A novel method for financial distress prediction based on sparse neural networks with $L_{1/2}$ regularization

Ying Chen · Jifeng Guo · Junqin Huang · Bin Lin

Received: 20 December 2021 / Accepted: 6 April 2022 / Published online: 27 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Corporate financial distress is related to the interests of the enterprise and stakeholders. Therefore, its accurate prediction is of great significance to avoid huge losses from them. Despite significant effort and progress in this field, the existing prediction methods are either limited by the number of input variables or restricted to those financial predictors. To alleviate those issues, both financial variables and non-financial variables are screened out from the existing accounting and finance theory to use as financial distress predictors. In addition, a novel method for financial distress prediction (FDP) based on sparse neural networks is proposed, namely FDP-SNN, in which the weight of the hidden layer is constrained with $L_{1/2}$ regularization to achieve the sparsity, so as to select relevant and important predictors, improving the predicted accuracy. It also provides support for the interpretability of the model. The results show that non-financial variables, such as investor protection and governance structure, play a key role in financial distress prediction than those financial ones, especially when the forecast period grows longer. By comparing those classic models proposed by predominant researchers in accounting and finance, the proposed model outperforms in terms of accuracy, precision, and AUC performance.

Keywords Financial distress prediction · Features selection · Sparse neural networks · $L_{1/2}$ regularization

1 Introduction

Corporate financial distress is one of the important research issues internationally. Both theoretical researchers and practical experience show that failures and bankruptcy filings are a result of financial or economic distress. Even if the firms survive from corporate failures, financial distress can still cause significant direct and indirect costs to them and their stakeholders [3, 5, 25, 35]. Therefore, how to detect and prevent financial distress on a timely basis would offer great attention and significant value to firms, regulators, investors, and other interest-related parties.

Researchers have devoted great efforts to find efficient and effective methods for corporate financial distress prediction, which mainly contains two research streams. One is the expansion and supplement of the predictive factors based on classic statistical methods, e.g., multivariate discriminant analysis (MDA) and logit regression model (Logit) or probit regression models (Probit). In this research stream, various financial and non-financial variables had been explored [2, 7, 11, 19, 52]. However, these classic statistical models were restricted by strict assumptions, such as variables being normally distributed, equal variance covariance matrices across treating and control groups and the absence of multi-collinearity etc. [6, 48]. These assumptions had greatly limited the number of predictors in the models, which made it hard to deal with a large number of predictors and improve accuracy.

The other stream in financial distress prediction is the choice and innovation of the method. To overcome the limitations of classic statistical methods, some researchers started to apply machine learning methods into FDP, among which support vector machine (SVM), decision tree (DT) and neural networks (NN) has been widely used [53, 62, 71]. Compared with classic statistical methods, machine learning methods increase the quantity of variables in the models,
enabling features selection and accuracy improvement. Especially, as an efficient method, neural network has been proved to possess the ability to approximate any nonlinear functions, and has been successfully applied to and exhibited excellent performance in FDP [12, 20]. However, existing methods for FDP always utilize only financial variables as predictors, ignoring non-financial variables. Actually, the related researchers have found that financial variables are the reflection of corporate financial situation, while non-financial variables including strategy and governance structure indeed determine the financial situation [38, 44]. Therefore, non-financial variables may be more powerful than financial variables in FDP.

Considering the above issues, this paper adopts more predictors including financial variables, non-financial variables, and also proposes a novel feature selection and prediction method for financial distress using sparse neural networks with $L_{1/2}$ regularization, in which the weights connect the input and hidden layer are designed in sparse coding, achieving the purpose of feature selection. Meantime, based on the simplified data, the recognition networks can act out better classification effects. The contributions of this paper are listed as follows:

- A novel prediction method for financial distress is proposed, which adopts the sparse neural networks with $L_{1/2}$ regularization to simplify features, and further improves recognition accuracy. Besides, it is extremely beneficial to the interpretability of the model.
- This paper considers extensive predictors including financial and non-financial variables, which greatly improves the accuracy of the FDP model. Besides, the results also show that non-financial variables are more important in the FDP.

The organization of the paper is described below. Sect. 2 reviews the related work. Section 3 introduces the related technologies and the proposed method. Section 4 exhibits and analyses experiments. Section 5 is conclusion and future work.

2 Related works

In this section, the related works on the financial distress and the neural networks, especially the technique in neural networks and sparsity.

2.1 Financial distress prediction

The FDP has been extensively researching areas since the late 1960s. Various statistical and intelligence techniques have been used in this area. However, the most widely used methods are still those classic statistical methods, e.g., MDA, Logit and Probit. Because the most important issue for finance and accounting researchers is to explore new financial distress predictors to build and verify the FDP theory and those methods are qualified for those tasks. Since Altman [2] have innovatively used financial variables to predict financial bankruptcy, researchers started their work on the expansion and supplement of financial predictors [9, 19, 47]. Hereafter, both accounting and finance researchers found that except for financial variables, the non-financial variables, such as government structure [11, 57], information disclosure [34, 52], investor protection [8] and strategy [23] can be used to predict financial distress. During exploring new factors, classic statistical methods had been widely used. However, those statistical models were questioned and criticized by strict assumptions of variation homogeneity of data [48], which makes them being sensitivity to multicollinearity and limited in the number of predictors. Kumar and Ravi [43] found that the maximum number of significant predictors in those models is 20 variables and more predictors would not improve the prediction performance anymore. That is, limitation for variables have restricted the information content in prediction models. As a result, it has been hard to select key features from a large number of predictors and improve accuracy.

