https://doi.org/10.31449/inf.v48i5.5392 Informatica **48** (2024) 111–120 **111**

Comparison of Model Performance in Forewarning Financial Crisis of Publicly Traded Companies: Different Algorithmic Models

Jingzheng Guo,[1] Yan Ding[2*]
[1]Finance Office, Hebei Sport University, Shijiazhuang, Hebei 050031, China
[2]School of Management Engineering, Tangshan Polytechnic College, Tangshan, Hebei 063299, China
E-mail: dm43154@163.com
[*]Corresponding author

Keywords: publicly traded company, financial crisis, early warning model, indicator selection, accuracy

Received: October 31, 2023

The financial crisis can have adverse effects on a company's development and even on the entire industry. Early warning and prevention of such crises through specific methods holds significant importance. This paper focuses on the prewarning of financial crises in publicly traded companies. Samples were selected from the CSMAR database to analyze data from the T-2 year and T-3 year. Thirty indicators were screened from perspectives such as levels of debt repayment. The performance of six different algorithmic models, including support vector machine, XGBoost, long short-term memory (LSTM), gate recurrent unit (GRU), bi-directional LSTM, and bi-directional GRU (BiGRU), were compared using the indicators screened by significance tests. The results indicated that the T-2-year data outperformed the T-3-year data in early warning. Among the various algorithmic models, BiGRU exhibited the best early warning performance, with an accuracy of 0.934, a true positive rate of 0.975, a true negative rate of 0.82, and an area under the curve of 0.986. Furthermore, the inclusion of non-financial indicators effectively enhanced model performance. These findings highlight the advantages of utilizing BiGRU for early detection of financial crisis, offering practical applications.

Povzetek: Razvit je bil model zgodnjega opozarjanja na finančne krize podjetij, testiran z bazo podatkov CSMAR in šestimi algoritmi: s podpornimi vektorji, XGBoost, LSTM, GRU, dvosmernim LSTM in dvosmernim GRU (BiGRU); slednji je najboljši s točnostjo 0,934 in AUC 0,986.

1 Introduction

Early warning of a financial crisis involves processing and analyzing financial data to identify potential difficulties and crises that a company may face. It is of significant importance to the company's management, investors, and financial institutions. It enables management to take timely remedial actions to avert crises, helps investors identify risks in advance to prevent investment losses [1], and assists financial institutions in avoiding the emergence of non-performing loans. As technology advances and richer financial data becomes available, research in early warning systems for financial crises has made substantial progress through the application of various algorithmic models, including statistical analysis and machine learning [2].

2 Related works

According to the literature in Table 1, currently, research on financial crisis prediction mainly relies on financial treatment and lacks analysis of non-financial indicators. This leads to insufficient comprehensiveness and authenticity in financial crisis prediction, and most of the studied models are based on machine learning methods with limited research on deep learning. Whether it is machine learning or deep learning, financial indicators or non-financial indicators, they all have significant research value in financial crisis prediction. Through comparative studies of different algorithm models, we can better identify the strengths and weaknesses of each model, providing more reliable decision-making basis for relevant stakeholders. Therefore, this article compares several different algorithm models in an attempt to find a more effective model for predicting financial crises of listed companies. This not only advances the theoretical research on financial crisis prediction but also provides decision-makers with new insights, promoting the healthy development of the financial industry.

Table 1: A table of related works

	Approach	Results/Findings
Jemovi et al. [3]	Panel logit regression	The dynamic discrete choice (binary) early warning model clearly outperformed the static model. The set of significant explanatory variables changed relative to the findings of the static model. The most significant predictor of the crises in the better performing model is

		deposit insurance system, followed by international reserves, M2-to-international reserves ratio, M2 multiplier, bank deposits, and bank reserves ratio.
Sun et al. [4]	Back-propagation neural network	The back-propagation neural network financial early warning model constructed in this paper has high prediction accuracy, which can be well used in the practice of financial early warning of mining listed companies
Liu [5]	The wavelet neural network improved by the fish swarm algorithm	The predicting correctness of samples is 100%, and results show that the fish swarm algorithm is an effective method for improving the financial risk early warning system.
Ashraf et al. [6]	The three-variable probit model and the Z-score model	The Z-score model more accurately predicts insolvency for both types of firms, i.e., those that are at an early stage as well as those that are at an advanced stage of financial distress.

