Xu, Zhang, & Liang, 2007; Sun & Li, 2008). Although financial factors do contribute significantly to FDP, they only cover quantitative information, which cannot comprehensively represent the status of companies. In light of this, some of the latest studies have highlighted the limited predictive capacity of financial variables, and have attempted to utilize certain nonfinancial factors in predicting financial distress (Cecchini, Aytug, Koehler, & Pathak, 2010; Hajek, Olej, & Myskova, 2014). Specifically, using textual analysis methods such as word frequency statistics and sentiment analysis, qualitative nonfinancial predictive factors can be extracted from unstructured texts, such as audit reports and annual reports; these textual factors can then supplement the traditional financial factors used in FDP (du Jardin, 2016; Wang et al., 2018).

The prediction method of financial distress can, typically, be divided into two categories: statistical methods and machine learning methods. Early research focused on statistical methods, which mainly include linear discriminant analysis (LDA), multiple discriminant analysis (MDA), factor analysis (FA), and probabilistic modeling (Chen, Ribeiro, & Chen, 2016; Kumar & Ravi, 2007; Suntraruk, 2010). While these statistical methods play a significant role in studies on FDP, when the restrictive assumptions that these models depend on are not satisfied, their validity and applicability become limited (Chen et al., 2016; Wang et al., 2018). Therefore, studies in recent years have focused on using machine learning methods for the purposes of FDP, such as the support vector machine (SVM), decision tree (DT), and artificial neural networks (ANN; Zhou et al., 2014; Geng et al., 2015; Olson, Delen, & Meng, 2012). With the exception of these single classifiers, studies have also constructed FDP models based on ensemble classifiers (Liang, Tsai, Dai, & Eberle, 2018; Wang, Hao, Ma, & Jiang, 2011). Moreover, deep learning models have also received attention in the field of FDP (Alexandropoulos, Aridas, Kotsiantis, & Vrahatis, 2019).

While previous studies have clearly made contributions to the area of FDP, optimizing the predictive factors and prediction models is still arguably in need of improvement. In order to achieve this with regard to predictive factors, this study combines traditional financial factors with management and textual factors; for prediction models, this study constructs single classifiers, ensemble classifiers, and deep learning models to predict financial distress. To the best of our knowledge, this is the first study to utilize three types of factors and both traditional machine learning and deep learning models to predict financial distress in listed Chinese companies. In so doing, this study attempts to answer the following research questions:

RQ1. Which are the key influencing factors in FDP of listed Chinese companies?
RQ2. Which models perform better than others in FDP of listed Chinese companies?
RQ3. How early can signs of financial distress be predicted in listed Chinese companies?

In order to engage with these research objectives, this study utilizes data from 3, 4, and 5 years ahead of the predicted benchmark year, revealing changes in the key predictive factors of different years, and constructing multiple machine learning models to predict financial distress. It is hoped that the results of this research can provide a template for early warning mechanisms for relevant economic agents, so that they can make the corresponding efforts to avoid financial losses.

The main contributions of this study are as follows. First, it provides a novel FDP method that integrates multiple predictive factors, time spans, and classifier models. Second, it highlights the discovery that management and textual features can play a significant role in FDP for listed Chinese companies; specifically, the study finds that these features can achieve optimum performance 4 years prior to the predicted benchmark year, and reveals the superiority of ensemble classifiers and deep learning models. Finally, the methods described in this paper may be seen as potential approaches that can concretely be applied in companies or at the level of economic policy towards avoiding the financial losses.

The remainder of this paper is organized as follows. Section 2 outlines the relevant studies on the prediction of financial distress. Section 3 describes the study framework and explains the analysis approaches in detail, followed by a presentation of the main research results in Section 4. Following this, Section 5 provides a discussion of the empirical results and answers to the research questions. The final section concludes this study and provides directions for future research.

2 | RELATED STUDIES

This section describes the different factors and models used in studies of FDP, as summarized in Table 1,

TABLE 1 Overview of related works

Research	Factors	Feature selection methods	Models
Alfaro, García, Gámez, and Elizondo (2008)	Financial factors	Filter	Adaboost, NN
Li and Sun (2009)	Financial factors	Wrapper	SVM
Xie, Luo, and Yu (2011)	Financial factors, management factors	Filter	SVM, MDA
Martin, Gayathri, Saranya, Gayathri, and Venkatesan (2011)	Financial factors	Wrapper	Fuzzy c-means/MARS
Hájek and Olej (2013)	Financial factors, textual factors	Filter	LR, NN, SVM
L. Zhou, Tam, and Fujita (2016)	Financial factors	Filter	LDA, LR, NN, DT, KNN, Adaboost, SVM
Liang, Lu, Tsai, and Shih (2016)	Financial factors, management factors	Filter, wrapper	SVM, KNN, CART, MLP, NB
Wang et al. (2018)	Financial factors, textual factors	Filter	SVM, bagging, boosting
Farooq and Qamar (2019)	Financial factors, management factors	Wrapper	Bayesian, SVM, DT, NN, KNN
Jo and Shin (2016)	Financial factors, textual factors	Filter	MLP, ANN

focusing on the research pertaining to listed Chinese companies.

2.1 | Factors used in predictions of financial distress

In the context of FDP, studies regard information that can represent the status of the companies as predictive factors; typically, the latter contain financial and non-financial aspects. Early studies on FDP mainly focused on financial factors. For example, Beaver (1966) selected financial ratios relating to six aspects of FDP, including cashflow, net income, debt to total asset, liquid asset to total asset, liquid asset to current debt, and turnover ratios. Since then, many studies have made efforts to optimize the financial predictive factors used in FDP. For example, Alfaro et al. (2008) selected 16 financial ratios for the prediction of financial failure, pointing out that earnings before taxes to capital ratio, and liabilities to total debt ratio, had the strongest predictive power in FDP. In general, financial factors are extracted by converting firms' accounting information into financial ratios; given that these factors play a significant role in understanding a firm's financial status, they can thus be used in predicting the financial distress (Manzaneque, García-Pérez-De-Lema, & Antón Renart, 2015; Sun, Li, Huang, & He, 2014; L. Zhou et al., 2016).

However, the predictive capability of financial factors is limited, as these factors only cover quantitative information, which cannot represent companies' qualitative aspects (du Jardin, 2016; Hajek et al., 2014). In light of this, recent studies have attempted to combine non-financial factors with financial factors in order to bolster FDP. Specifically, nonfinancial factors can be categorized into management factors and textual factors. Management factors can represent the internal and external governance of companies, such as the distribution of shares, the equity ratio of executives, board size, CEO duality, macroeconomic indicators, stock liquidity, auditors' qualified opinions, and executive compensation variables (Farooq & Qamar, 2019; Y. Jiang & Jones, 2018; Manzaneque, Priego, & Merino, 2016; Xie et al., 2011). These management factors are structured indicators, which can be directly extracted from companies' annual reports or other relevant publications. However, since textual information is unstructured, studies need to employ text analysis approaches to extract these particular factors from diverse sources.

The existing research has mainly used term frequency counts, readability analysis, and sentiment analysis methods (Cecchini et al., 2010; Hajek et al., 2014;

Loughran & McDonald, 2016) to explores companies' relevant financial texts, such as annual reports, financial news, and related social media information (Bollen, Mao, & Zeng, 2011; Kraus & Feuerriegel, 2017; Wang et al., 2018). The analysis results of these texts can be converted into textual factors for the application in question. Specifically, term frequency counts show the basic characteristics of financial texts, such as the length of financial documents and the frequency of certain keywords, while readability and sentiment analyses can reveal the semantic features of texts. Existing studies have implemented the Fog index, Coleman–Liau index, and automated readability index to evaluate the readability of company annual reports (Dong, Liao, & Liang, 2016; Malafronte, Porzio, & Starita, 2013). In the financial textual analysis field, studies primarily adopt the lexicon-based sentiment analysis method, with the most popular financial sentiment lexicon having been created by Loughran and McDonald (2011), which identifies positive, uncertain, litigious, strong model, and weak model sentiment words. To date, the relevant body of research has adopted sentiment analysis in identifying financial fraud, predicting financial distress, and forecasting the stock market (Bollen et al., 2011; Hajek & Henriques, 2017; Wang et al., 2018).