With the development of statistical methods and computer technology, machine learning methods started to be applied in FDP, including support vector machine, decision tree and neural networks. For example, Min and Lee [53] constructed a FDP model based on SVM with 38 financial variables as predictors. Sun and Li [62] design a FDP model based on DT with 35 financial variables as predictors. Chen and Du [12] test a FDP model based on NN with 37 financial variables as predictors. Zhou et al. [71] combined multiple machine learning approaches to select 20 features from 338 financial variables for FDP. Compared with classic statistical methods, the machine learning methods have advantages in their capability of modeling complex relationships between independent and dependent features without strong model assumptions, which makes it possible to put more predictors in a model so as to select features and improve accuracy. However, existing machine learning-based methods for FDP always use only financial variables without considering non-financial variables. Detailed comparison between classic statistical methods and machine learning methods for FDP is shown in Table 1.

2.2 Neural networks

Machine learning methods, especially neural networks, have been proved can fit linear and nonlinear relationships [31] and have been applied in various fields [1, 13, 30]. Kiran et al. [42] applied artificial neural networks (ANN) to predict
the number of students taking make-up examinations. Singh [61] used it to determine the length of intervals in fuzzy time series (FTS) forecasting. Singh et al. [65] adopted the backpropagation neural network (BP-NN) to reconstruct the missing color-channel data. Namasudra, Dhamodharavadhani, and Rathipriya [54] proposed a neural network-based tool to predict the confirmed, recovered, and death of COVID-19. Goel, Murugan, Mirjalili, and Chakrabartty [29] also achieved its automatic diagnosis by using convolutional neural network.

Especially, Chen and Du [12] applied data mining techniques in the form of neural networks to build and test financial distress prediction models. Meantime, they also demonstrated its feasibility and validity. Hereafter, many researchers supported the neural networks approach and found that neural networks performed better in financial distress predicting than decision trees and other alternative approaches such as SVM [26]. Despite its excellent performance, neural networks still face great challenges in dealing with high-dimensional data.

### 2.3 Sparsity regularization

The redundant information in the high-dimensional data [46] will seriously influence the performance of classifiers, especially those methods that have no feature extraction or selection [69]. Sparsity provides an effective method to reduce features and improve performance, and plays an increasingly important role in fields such as machine learning and image processing [55].

The sparsity approach removes a large number of redundant variables and retains only the explanatory variables that are most relevant to the response variables, simplifying the model and effectively solving many problems in modeling high-dimensional datasets [45, 66]. It has better explanatory power and facilitates data visualization, reduced computational effort and transmission storage.

$L_0$ regularization is the first sparse regularization method applied to variable selection and extraction, which can give the optimal variable selection constrained by the number of parameters. However, it needs to solve a difficult combinatorial optimization problem. The $L_1$ regularization proposed by Tibshirani [63] provides a powerful tool that only needs to solve a quadratic programming problem. However, its sparsity is lower than $L_0$. The $L_1/2$ regularization between them had proved to have better feature selection ability and compression representation ability than $L_1$, which has a wide range of value and significance [70]. M. Chen, Mi, He, Deng, and Wei [14] replaced the $L_1$ regularization with $L_1/2$ regularization in the reconstruction of the CT images, achieving great unbiasedness and acceleration. Liu et al. [50] proved its effectiveness in the variable selection. Wu et al. [68] investigated gene selection in cancer classification using the $L_1/2$ regularized logistic regression, which outperforms the other sparse methods.

### 3 Methodology

In this section, the basic methods and the proposed FDP-SNN will be described. In addition, the basic methods include neural networks and $L_1/2$ regularization.

#### 3.1 Basic methods

##### 3.1.1 Neural networks

The neural networks is a mathematical or computational model that mimics the structure and function of a biological neural network. It consists of a large number of neurons linked together for computation. In most cases, neural networks can change their internal structure on the basis of

| Streams | Methods | Examples | Achievements | Shortcoming |
|---------|---------|----------|--------------|-------------|
| Classic Statistical Methods | MDA | Altman [2] | New financial distress predictors can be explored to build and verify FDP theory | The maximum number of predictors in model is limited so that it is hard to select key features and improve accuracy |
| | Logit | Deakin [19] | | |
| | Probit | Martin [51] | | |
| | | Ohlson [56] | | |
| | | Casey et al. [11] | | |
| | | Zmijewski [72] | | |
| Machine Learning Methods | SVM | Hua et al. [37] | More predictors can be included in the model so that key features can be selected and accuracy can be improved | Only financial variables are considered as predictors and non-financial variables are ignored |
| | DT | Min and Lee [53] | | |
| | | Frydman et al. [24] | | |
| | | Sun and Li [62] | | |
| | NN | Altman [2] | | |
| | | Chen and Du [12] | | |

| Streams | Methods | Examples | Achievements | Shortcoming |
|---------|---------|----------|--------------|-------------|
| Classic Statistical Methods | SVM | Hua et al. [37] | More predictors can be included in the model so that key features can be selected and accuracy can be improved | Only financial variables are considered as predictors and non-financial variables are ignored |
| | DT | Min and Lee [53] | | |
| | | Frydman et al. [24] | | |
| | | Sun and Li [62] | | |
| | NN | Altman [2] | | |
| | | Chen and Du [12] | | |

