



Research article

A deep learning approach of financial distress recognition combining text

Jiawang Li* and **Chongren Wang***

School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan 250014, Shandong, China

* **Correspondence:** Email: wang_earnest@163.com, chongrenw@163.com.

Abstract: The financial distress of listed companies not only harms the interests of internal managers and employees but also brings considerable risks to external investors and other stakeholders. Therefore, it is crucial to construct an efficient financial distress prediction model. However, most existing studies use financial indicators or text features without contextual information to predict financial distress and fail to extract critical details disclosed in Chinese long texts for research. This research introduces an attention mechanism into the deep learning text classification model to deal with the classification of Chinese long text sequences. We combine the financial data and management discussion and analysis Chinese text data in the annual reports of 1642 listed companies in China from 2017 to 2020 in the model and compare the effects of the data on different models. The empirical results show that the performance of deep learning models in financial distress prediction overcomes traditional machine learning models. The addition of the attention mechanism improved the effectiveness of the deep learning model in financial distress prediction. Among the models constructed in this study, the Bi-LSTM+Attention model achieves the best performance in financial distress prediction.

Keywords: financial distress prediction; deep learning; attention mechanism; text analytics; LSTM

1. Introduction

Corporate financial distress is a financial risk, indicating a high probability of corporate bankruptcy, affecting the stability of financial markets and social and economic systems and bringing adverse effects to the global economy. Therefore, financial distress prediction (FDP) has attracted significant attention from stakeholders such as corporate managers, governments and investors. FDP can provide early warning information for corporate risks, help corporate managers take risk control measures to avoid the deterioration of the situation and help investors grasp the profitability of listed companies and adjust investment strategies to reduce expected investment losses [1–3].

FDP is a binary classification problem. Most early studies used corporate financial indicators combined with traditional machine learning models to detect corporate financial distress. When analyzing financial indicators for financial distress classification, statistical methods are used for feature engineering and combined with machine learning models. The commonly used models are Bayesian network, support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), and artificial neural networks (ANN). Deep learning technology has developed rapidly in recent years with the improvement of hardware computing power. Gradually, researchers began to introduce deep learning models into FDP research, such as convolutional neural networks (CNN) and recurrent neural networks (RNN).

The financial indicators used in FDP are usually extracted from the data of the company's annual financial statements. Companies in financial distress often manipulate information disclosure, thereby preventing the negative impact of financial distress on the enterprise [4, 5]. In addition, using financial indicator data to detect financial distress, the information source is single and limited, and it is challenging to identify enterprise managers' attitudes and manipulation traces. With the application of deep learning technology in natural language processing (NLP) research, deep learning text classification models have been introduced into FDP research. Experts in finance and accounting have begun to use powerful computing hardware and artificial intelligence technology to extract information from text data to identify financial distress [6, 7]. An enterprise's disclosure text comprises corporate financial reports, annual reports and prospectuses, of which the management discussion and analysis (MD&A) section of the company's annual report is the most studied MD&A text and is the primary channel for the disclosure of business information and financial information. It expresses the company management's views on potential opportunities and challenges in development to the outside world. So, the disclosed text features can improve the effectiveness of financial analysis [8–11]. However, the MD&A text content is long and unstructured. Determining how to deal with the long text and effectively extract the semantic features of the text is the main problem.

This study combines a deep learning text classification model with an attention mechanism to adapt to MD&A Chinese long text classification in FDP. Usually, the text sequence of the MD&A part of the annual reports of listed companies in China is very long, and its length is usually more than several thousand Chinese characters. In the MD&A Chinese text, the second half is mostly a numerical description, and the textual content that contains managers' views on enterprise development is in the first half. In our experiments, first, the text is preprocessed, and then the first 2000 words are selected as the MD&A Chinese text input of fixed length through the experiment, which covers all the keywords of the original Chinese text content. Second, we combine the attention mechanism with the gated recurrent unit (GRU) and bidirectional long-short term memory (Bi-LSTM) models to construct the GRU+Attention and Bi-LSTM+Attention models, capture time series information by using multiple hidden layer structures of GRU and LSTM and then use the attention mechanism further to summarize the time series key information in the message. Finally, the entire MD&A Chinese long text is extracted as a vector containing the critical information of the text, which is combined with the financial indicator vector to predict whether the company is facing financial distress. This paper constructed a deep learning text classification model adding an attention mechanism. Financial indicators were added for FDP research, and the effects of traditional machine learning models were compared. The experimental results show that the performance of the proposed Bi-LSTM+Attention model overcomes other deep learning and traditional machine models in area

under receiver operating characteristic curve (AUC-ROC) and area under precision-recall curve (PR-AUC).

The main contributions of this study to the research field of FDP are the following: For textual data, we validate the effectiveness of MD&A textual data in FDP. For model construction, the model with attention mechanism has improved the ability to extract critical information from Chinese long text sequences, and the prediction effect of financial distress is enhanced. For the prediction model, the Bi-LSTM+Attention model has a better binary classification effect than other models in FDP. It extracts Chinese text sequence information through the hidden layer in the bidirectional LSTM layer and uses the attention mechanism to increase the weight of important information in the hidden layer. As the model recognizes the critical information of the MD&A Chinese text, the prediction effect of financial distress is improved.

The rest of the paper is as follows. We review related research on FDP in Section 2. We present the flow of the study and the main model structure in Section 3. We describe the selection of samples, data preprocessing process, model settings and evaluation metrics in Section 4. We present the experimental results of each model in Section 5. We conclude and summarize the main findings in Section 6.

2. Literature review

Financial distress occurs when a company faces financial difficulties and faces financial risks. In the research on financial distress, different scholars have given different views. Beaver et al. defined the default of preferred dividends and debt default as financial distress [12]. Deakin et al. believed that the sign of financial distress is that the enterprise is insolvent or liquidated for the benefit of creditors [13]. Carmichael et al. believed that financial distress is a form of corporate debt default or insufficient funds [14]. Zmijewski et al. defined financial distress as filing for bankruptcy [15]. Altman et al. described financial distress in their research and believed that corporate bankruptcy was the most suitable scenario for financial distress [16]. Dimitras et al. defined financial distress as a situation in which a company cannot pay suppliers and preferred shareholders, overdraft bills and the company is legally bankrupt [17]. Ross et al. pointed out that financial distress is when a company's operations are forced to take corrective measures due to insufficient cash flow [18].

Most scholars choose to study special treatment (ST) companies as samples of financial distress. Ding et al. used statistical methods to compare and verify the degree of correlation between ST companies and the samples of financial distress. They found that companies classified as ST had an increased probability of financial deterioration and financial distress in the next year. [19]. Geng et al. found that ST companies are a good representative of financial distress, as they are more likely to go bankrupt in the future [20]. Ruan et al. used the ST labels of listed companies to indicate whether a firm is in financial distress [21]. In general, companies are marked as ST for two reasons: The listed company has incurred losses for two consecutive years after being audited by the accounting company. The other is that the net earnings per share of public companies are less than the par value of their shares. Typically, public companies marked as ST face severe financial deterioration, consistent with financial distress.

In traditional FDP research, most researchers first extract features from financial indicators and then combine statistical models or machine learning models to predict the financial distress of enterprises. The financial data of enterprises are often very related to their actual financial situation [22]. In the early

research, Beaver et al. pioneered the use of financial ratios to predict the financial distress of enterprises and combined them with a univariate discriminant model for analysis. The results showed that financial characteristics are effective in FDP research [12]. Altman et al. used the multivariate discriminant statistical method to study FDP and converted the accounting information of manufacturing enterprises into financial ratios. The results showed that using accounting information to calculate financial ratios can perform well in FDP [23].

Later, scholars paid more attention to sensitive indicators and conducted combined analyses of the company's open-access quantitative data. Some scholars focus on designing models to capture better features. For example, Wang et al. injected feature selection strategies into traditional models. They constructed the FS Boosting ensemble learner method, which can automatically capture the feature diversity of samples to obtain better performance [24]. Huang et al. studied the effect of feature preprocessing of financial indicators on the prediction effect. They constructed a least absolute shrinkage and selection operator (LASSO) selection technique to screen critical financial indicators and found that fewer essential variables can achieve better prediction performance [25]. There are also studies using financial features extracted by financial experts. For example, Zhou et al. combined classifiers such as LDA to test the features of FDP proposed by financial experts. The results showed that these features positively impact the FDP of listed companies in China [26]. For financial risk forecasts in other regions, Alaminos et al. summarized ten financial variables through previous research and used logistic regression to construct a general model that could explain global bankruptcy forecasts [27]. Huang et al. calculated 16 financial ratios based on four basic financial statements of listed companies in Taiwan and compared the performances of six machine learning methods in FDP [28].