2.2 | Methods of predicting financial distress

In order to accurately predict financial distress, studies have implemented various methods that can be divided into two categories: statistical methods and machine learning methods. Prior to the dissemination of machine learning, statistical methods were often applied to FDP, specifically, LDA, MDA, FA, as well as probabilistic modeling (Kumar & Ravi, 2007; H. Li & Sun, 2009; Suntraruk, 2010). However, many of these statistical models depend on restrictive assumptions, such as the normality, the independence of variables, and linearity (Chen et al., 2016; Wang et al., 2018). This means that when the analysis objects cannot satisfy these assumptions, the validity and applicability of such statistical models are limited. For this reason, subsequent studies have attempted to integrate less limited machine learning methods in FDP.

The process of applying machine learning methods generally involved two stages, namely, feature selection and model construction. The automatic feature selection process extracts valuable features from the original financial and nonfinancial factors, which can promote the efficiency and accuracy of the FDP. This feature selection can be broadly divided into the filter, wrapper, and embedded approaches (Chandrashekar & Sahin, 2014). In particular, many FDP studies apply filter and wrapper approaches to extract valuable features (Liang, Tsai, & Wu, 2015), such as the *t*-test, *f*-test, recursive feature elimination (RFE), and wrapper DT (Fallahpour, Lakvan, & Zadeh, 2017; Farooq & Qamar, 2019; Geng et al., 2015; J. Li, Qin, Yi, Li, & Shen, 2014; Liang et al., 2016). In addition, some studies use multiple feature selection methods to extract valuable features and compare the prediction performance of the different selected feature subsets (Cho, Hong, & Ha, 2010; Liang et al., 2016).

After feature selection, the extracted feature subsets can act as the input of the machine learning models. As previously mentioned, single classifiers, ensemble classifiers, and deep learning models were applied to the earlier studies of FDP. To be specific, single classifiers, such as SVM, DT, ANN, KNN, CART and NB, have been broadly applied to predicting financial distress (Geng et al., 2015; Hájek & Olej, 2013; Jo & Shin, 2016; Lorca, Landajo, & Andrés, 2014; Olson et al., 2012). Some studies have also used ensemble classifiers, including voting, bagging, boosting, and stacking, which have achieved high accuracy in FDP (Liang et al., 2018; Wang et al., 2018; L. Zhou et al., 2016). Moreover, the latest studies have shown that deep learning models are highly efficient predictive models in FDP, such as deep neural network (DNN) and deep dense multilayer perceptron (DDMP) (Alexandropoulos et al., 2019; Mai, Tian, Lee, & Ma, 2019).

2.3 | Prediction of financial distress in listed Chinese companies

Given that different countries have different accounting procedures and rules, to date there has been no unified definition of financial distress. In general, financial distress leads to the weakening of a company's profitability. As such, numerous studies in this field have taken bankruptcy as the outcome of financial distress; however, the bankruptcy-related data of listed Chinese companies is hard to acquire. As an alternative, the definition of the "Special Treatment" (ST) of a company, as presented by the China Securities Regulatory Commission (CSRC), may be seen as close to describing financial distress (Geng et al., 2015). According to the Chinese Stock Listing Exchange Rule, there are three main reasons for listed companies to be assigned the ST status: (1) negative earnings in two consecutive years; (2) net assets per share being less than the face value per share; and (3) abnormal financial behavior, as identified and claimed by the CSRC or stock exchanges (Y. Jiang & Jones, 2018). It is for these reasons that prior studies have regarded the ST as the symbol for financial distress of listed Chinese companies (Y. Jiang & Jones, 2018; H. Li & Sun, 2009; Xie et al., 2011).

Based on these definitions of ST for listed Chinese companies, studies have collected information related to both ST and healthy firms as the raw data with which to calculate FDP. Early research used the data of companies 1 and 2 (T-1, T-2) years before the ST assignment year in FDP (Ding, Song, & Zen, 2008; H. Li & Sun, 2009). However, according to the ST warning mechanism, the relevant authorities can assign an ST label to a given company based on its financial status in the 2 years previous to actual financial risk showing up. This means that predicting financial distress 1 or 2 years ahead of such ST assignment is redundant; hence studies have attempted to collect the relevant data 3, 4, and 5 (T-3, T-4, T-5) years ahead of the ST assignment year (Geng et al., 2015; Wang et al., 2018).

Given that the FDP of listed Chinese companies has received attention from scholars in this field, it may be seen as meaningful to explore the key predictive factors and improve the model performance of FDP. With regard to predictive factors, the prior literature has pointed out that both financial and nonfinancial (i.e., management and textual factors) factors contain information about a company's financial prospects, thus making it necessary to incorporate both in the prediction of financial distress. Regarding predictive methods, it is noteworthy that previous studies have made efforts to compare and optimize the performance of different machine learning models, while little research has comprehensively applied single classifiers, ensemble classifiers, and deep learning models to FDP. It can therefore be argued that comparing these diverse models is essential in this research field.

3 | METHODOLOGY

This study took financial, management, and textual factors as the predictive features, constructing 12 models for the FDP of listed Chinese companies. Based on the FDP processes outlined in earlier research (Geng et al., 2015; Wang et al., 2018), this study developed the framework shown in Figure 1, involving four key steps: data collection, data preparation and feature selection, model construction, and model evaluation. As shown in Figure 1,

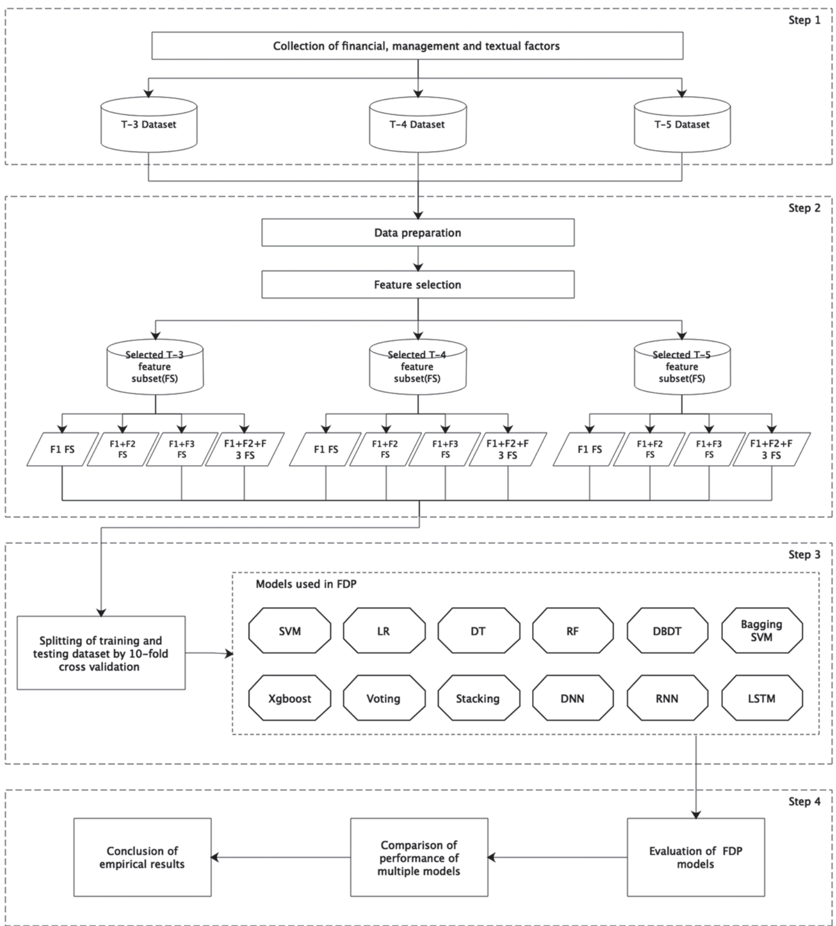

FIGURE 1 The framework of this study

step 1 comprises the preparation stage of this study, namely, collecting the financial and nonfinancial factors of FDP; step 2 focuses on selecting the key features from the T-3, T-4, and T-5 data sets; and steps 3 and 4 construct multiple models and evaluate their predictive performance. This study applied Python programming to undertake the feature selection and model construction.