Table 1 Comparison between classic statistical methods and machine learning methods for FDP
are calculated by Eqs. (4) and (5), respectively. For the predicted output are trained by using gradient descent optimization algorithms [59]. For the predicted output, the loss function can be established when using the Cross-entropy function, and is represented by Eq. (3).

\[
L = \frac{1}{N} \sum_i^{N} \left[ y_i \cdot \log(z_i) + (1 - y_i) \cdot \log(1 - z_i) \right]
\] (3)

The gradients of loss function \( L \) with respect to \( W \) and \( b \) are calculated by Eqs. (4) and (5), respectively.

\[
\frac{\partial L}{\partial W_j} = \frac{1}{N} \sum_i x_i (z_i - y_i)
\] (4)

\[
\frac{\partial L}{\partial b} = \frac{1}{N} \sum_i (z_i - y_i)
\] (5)

Then the weight \( W_j \) is updated iteratively by Eq. (6).

\[
W'_j = W_j - \eta \cdot \frac{\partial L}{\partial W_j}
\] (6)

where \( \eta \in (0, 1) \) is the learning rate, while the bias \( b \) is updated iteratively by Eq. (7).

\[
b' = b - \eta \cdot \frac{\partial L}{\partial b}
\] (7)

During the optimization, an iteration termination condition is set, either by terminating the recursion when the error is less than a certain value, or by setting the number of iterations. When it is finished, the neural network with optimal parameters can be obtained.

3.1.2 \( L_{1/2} \) regularization

Variable selection and feature extraction are the basic problems when processing the high-dimensional and massive data. If there are redundant variables in the data, identifying the real variables while eliminating the redundant ones is called the sparse problem.

Since the \( L_{1/2} \) regularization can produce a sparser solution than the \( L_1 \) regularization and is easier to solve than the \( L_0 \) regularization, it has been widely applied in sparse problem [45]. For the data \( \{X, Y\} \), assuming there is a unknown but definite dependencies \( f^*(x) \). Based on the training data, the variable selection aiming at prediction accuracy can be achieved by minimizing the expected risk as Eq. (8).

\[
\lim_{\beta} L(\beta) = E_{x,y,l}(y,f(x,\beta))
\] (8)

where \( \beta \) is the obtained parameter finally.

Since the distribution function of \( \{X, Y\} \) is unknown, the expected risk is replaced by empirical risk and calculated by minimizing empirical risk, as Eq. (9).

\[
\lim_{\beta} L_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} l(y_i, f(x_i, \beta))
\] (9)

Generally, over fitting occurs when solving the Eq. (9). To avoid this issue, it is solved by imposing some constraints on Eq. (9), such as Eq. (10).

\[
\lim_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} l(y_i, f(x_i, \beta)) + \lambda P(\beta) \right\}
\] (10)

where \( \lambda P(\beta) \) is the sparse regularization term, and \( \lambda \) is its coefficient. When \( L_{1/2} \) regularization is adopted, the parameter estimates \( \beta_{L_{1/2}} \) can be calculated using Eq. (11).
\[ \beta^*_l = \lim_{\mu \to 0} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 + \lambda \sum_{p} \| \beta_p \|^2 \right\} \quad (11) \]

### 3.2 FDP-SNN

The framework of the proposed FDP-SNN is shown in Fig. 2. First, the original data are input into the neural network with sparse regularization for optimization. After learning, the neural network can predict whether the company is facing financial distress. Meanwhile, the importance of various variables can be sorted by the weight in sparse coding, achieving the purpose of variables selection. The details about the method are described in the following subsections.

Since the financial distress in this work is high-dimensional data including financial variables and non-financial variables, it is necessary to reduce the redundant features information. Aiming at the characteristics of mutual correlation and nonlinearity between the characteristics of financial data, the advantages of neural network in solving nonlinear problems are combined with sparsity norm to solve the problem of feature selection and classification. In addition, the \( L_{1/2} \) norm is adopted to sparse the weights in the neural network, because it has better sparsity than \( L_1 \) and \( L_2 \) norm, and even their combination.

As shown in Fig. 2, the original financial distress data \( \{X, Y\} \) are fed into the single hidden neural network with input nodes, \( q \) hidden nodes and 2 output nodes. Its initial weight of each layer are \( W_{1,q} \in \mathbb{R} \) and \( W_{2,q} \in \mathbb{R} \). The transfer function from hidden layer to output layer is \( f : \mathbb{R} \to \mathbb{R} \). Particularly, it adopts a sigmoid function as an example. \( F: \mathbb{R} \to \mathbb{R} \), \( F = (f(x_1), f(x_2), ..., f(x_q)) \) is a defined energy function. Its final predict output \( Z \) is calculated by Eq. (12).

\[ Z = f(W_2 : F(X \cdot W_1 + b_1) + b_2) \quad (12) \]

To achieve the purpose of the feature selection, the weights \( W_1 \) between the input and hidden layer are restricted by \( L_{1/2} \) regularization, and the loss function of neural network is modified to Eq. (13).

\[ L(W) = \frac{1}{N} \sum_{i} [y_i \cdot \log(z_i) + (1 - y_i) \cdot \log(1 - z_i)] + \lambda ||W_1||_{1/2} \quad (13) \]

Then in the process of back propagation, its gradients are represented by Eq. (14) and Eq. (15).

\[ \frac{\partial L}{\partial W_2} = \frac{1}{n} \sum F(X \cdot W_1 + b_1)(z - y) \quad (14) \]

\[ \frac{\partial L}{\partial W_1} = \frac{1}{n} \sum X \cdot (z - y) \cdot W_2 \left[ 1 - F^2(X \cdot W_2 + b_2) \right] + \frac{\lambda \text{sgn}(W_1)}{2||W_1||} \quad (15) \]

The weight \( W_1 \) is updated iteratively by Eq. (16).

\[ W_1' = W_1 - \eta \cdot \frac{\partial L}{\partial W_1} \quad (16) \]

where the weight \( W_2 \) is updated iteratively by Eq. (17).

\[ W_2' = W_2 - \eta \cdot \frac{\partial L}{\partial W_2} \quad (17) \]

The algorithm runs \( \text{IteraMax} \) iterations to obtain the optimal classification model. In addition, the other purpose of the method is to select the influential features to further explained and analyzed the model.