3 Selection of prewarning indicators for financial crises

3.1 Study sample selection

According to current situation in China, publicly traded companies that have been undergone special treatment (including ST and ST*) were regarded as samples with financial crisis. Those who have never experience ST or ST* since being listed were regarded as normal samples. Samples were selected from the CSMAR database, and the selection criteria are as follows.

(1) Companies belonging to non-financial industries were selected. Financial industry companies are relatively unique in terms of their operational structure, making financial indicators non-comparable.

(2) A-share publicly traded companies were selected. There are more publicly traded companies in A-shares, and the data is more comprehensive.

(3) Only companies that have been specially treated due to abnormal financial condition were selected.

(4) Publicly traded companies that were specially treated for the first time during 2014-2020 were selected, excluding companies that were specially treated multiple times in the short term.

In the selection of paired samples, normal companies were screened from the database according to a 1:1 ratio, and the paired samples were required to be in the same industry and have comparable assets. Finally, 121 ST companies, along with their paired 121 non-ST companies, were selected as the study samples.

In the early warning research, the year defined as the year of being ST is year T. This paper mainly analyzed the early warning effect of the data from the T-2 and T-3 years. This is because the time span of year T-4 is relatively large, and all the data are affected by the environment, technology and so on, which is not informative. Generally, the financial statements of a company for year T-1 are published in March or April of year T. By that time, the financial crisis has already occurred, rendering any warning research meaningless. Therefore, the data from the T-2 and T-3 years were chosen for the study.

3.2 Selection of indicators

The data disclosed by listed companies contains a wealth of information related to financial crises, from which indicators can be selected for analysis in order to achieve crisis early warning. At present, there is no unified result in the selection of indicators. Based on the reference of existing research, this paper considered the following aspects in the selection of indicators.

(1) Earning level

It refers to the ability of a publicly traded company to gain earnings through operation. Long-term and stable earnings is the solid foundation of the company's development and also represents its resilience in the face of crisis. In the case of a good level of earnings, it means that the company has a stronger ability to create earnings, good development prospect, and high investability.

(2) Development level

It refers to a company's ability to continue to expand production and improve earnings on the current basis. It can determine whether a company has the possibility of long-term stable development. If the development level is good, it means that the company can maintain a high level of operation, investment, and financing, and has broad prospects for development.

(3) Debt service level

It refers to the ability of a publicly traded company to utilize its own assets to repay its debts. This ability can be considered from both short-term and long-term aspects. The short-term aspect reflects the company's current financial capacity. The long-term aspect reflects the company's long-term financial security. In the case of a poor debt service level, it indicates that the company's ability to repay its debts is weak and its capital chain may be unstable.

(4) Operating level

It refers to the ability of a company to utilize its existing assets to generate revenue, and it is a reflection of capital turnover. With a high operating level, a company utilizes its assets more fully and generates revenue at a faster rate.

(5) Cash flow level

It refers to the percentage of a company's cash, which can intuitively reflect the company's financial level. In the case of a poor cash flow level, the company is more likely to have a shortage of funds and a financial crisis.

(6) Equity structure

Unlike the first five aspects, equity structure is a non-financial factor. However, non-financial indicators can also reflect the current financial situation of a company to a certain extent and can further improve the effectiveness of early warning. Equity structure can reflect the distribution of a company's stock equity, and excessive dispersion or concentration is not conducive to the healthy operation of the company.

Combining the above six aspects, the preliminary selection of indicators is presented in Table 2.

Table 2: Preliminary selection of prewarning indicators.

Perspective	Serial number	Indicator
Earning level	X1	Earnings per share
	X2	Net sales margin
	X3	Return on net assets
	X4	Net interest rate on total assets
	X5	Return on current assets
	X6	Return on fixed assets
	X7	Ratio of net asset to cash flow
Development level	X8	Net profit growth rate
	X9	Net asset growth rate
	X10	Total asset growth rate
Debt service level	X11	Current ratio
	X12	Quick ratio
	X13	Cash flow ratio
	X14	Current liability ratio
	X15	Non-current liabilities ratio
	X16	Current assets ratio
	X17	Fixed asset ratio
Operating level	X18	Inventory turnover
	X19	Cash turnover ratio
	X20	Current asset turnover
	X21	Non-current asset turnover
	X22	Total asset turnover
	X23	Accounts receivable turnover ratio
Cash flow level	X24	Ratio of income to cash
	X25	Cash coverage ratio
	X26	Net cash flow per share
Equity structure	X27	Share proportion of the largest shareholder
	X28	Share proportion of the top three shareholders
	X29	Proportion of the top ten shareholders
	X30	Number of independent directors

In order to avoid the influence of the dimension and numerical values on the subsequent early warning model, all the indicators in Table 2 were normalized with the following formula:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}},$$

where X is the original data, X_{max} and X_{min} are the maximum and minimum values of the indicator.