With the rapid development of artificial intelligence today, more and more scholars are combining deep learning models in traditional research. Deep learning has many hidden layers for feature transformation. The layer-by-layer feature conversion converts the feature of the sample in the original space to the new feature space, making classification or prediction easier. Deep learning techniques make fewer assumptions about the data [10], and they have solid learning ability, comprehensive coverage, strong adaptability and good portability. Therefore, scholars have gradually introduced deep learning algorithms into various financial market studies, such as stock market prediction [29], bank bankruptcy prediction [30] and customer credit scoring [31]. These studies not only improve the research effectiveness of the problem but also can be used to extract features of supplementary data in addition to financial indicators. FDP is a classic research direction in financial market research and many recent studies in FDP have used deep learning models compared to statistical studies. For example, Halim et al. investigated the performance of deep learning models such as RNN, LSTM and GRU in the FDP of listed companies in Malaysia and found that deep learning methods can achieve better results than traditional machine learning models [3]. Li et al. constructed a sentiment dictionary based on a deep learning framework in the financial domain. The empirical results showed that the sentiment features of annual reports extracted through the dictionary significantly impact FDP [32]. Table 1 summarizes the relevant research on FDP according to six dimensions: author, year, sample size, main models, evaluation metrics and variable types.

Table 1. Analysis of comparisons with FDP models.

Study	Year	Sample size	Main Models	Used metrics	Variables type
Ding et al. [19]	2008	56/194	SVM, MDA, Logit, BPNN	Accuracy	FIN
Alaminos et al. [27]	2016	220/220	Logit	Accuracy	FIN
Huang et al. [25]	2017	156/156	RF, NN, SVM, C5.0	AUC, Accuracy, F1-Score	FIN
Wang et al. [1]	2018	129/1597	IST-RS, RS, SVM	AUC, F1-score, F2-score	FIN+TXT
Mai et al. [10]	2019	477/11350	DL-Embedding, DL-CNN	Accuracy, AUC	FIN+TXT
Jing et al. [33]	2020	38/1914	ZPP-LSTM	AUC, PR-AUC	FIN
Sun et al. [34]	2020	438/2190	Adaboost, SVM	Accuracy, F-value, G-value	FIN
Wang et al. [2]	2020	261/1574	RS2-ER, RS, SVM, AdaBoost	AUC	FIN+TXT
Du et al. [35]	2020	256/6898	CUS-GBDT, XGBoost	AUC, ACC, PPV	FIN
Ruan et al. [21]	2021	862/2978	BERT+HAN	AUC, Precision, Recall, F1-score, F2-score	FIN+TXT
Halim et al. [3]	2021	98/98	RNN, LSTM, GRU	Accuracy, Precision, Recall	FIN
Sun et al. [36]	2021	2235/352 /143/163	SVM	Accuracy	FIN
Wang et al. [37]	2021	220/6599	LDA, NB, XGB	F2-score	FIN
W. et al. [38]	2022	244/5130	CNN, LSTM, GRU	Acc, AUC, F1-Score, F2-Score	FIN+TXT
Liu et al. [39]	2022	380	ResNet, LSTM	RMSE, MA	FIN

Text mining and big data analysis have become hot topics in academia, and the development of text analysis research has promoted the study of traditional accounting and financial issues. Du Jardin developed an improved financial fraud detection system by combining textual features from company annual reports [4]. MD&A text is one of the commonly used text types in research. Qian et al. found text features unique to MD&A, such as vocabulary size, specialized vocabulary, readability and emotional tendencies. Thus, MD&A text promotes interdisciplinary research on the link between accounting information and corporate textual disclosure [11]. In addition, many studies analyze the textual information in MD&A as a research object and use it as a supplement to corporate financial data [7, 9, 10, 38, 40].

There are methods for text information extraction in text analysis research. Word2Vec is a commonly used model based on ANN. It encodes each word by building a bag-of-words model that maps the text to a vector containing the contextual lexical order [41]. Mai et al. used the word vector method to extract the text features of the MD&A part in their study, and the study showed that the deep learning model has superior prediction performance after adding text disclosure indicators. Then, combining accounting ratio and text features can further improve the prediction performance of deep learning models [10]. Xiuguo W et al. extracted the word vectors of the text features of the MD&A part through the Word2Vec method and developed an enhanced financial fraud detection system by using deep learning models such as LSTM and GRU. Their results showed that deep

learning methods outperform traditional machine learning methods [38]. In other studies on FDP, Wang et al. used an ensemble learning approach by combining accounting-based features and textual disclosure information and found that a model containing real disclosed textual features is more efficacious [38]. Ruan et al. found that few scholars in FDP research use pre-trained end-to-end models to process text, so they introduced the Bidirectional Encoder Representations from Transformer (BERT) Chinese pre-training model for word embedding processing in the research. They have added the hierarchical attention neural network (HAN) to alleviate the characteristics of long text feature extraction issues [21].

The structure of RNN gives it advantages in processing sequence information. RNN stores past information and current input by introducing state variables and simultaneously generates the output passed to the next node according to the past information. However, the information update and preservation of long-term and short-term dependencies of latent variables in the RNN model are unstable, and some studies have constructed the LSTM model to solve this problem [42]. Furthermore, due to the gradient vanishing and exploding problems, the effect of semantic features extracted by ordinary RNN models will gradually deteriorate with the increase of text sequence length. Studies have shown that adding an attention mechanism to the problem of super-long text classification can improve the model's ability to identify critical information [43, 44]. The origin of the attention mechanism is to simulate the human brain's attention to things, and it was first used in image research [45]. The attention mechanism calculates the importance of each part by weighted summation of different parts of the output of the time series model and then captures more critical information. Later studies applied the attention mechanism to the field of NLP. Bahdanau et al. constructed a machine translation model based on the attention mechanism by adding the attention mechanism to the language codec to calculate the importance of the input and output of the translation model [46].

3. Methodology

The main purpose of our research is to construct a model suitable for long-text classification for FDP based on the combination of financial indicators. Typically, the long MD&A text of corporate disclosure contains several thousand characters, and the narrative structure is not uniform, which greatly complicates text feature extraction and risk information identification.

3.1. Deep learning architectures for long text

In the commonly used word embedding methods for converting unstructured text data into structured text features, there are two main approaches: One is the word embedding method that does not consider the context, such as TF-IDF and bag-of-words model, by taking the words in the text as a set and representing the text features by the word frequency or the number of times the words appear. This approach is simple to extract the subject of the text, ignoring the timing and context information of the vocabulary, and the representation effect of long text features is poor. The other is the word embedding method that considers the context, such as the Word2Vec word vector and deep learning word embedding models. The Word2Vec word vector model is calculated by a shallow neural network and maps the text to a vector containing the order of context words. The length of the context words involved in the calculation depends on the length of the set sliding window. The deep learning

word embedding model usually refers to the vocabulary encoding layer in deep learning models, such as the word encoding layer based on the RNN model and the word encoding layer based on the Transformer. This research introduces and compares the Word2Vec word vector model and two deep learning word embedding models.

Because the MD&A text content is complex and lengthy, the dimension of the text vector after word embedding is very high, which brings difficulty to text feature extraction. For using Word2Vec word vectors to represent text, this method ignores the time sequence features in sentences and is less compatible with long texts. For RNN models based on time series data, such as LSTM and GRU, the gradient vanishing and exploding problems will also occur when the text is too long. For the Transformer-based BERT pre-training model to embed text as a vector, the BERT model limits the length of the input text sequence to no more than 512 words, which may cause some semantic information not to be calculated. In order to improve these problems, studies have shown that adding an attention mechanism can enhance the critical information extraction ability of time series models when processing long sequences [47]. The attention mechanism updates the weight by calculating the information distribution in the text vector, assigning a higher weight to the vital information in the sentence, and reducing the weight of the irrelevant information so that the critical information in the sentence is summarized in the output. Therefore, this paper proposes introducing an attention mechanism layer into the deep learning text classification model by referring to the attention mechanism principle and constructing a deep learning text classification model based on the attention mechanism.