Based on the obtained weights \( W_1 \) connecting the input layer and the hidden layer, the feature selection process can be implemented. Their absolute values are used as the
ranking basis to get the feature ordering for the financial distress, which can be called predictive power weight $W_p$. For a variable $x_i$, its predictive power weight $W_{pi}$ can be computed by Eq. (18).

$$W_{pi} = \sum_{j=1}^{q} W_{1ij}$$

(18)

For all variables, a sorted list can be obtained, it is represented by Eq. (19).

$$L = \text{Rank}\{W_{p1}, W_{p2}, W_{p3}, \ldots, W_{pn}\}$$

(19)

In which list, the top $N_f$ features can be selected and analyzed to explain the model. The entire processes are described in Algorithm 1.

### Algorithm 1 Sparse Neural Networks with $L_{1/2}$ norm for FDP

Require: Training Set $\{X, Y\}$, Learning Rate $\eta$, Regularization Coefficient $\lambda$, Iteration $\text{IteraMax}$;  
Ensure: $W_1, W_2$:

1. Randomly initialize all connection weights and bias in the network within the range of $(0, 1)$;
2. for $i < \text{IteraMax}$ do
3. Calculating the output via forward propagation and Eq. (15);
4. Calculating the loss value by using Eq. (13);
5. Calculating the gradients of $W_1$ and $W_2$ by Eq. (14) and Eq. (15);
6. Updating $W_1$ and $W_2$ by Eq. (16) and Eq. (17);
7. end for
8. Calculating $W_p$ by using $W_1$, and sorting variables by Eq. (19);
9. Selecting the top $N_f$ features;
10. return $W_1, W_2, N_f$ features

### 4 Experiments and results

In this section, the FDP data, evaluation standard, parameter analysis, model performance and its interpretability will be introduced.

#### 4.1 FDP data and evaluation standard

Previous research set many criteria to define and distinguish whether the company is or will be in financial distress, among which firm bankruptcy [2, 32, 52, 56], debt restructuring [4, 11, 27], and debt default [10, 28, 33] are most widely used. In China, the sample size of listed firms filing for bankruptcy is extremely small [8, 15], thus we use the other two criteria to define corporate financial distress: (1) whether the firm is experiencing a debt restructuring in a given year (debt restructuring), and (2) whether the firm has a debt default in a given year (debt default). In the China stock market, the listed companies always disclose their financial statements for the last fiscal year around April in a year [71]. To predict the financial distress for a company, we used predictor data obtained 2 and 3 years before the companies met the financial distress criteria. Detailed definitions of financial distress variables are shown in Table 2.

| Definitions | Variables | Definitions |
|-------------|-----------|-------------|
| Debt        | F2-ds     | Dummy variable indicating debt restructuring in next 2 years |
| Restructuring | F3-ds    | Dummy variable indicating debt restructuring in next 3 years |
| Debt        | F2-df     | Dummy variable indicating debt default in next 2 years |
| Default     | F3-df     | Dummy variable indicating debt default in next 3 years |

1 We referred about 30 papers published in A+ journals (e.g., Journal of Accounting Research, The Accounting Review, Journal of Financial Economics etc.) and 13 papers published in Chinese top journals (e.g., Economic Research Journal, Accounting Research, Nankai Business Review etc.)
liquidity, profitability, capacity of corporate governance and strategy. Detailed descriptions of predictor variables are shown in Table 3.

We collect data of financial distress variables and predictor variables from two commonly used database: China Stock Market and Accounting Research Database (CSMAR) and Chinese Research Data Services Platform (CNRDS). The sample period is from 2007 to 2019 and there are 10,731 company-year observations. The financial distress variables whose missing values take more than 10% of the total company-year observations are excluded. After that, there are 163 financial variables measuring capital structure, cash management, development capability, liquidity, profitability, shareholder benefit, size, turnover, variability and 36 non-financial variables measuring governance structure, information disclosure, investor protection and strategy. Table 4 is the sample distribution of financial distress firms (FSMs) by year. We can see that the absolute number of FSMs is increasing over time with a transitory decline in year 2010-2011. However, the proportion of FSMs is decreasing by year before 2015.

Figure 3 describes the distribution of predictor variables. We can see that the number of financial predictor variables is much larger than those non-financial ones, manifesting those previous researchers have focused on financial variables to predict financial distress. Among those prediction variables, capital structure, profitability and variability are the most commonly used financial variables, while governance structure is the most frequently used non-financial variables.

In the experiments, the samples are consisted of 10,533 firm-year observations to predict financial distress in the next 2 years and 9189 firm-year observations to predict financial distress in the next 3 years. For each one, 80%
of them is randomly selected as the training set, and the remaining 20% as the test set.