In the 30 preliminary indicators, there may be some that are not significantly related to financial crises. If all these indicators are used as inputs for subsequent early warning models, it will result in longer training time and a decrease in accuracy. Therefore, in order to ensure the reliability of the indicators and improve the effectiveness of subsequent financial warning models, it is necessary to conduct significance tests on the selected prewarning indicators. Therefore, before the significance test, the Shapiro-Wilk test [7] was conducted to examine the normality of the variables. A p value greater than 0.05 was used as the criterion to determine whether the indicators followed a normal distribution.

Table 3: Results of the normal distribution test (note: bolded indicates p > 0.05).

Indicator	P value at the T-2 year	Indicator	P value at the T-3 year
X1	0.000	X1	0.000
X2	0.000	X2	0.005
X3	0.000	X3	0.000
X4	0.001	X4	0.000
X5	0.000	X5	0.000
X6	0.000	X6	0.004
X7	**0.341**	X7	**0.264**
X8	0.000	X8	0.000
X9	0.000	X9	0.000
X10	0.000	X10	0.000
X11	0.000	X11	0.005
X12	0.000	X12	0.000
X13	**0.511**	X13	**0.425**
X14	**0.323**	X14	**0.552**
X15	**0.125**	X15	0.001
X16	0.000	X16	0.00

X17	0.005	X17	0.000
X18	0.000	X18	0.000
X19	0.000	X19	0.000
X20	0.000	X20	0.000
X21	0.004	X21	0.000
X22	**0.198**	X22	**0.201**
X23	0.000	X23	0.000
X24	0.000	X24	0.000
X25	0.000	X25	0.000
X26	0.005	X26	0.000
X27	0.000	X27	0.000
X28	0.000	X28	0.000
X29	**0.501**	X29	**0.263**
X30	0.001	X30	0.000

From Table 3, it can be observed that in year T-2, all indicators do not obey normal distribution except X7, X13, X14, X15, X22, and X29, and in year T-3, all indicators do not obey normal distribution except X7, X13, X14, X22, and X29.

The T-test is a commonly used method for comparing whether there is a significant difference between the means of two sample groups. A T-test was performed on indicators that adhere to a normal distribution [8]:

$$T = \frac{(\overline{X_1}-\overline{X_2})-(m_1-m_2)}{\sqrt{\frac{\sigma_1^2}{N_1}+\frac{\sigma_2^2}{N_2}}},$$

where $\overline{X_1}$ and $\overline{X_2}$ are sample means of indicators for ST and non-ST companies, m_1 and m_2 are population means, σ_1 and σ_2 are population variances, N_1 and N_2 are sample sizes.

$P < 0.05$ was taken as a criterion to determine whether the indicators have significance. The results are presented in Table 4.

Table 4: T-test results.

Indicator	P value at the T-2 year	Indicator	P value at the T-3 year
X7	0.000	X7	0.000
X13	0.000	X13	0.000
X14	0.000	X14	0.000
X15	0.000	X22	0.000
X22	0.000	X29	0.000
X29	0.000		

According to Table 4, all indicators satisfied $p < 0.05$, i.e., they could help effectively distinguish between ST and non-ST companies; therefore, they were retained.

In cases where the data does not follow a normal distribution, T-test is not applicable, whereas Mann-Whitney U test does not rely on data distribution. Therefore, the Mann-Whitney U test [9] was conducted on indicators that do not obey the normal distribution:

$$U_1 = N_1N_2 + \frac{N_1(N_1+1)}{2} - \sum_{i=1}^{N_1} R_i,$$
$$U_2 = N_1N_2 + \frac{N_2(N_2+1)}{2} - \sum_{i=1}^{N_2} R_i,$$

where N_1 and N_2 are sample sizes, and R_i is the rank of each set of data.