3.1 | Data collection and predictive factors

This study selected 212 listed Chinese companies that had received the ST label from 2014 to 2018 as examples of financial distress. Specially, there were 37 instances of ST in 2014, 42 in 2015, 61 in 2016, 59 in 2017, and 54 in 2018. According to the principle of matching companies in the same industry and with a similar asset size (Liu, Wu, & Li, 2019; Sun et al., 2014), 212 nondistressed companies were selected as healthy instances in the benchmark year of 2018. Moreover, this study implemented the FDP models based on predictive factors collected 3, 4, and 5 years before the companies received the ST label. This means that if a given company received the ST label in 2017, this research obtained information about this company in 2014 (T-3), 2013 (T-4), and 2012 (T-5). Similarity, the predictive factors of the healthy companies, as the control group, were collected in 2015 (T-3), 2014 (T-4), and 2013 (T-5). Therefore, the data obtained were stored in the T-3, T-4, and T-5 data sets separately.

After both the control and ST group instances were specified, the financial factors (F1), management factors (F2), and textual factors (F3) were collected, totaling 74 indicators. In line with the work of Alfaro et al. (2008), this study selected the predictive factors based on three principles: that the selected indictors were utilized in previous studies; that the chosen indicators were obtainable; and that the selected indicators met the needs of this study. The following subsection details the predictive factors used this study.

3.1.1 | Financial factors of FDP

This study chose 46 financial indicators as the predictive financial factors of FDP, based on existing research (Farooq & Qamar, 2019; Geng et al., 2015; Wang et al., 2018), which is shown in Table 2. These factors can be divided into six categories, representing a company's profitability, solvency, financial structure, operational capabilities, development capacity, and stock ratios. All of the financial factors were obtained from the China

TABLE 2 List of financial factors

Type	Factor	Definition	Type	Factor	Definition
Profitability	x1	Earnings before income tax/average total assets	Operational capabilities	x26	Sales revenue/average balance of accounts receivable
	x2	Net profit/average total assets		x27	Sales cost/average inventory
	x3	Net profit/average current assets		x28	Accounts receivable turnover days + inventory turnover days
	x4	Net profit/average fixed assets		x29	Sales cost/average balance of accounts payable
	x5	Net profit/average shareholders' equity		x30	Sales revenue/average current assets
	x6	Earnings before income tax/invested capital		x31	Sales revenue/average fixed assets
	x7	(sales revenue − sales cost)/sales revenue		x32	Sales revenue/average total assets
	x8	Sales profit/sales revenue	Development capacity	x33	Total assets growth of this year/total assets of last year
	x9	Net profit/sales revenue		x34	Net assets growth of this year/net assets of last year
	x10	Sales cost − sales revenue		x35	Net profit growth of this year/net profit of last year

(Continues)

TABLE 2 (Continued)

Type	Factor	Definition	Type	Factor	Definition
Solvency	x11	Current assets/current liabilities		x36	Total profit growth of this year/net profit of last year
	x12	(current assets − inventory)/current liabilities		x37	Sales revenue growth of this year/sales revenue of last year
	x13	Current assets − current liabilities		x38	Total shareholders' equity growth of this year/total shareholders' equity of last year
	x14	Net operating cash flow/current liabilities	Stock ratios	x39	Net profit/number of ordinary shares at the end of the year
	x15	Total liabilities/total assets		x40	Net increase in cash and cash equivalents/number of ordinary shares at the end of the year
	x16	Total liabilities/(total assets-net intangible assets)		x41	Sales revenue/number of ordinary shares at the end of the year
	x17	Total assets/total shareholders' equity		x42	Sales profit/number of ordinary shares at the end of the year
	x18	Total liabilities/total shareholders' equity		x43	Total shareholders' equity/number of ordinary shares at the end of the year
	x19	Total shareholders' equity/total liabilities		x44	Capital reserves/number of ordinary shares at the end of the year
	x20	Net operating cash flow/total liabilities		x45	(surplus reserves + undistributed profit)/number of ordinary shares at the end of the year
Financial structure	x21	Current assets/total assets		x46	Net operating cash flow/number of ordinary shares at the end of the year
	x22	Cash flow/total assets			
	x23	Fixed assets/total assets			
	x24	Shareholders' equity/fixed assets			
	x25	Shareholders' equity/fixed assets			

Stock Market & Accounting Research (CSMAR) database.

3.1.2 | Management factors of FDP

Based on prior research (Ghazali, Shafie, & Sanusi, 2015; Y. Jiang & Jones, 2018; Manzaneque et al., 2016; Tykvová & Borell, 2012), this study selected 13 indicators as the predictive management factors of FDP. Specifically, these factors contained information pertaining to four dimensions: board structure, equity structure, internal control information, and auditors' opinions. Table 3 presents the list of management predictive factors of this study, with all of the information also having been collected from CSMAR.

3.1.3 | Textual factors of FDP

As previously mentioned, annual reports contain information that plays a significant role in reflecting the financial status of listed companies. Thus this study extracted textual factors from the annual reports of listed Chinese companies, including the length of these reports and their sentiment features. Specifically, this study separately applied the L&M dictionary (created by Loughran & McDonald, 2011) and the National Taiwan University Sentiment Dictionary (NTUSD) to undertake a sentiment analysis and obtain the relevant predictive sentiment factors. These two dictionaries have been applied in existing research to analyze the sentiment features of listed Chinese companies' annual reports (X. Li, Liu, & Wang, 2019; B. Zhou, Zhang, & Zeng,

TABLE 3 List of management factors

Type	Factor	Definition
Board structure	x47	Whether the top 10 shareholders are associated (1 = not associated, 2 = associated, 3 = not sure)
	x48	Whether the roles of chairman and CEO are separate (1 = same person, 2 = different person)
	x49	Number of directors
	x50	Number of independent directors
	x51	Number of shares held by executives
Equity structure	x52	Percentage of shares owned by company's largest shareholder
	x53	Percentage of shares owned by company's top 10 shareholders
	x54	Percentage of shares owned by company's largest shareholder/percentage of shares owned by company's second largest shareholder
Internal control information	x55	Whether internal control evaluation report is disclosed (1 = yes, 2 = no)
	x56	Whether the conclusion of the internal control evaluation report was issued (1 = yes, 2 = no)
	x57	Whether or not the internal control is effective (1 = yes, 2 = no)
	x58	Whether or not there is a shortcoming in internal control (1 = yes, 2 = no)
Auditors' opinions	x59	Audit opinion (1 = standard unqualified opinion, 2 = reserved opinion, 3 = negative opinion, 4 = unable to comment, 5 = unqualified opinion plus paragraph, 6 = reserved opinion plus paragraph)

2018). In addition, as pointed out by Myšková and Hájek (2017), the Management Discussion and Analysis (MD&A) section is the most important part of 10-K, that is, the annual reports of US-listed companies. Thus this research also extracted the textual features from the MD&A of listed Chinese companies' annual reports, in order to reveal whether these companies' MD&A could contribute to FDP.

The textual factors examined in this study were collected from the annual reports of listed Chinese companies and the database of Chinese Research Data Services (CNRDS) database, which provides valuable sentiment analysis results of these companies' annual reports (Chen, Kim, Wei, & Zhang, 2018; F. Jiang, Lee, Martin, & Zhou, 2019). Table 4 lists the predictive textual factors included in this study and the calculation process used to extract the sentiment factors.

3.2 | Data preparation and feature selection

After collecting the raw data, this study applied the data preparation and feature selection process to obtain valuable features for the FDP. The preprocessing steps for the experimental raw data were as follows: First, the missing data were removed and the variables normalized. The processed data records of healthy companies and ST companies in T-3, T-4, and T-5 were then combined as the initial data set.