In the experiment, the performance of the models is measured in terms of the accuracy and precision, which are calculated by the Eqs. (20) and (21) [18].

\[ \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \]  
\[ \text{Precision} = \frac{TP}{TP + FP} \]  

where the variables, e.g. TP, are listed in the Table 5.

4.2 Parameter analysis

In order to capture the optimal performance of the model, the parameters of the FDP-SNN are analyzed and shown in Fig. 4. For the regularization coefficient \( \lambda \), a small or large value both are not conducive to the improvement of model performance, and their accuracy is only about 50%. When \( \lambda = 0.0001 \), the accuracy reaches the highest value 86.48%. With regard to hidden nodes, there are some upward and downward trends in a certain accuracy in a small range.

### Table 4: Distribution of financial distress variables by year

| Year | Debt restructuring | Debt default |
|------|--------------------|--------------|
|      | Non-FSMs | FSMs | Proportion of FSMs | Non-FSMs | FSMs | Proportion of FSMs |
| 2007 | 369 | 82 | 0.18 | 347 | 104 | 0.23 | 451 |
| 2008 | 429 | 87 | 0.17 | 402 | 114 | 0.22 | 516 |
| 2009 | 426 | 81 | 0.16 | 402 | 105 | 0.21 | 507 |
| 2010 | 407 | 66 | 0.14 | 385 | 88 | 0.19 | 473 |
| 2011 | 523 | 68 | 0.12 | 492 | 99 | 0.17 | 591 |
| 2012 | 629 | 75 | 0.11 | 598 | 106 | 0.15 | 704 |
| 2013 | 762 | 97 | 0.11 | 737 | 122 | 0.14 | 859 |
| 2014 | 805 | 104 | 0.11 | 778 | 131 | 0.14 | 909 |
| 2015 | 844 | 109 | 0.11 | 827 | 126 | 0.13 | 953 |
| 2016 | 917 | 129 | 0.12 | 905 | 141 | 0.13 | 1046 |
| 2017 | 992 | 148 | 0.13 | 975 | 165 | 0.14 | 1140 |
| 2018 | 1072 | 166 | 0.13 | 1030 | 208 | 0.17 | 1238 |
| 2019 | 1178 | 166 | 0.12 | 1128 | 216 | 0.16 | 1344 |
| Total | 9353 | 1378 | 0.13 | 9006 | 1725 | 0.16 | 10,731 |

### Table 5: The definitions of variables in the Eqs. 20 and 21

| Variables       | Definitions                                                                 |
|-----------------|-----------------------------------------------------------------------------|
| TP (true positive) | An instance is positive class and is also judged to be a positive class    |
| FN (false negative) | An instance is originally positive class while is judged to be false class |
| FP (false positive) | An instance is originally a false class while is judged to be positive one |
| TN (true negative) | An instance is a false class and is also determined to be a false class   |
Relatively speaking, more neural nodes can improve the accuracy of the model. Finally, 45 hidden nodes are adopted and its accuracy is 86.48%.

Besides, the iteration is also analyzed. With the increase of the iteration, the training accuracy has been rising, while the test accuracy has decreased sharply when iteration is 50,000. This is because the model with large iteration will be overfit, and further led to a significant reduction of the test accuracy. For the learning rate, despite large fluctuations, its overall trend is also rising first and then dropping. When the learning is set to 0.4, the FDP-SNN has the highest accuracy. Based on the above parameter analysis, the parameter settings of all experiments are determined, which have displayed in Table 6.

| Instances | Hidden node | $\lambda$ | $\eta$ | Iterations |
|-----------|-------------|-----------|--------|------------|
| F2-ds     | 45          | 0.0001    | 0.4    | 40000      |
| F3-ds     | 40          | 0.0001    | 0.4    | 20000      |
| F2-df     | 35          | 0.0001    | 0.4    | 30000      |
| F3-df     | 55          | 0.0001    | 0.4    | 40000      |

4.3 Verification on sparsity

In this paper, $L_{1/2}$ regularization is adopted to sparse the weight in neural networks to select the effective features and make it correct decisions, improving performance. To verify its effectiveness, a verification experiment is set up. It compares the accuracy on two predictors (neural networks) that with or without the sparse regularization.

As shown in Fig. 5, the predictor that with sparse regularization accurately takes advantage. Despite different predicted targets, its all accuracy is higher than those without one. Particularly, the maximum promoting value
reaches 6.53% on the F3-ds. These phenomena prove that selecting more effective feature using sparse regularization is indeed helpful to improve the ability of model recognition.

### 4.4 Results and analysis

Table 7 compares test accuracy of the proposed sparse neural networks and other intelligent methods on the four indexes. It can be observed that sparse neural networks are superior to other methods in accuracy. Regardless of the four indexes, the effect of naive bayes is the worst because it assumes that attributes are independent of each other. However, both the number and the correlation of attributes in this work are large, which makes classification effect poor. Contrast to other methods, the neural networks successfully improve the classification effect, and their accuracy on the F2-ds has up to 82.20%. Importantly, the values have been improved further when introducing the sparse regularization, where it has up to 87.30% on the F3-ds, which has proved that selecting valid features is helpful for classification.