$P < 0.05$ was taken as a criterion to determine whether the indicators have significance. The results are presented in Table 5.

Table 5: Results of the Mann-Whitney U test (note: bolded indicates $p > 0.05$).

Indicator	P value at the T-2 year	Indicator	P value at the T-3 year
X1	0.000	X1	0.000
X2	0.000	X2	0.000
X3	0.001	X3	0.000
X4	**0.056**	**X4**	**0.057**
X5	**0.072**	**X5**	**0.051**
X6	**0.068**	**X6**	**0.062**
X8	0.000	**X8**	**0.053**
X9	**0.074**	**X9**	**0.062**
X10	0.000	X10	0.000
X11	**0.076**	**X11**	**0.059**
X12	**0.074**	**X12**	**0.067**
X16	**0.055**	X15	0.000
X17	**0.057**	**X16**	**0.054**
X18	**0.068**	**X17**	**0.062**
X19	**0.057**	**X18**	**0.072**
X20	0.000	**X19**	**0.061**
X21	**0.064**	**X20**	**0.067**
X23	0.000	**X21**	**0.058**

X24	**0.067**	**X23**	**0.071**
X25	**0.075**	**X24**	**0.072**
X26	0.000	**X25**	**0.062**
X27	0.000	X26	0.000
X28	**0.056**	X27	0.000
X30	0.000	**X28**	**0.052**
		X30	0.000

Excluding the indicators that are not significant in Table 5, the final prewarning indicators obtained are listed in Table 6.

Table 6: Indicators that passed the test.

Aspect	Indicator for year T-2	Aspect	Indicator for year T-3
Earning level	X1	Earning level	X1
	X2		X2
	X3		X3
	X7		X7
Development level	X8	Development level	X10
	X10		
Debt service level	X13	Debt service level	X13
	X14		X14
	X15		X15
Operating level	X20	Operating level	X22
	X22		
	X23		
Cash flow level	X26	Cash flow level	X26
Equity structure	X27	Equity structure	X27
	X29		X29
	X30		X30

From Table 6, it is evident that 16 indicators were retained for year T-2, while 13 were retained for year T-3. This suggested that as the ST year approached, anomalies in the indicators became more pronounced. Among the various aspects considered, the earning level retained the most indicators, indicating that the earning level played a crucial role in predicting financial crises. Furthermore, Table 6 suggests that only one of the non-financial indicators, namely equity structure, was excluded. This demonstrated the importance of including non-financial indicators in earning warnings.

4 Different algorithmic models

4.1 Support vector machine

The support vector machine (SVM) is a single machine learning model widely used for data classification and recognition [10]. It exhibits excellent performance in handling nonlinear relationships, robustness to noise, and good generalization ability. SVM can capture potential nonlinear relationships in financial data and handle noisy financial data better. Therefore, a financial crisis prewarning model for listed companies based on SVM can be established. It works by creating a hyperplane to categorize data into two or more classes. For datasets $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$, $y_i \in (-1,1)$, its optimal hyperplane can be written as:

$$wx + b = 0,$$

where w is a weight and b is a bias. To solve the optimal hyperplane, it is converted to a dual problem and solved using the Lagrange transform. The equations are:

$$maxW(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j),$$

$$s.t. \sum_{i=1}^{n} \alpha_i y_i = 0,$$

where α represents the Lagrange multiplier and $k(x_i \cdot x_j)$ is the kernel function. The decision function of SVM can be written as:

$$f(x) = sign\{\sum_{i=1}^{n} \alpha_i\, y_i K(x_i, x_j) + b\}.$$

4.2 XGBoost

XGBoost is an integrated machine learning model [11], which is based on the principle of integrated learning through multiple decision trees for better classification performance. XGBoost automatically selects the most important features and performs well in handling imbalanced data. Financial crisis data often exhibits imbalance and high complexity, but XGBoost can effectively capture nonlinear relationships and identify key indicators for predicting financial crises. Even in situations where there is an imbalance in the samples of financial crises, it maintains good performance. Therefore, XGBoost can be utilized to build a financial crisis prewarning model. For dataset $D = (x_i, y_i)$, the classification result of the l-th tree can be written as $\hat{y}_l = \sum_{k=1}^{k} f_k(x_i)$, where $f_k(x_i)$ is the classification result of the k-th tree. The classification result of the tree and the loss function is written as:

$$loss = \sum_i l(\hat{y}_l, y_i) + \sum_k \Omega(f_l),$$

where $\hat{y}_l$ is the predicted value, y_i is the actual value, and $\Omega(f_l)$ is a penalty term. The objective function at the t-th round of iteration can be written as:

$$loss^{(t)} = \sum_{i=1}^{n} l\left(\hat{y}_l^{(t-1)}, y_i + f_t(x_i)\right) + \Omega(f_t),$$

where $\hat{y}_l^{(t)}$ refers to the predicted value of the t-th round.