These three data sets comprised the input of the feature selection algorithm, using which this study was able to obtain valuable features for the FDP in three different years. As previously mentioned, numerous studies have applied the filter or wrapper methods in order to select features. This study thus adopted the wrapper-based RFE-SVM, developed by Guyon, Weston, Barnhill, and Vapnik (2002), as the feature selection approach; this is a greedy algorithm that first trains the SVM and then eliminates those features, whose removal leads to the largest margin of class separation in an iterative way. Previous research has highlighted that RFE-SVM is an effective method for selecting features in the financial field (J. Li et al., 2014; Maldonado, Bravo, López, & Pérez, 2017).

TABLE 4 List of textual factors

Type	Factor	Definition
MD&A	x60	Number of positive terms in MD&A
	x61	Number of negative terms in MD&A
	x62	Number of sentences in MD&A
	x63	Number of text words in MD&A
Annual reports	x64	Number of words in annual report
	x65	Number of text words in annual report
	x66	Number of sentences in annual report
	x67	Number of positive vocabularies in annual report based on L&M dictionary
	x68	Number of negative vocabularies in annual report based on L&M dictionary
	x69	Number of positive vocabularies in annual report based on NTUSD
	x70	Number of negative vocabularies in annual report based on NTUSD
	x71	(x67 − x68)/x65
	x72	(x67 − x68)/(x67 + x68)
	x73	(x69 − x70)/(x69 + x70)
	x74	(x69 − x70)/x65

3.3 | Model construction

This study compared the predictive performance of single classifiers, ensemble classifiers, and deep learning models, including logistic regression (LR), SVM, DT, random forest (RF), gradient boosting decision tree (GBDT), Xgboost, bagging SVM), voting, stacking, DNN, recurrent neural networks (RNN), and long short-term memory (LSTM).

1. *Single classifiers.* LR, SVM, and DT were the single classifiers used in this study. LR is a discriminative classifier that is linear in its parameters, and serves to explain the relationship between one dependent binary variable and one or more independent variables. SVM can solve linear and nonlinear problems through creating a line or hyperplane to separate the data into classes. DT uses a tree-like graph to divide samples into subsets until all of the training data are correctly classified. These three classifiers have been employed in numerous FDP studies (Hájek & Olej, 2013; Olson et al., 2012; Zhou et al., 2016), and some ensemble classifiers have been developed from these single classifiers.
2. *Ensemble classifiers.* This study adopted voting, bagging (RF and bagging SVM), boosting (GBDT and Xgboost), and stacking ensemble methods to predict financial distress. Ensemble models seek to extract the individual benefits of each classifier and combine them to obtain a better solution (Hajek & Henriques, 2017). Specifically, voting combines

multiple algorithms, whereby the one that receives the largest number of votes is selected as the final classification decision. Bagging is a method for reducing the variance of a classifier to average together multiple classifiers. Boosting can build several incremental models to decrease the bias while keeping the variance small. Stacking is based on constructing multilevel classifiers in a hierarchical way, which method is effective in increasing the predictive force of the classifier. These ensemble classifiers have achieved a satisfactory predictive performance in the context of FDP (Alfaro et al., 2008; Liang et al., 2018; Wang et al., 2018; L. Zhou et al., 2016).

3. *Deep learning models.* DNN, RNN, and LSTM were the deep learning models utilized in this study. DNN can be regarded as a multilayer perceptron (MLP), with many hidden layers comprising a fully connected network. RNN is a special type of neural network that contains at least one feedback connection as an internal state from the neurons' outputs to the inputs. LSTM extends RNN by adding three gates (forget gate, input gate, and output gate) in order to learn long-term dependency in a sequence. These deep learning models have been seen to yield reasonable results for classification prediction problems, although only a few recent studies have applied deep learning methods to predict financial distress (Alexandropoulos et al., 2019; Mai et al., 2019).

This study employed the Python library Scikit-learn (Pedregosa et al., 2011) to construct single and ensemble classifiers, and applied the Python library Keras and Tensorflow (Abadi et al., 2015) to build the deep learning models. These classifiers were trained using the parameter settings presented in Table 5. Prior to model construction, this study adopted a 10-fold cross-validation strategy in order to divide the training and testing data sets, meaning that the data set was divided into 10 distinct subsets, in which nine subsets were taken as the training set and the remaining one was the testing set. This, in turn, meant that a model could be trained and tested 10 times over 10 different combinations of training and testing sets, whereby averaging the 10 results yielded the model's performance.

3.4 | Model evaluation

Evaluating the performance of the machine learning models was also an important step, for which this study employed five evaluation metrics: accuracy, F-measure, Type I error, Type II error, and area under the receiver operating characteristic curve (AUC). The machine learning models generated four types of possible result by which the samples could be classified, namely, true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which could be represented by a confusion matrix. In this study, financially distressed samples were regarded as the positive class and the others were treated as the negative. The evaluation metrics were defined as follows:

Accuracy denotes the percentage of correctly classified samples obtained:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{FN}+\text{TN}}. \tag{1}$$

TABLE 5 Parameter settings of machine learning methods

Method	Parameter
LR	Solver used in the optimization problem = liblinear; norm used in the penalization = l2
SVM	Kernel type used in the algorithm = linear; penalty parameter C of the error term = 1; class weight = balanced
DT	Function used to measure the quality of a split = gini; maximum depth of trees unlimited
RF	Number of trees in the forest = 100; the function to measure the quality of a split = gini; maximum depth of trees unlimited
Bagging SVM	Base estimator to be fitted on random subsets of the data set = SVM; number of base estimators in the ensemble = 20
GBDT	Loss function to be optimized = exponential; number of boosting stages to be perform = 100
Xgboost	Number of trees to be fitted = 100; maximum tree depth for base learners = 5; boosting learning rate = 0.1
Voting	Base estimators: LR, bagging SVM, GBDT; voting rule = soft
Stacking	Base estimators: LR, bagging SVM, GBDT; meta classifier = LR
DNN	Number of hidden layers = 3; activation function = sigmoid; batch size = 16; epoch = 100; model optimizer = Adam
RNN	Number of hidden layers = 3; activation function = sigmoid; batch size = 16; epoch = 100; model optimizer = Adam
LSTM	Number of hidden layers = 3; activation function = sigmoid; batch size = 16; epoch = 100; model optimizer = Adam

F-measure denotes the harmonic mean of precision and recall rate, which balances false positives and false negatives:

$$\mathrm{F-measure} = 2 * \frac{\frac{\mathrm{TP}}{\mathrm{TP+FP}} * \frac{\mathrm{TP}}{\mathrm{TP+FN}}}{\frac{\mathrm{TP}}{\mathrm{TP+FP}} + \frac{\mathrm{TP}}{\mathrm{TP+FN}}}. \tag{2}$$

Type I error and Type II errors are denoted as follows:

$$\mathrm{Type\,I\,error} = \frac{\mathrm{FP}}{\mathrm{FP+TN}}, \tag{3}$$

$$\mathrm{Type\,II\,error} = \frac{\mathrm{FN}}{\mathrm{TP+FN}}. \tag{4}$$

Moreover, AUC is also a widely used evaluation metric, which is independent of the classes' prior distribution, and can portray the probability of a classifier's performance (Hajek & Henriques, 2017; Wang et al., 2018).

4 | EXPERIMENTAL RESULTS

The experiment in this study was completed by implementing the process outlined in Section 3. As previously mentioned, feature selection and model construction comprised the major tasks of this experiment. This section first presents the selected features from the T-3, T-4, and T-5 data sets, and then evaluates the classification results of the multiple models.

4.1 | Feature selection results

This study employed RFE-SVM to select the valuable features for FDP; Table 6 presents the features selected across the three data sets. 25, 43, and 26 variables were selected from the T-3, T-4, and T-5 data sets, respectively, meaning that the T-4 data set contained more valuable information than the other two. It is also noteworthy that financial features accounted for the largest part of the selected features across all of the data sets, followed by textual features, and then management features. Specifically, the selected financial features covered all six aspects (profitability, solvency, financial structure, operational capabilities, development capacity, and stock ratios) that were here taken to comprise financial factors; the equity structure and auditors' opinions were found to play an important role in selected management features; and the selected textual features contained more variables drawn from annual reports than from MD&A data.