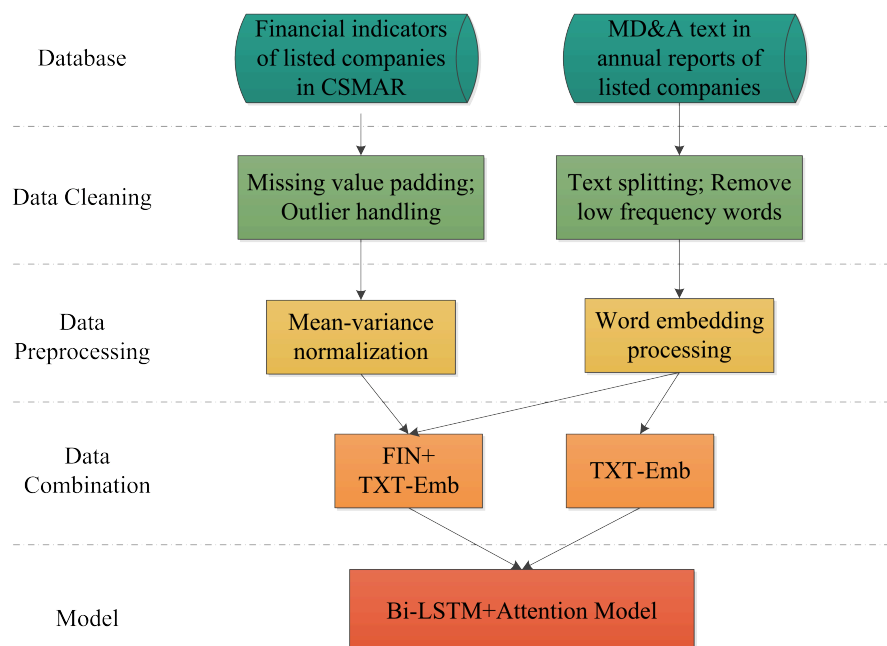


Figure 1. Research flow chart.

Our proposed attention-based framework for long text processing consists of three steps. The research process is shown in Figure 1. In the data cleaning step, the samples with missing MD&A text information were deleted, the collected MD&A long text data were segmented, and the low-frequency

words and stop words were deleted. In the word embedding step, the first 2000 words of the MD&A text are taken to fix the length of the text, and the deep learning word embedding model vectorizes the text. In the data combination step, the text vector and financial indicators are combined, and the plain text vector and the combination of the two are passed into the deep learning text classification model, respectively.

3.2. Model construction

We use the Bi-LSTM model's output as the attention mechanism's input to extract MD&A long text features. We concatenate the forward and reverse hidden layers of Bi-LSTM and use the output layer as the input of the attention layer. The attention layer calculates the importance of different time series of Bi-LSTM output sequence and finally calculates the weight and summary results.

The flowchart of our model is shown in Figure 2. The model structure is divided into five layers: input layer, word vector layer, Bi-LSTM layer, attention mechanism layer and output layer.

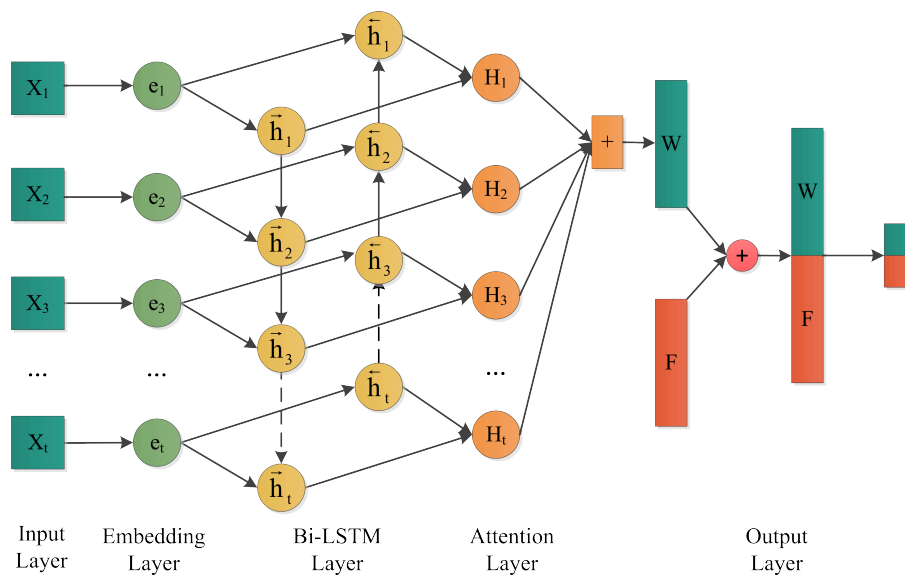


Figure 2. Structure of the Bi-LSTM+Attention model.

3.2.1. Input layer

The input layer is used to preprocess the input text data, which helps the model better focus on key information in MD&A text. We choose MD&A data that can represent the management concept of enterprise managers as text data and perform data cleaning and word segmentation preprocessing on MD&A data. Preprocessing makes the MD&A text shorter and more compact. Finally, we take the preprocessed MD&A Chinese text word segmentation sequence as the input model. X_t represents a single Chinese word. To display the text content more intuitively, we have chosen the MD&A Chinese text of a company with a stock code of '000007' in 2017 as an example. The MD&A Chinese text data before and after preprocessing is shown in the Appendix for reference.

3.2.2. Embedding layer

The word embedding layer converts preprocessed text into text vectors by encoding words into numbers. We constructed a Chinese encoding dictionary based on all preprocessed MD&A Chinese texts and set the text length to 2000 through truncation and zero padding. Finally, we encode the MD&A Chinese text passed to the input layer based on the constructed dictionary and convert the preprocessed MD&A Chinese long text into a numerical vector with a sequence length of 2000. e_t represents the encoded value of a single word.

3.2.3. Bi-LSTM layer

The Bi LSTM layer generates text semantic vectors related to financial distress by calculating MD&A Chinese long text vectors. The advantage of LSTM is that it can alleviate gradient vanishing and exploding problems so that LSTM can perform better in long-dependent sequences. The LSTM cell is shown in Figure 3.

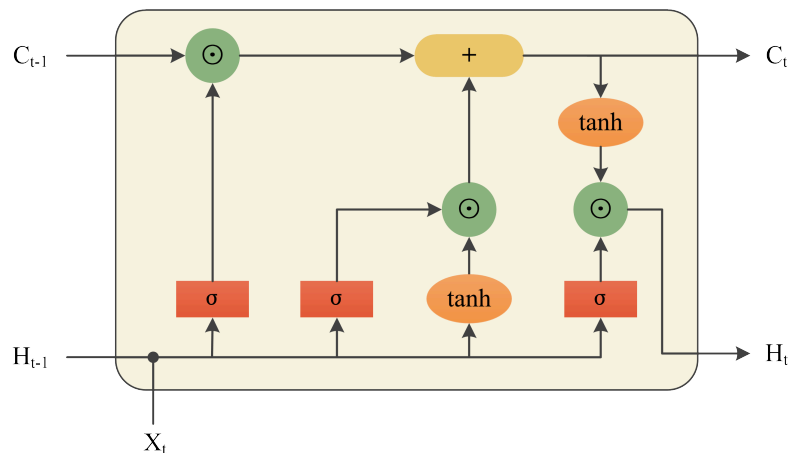


Figure 3. LSTM model structure.

There are three types of gates (σ) in LSTM networks: input gates, forget gates and output gates. The input gate determines when data is read into the cell; the forget gate resets the cell's contents; the output gate outputs entries from the cell. Each unit of LSTM is calculated as formulas (3.1)–(3.6). Among them, h represents the number of hidden units, X represents the sequence input of the current time step, W represents the weight parameter between two-time steps of the hidden layer, and b represents the bias parameter.

Structurally, the three gates are processed by three fully connected layers with a sigmoid activation function. I_t , O_t and F_t represent the values of the input gate, output gate and forget gate calculated by the sigmoid activation function, respectively. \tilde{C}_t represents the candidate memory cell and uses tanh as the activation function to update the memory cell. C_t represents a memory cell whose computation controls the learning of new data and the retention of memories. H_t represents the hidden state, which represents the output value of the current time step. In the Bi-LSTM layer, the information of the word vector layer is input to the forward hidden layer \vec{h}_t and the reverse hidden layer \overleftarrow{h}_t . Then, it outputs

the results of the forward and reverse hidden layers and the last memory cell.

$$I_t = \sigma(W_{xi}X_t + W_{hi}H_{t-1} + b_i) \quad (3.1)$$

$$O_t = \sigma(W_{xo}X_t + W_{ho}H_{t-1} + b_o) \quad (3.2)$$

$$F_t = \sigma(W_{xf}X_t + W_{hf}H_{t-1} + b_f) \quad (3.3)$$

$$\tilde{C}_t = \tanh(W_{xc}X_t + W_{hc}H_{t-1} + b_c) \quad (3.4)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (3.5)$$

$$H_t = \tanh(C_t) \odot O_t \quad (3.6)$$

3.2.4. Attention mechanism layer

The Attention mechanism layer is used to weight the text semantic vectors calculated by the Bi-LSTM model, highlighting key semantic information and thus better predicting financial distress. The purpose of the attention mechanism is to focus on the information that is more important to the target task in the input sequence information while reducing the weight of other irrelevant information and gradually filtering out non-critical information. It updates the weights by calculating the information distribution of the sequence so that the output can select some critical input information for summarization. In RNN-based encoder-decoder, Bahdanau attention [44] treats the decoder hidden state at the previous time step as a query and the encoder hidden state at all time steps as both keys and values. The calculation method of the attention mechanism in Bi-LSTM is shown in formulas (3.7)–(3.10).