4.3 Long short-term memory neural network

Long short-term memory (LSTM) is a deep learning model [12], which has better learning performance for features and better classification results compared to machine learning models. Financial data generally exhibits strong temporal patterns, and through LSTM, it is able to better learn the long-term dependencies within sequences, capturing trends and changes in the data more accurately, thus enabling more precise predictions of financial crises. LSTM regulates the cell state mainly through three gates. Firstly, the forgetting gate decides whether to retain the information of previous state c_{t-1}. The output is:

$$f_t = \sigma\left[W_f \cdot (h_{t-1}, x_t) + b_f\right].$$

The input gate determines how much information should be passed to update the state:

$$i_t = \sigma[W_i \cdot (h_{t-1}, x_t) + b_i].$$

After the calculation of c_{t-1}, f_t, and i_t, the new cell state is obtained:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh[W_c \cdot (h_{t-1}, x_t) + b_c].$$

The output gate determines the information passed from the current cell state to the hidden state:

$$o_t = \sigma[W_o \cdot (h_{t-1}, x_t) + b_o].$$

where W and b are the weight and threshold of each layer. Finally, the output of LSTM can be written as:

$$h_t = o_t \cdot \tanh(c_t).$$

4.4 Gated recurrent unit

Gated recurrent unit (GRU) is also a deep learning method [13], which uses only two gates compared to LSTM, thus increasing the speed of training. Compared to LSTM, GRU achieves a balance between long-term and short-term memory and has higher training efficiency. Therefore, in the processing of financial data, utilizing GRU can better capture short-term fluctuations and long-term trends in the data, resulting in more effective models trained in a shorter time. In GRU, let the current input be x_t and the output at the last moment be h_{t-1}, how much information needs to be abandoned is decided by the reset gate:

$$R_t = \sigma[W_r \cdot (h_{t-1}, x_t)].$$

The update gate is responsible for determining the amount of information that should be transmitted:

$$Z_t = \sigma[W_z \cdot (h_{t-1}, x_t)].$$

The forward propagation process of GRU can be written as:

$$h_t' = \tanh[W_h x_t + U_h(h_{t-1} r_t)],$$

$$h_t = (1 - z_t) h_t' + z_t h_{t-1},$$

where h_t' is the candidate hidden layer, h_t is the hidden layer, W and U are weights.

For both LSTM and GRU models, their performance can be further improved by using a bidirectional structure called bi-directional LSTM (BiLSTM) and bi-directional GRU (BiGRU) [14]. Taking BiLSTM as an example, it includes a forward LSTM layer and a backward LSTM layer to capture the forward and backward information of the sequences, which can be expressed as:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(h_{t-1}, x_t, c_{t-1}),$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(h_{t+1}, x_t, c_{t+1}),$$

$$H_t = [\overrightarrow{h_t}, \overleftarrow{h_t}].$$

BiGRU uses the same structure.

5 Results and analysis

5.1 Experimental setup

The machine learning model was built by sklearn library in Python, and the deep learning model was built by the PyTorch framework and tuned through grid search method [15]. The parameter ranges are presented in Table 7.

Table 7: Model parameter settings.

Model	Hyper-parameter	Range
SVM	Penalty coefficient	[100,10,1,0.1,0.01]
	gamma	[auto,0.1]
XGBoost	Number of weak learners	[10,20,30,40,50,60]
	Maximum depth	[1,3,5,7,9]
	Regularization parameter	[0,0.001,0.005,0.01,0.1]
	Learning rate	[0.1,0.01,0.001]
LSTM, GRU, BiLSTM and BiGRU	Number of neurons in hidden layer	[16,32,64,128]
	Epoch	[10,30,50,100,200,300]
	Batch size	[64,128,256,512,1024]
	Learning rate	[0.001,0.005,0.01,0.1]

The experiment adopted a five-fold cross-validation method to compare the performance of different algorithmic models after tuning. Based on the confusion matrix (Table 8), ST companies were set as 1, and non-ST companies were set as 0.