In addition, the features repeatedly selected by all three time-span data sets are presented in Table 7, and can be regarded as the most significant features. The financial features covering all six aspects mainly included the return on total assets, current assets ratio, current debt ratio, receivable turnover, net asset growth per share, net asset per share, etc.; these factors represent the assets, debt, and stock information. There was only one management factor selected across all three data sets, indicating that the percentage of shares owned by the companies' top 10 shareholders was the key element of information for measuring the financial status of the company. With regard to textual factors, the number of

TABLE 6 Selected features for from data sets

		T-3 data set	T-4 data set	T-5 data set
Financial factors	Profitability	x1, x8	x1, x2, x3, x8, x10	x1, x2, x6
	Solvency	x11, x12, x14	x11, x12, x19	x11, x12, x14, x19
	Financial structure	x21, x25	x21, x24, x25	x21, x23, x25
	Operational capabilities	x26, x28	x26, x28, x31, x32	x26, x28, x30
	Development capacity	x34, x35, x38	x34, x35, x37, x38	x36, x37, x38
	Stock ratios	x43, x45	x39, x41, x42, x43, x44, x45	x39, x42, x43
Management factors	Board structure		x48	
	Equity structure	x53	x52, x53, x54	x53
	Internal control information		x55	
	Auditors' opinions	x59	x59	
Textual factors	MD&A	x61, x62, x63	x60, x61, x62, x63	x61, x63
	Annual reports	x64, x66, x67, x69, x72, x73	x64, x65, x66, x67, x68, x69, x70, x72	x64, x65, x66, x69

TABLE 7 Selected features for from data set

	Types of factor	Selected features
Financial factors	Profitability	x1
	Solvency	x11, x12
	Financial structure	x21, x25
	Operational capabilities	x26, x28
	Development capacity	x38
	Stock ratios	x43
Management factors	Equity structure	x53
Textual factors	MD&A	x61, x63
	Annual reports	x64, x66, x69

words, sentences, and both negative and positive terms in MD&A or annual reports were found to play an important role in FDP.

4.2 | Classification results of multiple models

This study evaluated the experimental results pertaining to the different features, models, and time spans, and presented the results of evaluation metrics mentioned in Section 3.4.

Table 8 illustrates the findings with regard to accuracy (Acc), F-measure (F) and AUC, which represent the predictive performance of multiple models; the highest values of the models under each time span and feature are presented in bold font.

Many models yielded comparable results in terms of accuracy and F-measure. Specifically, RNN obtained the highest accuracy and F-measure for F1 + F3 and F1 + F2 + F3 in all three time spans, and the highest accuracy for F1 + F2 in the T-4 time span. It is also noteworthy that the highest accuracy (96.46%) and F-measure (96.01%) were obtained by RNN for F1 + F2 + F3 (T-4), followed by 94.57% (accuracy) and 94.52% (F-measure) for F1 + F3 (T-4), and 93.90% (accuracy) and 93.69% (F-measure) for F1 + F2 + F3 (T-5). In addition, RF, GBDT, and stacking achieved a satisfactory F1 performance in all three time spans. According to the experimental results, RF and GBDT obtained the highest accuracy and F-measure on F1 in time span T-3, while GBDT and stacking demonstrated the highest accuracy and F-measure for F1 in time spans T-4 and T-5. Moreover, voting and Xgboost obtained the highest accuracy and F-measure for F1 + F2 in time spans T-3 and T-5, respectively.

With AUC, both DNN and RNN were seen to achieve an outstanding performance for F1 + F2, F1 + F3, and F1 + F2 + F3 across all three time spans. Specifically, the top two highest AUC values were obtained by DNN, at 93.89% for F1 + F2 + F3 (T-4) and 93.84% for F1 + F2 + F3 (T-3). With the F1 + F3 data, RNN yielded the highest AUC of 93.73% (T-3), 93.56% (T-5), and 93.54% (T-4), while DNN performed better in obtaining AUC at F1 + F2, at 88.56% (T-4), 87.98% (T-5), and 87.84% (T-3). In the ensemble models, RF, GBDT and stacking played a significant role in analyzing financial features. To be specific, RF and GBDT achieved the highest AUC for F1 (T-3), while GBDT and stacking manifested the highest AUC for F1 in time spans T-4 and T-5.

In general, from the feature perspective, the financial features in this study were seen to achieve around 85% accuracy in predicting financial distress, which was similar to that found in earlier studies (Farooq & Qamar, 2019; Liang et al., 2018), thus proving that the financial features selected here are valuable and efficient. This study also highlights that combining financial, management, and textual features can achieve the highest accuracy in FDP. Moreover, compared with management features, textual features were here found to play a more important role in supplementing the traditional financial features of FDP. With regard to the models developed, the ensemble models achieved better FDP performance in terms of analyzing financial and management features, while the deep learning models showed certain advantages in predicting financial distress when incorporating textual features. From the perspective of the different time spans examined, the financial features were found to have a similar prediction ability across all time spans, with T-4 achieving better performance when management and/or textual features supplemented the financial features, followed by the performance of T-5 and T-3, in that order.

Table 9 presents the Type I and Type II errors, showing the performance of the different models and time spans in identifying healthy and ST companies, with the best values highlighted in bold font. Specially, RNN was

TABLE 8 Accuracy, F-measure and AUC results of multiple models

	Feature	F1			F1 + F2			F1 + F3			F1 + F2 + F3		
	Method	**Acc**	**F**	**AUC**	**Acc**	**F**	**AUC**	**Acc**	**F**	**AUC**	**Acc**	**F**	**AUC**
T-3	LR	76.19	75.61	74.31	78.57	78.22	77.08	88.09	87.90	86.81	88.09	87.90	86.81
	SVM	80.85	80.11	78.47	80.95	80.77	79.86	80.95	80.49	79.16	83.33	83.38	83.33
	DT	73.81	73.70	72.91	76.19	76.19	75.69	78.57	78.63	78.47	80.66	79.02	83.53
	RF	**85.70**	**85.57**	**84.72**	85.70	85.37	84.02	88.09	88.04	87.50	89.74	88.74	87.50
	Bagging SVM	80.95	80.49	79.16	80.95	80.77	79.86	88.09	87.90	86.81	90.47	90.38	89.58
	GBDT	**85.70**	**85.57**	**84.72**	85.70	85.57	85.41	86.32	85.61	86.26	88.09	87.90	86.81
	Xgboost	80.95	80.77	80.86	83.33	80.49	82.63	83.33	83.26	82.63	85.71	85.37	84.02
	Voting	83.33	83.26	82.63	**85.71**	**85.68**	85.71	88.09	88.17	87.77	90.47	90.51	90.97
	Stacking	80.85	80.11	78.47	83.33	83.38	83.33	85.71	85.71	85.41	88.09	87.90	86.80
	DNN	82.31	79.81	84.74	84.55	83.66	**87.84**	91.51	90.49	93.56	92.45	92.01	**93.84**
	RNN	82.43	80.37	84.91	84.07	82.57	87.53	**92.33**	**91.85**	**93.73**	**93.86**	**93.42**	93.59
	LSTM	81.66	79.02	83.53	83.02	81.77	86.28	89.50	88.16	90.65	89.97	89.14	92.10
T-4	LR	73.80	73.38	72.22	73.80	73.88	73.61	88.09	88.04	87.50	88.09	88.04	87.50
	SVM	76.19	75.96	75.00	76.19	76.19	75.69	85.70	85.57	84.72	85.71	85.71	85.41
	DT	73.80	73.70	72.91	80.95	80.77	79.86	83.33	83.26	82.63	88.09	88.09	88.04
	RF	80.95	80.95	80.55	83.33	83.06	81.94	83.33	83.26	82.63	88.09	87.90	86.81
	Bagging SVM	78.57	78.48	77.77	78.57	78.48	77.77	90.47	90.38	89.58	90.47	90.38	89.58
	GBDT	**85.71**	**85.71**	**85.41**	86.32	85.61	86.26	88.09	87.90	86.80	88.09	87.90	86.81
	Xgboost	80.95	80.95	80.55	83.33	83.26	82.63	83.33	83.26	82.63	85.71	85.71	85.41
	Voting	80.95	81.21	83.41	83.88	83.24	83.33	85.71	85.71	85.58	90.47	90.38	89.58
	Stacking	**85.71**	**85.71**	**85.41**	88.09	87.90	86.80	88.09	88.09	88.04	90.47	90.51	90.97
	DNN	83.13	83.22	84.53	88.32	87.72	**88.56**	93.63	93.09	93.21	95.87	95.34	**93.89**
	RNN	83.37	81.97	84.89	**90.44**	**90.34**	88.01	**94.57**	**94.52**	**93.54**	**96.46**	**96.01**	93.13
	LSTM	80.66	80.07	82.58	83.37	82.29	85.58	90.68	90.23	91.54	91.98	91.72	90.70
T-5	LR	78.57	78.22	77.08	80.95	80.77	79.86	80.95	81.09	81.25	80.95	81.03	81.25
	SVM	78.57	78.63	78.47	80.95	80.77	79.86	80.95	80.95	80.55	80.95	81.09	80.55
	DT	71.42	71.55	72.22	73.80	73.94	74.30	78.57	78.22	77.08	83.33	83.06	81.94
	RF	78.57	78.63	78.47	83.33	83.26	82.63	85.71	85.71	85.41	85.71	85.71	85.41
	Bagging SVM	78.57	78.63	78.47	80.95	80.77	79.86	80.95	81.03	81.25	83.33	83.38	83.33
	GBDT	**85.71**	**85.37**	**84.02**	85.71	85.68	85.71	88.09	87.90	86.80	88.09	87.90	86.81
	Xgboost	83.33	83.06	81.94	**88.09**	**88.04**	87.50	88.09	88.17	87.77	90.47	90.38	89.58
	Voting	80.95	81.03	81.25	83.33	83.30	83.63	85.71	85.68	85.71	85.71	85.64	86.13
	Stacking	**85.71**	**85.37**	**84.02**	85.71	85.71	85.41	88.09	87.90	86.80	88.09	88.04	87.50
	DNN	85.14	84.62	83.36	87.14	86.52	**87.98**	89.26	89.18	90.25	89.74	88.74	90.82
	RNN	83.13	82.22	84.53	86.43	86.04	86.26	**92.39**	**91.97**	**93.56**	**93.90**	**93.69**	**92.21**
	LSTM	82.30	80.79	85.01	85.54	84.36	86.60	87.02	85.95	88.99	87.38	87.35	89.02