$$u_i = \text{ReLU}(W_h h_i + b_h) \quad (3.7)$$

$$m_i = \tanh(o_i) \quad (3.8)$$

$$a_i = \text{softmax}(u_i m_i) \quad (3.9)$$

$$c = \sum_{i=1}^t a_i o_i \quad (3.10)$$

Among them, h_i represents the hidden layer vector of the i -th time step of Bi-LSTM, which is used as the query vector. The o_i represents the output vector of the i -th time step of Bi-LSTM as the key and value vector. The t represents the length of the text feature sequence, and u_i and m_i represent the hidden layer and output layer vectors after activation by ReLU and tanh. The attention weight a_i represents the attention weight calculated at the i th time step. The c_t is the attention-weighted sum of the outputs of all time steps of Bi-LSTM.

In the attention layer, the input information is the output of the forward and reverse hidden layers of Bi-LSTM, and the importance of the Bi-LSTM output at each moment is calculated in the attention mechanism. Then, the output results at each moment are weighted and summed to obtain H_t . Finally, the summarized text features are output.

3.2.5. Output layer

The fully connected neural network can learn the relationship between different features through the fully connected layer and use the activation function to increase the nonlinear ability of the model. In the output layer, we constructed a fully connected neural network to fuse the extracted text semantic features and financial indicators and calculate the binary output of financial distress. We concatenate the text feature vector extracted by the attention layer with the financial indicator vector to construct a merged vector and receive the merged vector through a fully connected neural network, in which several ReLU activation functions increase nonlinearity. When the research needs to check the prediction effect of the text feature vector, the text feature vector can be directly passed to the fully connected neural network. Finally, the softmax function calculates the binary classification output of the enterprise's financial distress.

4. Discussion

4.1. Dataset

Our experimental data comprises corporate financial data, annual report text data and financial distress labels. The main data comes from the China Stock Market and Accounting Research Database (CSMAR). For some missing data, we manually supplemented the missing data using the data from the CNIFO repository * and the official website of the listed company. For China's A-share stock market, the Shenzhen stock exchanges will announce the listed companies that have received ST every year. In the process of processing data, this study combines the label of year T with the enterprise data of year T-2 [10, 48, 49]. For example, Jinhua Group was warned of investment risks and became ST after the release of its 2019 annual report in 2020, and we use its ST label for 2020 and its financial and annual report data for 2018 to conduct financial distress prediction research. The dummy variable of the listed company's financial distress is set. If the listed company is marked as ST, the dummy variable of the year is 1. Otherwise, it is 0. The experimental data set consists of financial indicators and MD&A texts. In sample selection, we selected the research object of 1642 listed companies in the Shenzhen A-share market from January 2017 to December 2020, and we selected financial data and Chinese annual report texts of companies from January 2015 to December 2018 as the dataset. From the 1642 listed companies, we screened out 74 companies marked as ST in the middle of the four years. By removing banking, financial and utility enterprises such as banks from the sample, we obtained a dataset consisting of 296 ST enterprise annual report sample instances and 6272 regular enterprise annual report sample instances, a total of 6568 annual enterprise reports.

4.2. Text data

As a crucial part of an enterprise's annual report, MD&A data is the main output channel for the disclosure of business and financial information, expressing the management's views on potential opportunities and challenges in development to the outside world. Structurally, MD&A information is mainly text information, a channel for voluntary disclosure by listed companies. In order to facilitate the vectorization of unstructured text and use it as the input of the model, it is usually necessary to further preprocess the text before building the model. Unlike English, there are no space marks

*<http://www.cninfo.com.cn/new/index>

between Chinese terms. Therefore, in Chinese document preprocessing, word segmentation is difficult to perform with ambiguous words. Referring to the Chinese text processing methods commonly used in NLP research, we use the Jieba word segmentation tool to segment MD&A Chinese text data, remove low-frequency words through word frequency statistics and then delete stop words according to the artificially supplemented stop word dictionary of the Harbin Institute of Technology of China. The MD&A text vector must be converted into a digital format. We use Word2Vec text vectorization for machine learning and a text embedding method for deep learning models.

Regarding Chinese text length, the MD&A part of the corporate annual report exceeds several thousand words, which is super-long text content, which increases the difficulty of text feature extraction. The length distribution of MD&A Chinese text before and after preprocessing is shown in Figure 4. After we reviewed the text structure of MD&A, we found that critical information mainly exists in the first half of the text. We facilitate the word embedding processing of the model by controlling the text sequence to be the same length, so truncating longer texts, and zero-padding short texts are necessary. Here, we select 2000 words as the fixed length of the text sequence according to the experimental results and compare their effects on different models.

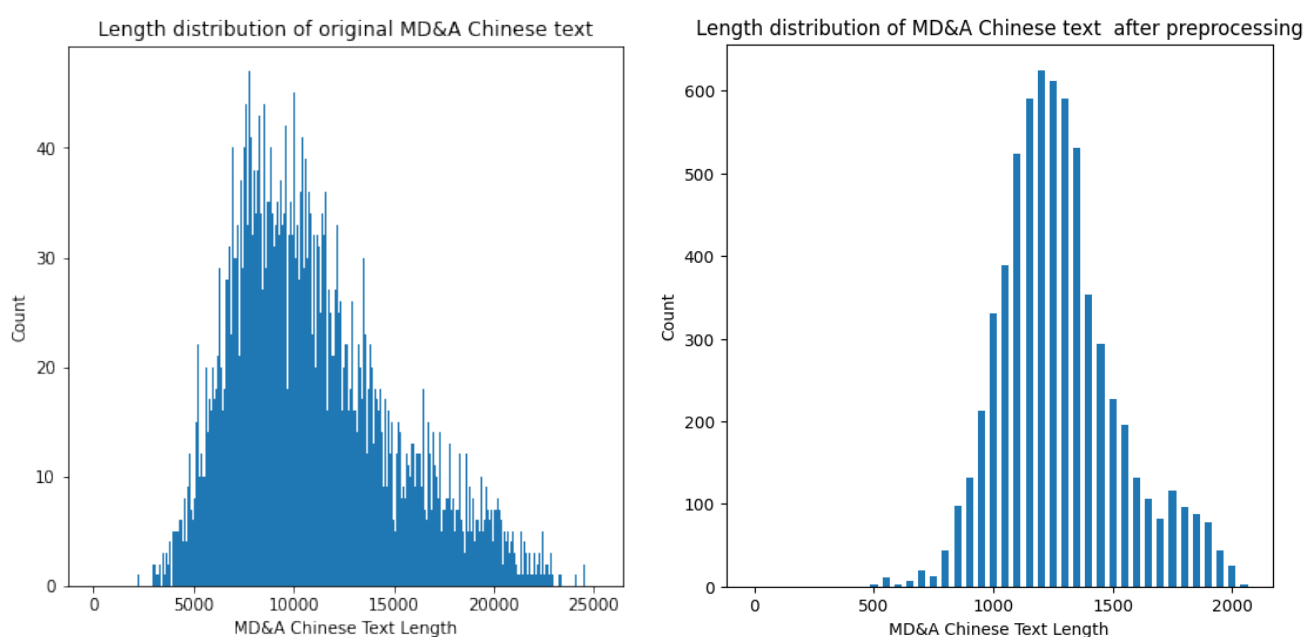


Figure 4. Length distribution of MD&A Chinese text before and after preprocessing.

After that, for the machine learning model, we refer to the practice of W. Xiuguo et al., adopt Word2Vec processing based on word vector embedding for the text and convert the word segmentation result into a 256-dimensional vector representation to facilitate the calculation of the algorithm [38]. The word embedding model we use is the word vector model designed by the China Institute of Space Science [†], and its training tool is Word2Vec in Gensim, using the content of the Baidu Encyclopedia as a corpus. The deep learning model does not require much feature processing

[†]<https://spaces.ac.cn/archives/4304>

on the data. After removing stop words and low-frequency words, we input the Chinese vocabulary sequence into the word embedding layer of the model. The Embedding layer converts the text into a digital vector according to the encoded sequence by labeling the Chinese vocabulary. The parameters of the Word2Vec model are shown in Table 2.