Additionally, to verify performance improvement of the proposed model, it is also compared with models proposed by predominant researches, which employ classic statistical methods. The Z-Score proposed by Altman [2] is the first study that uses MDA to predict corporate financial distress. O-Score proposed by Ohlson [56] is the first work that adopted Logit to predict the financial distress of listed companies. Campbell et al. [10] proposed a simplified financial distress prediction model, which combined the traditional financial variables with non-financial variables (the stock market variables), which is also supposed to be more predictive than Z-Score and O-Score. Their comparisons are detailed in Table 8.

As shown in Tables 9 and 10, the proposed model outperforms the benchmark models on all of the performance

### Table 7 The comparison of the test accuracy on different methods

| Methods                | F2-ds | F2-df | F3-ds | F3-df |
|------------------------|-------|-------|-------|-------|
| Naive bayes            | 73.25 | 71.84 | 75.44 | 68.53 |
| K-Nearest neighbor     | 77.92 | 73.70 | 78.23 | 73.41 |
| Support vector machine | 78.60 | 73.20 | 78.60 | 73.71 |
| Decision tree          | 78.63 | 74.20 | 79.13 | 74.20 |
| Decision stump         | 79.43 | 76.86 | 79.68 | 75.93 |
| Neural networks        | 82.20 | 79.92 | 81.83 | 81.15 |
| Random forest          | 84.86 | 81.40 | 84.92 | 81.19 |
| Sparse neural networks | **85.23** | **82.20** | **87.30** | **83.30** |

The significance of bold values represents the best result

### Table 8 Comparisons of our model with benchmark models

| Model   | Method   | References |
|---------|----------|------------|
| Our     | FDP-SNN  | 199 variables (163 financial variables and 36 non-financial variables) |
| Altman  | MDA      | 5 financial variables |
| Ohlson  | Logit    | 9 financial variables |
| Campbell| Logit    | 10 financial variables and 5 non-financial variables |

### Table 9 The performance of different models predicting financial distress in the next 2 years

| Methods | F2-ds  |        |        |
|---------|--------|--------|--------|
|         | Accuracy | Precision | AUC    |
| Altman  | 0.797  | 0.207   | 0.639  |
| Ohlson  | 0.802  | 0.227   | 0.653  |
| Campbell| 0.797  | 0.213   | 0.618  |
| Our     | **0.852** | **0.439** | **0.690** |

The significance of bold values represents the best result

### Table 10 The performance of different models predicting financial distress in the next 3 years

| Methods | F3-ds  |        |        |
|---------|--------|--------|--------|
|         | Accuracy | Precision | AUC    |
| Altman  | 0.804  | 0.261   | 0.639  |
| Ohlson  | 0.803  | 0.259   | 0.654  |
| Campbell| 0.819  | 0.273   | 0.640  |
| Our     | **0.873** | **0.450** | **0.700** |

The significance of bold values represents the best result
measures. Among them, the precision in our model is almost twice that of the benchmark model, attributing the success of the improvement to the effectiveness of the variables and the features selection in the sparse neural networks. On the one hand, our model allows the input of more variables. However, due to the limitation of the number of input variables, traditional models only use financial variables, ignoring non-financial variables with more predictive power, so the performance is weaker. On the other hand, sparse neural networks can focus on more effective features, making the model correct decisions, to improve performance. Experimental results proved that the sparse neural networks for feature selection and prediction is effective.

4.5 Interpretability of FDP-SNN

Based on the obtained FDP-SNN, each predictor variable can be given a predictive power weight, which is able to explain the model to some extent. As shown in Table 11, non-financial variables have greater predictive power than those financial variables. Besides, this difference is more significant when our model is predicting financial distress in the next 3 years rather than next 2 years, which indicates that non-financial variables become more important with the forecast period grows longer. The average weights of strategy predictors are instable among the groups of financial distress. They have greater prediction weights for F2-ds.

### Table 11 Average predictive power weights of all groups of financial distress variables

| Category       | Types               | Average Weights | F2-ds | F3-ds | F2-df | F3-df |
|----------------|---------------------|-----------------|-------|-------|-------|-------|
| Financial      | Capital structure   | 2.310           | 2.802 | 1.912 | 3.249 |       |
|                | Cash management     | 2.661           | 3.121 | 1.938 | 3.911 |       |
|                | Development capability | 0.967     | 1.891 | 0.495 | 1.218 |       |
|                | Liquidity           | 3.387           | 3.919 | 2.595 | 3.606 |       |
|                | Profitability       | 1.841           | 2.240 | 2.504 | 2.270 |       |
|                | Shareholder benefit | 1.661           | 3.001 | 1.592 | 2.871 |       |
|                | Size                | 1.060           | 1.369 | 1.113 | 2.129 |       |
|                | Turnover            | 1.611           | 2.749 | 1.136 | 2.479 |       |
|                | Variability         | 2.765           | 0.869 | 0.875 | 1.532 |       |
| Non-financial  | Governance structure| 2.826           | 5.487 | 2.123 | 7.205 |       |
|                | Information disclosure | 1.426     | 7.864 | 2.827 | 3.573 |       |
|                | Investor protection | 17.418          | 20.304| 15.037| 14.716|       |
|                | Strategy            | 20.408          | 2.735 | 2.005 | 35.882|       |
| Subtotal       |                     | 4.786           | 6.965 | 3.310 | 8.819 |       |
| Total          |                     | 2.646           | 3.244 | 1.963 | 3.734 |       |