found to have minor Type I and Type II errors for F1 + F3 and F1 + F2 + F3, while the ensemble models performed better for F1 and F1 + F2. Specifically, regarding Type I errors, with the F1 + F2 + F3 data RNN achieved the minimum Type I errors at 3.30% (T-4), followed by 5.54% (T-3) and 6.65% (T-5). Moreover, RNN also obtained a minor Type I error for F1 + F3, at 5.10% (T-4), 7.30% (T-3), and 8.22% (T-5). With the exception of RNN, ensemble models such as GBDT, RF and Xgboost obtained better results for F1 and F1 + F2. In terms of Type II errors, RNN also achieved the best results for F1 + F3 and F1 + F2 + F3 across all time spans, at 3.34%

TABLE 9 Type I and type II error of multiple models

	Feature / Method	F1		F1 + F2		F1 + F3		F1 + F2 + F3	
		Type I error	Type II error	Type I error	Type II error	Type I error	Type II error	Type I error	Type II error
T-3	LR	23.51	28.05	21.27	24.22	11.47	16.82	11.27	14.81
	SVM	18.23	26.33	18.75	19.23	18.68	21.42	16.17	13.85
	DT	25.41	24.03	23.79	21.16	20.86	18.83	18.70	16.34
	RF	14.10	16.53	**13.49**	17.85	11.37	18.51	9.49	8.78
	Bagging SVM	18.28	21.42	18.75	19.23	11.48	14.81	9.09	11.53
	GBDT	**13.53**	15.34	13.56	11.04	13.40	15.73	11.71	9.50
	Xgboost	18.97	17.03	16.03	14.28	16.64	16.00	13.92	13.04
	Voting	16.53	**13.11**	13.53	**10.52**	11.43	17.69	9.06	15.00
	Stacking	18.75	23.53	16.23	13.04	13.92	15.81	11.43	10.52
	DNN	17.41	17.56	14.77	15.18	9.09	10.58	6.97	7.19
	RNN	17.32	18.75	15.43	13.76	**7.30**	**8.21**	**5.54**	**7.06**
	LSTM	17.57	18.38	16.42	16.84	9.83	10.32	9.19	8.44
T-4	LR	25.52	24.92	26.08	21.73	11.69	12.00	11.69	18.53
	SVM	23.19	23.07	23.36	20.83	13.95	14.81	13.96	17.83
	DT	25.62	26.03	18.59	16.34	16.16	15.45	11.80	12.32
	RF	18.83	19.38	16.45	18.51	16.45	16.00	11.78	10.66
	Bagging SVM	21.09	20.70	21.10	20.00	9.31	11.53	9.42	11.53
	GBDT	**13.63**	12.50	13.55	12.67	11.78	14.81	11.58	10.20
	Xgboost	18.85	16.66	16.46	12.54	16.21	15.43	13.96	11.54
	Voting	18.87	**11.11**	15.90	**10.23**	14.05	15.78	9.09	6.25
	Stacking	13.87	12.53	11.77	11.56	11.67	10.16	9.21	8.50
	DNN	15.76	14.81	11.35	10.68	5.86	5.67	3.91	4.92
	RNN	16.41	15.61	**9.34**	10.55	**5.10**	**4.16**	**3.30**	**3.34**
	LSTM	19.12	16.06	15.89	14.05	9.10	10.52	7.58	7.35
T-5	LR	21.22	17.66	18.70	14.53	18.83	20.83	18.72	20.83
	SVM	20.92	16.79	18.93	12.50	18.28	16.66	18.61	17.39
	DT	28.47	30.51	25.87	26.32	20.88	19.04	16.45	15.06
	RF	16.66	12.50	16.45	12.50	14.18	13.23	14.07	13.39
	Bagging SVM	21.32	17.39	18.78	19.23	19.03	13.04	16.32	13.63
	GBDT	20.79	22.36	14.05	21.04	11.69	10.67	11.47	10.39
	Xgboost	**14.07**	18.51	11.90	12.00	11.58	19.23	9.18	8.68
	Voting	15.93	25.00	16.44	**10.52**	14.16	10.52	13.74	20.83
	Stacking	18.46	**12.33**	13.74	16.66	11.79	15.45	11.58	22.22
	DNN	14.22	13.50	**11.89**	11.30	10.50	9.19	9.60	9.54
	RNN	16.45	14.98	13.35	11.62	**8.22**	**7.47**	**6.65**	**5.71**
	LSTM	17.33	17.16	14.13	12.96	12.43	11.77	12.4	9.26

for F1 + F2 + F3 (T-4), 5.71% for F1 + F2 + F3 (T-5), 7.06% for F1 + F2 + F3 (T-3), 4.16% for F1 + F3 (T-4), 7.47% for F1 + F3 (T-5), and 8.21% for F1 + F3 (T-3). It should be noted that voting demonstrated a satisfactory performance for F1 and F1 + F2 across all time spans, except for stacking at F1 (T-5).

In general, the models obtained relatively balanced values in the Type I and Type II errors across different features combinations and time spans. To be specific, compared with T-3, T-4 and T-5 yielded minor Type II errors, indicating that the T-4 and T-5 information was more likely to provide a higher accuracy in identifying ST companies. In addition, most of the models were found to have larger Type I errors and minor Type II errors. Thus it may be concluded that the empirical results of this study can obtain relative low costs of failing to detect financially distressed companies (Type II error), which is more significant than predicting financial distress when the latter is not present (Type I error) (Bauer & Agarwal, 2014; Hajek & Henriques, 2017).

5 | DISCUSSION AND IMPLICATIONS

This study proposes a framework combining financial, management, and textual factors to predict which companies will become financially distressed, based on different time spans and comparing the prediction performance of multiple models. As previously mentioned, this study attempts to answer three questions: Which are the key influencing factors in the FDP of listed Chinese companies? Which models perform better than others in the FDP of listed Chinese companies? How early can signs of financial distress be predicted in listed Chinese companies?