Table 2. Parameter setting of Word2Vec model.

Parameter	Parameter description	Parameter value
size	The dimension of word vector	100
window	The maximum distance of words	5
min_count	Word frequency screening	2
negative	Set negative sampling	3
sample	The threshold for downsampling high-frequency words	0.001

4.3. Quantitative data

Table 3. Selection of financial indicators of listed companies.

Variable category	Secondary variables	Variable category	Secondary variables
Solvency	X1 Current ratio	Risk level	X24 Financial leverage
	X2 Quick ratio		X25 Operating leverage
	X3 Cash ratio	Operating ability	X26 Accounts receivable/revenue
	X4 Working capital to borrowing ratio		X27 Accounts receivable turnover
	X5 Operating income		X28 Inventory/revenue
	X6 Net cash flow/current debt ratio		X29 Inventory turnover
	X7 Assets and liabilities		X30 Accounts payable turnover
	X8 Long-term borrowings/total assets		X31 Working capital turnover
	X9 Tangible assets liability ratio		X32 Total asset turnover
	X10 Long-term gearing ratio		X33 Shareholders' equity turnover
	X11 Long-term debt-to-equity ratio		X34 Capital accumulation rate
	X12 Long-term debt/working capital		X35 Total asset growth rate
Cash flow analysis	X13 Net profit and cash content	Development ability	X36 Basic EPS growth
	X14 Operating income cash content		X37 Net profit growth rate
	X15 Operating profit net cash content		X38 Operating profit growth rate
	X16 Company cash flow		X39 Operating income growth rate
	X17 Operational index		X40 Owner's Equity Growth Rate
Per share indicators	X18 Gross operating income per share	Ratio structure	X41 The growth rate of net assets per share
	X19 Liabilities per share		X42 Current asset ratio
	X20 Earnings per share		X43 Cash-to-asset ratio
Relative value indicator	X21 P/E ratio		X44 Fixed asset ratio
	X22 P/B ratio		X45 Current debt ratio
	X23 Common stock yield		X46 Operating debt ratio

The financial indicators used in this research were downloaded from the CSMAR database. The choice of financial indicators has a significant impact on the outcome of the forecasting algorithm. In previous studies, the scenarios of financial risk predicted using financial indicators include financial fraud prediction, financial restatement prediction and FDP. These types of studies are all centered on the financial risks of enterprises. The selection of financial indicators has been repeatedly screened and verified, and the structure of financial indicators in financial risk prediction has been gradually summarized. This paper summarizes the financial indicators commonly used in previous financial distress prediction studies and integrates the conclusions of financial risk related studies [10,35,36,49]. Finally, we selected 46 typical financial indicators from eight dimensions: ratio structure, solvency, development ability, risk level, operating ability, relative value indicators, per share indicators and cash flow analysis. The financial indicators used in this study are shown in Table 3.

4.4. Model implementation

We chose PyTorch as the deep learning framework, an open-source Python machine-learning library that can utilize powerful GPUs to accelerate the computation of tensors. PyTorch was designed and open-sourced by the Facebook Artificial Intelligence Research Institute (FAIR) based on Torch.

We use the Adam (Adaptive Moment Estimation) optimizer as the gradient descent optimization algorithm in the deep learning training process. We add a decreasing learning rate setting to the model to help the model learn detailed information. In order to prevent the model from overfitting, we added a Dropout layer and a LayerNorm normalization method to the output layer of the model. The former makes some neurons stop working according to a certain probability, making the model more generalizable. The latter mitigates overfitting by normalizing each batch. Finally, we adopt a strategy of early stopping in model training.

We randomly selected 60% of the dataset as the training set, 20% as the validation set and the remaining 20% as the test set. The parameters of Bi-LSTM and Bi-LSTM+Attention models are shown in Table 4.

Table 4. Parameter setting of Bi-LSTM and Bi-LSTM+Attention models.

Parameter	Parameter description	Parameter value
lr	Learning rate	0.001
dropout	Dropout ratio	0.5
batch_size	Batch size	64
hidden_dims	Neurons number of LSTM model	16
max_length	MD&A sequence length	2000
embed_dim	Dimension of sequence	32

4.5. Metrics

FDP research is a typical binary classification problem, and we aim to build a model with good generalization performance. This research selected five evaluation indicators (Precision, Recall, F1-score, AUC, PR-AUC) to measure the classification and prediction performance of the FDP model. We use modules in the Scikit-learn library in Python to calculate Precision, Recall, F1-score, AUC and PR-AUC metrics to evaluate and select training models.

The precision rate refers to the proportion of all predicted distressed enterprises that experience distress, and the recall rate refers to the proportion of the distressed enterprises that are predicted correctly. The formulas for Precision and Recall are shown in Eqs (4.1) and (4.2):

$$Precision = \frac{TP}{TP + FP} \quad (4.1)$$

$$Recall = \frac{TP}{FN + FP} \quad (4.2)$$

Among them, true positives (TP) indicate that the predictions of those distressed companies are correct. False Negatives (FN) indicate that these distressed companies have incorrect predictions, classifying them as non-distressed companies. A true negative (TN) indicates that the predictions of non-distressed companies are correct. False positives (FP) indicate that those non-distressed companies have incorrect predictions, classifying them as distressed companies.

The F1 is often used to measure the model's classification performance. It can be represented by the weighted average of the Precision and Recall of the model as shown in Eq (4.3):

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (4.3)$$

In studies using imbalanced datasets, the AUC and PR-AUC metrics indicate the extent to which the model is independent of sample proportions. The AUC value is the area under the ROC curve. The ROC curve is also called the susceptibility curve. AUC value refers to the area under the ROC curve. The ROC curve, also known as the sensitivity curve, is a curve with recall as the ordinate and (1-specificity) as the abscissa. AUC stands for prediction effect, and the larger the AUC is, the better the model prediction performance. Since this research focuses more on positive samples, and both Precision and Recall measure the ability of the model to find positive samples, here we add the PR-AUC indicator to test the model's performance in unbalanced samples. PR-AUC is the average of Precision calculated based on each Recall threshold. Usually, larger PR-AUC values indicate better model performance [33]. In conclusion, the values of the above five indicators are between 0 and 1. The closer to 1 the value is, the better the prediction performance of the model.

4.6. Comparison model

To demonstrate the effectiveness of the Bi-LSTM+Attention model, we selected TextCNN, GRU, Transformer and traditional machine learning models for comparative analysis.

4.6.1. TextCNN

CNN has the advantage of having fewer parameters, sampling efficiently and being suitable for efficient parallel computation on GPUs. The TextCNN model is designed for text classification. TextCNN captures the local features of the text and extracts semantic segmentation information at different levels by automatically combining and filtering local features, similar to a sliding window containing multiple words [50]. The TextCNN model we constructed uses a one-dimensional convolutional layer and a temporal maximum pooling layer containing four components: word embedding, convolution, pooling and softmax. The parameters of the TextCNN model are shown in Table 5.

Table 5. Parameter setting of TextCNN model.

Parameter	Parameter description	Parameter value
lr	Learning rate	0.001
dropout	Dropout ratio	0.7
batch_size	Batch size	32
filter_sizes	The size of convolutional kernels	(2, 3, 4)
num_filters	The number of convolutional kernels	16
max_length	MD&A sequence length	2000

4.6.2. GRU

GRU can be regarded as a simplified variant of LSTM, but GRU is relatively faster. GRU can capture the dependencies on sequences with time step distance. It is designed to reduce the gradient disappearance problem and thus retain information on long sequences. The parameters of the GRU and GRU+Attention models are shown in Table 6.

Table 6. Parameter setting of GRU and GRU+Attention models.

Parameter	Parameter description	Parameter value
lr	Learning rate	0.001
dropout	Dropout ratio	0.5
batch_size	Batch size	32
hidden_dims	Neurons number of GRU model	16
max_length	MD&A sequence length	2000
embed_dim	Dimension of sequence	32

4.6.3. Transformer

Table 7. Parameter setting of Transformer model.

Parameter	Parameter description	Parameter value
lr	Learning rate	0.001
dropout	Dropout ratio	0.7
batch_size	Batch size	16
hidden_dims	Dimension of feedforward network	16
max_length	MD&A sequence length	2000
embed_dim	Dimension of sequence	16

Vaswani et al. proposed Transformer, which achieved the best results in translation tasks by building a model based on the attention mechanism, and demonstrated that Transformer could be generalized to learning tasks in other domains [51]. Compared to implementing the attention mechanism constructed by Bahdanau above, Transformer consists of multiple modules with attention combined to form an encoder and decoder. Since the FDP study is not a language-generating task, here we use the encoding

part of the Transformer (Encoder) for text feature extraction. The parameters of the Transformer model are shown in Table 7.