Table 8: Confusion matrix.

	Predicted value = 1	Predicted value = 0
Actual value = 1	TP	FN
Actual value = 0	FP	TN

The evaluation indicators are listed below.

(1) Classification accuracy (ACC): the proportion of correctly classified samples to the total number, $ACC = \frac{TP+TN}{TP+FN+TN+FP}$.

(2) True positive rate (TPR): how many positive samples were correctly classified, $TPR = \frac{TP}{TP+FN}$;

(3) True negative rate (TNR): how many negative samples were correctly categorized: $TNR = \frac{TN}{TN+FP}$;

(4) Area under the curve (AUC): the area under the receiver operator characteristic (ROC) curve; the higher

the value, the more superior the model's classification performance.

5.2 Comparison of results

Firstly, the ACC of different algorithmic models was compared in Table 9.

Table 9: Comparison of ACC between different algorithmic models.

	Year T-2	Year T-3
SVM	0.853	0.752
XGBoost	0.862	0.775
LSTM	0.897	0.794
GRU	0.912	0.807
BiLSTM	0.927	0.812
BiGRU	0.934	0.839

As shown in Table 9, firstly, both the SVM and XGBoost models had lower ACC values compared to the deep learning methods. Then, the comparison between single model and bidirectional model demonstrated that the ACC values of the BiLSTM and BiGRU models were higher than those of the LSTM and GRU models, and the ACC value of the BiGRU model was the highest among the six models that were compared. Taking the T-2 year as an example, the ACC of the BiGRU model was 0.934, which was 0.76% higher than the BiLSTM model, 2.41% higher than the GRU model, and 9.5% higher than the SVM model. Then, the comparison of the T-2 and T-3 years suggested that the ACC value of the models was not as good when using the data from the T-3 year compared to using the data from the T-2 year. This indicated that using the T-2 year data for classification resulted in better performance.

Comparison of TPR and TNR between different algorithms is given in Table 10.

Table 10: Comparison of TPR and TNR between different algorithmic models.

	TPR		TNR	
	T-2 year	T-3 year	T-2 year	T-3 year
SVM	0.742	0.633	0.766	0.651
XGBoost	0.792	0.657	0.821	0.687
LSTM	0.827	0.672	0.845	0.703
GRU	0.835	0.692	0.857	0.725
BiLSTM	0.857	0.712	0.876	0.747
BiGRU	0.875	0.733	0.892	0.762

From Table 10, it can also be observed that the TPR and TNR of the models were not as good when using data from T-3 year compared to when using data from T-2 year. This indicated that the closer the data used was to the year when the financial crisis occurred, the better the early warning performance. From the comparison of different algorithmic models, it can be concluded that the BiGRU model exhibited the best early warning performance. Taking year T-2 as an example, the TPR value of the BiGRU model was 0.875, which showed a 17.92% improvement compared to the SVM mode. Additionally, its TNR was 0.892, demonstrating a 16.45% improvement compared to the SVM model. These findings confirmed the effectiveness of the model.

The comparison of AUC between different algorithmic models is illustrated in Figure 1.

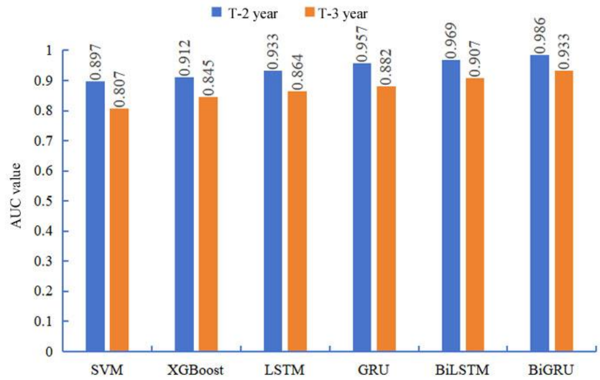

Figure 1: Comparison of AUC values between different algorithmic models.