For the first research question, in order to discern the key factors involved in FDP, this study applied a wrapper approach to select meaningful features. The results show that financial features accounted for the majority of the selected features, in line with previous research (Farooq & Qamar, 2019; Y. Jiang & Jones, 2018), since these ratios can directly reflect the financial status of the listed companies. Only a few management factors were selected, particularly for time spans T-3 and T-5, and, compared with board structure and internal control information, equity structure and auditors' opinions emerged as more significant in terms of the FDP of listed Chinese companies. These results support those found by Xie et al. (2011). While they may seem to contradict the notion that board structure, such as the number of independent board numbers and CEO duality, can play an important role in distinguishing between financially distressed and healthy companies in Spain (Manzaneque et al., 2016), these results can be explained by the difference in companies between China and Spain. Regarding textual factors, the text features from both the MD&A and annual reports were found to be important for use in FDP, which has also been proven by prior studies (du Jardin, 2016; Myšková & Hájek, 2017; Wang et al., 2018).

One striking result is that the basic textual features, including the number of words and sentences, show more advantages than other textual factors in terms of FDP; to be specific, sentiment features, such as the sentiment score of annual reports (features x71, x72, x73, x74) were found to be not fully regarded as key factors in all of the time spans under study, supporting the work of Wang et al. (2018). This may be due to the sentiment dictionary used in CNRDS, as this database employs an L&M dictionary and NTUSD to analyze the sentiment features of MD&A and annual reports. Although these two dictionaries are widely applied in analyzing the financial reports of listed Chinese companies (X. Li et al., 2019; B. Zhou et al., 2018), the translated L&M dictionary cannot entirely identify the sentiment drawn from Chinese financial reports, owing to the linguistic descriptive discrepancies between Chinese and English. Moreover, the NTUSD is the general Chinese emotion dictionary, which cannot accurately recognize the sentiment of exclusive financial vocabulary. Thus, building the proprietary Chinese financial sentiment dictionary is significant for analyzing Chinese financial texts.

After integrating three different types of selected features into four groups (F1, F1 + F2, F1 + F3, F1 + F2 + F3) as inputs for the multiple models, it was noteworthy that F1 + F2 + F3 achieved a better performance than other feature sets in FDP; in comparison with F1 + F2, F1 + F3 yielded a higher accuracy in predicting financial distress, which is a unique finding of this study. This indicates that not only financial factors, but also management and textual factors, have a significant effect on FDP. This is because, given that financial distress is a dynamic and long-term process, management and textual features contain external information that cannot be covered by financial features alone. It also emerged that the textual information from financial reports can play a more important role than management factors in supplementing the traditional financial factors involved in FDP.

For the second research question, in interpreting the prediction results of the multiple models in this research, to the best of our knowledge this is the first study that applies single, ensemble, and deep learning models for the FDP of listed Chinese companies. In general, ensemble and deep learning models outperformed the single classifiers. Ensemble models such as GBDT, RF, and stacking achieved higher predictive accuracy for F1 and F1 + F2, whereas the deep learning models performed better in analyzing F1 + F3 and F1 + F2 + F3. This suggests that, in comparison with ensemble classifiers, deep learning models may have advantages in analyzing data with relatively high dimensionality. In addition, the current results make it possible to state that deep learning models can achieve satisfactory results in analyzing small data sets, as supported by the work of Alexandropoulos et al. (2019). Moreover, when comparing the performance of three deep learning models, interestingly, simpler models (DNN and RNN) were found to be more effective than the complicated one (LSTM), which is similar to the last study to find that simpler deep learning models are more likely to achieve higher predictive accuracy in bankruptcy predictions (Mai et al., 2019).

The evaluation metrics of the Type I and II errors suggest that most of models utilized in this study were able to achieve a low rate of ST companies incorrectly being classified as healthy ones. For the results of accuracy, F-measure, and AUC results, the deep learning models also performed better in dealing with data sets (F1 + F3 and F1 + F2 + F3), while ensemble classifiers were good at analyzing F1 and F1 + F2. This indicates that the models and features used in this study were able to attain a high prediction accuracy in detecting financially distressed firms, which can be regarded as the central task of intelligent FDP (Bauer & Agarwal, 2014; Hajek & Henriques, 2017).

For the third research question, considering the predictive performance in terms of different time spans, the T-4 data outperformed other data sets. To be specific, there was little difference found between the three time spans when only utilizing financial features for FDP, although some models achieved higher accuracy in the T-3 time span, since it is easier to predict financial distress when closer to the ST labeling years (Geng et al., 2015). However, after integrating management and/or textual features with financial features, T-4 and T-5 showed advantages in predicting financial distress, which is similar to previous research (Wang et al., 2018). Given that it takes an extended period of time for healthy firms to turn into ST-labeled ones, management and textual features may contain the implicit signals of financial distress earlier than financial features. A striking result here is that the T-4 data achieved the better accuracy in FDP than T-5, which is a unique finding of this study. For instance, the best predictive performance was obtained by RNN for F1 + F2 + F3 in the time span T-4. This indicates that the earlier management and textual features cannot totally maintain the better predictive performance in FDP—such features can only achieve the best effect within a certain duration.

Moreover, this study may be seen as significant in terms of both theoretical and practical aspects. From a theoretical perspective, this research creates a framework to predict financial distress that integrates financial, management, and textual factors in three time spans, based on 12 machine learning models; this integration framework is developed from earlier literature (Geng et al., 2015; Wang et al., 2018). This mixed approach has not gained much attention in FDP and can potentially be expanded to analyze other financially related questions, such as predictions of fraud and bankruptcy. This study also designs an index for FDP that covers financial, management, and textual information, thus building on earlier studies (Y. Jiang & Jones, 2018; Liang et al., 2016; Myšková & Hájek, 2017; Xie et al., 2011). From the practical perspective, this study reveals that management and textual factors can supplement the traditional financial factors involved in FDP, especially textual factors, as well as highlighting that earlier relevant data possess advantages in terms of the prediction of financial distress. Therefore, the results of this study can also be harnessed to construct useful early warning signals to the relevant economic agents, which, to some extent, can have a positive effect on avoiding financial losses.

6 | CONCLUSION AND FUTURE WORK

This study develops an FDP framework that integrates multiple predictive factors, time spans, and classifier models, and offers insight into the financial distress features of listed Chinese companies. The study also constructs an index containing 74 factors covering the financial, management and textual features of these listed companies, employing the wrapper-based method to select valuable features, then regarding the selected features as inputs for the multiple classifier models, and, finally, analyzing the results of the experiment and answering three research questions. The findings of this study present a different picture to those of previous studies. Notably, this study combines financial features with management and textual features, which is an expansion of earlier studies (Xie et al., 2011; Liang et al., 2016; Myšková & Hájek, 2017; Y. Jiang & Jones, 2018), and highlights that textual features can play a more significant role in supplementing the traditional financial features of FDP than management ones. This study also finds that deep learning models can promote the predictive performance of FDP. In addition, this study suggests that earlier management and textual features can achieve the best effect within a certain duration (T-4 in this study). These new findings extend the existing research on the FDP of listed Chinese companies and, it is hoped, will have a positive effect on identifying financial distressed companies and reducing the potential financial losses of concerned investors.

Several limitations also should be taken into consideration when generalizing the results of this study. This study applied the widely used L&M and NTUSD dictionaries to analyze the sentiment of listed Chinese companies' financial reports, whereas many of such features were not selected as valuable features in FDP. Future studies could therefore construct a Chinese financial sentiment lexicon that would be significant for analyzing Chinese financial texts. This study was also limited by the utilization of a balanced sample of ST and healthy firms. Future research could assess the effects of increasing the number of healthy firms in the training data, whereby some models designed for imbalanced data could be added into the FDP framework. In addition, this study only collected financially distressed samples from 2014 to 2018, and criticized the ST risk prediction models based on data for T-3, T-4, and T-5 time spans. Future studies could benefit from collecting ST samples over a longer duration of time; it is also possible to use the data of T-6 to T-10 time spans in order to identify the signs of financial deterioration at an earlier time, which is something for future research to consider.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NSFC; 71673209).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the Chinese Research Data Services (CNRDS) database and the China Stock Market & Accounting Research (CSMAR) database. These data were derived from the following resources available in the public domain: http://www.gtarsc.com and http://www.cnrds.com.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Retrieved from http://download.tensorflow.org/paper/whitepaper2015.pdf

Alexandropoulos, S. A. N., Aridas, C. K., Kotsiantis, S. B., & Vrahatis, M. N. (2019, May). A deep dense neural network for bankruptcy prediction. In *International Conference on Engineering Applications of Neural Networks* (pp. 435–444). Berlin, Germany: Springer.