4.6.4. Other benchmark models

We also selected several traditional machine learning classifiers as comparison models for deep learning models. Past studies have shown that SVM performs well on complex financial data. SVM projects the original data to higher dimensions by nonlinear mapping and then builds the optimal decision hyperplane to maximize the distance between the two closest samples on either side of the plane. SVM works better when the sample size is small compared to other nonlinear methods. Inspired by the central nervous system of animals, the ANN model applies neuronal information processing and excels in self-learning and self-organization. The decision tree model is a tree structure describing instance classification, consisting of nodes and directed edges. RF is an ensemble algorithm that adds a voting mechanism to the ensemble of multiple decision trees, which can alleviate the balance error of unbalanced classification data sets. XGBoost is an algorithm implementation based on gradient boosting decision tree (GBDT). It has the advantages of high efficiency, flexibility and portability. It is often used to solve large-scale data problems in industrial fields.

5. Empirical results

This research compares the effects of machine learning and deep learning models combined with financial and text indicators. Our empirical results can be divided into two parts. One part is the experimental results of financial indicators and text vectors on traditional machine learning models (RF, SVM, XGBoost, and ANN). The other part is the experimental results of financial metrics and text data on deep learning models (TextCNN, Bi-LSTM, GRU, Transformer, GRU+Attention, and Bi-LSTM+Attention). In order to verify the improvement effect of the attention mechanism on semantic extraction of MD&A Chinese long texts, this study added ablation experiments and constructed Bi LSTM and GRU models without attention mechanism, respectively. We will focus on these two parts to start the analysis.

5.1. Traditional machine learning models

In the traditional machine model, we tested the single financial indicator and the financial indicator combined with the text word vector indicator. The classification results of a single financial index are shown in Table 8. The results show that the SVM model performs better in Precision metrics, while the AUC and PR-AUC values of the XGBoost, RF and ANN models perform better.

Table 8. Results of financial indicators in machine learning models.

	AUC	PR-AUC	Precision	Recall	F1-score
SVM	0.6713	0.4442	0.5000	0.3596	0.4183
XGBoost	0.7217	0.4029	0.2778	0.5056	0.3586
RF	0.7107	0.4012	0.3066	0.2135	0.4719
ANN	0.6930	0.4200	0.3978	0.4157	0.4066

The results of the machine learning model after adding text vectorization indicators are shown in Table 9.

The empirical results show that the AUC metrics of XGBoost, RF and ANN models are still excellent after adding text vectorization indicators, and the AUC metrics have increased by 1.27%, 1.86% and 4.31%. SVM and ANN also perform well on the PR-AUC indicator, and adding the text vectorization indicator improves by 5.83% and 13.95%. The recall effect of the XGBoost model performs better when the text vectorization index is added, increasing by 6.67% after adding it. Overall, we found that after adding text vectorization features, all traditional machine learning models improved slightly, proving that text data provided certain incremental information for machine learning models. From the above results, it can be seen that due to the high requirements for feature engineering and the limited processing ability of the Word2Vec method for long text vectors, the performance of machine learning models in FDP research combining long text is limited.

Table 9. Results of financial indicators combined with text indicators in machine learning models.

	AUC	PR-AUC	Precision	Recall	F1-score
SVM	0.6780	0.4701	0.5410	0.3708	0.4400
XGBoost	0.7309	0.4038	0.2474	0.5393	0.3392
RF	0.7239	0.4101	0.2922	0.5056	0.3704
ANN	0.7229	0.4786	0.4615	0.4719	0.4667

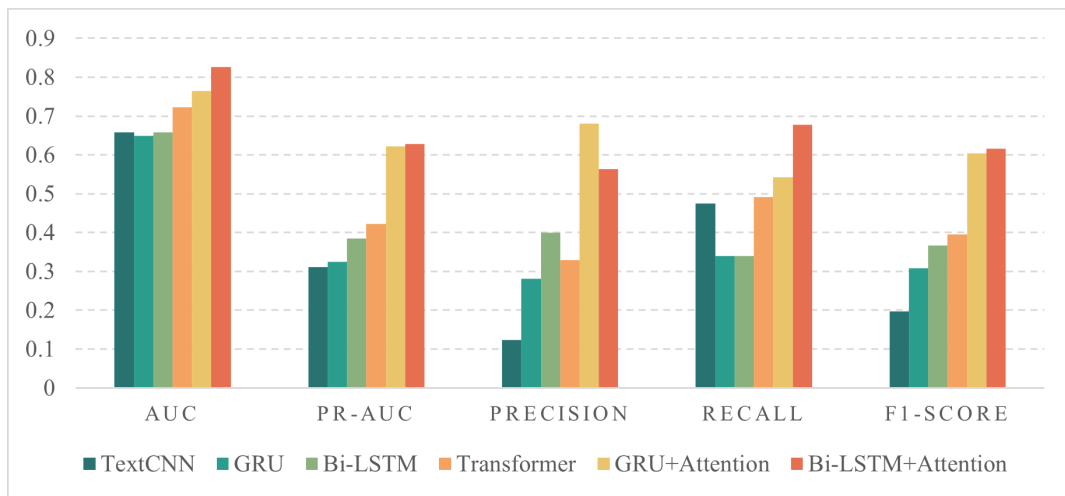
5.2. Deep learning models

In the deep learning model, we tested the performance of two types of models based on text indicators before and after adding financial indicators. The classification results of the single-text indicators are shown in Table 10 and Figure 5. The experimental results indicate that various deep learning models have certain effects on FDP research when only using MD&A text data, which verifies the reliability and effectiveness of MD&A Chinese long text data for FDP research. The empirical results show that the AUC and PR-AUC of the attention-based Transformer model are slightly better than Text-CNN, GRU and Bi-LSTM models in frequently-used text classification models.

In the ablation experiment, comparing the GRU+Attention and Bi-LSTM+Attention models with GRU and Bi-LSTM without attention mechanism, it is evident that the effectiveness of each indicator has been improved to a certain extent. Their AUC metrics are improved by 17.87% and 25.72%, which means that the attention mechanism enhances the deep learning model's ability to capture long text information. In terms of model effect, the performance of the Bi-LSTM+Attention model is better than that of the GRU+Attention model, and the AUC index is 8.02% higher than the latter, indicating that the Bi-LSTM+Attention model has a more significant classification effect than other models in the single-text index.

Table 10. Results of single-text data in deep learning.

	AUC	PR-AUC	Precision	Recall	F1-score
TextCNN	0.6584	0.3110	0.1239	0.4746	0.1965
GRU	0.6492	0.3252	0.2817	0.3390	0.3077
Bi-LSTM	0.6575	0.3843	0.4000	0.3390	0.3670
Transformer	0.7223	0.4220	0.3295	0.4915	0.3946
GRU+Attention	0.7652	0.6219	0.6809	0.5424	0.6038
Bi-LSTM+Attention	0.8266	0.6279	0.5634	0.6780	0.6154

**Figure 5.** Results of single-text data in deep learning.

Usually, when the AUC metric result of the test set is more significant than 0.75, it indicates that the model's discriminative ability is good [52]. Therefore, our experimental results indicate that deep learning models based on attention mechanisms can achieve better discrimination ability, and the overall performance of the Bi LSTM+attention model is the best.

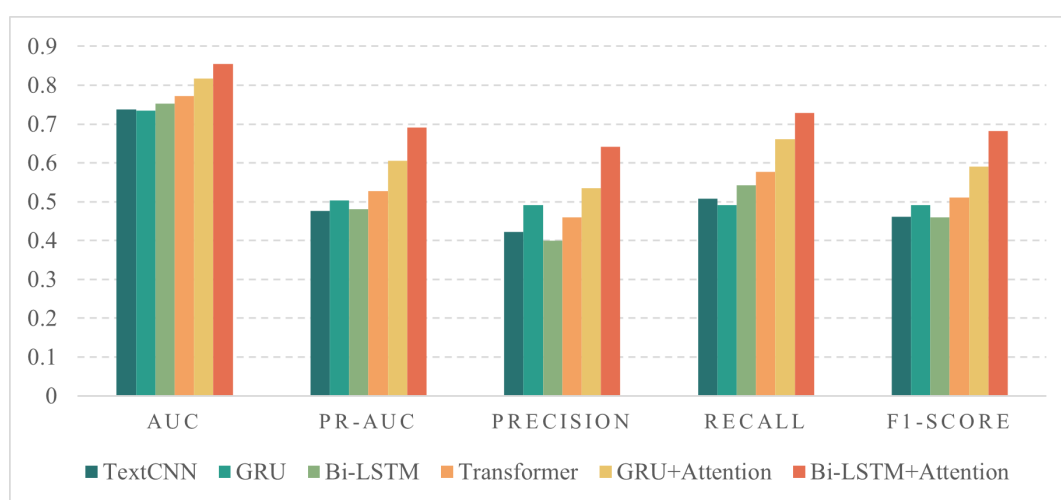
The results of adding both textual and financial indicators to the deep learning model are shown in Table 11 and Figure 6. The empirical results show that, on the whole, the effect of each deep learning model is improved after adding financial indicators, indicating that the addition of financial indicators also provides incremental information for the text classification model. Based on the attention mechanism in the commonly used text classification models, the AUC and PR-AUC effects of the Transformer model are slightly better than those of the Text-CNN, GRU and Bi-LSTM models.