According to Figure 1, the BiGRU model exhibited the best performance compared to the other models. At year T-2, the BiGRU model achieved an AUC value of 0.986, demonstrating a 1.75% improvement compared to the BiLSTM model, a 3.03% increase compared to GRU, and a 9.92% increase compared to the SVM model. When the data from year T-3 was used, the AUC value of the BiGRU model was 0.933, which was 5.38% lower compared to the result obtained when the data from year T-2 was used. This further supported the effectiveness of the BiGRU model in predicting financial crises in publicly traded companies.

Then, the choice of indicators was analyzed using the BiGRU model. Taking year T-2 as an example, the specific values are displayed in Table 11.

Table 11: Impact of indicator selection on early warning performance.

	30 preliminary indicators in Table 2	Indicators passing the significance test in Table 6	Indicators after excluding equity structure (X27, X29, X30)
ACC	0.812	0.934	0.907
TPR	0.752	0.875	0.841
TNR	0.761	0.892	0.851
ACC	0.874	0.986	0.941
AUC	0.812	0.934	

From Table 11, it can be found that if all indicators without screening were used as inputs for the model, the resulting ACC value was only 0.812, which showed a decrease of 13.06% compared to the screened indicators. The TPR value was 0.752, showing a decrease of 14.06%, and the TNR value was 0.761, showing a decrease of 14.69%. The AUC value also decreased by 11.36% to reach 0.874. This result demonstrated the impact of significance testing on warning performance; poor quality indicators that have not been screened could actually lead to a decline in warning performance. The early warning performance of the BiGRU model showed a significant decrease after the exclusion of the equity structure indicators, where the ACC was 0.907, which was decreased by 2.89%, the TPR was 0.841, which was decreased by 3.89%, the TNR was 0.851, which was decreased by 4.6%, and the ACC was 0.941, which was decreased by 4.56%. These results proved the importance of non-financial indicators in early warning research. More non-financial indicators can be considered in future work to further enhance early warning performance.

6 Discussion

Early warning of financial crises is an important aspect of company management, and with the advancement of technology, more and more methods have been applied. However, current research mostly focuses on in-depth analysis of a single method and limited feature selection to financial indicators. Therefore, further research is needed regarding the issue of early warning for financial crises in listed companies. This article compared the effectiveness of different algorithm models in predicting financial crises for listed companies. It not only analyzed machine learning methods, but also studied deep learning methods. Additionally, non-financial indicators were added to the indicator selection process to better understand the strengths and weaknesses of different models, providing new insights for research on financial crisis prediction.

From the experimental results, it can be seen that among the six compared algorithm models, BiGRU performed the best in financial crisis prediction with high values of ACC and other indicators. BiGRU is a deep learning method that can simultaneously learn from both forward and backward data. It also exhibits better performance in handling non-linear relationships and temporal features, thus achieving superior results compared to methods such as machine learning. From the perspective of data selection, the predictive power of the T-3 year was inferior to that of the T-2 year. This result indicated that in financial crisis prediction, the closer the selected data was to the occurrence of a financial crisis, the more relevant features it contained, and the better its predictive effect was. Finally, the analysis of indicator selection showed that input data quality had a certain impact on the results of predictive models and indicators not subjected to significance testing could lead to decreased predictive effectiveness. Furthermore, after excluding equity structure (non-financial indicators), there was also a decrease in explanatory power for financial crises - demonstrating the important role non-financial indicators play in predicting financial crises.

In practical applications, suitable financial crisis warning models can be chosen based on different factors such as industry, company size, and actual needs. Due to varying requirements for real-time performance and computational resources in different scenarios, it is also possible to select models that are more closely aligned with real-world applications. Although this study provides a new approach to the financial crisis warning problem for listed companies, there are still some limitations. For example, there is a limited selection of non-financial indicators and a small scope of research data. The real-time aspect of the model has not been fully considered. In future work, it is possible to gather more financial data from different industries to analyze the applicability of financial crisis warning models. Additionally, consideration can be given to a lightweight design of the model in order to further enhance its computational efficiency and predictive performance.

7 Conclusion

This paper focuses on the prewarning of financial crises in publicly traded companies and conducts a comparative analysis of six different algorithmic models using selected indicators. The results revealed that among these models, the BiGRU model demonstrated the most robust performance. It achieved an ACC of 0.934, a TPR of 0.975, a TNR of 0.82, and an AUC of 0.986. The research results confirm the effectiveness of the BiGRU model in providing prewarning for financial crises in publicly traded companies and its potential for further promotion and application in the real world.