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, *45*(1), 110–122.

Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance*, *40*, 432–442.

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

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision Support Systems*, *50*(1), 164–175.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, *40*(1), 16–28.

Chen, C., Kim, J. B., Wei, M., & Zhang, H. (2018). Linguistic information quality in customers' forward-looking disclosures and suppliers' investment decisions. *Contemporary Accounting Research*, *36*(3), 1751–1783.

Chen, N., Ribeiro, B., & Chen, A. (2016). Financial credit risk assessment: A recent review. *Artificial Intelligence Review*, *45*(1), 1–23.

Cho, S., Hong, H., & Ha, B. C. (2010). A hybrid approach based on the combination of variable selection using decision trees and case-based reasoning using the Mahalanobis distance: For bankruptcy prediction. *Expert Systems with Applications*, *37*(4), 3482–3488.

Ding, Y., Song, X., & Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*, *34*(4), 3081–3089.

Dong, W., Liao, S., & Liang, L. (2016). Financial statement fraud detection using text mining: A systemic functional linguistics theory perspective. *In 20th Pacific Asia Conference on Information Systems* (p. 188), Chiayi, Taiwan.

du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, *254*(1), 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.

Farooq, U., & Qamar, M. A. J. (2019). Predicting multistage financial distress: Reflections on sampling, feature and model selection criteria. *Journal of Forecasting*, *38*(7), 632–648. https://doi.org/10.1002/for.2588

Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236–247.

Ghazali, A. W., Shafie, N. A., & Sanusi, Z. M. (2015). Earnings management: An analysis of opportunistic behaviour, monitoring mechanism and financial distress. *Procedia Economics and Finance*, *28*, 190–201.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1–3), 389–422.

Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud: A comparative study of machine learning methods. *Knowledge-Based Systems*, *128*, 139–152.

Hájek, P., & Olej, V. (2013). Evaluating sentiment in annual reports for financial distress prediction using neural networks and support vector machines. In *International Conference on Engineering Applications of Neural Networks* (pp. 1–10). Berlin, Germany: Springer.

Hajek, P., Olej, V., & Myskova, R. (2014). Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making. *Technological and Economic Development of Economy*, *20*(4), 721–738.

Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, *33*(2), 434–440.

Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, *132*(1), 126–149.

Jiang, Y., & Jones, S. (2018). Corporate distress prediction in China: A machine learning approach. *Accounting and Finance*, *58*(4), 1063–1109.

Jo, N. O., & Shin, K. S. (2016). Bankruptcy prediction modeling using qualitative information based on big data analytics. *Journal of Intelligence and Information Systems*, *22*(2), 33–56.

Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, *104*, 38–48.

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), 1–28.

Li, H., & Sun, J. (2009). Gaussian case-based reasoning for business failure prediction with empirical data in China. *Information Sciences*, *179*(1–2), 89–108.

Li, J., Qin, Y., Yi, D., Li, Y., & Shen, Y. (2014). Feature selection for support vector machine in the study of financial early warning system. *Quality and Reliability Engineering International*, *30*(6), 867–877.

Li, X., Liu, J., & Wang, K. (2019). Pledgee competition, strategic disclosure, and future crash risk. *China Journal of Accounting Research*, *12*(3), 271–291.

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*(2), 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*(2), 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.

Liu, J., Wu, C., & Li, Y. (2019). Improving financial distress prediction using financial network-based information and GA-based gradient boosting method. *Computational Economics*, *53*(2), 851–872.

Lorca, P., Landajo, M., & Andrés, J. D. (2014). Nonparametric quantile regression-based classifiers for bankruptcy forecasting. *Journal of Forecasting*, *33*(2), 124–133.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, *66*(1), 35–65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187–1230.

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, *274*(2), 743–758.

Malafronte, I., Porzio, C., & Starita, M. G. (2013). Disclosure practices and financial crisis: Empirical evidences in the European insurance industry. *The British Accounting Review*, *38*(4), 387–404.

Maldonado, S., Bravo, C., López, J., & Pérez, J. (2017). Integrated framework for profit-based feature selection and SVM classification in credit scoring. *Decision Support Systems*, *104*, 113–121.

Manzaneque, M., García-Pérez-De-Lema, D., & Antón Renart, M. (2015). Bootstrap replacement to validate the influence of the economic cycle on the structure and the accuracy level of business failure prediction models. *Journal of Forecasting*, *34*(4), 275–289.

Manzaneque, M., Priego, A. M., & Merino, E. (2016). Corporate governance effect on financial distress likelihood: Evidence from Spain. *Revista de Contabilidad*, *19*(1), 111–121.

Martin, A., Gayathri, V., Saranya, G., Gayathri, P., & Venkatesan, P. (2011). A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy C-means and MARS. *International Journal on Soft Computing*, *2*(1). https://doi.org/10.5121/ijsc.2011.2102

Myšková, R., & Hájek, P. (2017). Comprehensive assessment of firm financial performance using financial ratios and linguistic analysis of annual reports. *Journal of International Studies*, *10*(4), 96–108.

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(October), 2825–2830.

Sun, J., & Li, H. (2008). Data mining method for listed companies' financial distress prediction. *Knowledge-Based Systems*, *21*(1), 1–5.

Sun, J., Li, H., Huang, Q. H., & He, K. Y. (2014). Predicting 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.

Suntraruk, P. (2010). A review of statistical methods in the financial distress literature. *AU Journal of Management*, *8*(2), 31–41.

Tykvová, T., & Borell, M. (2012). Do private equity owners increase risk of financial distress and bankruptcy? *Journal of Corporate Finance*, *18*(1), 138–150.

Wang, G., Chen, G., & Chu, Y. (2018). A new random subspace method incorporating sentiment and textual information for financial distress prediction. *Electronic Commerce Research and Applications*, *29*, 30–49.

Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, *38*(1), 223–230.

Wanke, P., Barros, C. P., & Faria, J. R. (2015). Financial distress drivers in Brazilian banks: A dynamic slacks approach. *European Journal of Operational Research*, *240*(1), 258–268.

Xie, C., Luo, C., & Yu, X. (2011). Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies. *Quality and Quantity*, *45*(3), 671–686.

Zhou, B., Zhang, C., & Zeng, Q. (2018). Does the rhetoric always hide bad intention: Annual report's tone and stock crash risk. *China Journal of Accounting Studies*, *6*(2), 178–205.

Zhou, L., Lai, K. K., & Yen, J. (2014). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science*, *45*(3), 241–253.

Zhou, L., Tam, K. P., & Fujita, H. (2016). Predicting the listing status of Chinese listed companies with multi-class classification models. *Information Sciences*, *328*, 222–236.

AUTHOR BIOGRAPHIES

Xiaobo Tang is a Professor of Information Science at the School of Information Management, Wuhan University, China. He holds a PhD in Management Science and Engineering from Wuhan University. His primary research interests include knowledge organization and intelligence analysis.

Shixuan Li is a Doctoral student at the School of Information Management, Wuhan University, China. She received an MS degree in Management and Information System from The University of Manchester, UK. Her primary research interests include social media analysis, financial textual analysis and knowledge organization. Shixuan Li is the corresponding author and can be contacted at: shixuan.li@hotmail.com

Mingliang Tan is a Doctoral student at the School of Information Management, Wuhan University, China. His primary research interests include business intelligence, financial textual analysis and knowledge organization.

Wenxuan Shi is a Doctoral student at the School of Information Management, Wuhan University, China. Her primary research interests include knowledge organization and financial textual analysis.

How to cite this article: Tang X, Li S, Tan M, Shi W. Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of Forecasting*. 2020;39:769–787. https://doi.org/10.1002/for.2661