### Table 12 The top 10 features with the largest weight from models predicting debt restructuring

| Feature          | Types            | Financial predictor | Weighs |
|------------------|------------------|---------------------|--------|
| F2-ds            | Develop          | Investor protection | No     | 47.397 |
|                  | HPAINV           | Strategy            | No     | 35.725 |
|                  | TBI              | Variability         | Yes    | 34.548 |
|                  | ARTA             | Liquidity           | Yes    | 11.860 |
|                  | QATA             | Liquidity           | Yes    | 9.605  |
|                  | SD-FANW          | Variability         | Yes    | 8.978  |
|                  | APA              | Capital structure   | Yes    | 8.430  |
|                  | MPNMP            | Profitability       | Yes    | 8.274  |
|                  | OCNF             | Cash management     | Yes    | 7.872  |
|                  | STA              | Turnover            | Yes    | 7.402  |
| F3-ds            | SOE              | Governance structure| No     | 52.579 |
|                  | Develop          | Investor protection | No     | 50.813 |
|                  | Over-predict     | Information disclosure | No  | 33.576 |
|                  | EGR              | Development capability | Yes | 15.199 |
|                  | QATA             | Liquidity           | Yes    | 13.720 |
|                  | CTA              | Liquidity           | Yes    | 11.101 |
|                  | IRRI             | Profitability       | Yes    | 10.232 |
|                  | ARTA             | Liquidity           | Yes    | 9.429  |
|                  | APA              | Capital structure   | Yes    | 9.188  |
|                  | STA              | Turnover            | Yes    | 8.548  |

### Table 13 The top 10 features with the largest weight from models predicting debt default

| Feature          | Types            | Financial predictor | Weighs |
|------------------|------------------|---------------------|--------|
| F2-df            | Develop          | Investor protection | No     | 38.899 |
|                  | TBQAI            | Profitability       | Yes    | 35.030 |
|                  | APA              | Capital structure   | Yes    | 7.509  |
|                  | QATA             | Liquidity           | Yes    | 6.388  |
|                  | ARTA             | Liquidity           | Yes    | 6.359  |
|                  | BDTL             | Capital structure   | Yes    | 6.356  |
|                  | Fraud            | Investor protection | No     | 5.592  |
|                  | NS-opinions      | Information disclosure | No  | 5.481  |
|                  | INNWC            | Liquidity           | Yes    | 5.463  |
|                  | OCFA             | Cash management     | Yes    | 5.265  |
|                  | HPAINV           | Strategy            | No     | 62.902 |
|                  | Exec-turn        | Governance structure | No   | 37.333 |
|                  | Develop          | Investor protection | No     | 30.642 |
|                  | Man-hold         | Governance structure | No   | 12.823 |
|                  | CM-hold          | Governance structure | No   | 12.795 |
|                  | SOE              | Governance structure | No   | 12.560 |
|                  | MPNMP            | Profitability       | Yes    | 11.592 |
|                  | QATA             | Liquidity           | Yes    | 11.057 |
|                  | ARTA             | Liquidity           | Yes    | 10.231 |
|                  | OCFA             | Cash management     | Yes    | 9.878  |
and F3-df, but are relatively lower for F3-ds and F2-df. One possible reason is that there are too few variables in the strategy group. Investor protection group, by contrast, is consistently and highly predictive among all financial distress groups. This result consists with the theory of law and finance, which suggests that investor protection is a key factor affecting corporate finance [58, 60].

Tables 12 and 13 demonstrate the top 10 features with the largest weights extracting from models predicting debt restructuring and debt default. The definitions of these features are summarized in Table 15. Consist with our findings that non-financial variables have greater predictive power, all the features with the largest weight for F2-ds, F3-ds, F2-df and F3-df are non-financial variables. Specifically, the feature Develop a dummy variable indicating whether the company registered in the developed provinces\(^2\) has the greatest predictive power for F2-ds and F2-df. This feature was proposed by Hu and Jin [36], who believed that the theory of political tournaments implied local governments had a strong incentive to internalize social burdens in listed companies in their jurisdictions. Thus, it is reasonable to participate that the level of development where a company is located would affect the firms financial positions. The feature SOE, a dummy variable indicating whether the firm is state-owned enterprise or not, has the greatest power to predict F3-ds. Consist with Wu and Wu [67], we believed that state-owned enterprises had strong support from the government and were less likely to run into financial distress. The feature HPAINV, a dummy variable indicating whether the company is investment-oriented or not, has the greatest power to predict F3-df. This feature was put forward by Wang et al. [64], who suggested that compared with operation-oriented, investment-oriented company suffer less financial risk.

Moreover, non-financial predictor variables are more important with the forecast period grows longer. As shown in Table 12, non-financial variables occupy the top 2/3 features when predicting F2-ds/F3-ds. Most importantly, when we use debt default to proxy financial distress (results are shown in Table 13), there are only 3 non-financial variables in top 10 when predicting F2-df. However, when predicting F3-df, the number of non-financial variables raise up to 6 with top positions.

Based on the comparison between the weights of financial and non-financial predictor variables, we may conclude that non-financial predictor variables are more important in predicting financial distress. However, the number of financial predictor variables is 163, while the number of non-financial predictor variables is 36, which means that the difference of weight may be driven by the difference of variables number.

Thus, we carry out a robust test. Specifically, 36 financial indicators with the highest weight are selected and put into the model together with 36 non-financial indicators to re-compare the weights.

As shown in Table 14, except for predicting F2-ds, non-financial indicators have strong predictive ability in predicting other financial distress variables, including higher average weight and higher proportion in the ten variables with the highest weight. This suggests that financial variables are only better at predicting short-term debt restructurings. As the scope of financial distress expands and the predicting period becomes longer, the predictive power of non-financial variables increases.