Similarly, in the ablation experiment, the AUC index values of GRU and LSTM increased by 6.88% and 13.66%, respectively. It can be seen that the introduction of the attention mechanism has improved the effectiveness of the model.

Regarding model performance, the Bi-LSTM+Attention model outperformed the GRU+Attention model, and the AUC index was 8.99% higher than the latter. It shows that the classification effect of the Bi-LSTM+Attention model is more significant than other models when text indicators are combined with financial indicators.

Table 11. Results of text indicators combined with financial indicators in the deep learning model.

	AUC	PR-AUC	Precision	Recall	F1-score
TextCNN	0.7379	0.4765	0.4225	0.5085	0.4615
GRU	0.7338	0.5029	0.4915	0.4915	0.4915
Bi-LSTM	0.7521	0.4815	0.4000	0.5424	0.4604
Transformer	0.7722	0.5274	0.4595	0.5763	0.5113
GRU+Attention	0.8170	0.6052	0.5342	0.6610	0.5909
Bi-LSTM+Attention	0.8548	0.6914	0.6418	0.7288	0.6825

**Figure 6.** Results of text indicators combined with financial indicators in the deep learning model.

6. Conclusions

Prediction of financial distress of listed companies can provide enterprise managers with early warning of enterprise risks and help investors make decisions to avoid financial losses. In traditional research, researchers usually predict financial distress by using corporate financial data for feature extraction or combining text word vector features. Due to the application of deep learning in natural language processing, various deep learning based text classification models are utilized by FDP researchers.

This study took 1642 listed companies in China's Shenzhen A-share market from 2017 to 2020 as the research object, constructed a deep learning text classification model and used MD&A Chinese text data and financial indicators to predict the financial distress of listed companies. However, the text in the MD&A section of listed companies in China was several thousand words long, making the text analysis work extremely difficult. To cope with the long text sequence classification problem, we combined the attention mechanism with GRU and LSTM models to construct GRU+Attention and Bi-LSTM+Attention models. This study captured the time series information by using multiple hidden

layer structures of GRU and LSTM and then used the attention mechanism to summarize the time series critical information in the MD&A long text.

By comparing the experimental data with traditional machine learning models and deep learning models, the research results are as follows:

1). We verify that the MD&A Chinese text of listed companies can provide incremental information for the FDP model. The deep learning models we constructed are more effective in identifying corporate financial distress by combining text and financial data than traditional machine learning models.

2). We found that the attention mechanism can improve the long text classification performance of the deep learning model in FDP research.

3). Compared with the ordinary deep learning model, the Bi-LSTM+Attention model we constructed has different degrees of improvement in the measurement indicators of FDP research.

Our research has several limitations waiting for future research. First, the Transformer model in this research experiment may not show the best effect due to hardware limitations, and future research can try to debug it under a better memory configuration. Second, the dataset of this study uses the MD&A Chinese texts in the annual reports of listed companies as text data. Future research can test this model when the hardware is improved. Third, the deep learning model based on the attention mechanism constructed relies on the Chinese text training of the annual reports of listed companies in China, and the applicability of FDP of listed companies in other countries needs further research. Lastly, we built two deep learning models combined with an attention mechanism for long text classification: GRU+Attention and Bi-LSTM+Attention. In future research, we can try to combine the attention mechanism in other text classification algorithms to explore the effect of different models based on the attention mechanism in FDP research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work has been supported by grants from the National Social Science Foundation Key Projects of China (21AZD022).

Conflict of interest

The authors declare there is no conflicts of interest.

References

1. G. Wang, G. Chen, Y. Chu, A new random subspace method incorporating sentiment and textual information for financial distress prediction, *Electron. Commer. Res. Appl.*, **29** (2018), 30–49. <https://doi.org/10.1016/j.elerap.2018.03.004>

2. G. Wang, J. L. Ma, G. Chen, Y. Yang, Financial distress prediction: Regularized sparse-based random subspace with er aggregation rule incorporating textual disclosures, *Appl. Soft Comput.*, **90** (2020), 106152. <https://doi.org/10.1016/j.asoc.2020.106152>
3. Z. Halim, S. M. Shuhidan, Z. M. Sanusi, Corporation financial distress prediction with deep learning: analysis of public listed companies in malaysia, *Bus. Process Manage. J.*, **27** (2021), 1163–1178. <https://doi.org/10.1108/Bpmj-06-2020-0273> .
4. P. du Jardin, A two-stage classification technique for bankruptcy prediction, *Eur. J. Oper. Res.*, **254** (2016), 236–252. <https://doi.org/10.1016/j.ejor.2016.03.008>
5. D. Campa, M. D. Camacho-Minano, The impact of SME's pre-bankruptcy financial distress on earnings management tools, *Int. Rev. Financ. Anal.*, **42** (2015), 222–234. <https://doi.org/10.1016/j.irfa.2015.07.004>
6. J. Bertomeu, E. Cheynel, E. Floyd, W. Pan, Using machine learning to detect misstatements, *Rev. Accounting Stud.*, **26** (2021), 468–519. <https://doi.org/10.1007/s11142-020-09563-8>
7. J. Donovan, J. Jennings, K. Koharki, J. Lee, Measuring credit risk using qualitative disclosure, *Rev. Accounting Stud.*, **26** (2021), 815–863. <https://doi.org/10.1007/s11142-020-09575-4>
8. W. Ben-Amar, I. Belgacem, Do socially responsible firms provide more readable disclosures in annual reports, *Corporate Social Responsib. Environ. Manage.*, **25** (2018), 1009–1018. <https://doi.org/10.1002/csr.1517>
9. P. Hajek, R. Henriques, Mining corporate annual reports for intelligent detection of financial statement fraud—a comparative study of machine learning methods, *Knowledge-Based Syst.*, **128** (2017), 139–152. <https://doi.org/10.1016/j.knosys.2017.05.001>
10. F. Mai, S. N. Tian, C. Lee, L. Ma, Deep learning models for bankruptcy prediction using textual disclosures, *Eur. J. Oper. Res.*, **274** (2019), 743–758. <https://doi.org/10.1016/j.ejor.2018.10.024>
11. Y. B. Qian, A critical genre analysis of mda discourse in corporate annual reports, *Discourse Commun.*, **14** (2020), 424–437. <https://doi.org/10.1177/1750481320910525>
12. W. H. Beaver, Financial ratios as predictors of failure, *J. Accounting Res.*, **4** (1966), 71–111. <https://doi.org/10.2307/2490171>
13. E. B. Deakin, A discriminant analysis of predictors of business failure, *J. Accounting Res.*, **10** (1972), 167–179. <https://doi.org/10.2307/2490225>
14. D. Carmichael, *Auditor's Reporting Obligation: The Meaning and Implementation of the Fourth Standard of Reporting; Auditing Research Monographh, I*, American Institute of Certified Public Accountants, 1978.
15. M. E. Zmijewski, Methodological issues related to the estimation of financial distress prediction models, *J. Accounting Res.*, **22** (1984), 59–82. <https://doi.org/10.2307/2490859>
16. E. I. Altman, *The Prediction of Corporate Bankruptcy: A Discriminant Analysis*, University of California, Los Angeles, 1967.
17. A. I. Dimitras, S. H. Zanakis, C. Zopounidis, A survey of business failures with an emphasis on prediction methods and industrial applications, *Eur. J. Oper. Res.*, **90** (1996), 487–513. [https://doi.org/10.1016/0377-2217\(95\)00070-4](https://doi.org/10.1016/0377-2217(95)00070-4)