References

[1] Alamsyah A, Kristanti N, Kristanti F T (2021). Early warning model for financial distress using Artificial Neural Network. *IOP Conference Series: Materials Science and Engineering*, 1098, pp. 1-6. https://doi.org/10.1088/1757-899X/1098/5/052103.

[2] Duprey T, Klaus B (2022). Early warning or too late? A (pseudo-)real-time identification of leading indicators of financial stress. *Journal of Banking & Finance*, 138, pp. 106196. https://doi.org/10.1016/j.jbankfin.2021.106196.

[3] Jemovi M, Marinkovi S (2021). Determinants of financial crises—An early warning system based on panel logit regression. *International Journal of Finance & Economics*, 26, pp. 103-117. https://doi.org/10.1002/ijfe.1779.

[4] Sun X, Lei Y (2021). Research on financial early warning of mining listed companies based on BP neural network model. *Resources Policy*, 73, pp. 1-10. https://doi.org/10.1016/j.resourpol.2021.102223.

[5] Liu YS (2018). A Fish Swarm Algorithm for Financial Risk Early Warning. *International Journal of Enterprise Information Systems (IJEIS)*, 14, pp. 54-63. https://doi.org/10.4018/IJEIS.2018100104.

[6] Ashraf S, Félix EGS, Serrasqueiro Z (2019). Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress?. *Journal of Risk and Financial Management*, 12, pp. 1-17. https://doi.org/10.3390/jrfm12020055.

[7] Cacilda D S S, Cláudia GS, Sonia Maria AALL, Natali SG, Renata FM, Roberto R (2022). Clinical Specialty Setting as Determinant of Management of Psoriatic Arthritis: A Cross-Sectional Brazilian Study. *Journal of Clinical Rheumatology*, 28, pp. 120-125. https://doi.org/10.1097/RHU.0000000000001812.

[8] Zhou J Y, Mccomish S L, Tolmachev S Y (2019). A Monte Carlo t Test to Evaluate Mesothelioma and Radiation in the United States Transuranium and Uranium Registries. *Health Physics*, 117, pp. 187-192. https://doi.org/10.1097/HP.0000000000001075.

[9] Alzoraigi U, Shubbak F (2021). Effectiveness of preoperative tour to a simulated anaesthesia induction at operating theatre in reducing preoperative anxiety in children and their parents: a pragmatic, single-blinded, randomised controlled trial/ King Fahad Medical City. *BMJ Simulation and Technology Enhanced Learning*, 7, pp. 397-403. https://doi.org/10.1136/bmjstel-2020-000707.

[10] Sretenovic A, Jovanovic R, Novakovic V, Nord N (2018). Support vector machine for the prediction of heating energy use. *Thermal Science*, 2018, pp. 126-126. https://doi.org/10.2298/TSCI170526126S.

[11] Li CB, Zheng XS, Yang ZK, Kuang L (2018). Predicting Short-Term Electricity Demand by Combining the Advantages of ARMA and XGBoost in Fog Computing Environment. *Wireless Communications & Mobile Computing*, 2018, pp. 1-18. https://doi.org/10.1155/2018/5018053.

[12] Zhang C, Wang WZ, Zhang C, Fan B, Wang JG, Gu FS, Xue Y (2022). Extraction of local and global features by a convolutional neural network–long short-term memory network for diagnosing bearing faults. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 236, pp. 1877-1887. https://doi.org/10.1177/09544062211016505.

[13] Yevnin Y, Chorev S, Dukan I, Toledo Y (2023). Short-term wave forecasts using gated recurrent unit model. *Ocean Engineering*, 268, pp. 1-8. https://doi.org/10.1016/j.oceaneng.2022.113389.

[14] Ren Y, Liao F, Gong Y (2020). Impact of News on the Trend of Stock Price Change: an Analysis based on the Deep Bidirectiona LSTM Model. *Procedia Computer Science*, 174, pp. 128-140. https://doi.org/10.1016/j.procs.2020.06.068.

[15] Belete D M, Huchaiah M D (2022). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *International Journal of Computers & Applications*, 44, pp. 875-886. https://doi.org/10.1080/1206212X.2021.1974663.