18. S. A. Ross, R. Westerfield, J. F. Jaffe, *Corporate Finance*, Irwin/McGraw-Hill, 1999.
19. Y. Ding, X. Song, Y. Zen, Forecasting financial condition of chinese listed companies based on support vector machine, *Expert Syst. Appl.*, **34** (2008), 3081–3089. <https://doi.org/10.1016/j.eswa.2007.06.037>
20. R. B. Geng, I. Bose, X. Chen, Prediction of financial distress: An empirical study of listed chinese companies using data mining, *Eur. J. Oper. Res.*, **241** (2015), 236–247. <https://doi.org/10.1016/j.ejor.2014.08.016>
21. S. Ruan, X. Sun, R. Yao, W. Li, Deep learning based on hierarchical self-attention for finance distress prediction incorporating text, *Comput. Intell. Neurosci.*, **2021** (2021), 1165296. <https://doi.org/10.1155/2021/1165296>
22. F. Y. Lin, D. R. Liang, E. C. Chen, Financial ratio selection for business crisis prediction, *Expert Syst. Appl.*, **38** (2011), 15094–15102. <https://doi.org/10.1016/j.eswa.2011.05.035>
23. E. I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *J. Finance*, **23** (1968), 589–609. <https://doi.org/10.2307/2978933>
24. G. Wang, J. Ma, S. L. Yang, An improved boosting based on feature selection for corporate bankruptcy prediction, *Expert Syst. Appl.*, **41** (2014), 2353–2361. <https://doi.org/10.1016/j.eswa.2013.09.033>
25. J. Huang, H. B. Wang, G. Kochenberger, Distressed chinese firm prediction with discretized data, *Manage. Decis.*, **55** (2017), 786–807. <https://doi.org/10.1108/Md-08-2016-0546>
26. L. G. Zhou, K. P. Tam, H. Fujita, Predicting the listing status of chinese listed companies with multi-class classification models, *Inf. Sci.*, **328** (2016), 222–236. <https://doi.org/10.1016/j.ins.2015.08.036>
27. D. Alaminos, A. Del Castillo, M. A. Fernandez, A global model for bankruptcy prediction, *PLoS One*, **11** (2016), e0166693. <https://doi.org/10.1371/journal.pone.0166693>
28. Y. P. Huang, M. F. Yen, A new perspective of performance comparison among machine learning algorithms for financial distress prediction, *Appl. Soft Comput.*, **83** (2019), 105663. <https://doi.org/10.1016/j.asoc.2019.105663>
29. K. Olorunnimbe, H. Viktor, Deep learning in the stock market-a systematic survey of practice, backtesting, and applications, *Artif. Intell. Rev.*, **56** (2023), 2057–2109. <https://doi.org/10.1007/s10462-022-10226-0>
30. S. Ben Jabeur, V. Serret, Bankruptcy prediction using fuzzy convolutional neural networks, *Res. Int. Bus. Finance*, **64** (2023), 101844. <https://doi.org/10.1016/j.ribaf.2022.101844>
31. Y. D. Wang, Y. L. Jia, Y. H. Tian, J. Xiao, Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring, *Expert Syst. Appl.*, **200** (2022), 117013. <https://doi.org/10.1016/j.eswa.2022.117013>
32. S. X. Li, W. X. Shi, J. C. Wang, H. S. Zhou, A deep learning-based approach to constructing a domain sentiment lexicon: A case study in financial distress prediction, *Inf. Process. Manage.*, **58** (2021), 102673. <https://doi.org/10.1016/j.ipm.2021.102673>

33. J. Jing, W. Yan, X. Deng, A hybrid model to estimate corporate default probabilities in china based on zero-price probability model and long short-term memory, *Appl. Econ. Lett.*, **28** (2020), 413–420. <https://doi.org/10.1080/13504851.2020.1757611>
34. J. Sun, H. Li, H. Fujita, B. B. Fu, W. G. Ai, Class-imbalanced dynamic financial distress prediction based on adaboost-svm ensemble combined with smote and time weighting, *Inf. Fusion*, **54** (2020), 128–144. <https://doi.org/10.1016/j.inffus.2019.07.006>
35. X. D. Du, W. Li, S. M. Ruan, L. Li, Cus-heterogeneous ensemble-based financial distress prediction for imbalanced dataset with ensemble feature selection, *Appl. Soft Comput.*, **97**, (2020), 106758. <https://doi.org/10.1016/j.asoc.2020.106758>
36. J. Sun, H. Fujita, Y. J. Zheng, W. G. Ai, Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods, *Inf. Sci.*, **559** (2021), 153–170. <https://doi.org/10.1016/j.ins.2021.01.059>
37. H. Wang, X. Liu, Undersampling bankruptcy prediction: Taiwan bankruptcy data, *PLoS One*, **16** (2021), e0254030. <https://doi.org/10.1371/journal.pone.0254030>
38. X. Wu, S. Du, An analysis on financial statement fraud detection for chinese listed companies using deep learning, *IEEE Access*, **10** (2022), 22516–22532. <https://doi.org/10.1109/ACCESS.2022.3153478>
39. J. Liu, J. Li, Risk analysis of textile industry foreign investment based on deep learning, *Comput. Intell. Neurosci.*, **2022** (2022), 3769670. <https://doi.org/10.1155/2022/3769670>
40. P. Craja, A. Kim, S. Lessmann, Deep learning for detecting financial statement fraud, *Decis. Support Syst.*, **139** (2020), 113421. <https://doi.org/10.1016/j.dss.2020.113421>
41. T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, *arXiv preprint*, (2013), arXiv:1301.3781. <https://doi.org/10.48550/arXiv.1301.3781>
42. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>
43. S. J. Yu, D. L. Liu, W. F. Zhu, Y. Zhang, S. M. Zhao, Attention-based lstm, gru and cnn for short text classification, *J. Intell. Fuzzy Syst.*, **39** (2020), 333–340. <https://doi.org/10.3233/Jifs-191171>
44. A. Galassi, M. Lippi, P. Torrioni, Attention in natural language processing, *IEEE Trans. Neural Networks Learn. Syst.*, **32** (2021), 4291–4308. <https://doi.org/10.1109/TNNLS.2020.3019893>
45. V. Mnih, N. Heess, A. Graves, Recurrent models of visual attention, *arXiv preprint*, (2014), arXiv:1406.6247. <https://doi.org/10.48550/arXiv.1406.6247>
46. D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, *arXiv preprint*, (2014), arXiv:1409.0473. <https://doi.org/10.48550/arXiv.1409.0473>
47. Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ACL, San Diego, USA, (2016), 1480–1489.
48. J. Sun, H. Fujita, P. Chen, H. Li, Dynamic financial distress prediction with concept drift based on time weighting combined with adaboost support vector machine ensemble, *Knowledge-Based Syst.*, **120** (2017), 4–14. <https://doi.org/10.1016/j.knosys.2016.12.019>

49. S. Zhao, K. Xu, Z. Wang, C. Liang, W. Lu, B. Chen, Financial distress prediction by combining sentiment tone features, *Econ. Modell.*, **106** (2022), 105709. <https://doi.org/10.1016/j.econmod.2021.105709>
50. J. B. Jing, W. W. Yan, X. M. Deng, A hybrid model to estimate corporate default probabilities in China based on zero-price probability model and long short-term memory, *Appl. Econ. Lett.*, **28** (2021), 413–420. <https://doi.org/10.1080/13504851.2020.1757611>
51. Y. Chen, *Convolutional Neural Network for Sentence Classification*, Master's thesis, University of Waterloo, 2015.
52. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, *arXiv preprint*, (2017), arXiv:1706.03762. <https://doi.org/10.48550/arXiv.1706.03762>

Appendix

Original MD&A Chinese text:

“一概述报告期内公司主营业务为物业管理和房屋租赁业该块业务经营业绩一直较为稳定是公司经营工作的基础和有益补充公司新涉足的信息技术服务业尚处于前期投入阶段未形成有效收入需进一步的投入与培育但就公司涉足的新业务项目来看新业务已展现出的良好发展势头及高效创收能力公司转型可谓初见成效报告期内公司进一步... ..”

Preprocessing results:

“概述 报告 期 内 公 司 主 营 业 务 物 业 管 理 房 屋 租 赁 业 该 块 业 务 经 营 业 绩 较 为 稳 定 公 司 经 营 工 作 基 础 有 益 补 充 公 司 新 涉 足 信 息 技 术 服 务 业 尚 处 于 前 期 投 入 阶 段 未 收 入 需 进 一 步 投 入 培 育 公 司 涉 足 新 业 务 项 目 来 看 新 业 务 已 展 现 出 发 展 势 头 高 效 创 收 能 力 公 司 转 型 可 谓 初 见 成 效 报 告 期 内 公 司 进 一 步”



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)