Those results are consistent with the new finding that non-financial variables are more powerful than financial ones in explaining and predicting corporate financial conditions. There are two main reasons. Firstly, non-financial variables are usually the determinants of the corporate financial situation, while financial variables are its reflection. The current financial situation of a company is the result of its past operation and governance, and the future financial situation depends on its current management model ([44]). In summary, financial variables measure what a company “has done,” while non-financial variables measure what a company “is doing.” Thus, non-financial variables, measuring a company’s operation and governance, are more future-orientated.

Secondly, financial variables are generated by the financial accounting information disclosed by the company, which is easily manipulated by the management [38]. It has the following three attributes: (1) the production process is complicated, (2) can be affected by management accounting policies, and (3) often used to evaluate management performance, and thus management has the motivation and ability to manipulate financial accounting information. In contrast, non-financial accounting information does not possess these attributes. To a certain extent, non-financial variables are also more reliable than financial variables.

2 Developed provinces in China include Jiangsu, Zhejiang, Shanghai, Guangdong and Beijing.
5 Conclusion

In this study, a novel prediction method for financial distress is proposed, which is based on sparse neural networks and whose hidden layer with $L_{1/2}$ regularization can select the efficient feature, so as to improve the performance on prediction. Based on the existing accounting and finance theory, we identify 163 financial variables and 36 non-financial variables that might affect financial distress and then select the top 10 predictors with the largest weights. The empirical results show that non-financial predictor variables are more important in financial distress prediction especially when the forecast period grows longer. Besides, the performance of FDP-SNN is assessed by comparing it with three benchmark models and find that FDP-SNN outperforms these benchmark models in accuracy, precision, and AUC performance by a large margin.

From this study, we can get the following inspirations: First, sparse neural networks with $L_{1/2}$ regularization can be used to select features and build a better model for predicting financial distress. Second, the neural networks model enables us to consider the financial distress predictors from multiple aspects comprehensively without considering the limitation of input variables. Finally, some future-oriented non-financial variables play a key role in the prediction of...
financial distress, which is consistent with accounting and finance theory.

Future research might also include other variables which have not been mentioned by these existing papers but are important in finance and accounting, such as managers’ personalities, corporate cultures, and organizational identification, etc. Besides, whether the correlations of these predictors are determinant factors in the financial distress prediction will also be our future research.

Appendix A Supplementary variables

This appendix table provides the supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research. All Variables are from the related references.3

Acknowledgements This work was funded in part by The National Natural Science Foundation of China (No. 71802199), National Social Science Foundation of China (NO. 18VSJ082) and Philosophy and Social Science Foundation of Guangdong Province, China (No. GD21TW07-01).

Data availability The datasets generated and analysed during this study are available in the CSMAR repository: https://www.gtarsc.com/.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Alfaro-Ponce M, Argüelles A, Chairez I, Pérez A (2019) Automatic electroencephalographic information classifier based on recurrent neural networks. Int J Mach Learn Cybern 10(9):2283–2295
2. Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Financ 23(4):589–609
3. Andrade G, Kaplan SN (1998) How costly is financial (not economic) distress? Evidence from highly leveraged transactions that became distressed. J Financ 53(5):1443–1493
4. BKD, HSW (1991) Evaluating financial distress resolution using prior audit opinions. Contemp Account Res 8(1):97–114
5. Babina T (2020) Destructive creation at work: how financial distress spurs entrepreneurship. Rev Financ Stud 33(9):4061–4101
6. Balcaen S, Ooghe H (2006) 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. Br Account Rev 38(1):63–93
7. Beaver WH, McNichols MF, Rhie JW (2005) Have financial statements become less informative? evidence from the ability of financial ratios to predict bankruptcy. Rev Acc Stud 10(1):93–122
8. Bhattacharjee A, Han J (2014) Financial distress of Chinese firms: microeconomic, macroeconomic and institutional influences. China Econ Rev 30:244–262
9. Blum M (1974) Failing company discriminant analysis. J Account Res 1–25
10. Campbell JY, Hilscher J, Szilagyi J (2008) In search of distress risk. J Financ 63(6):2899–2939
11. Casey CJ, McGee VE, Stickney CP (1986) Discriminating between reorganized and liquidated firms in bankruptcy. Account Rev 249–262
12. Chen WS, Du YK (2009) Using neural networks and data mining techniques for the financial distress prediction model. Expert Syst Appl 36(2):4075–4086
13. Chen H, Yao M, Gu Q (2020) Pothole detection using location-aware convolutional neural networks. Int J Mach Learn Cybern 11(4):899–911
14. Chen M, Mi D, He P, Deng L, Wei B (2014) A CT reconstruction algorithm based on L1/2 regularization. Comput Math Methods Med 2014
15. Chen X, CZ (2000) Predicting financial distress in Chinese listed firms. China Account Financ Rev 2(3):55–92 (in Chinese)
16. Cui X-G, Wang L-Y (2007) Over-speed growth, financial crisis and risk forecasting. Account Res 12:55–62
17. Damolina IG, Khoury SJ (1980) Ratio stability and corporate failure. J Financ 35(4):1017–1026
18. Davis J, Goadrich M (2006) The relationship between precision-recall and roc curves. In: Proceedings of the 23rd international conference on machine learning, pp 233–240
19. Deakin EB (1972) A discriminant analysis of predictors of business failure. J Account Res 167–179
20. Fanning KM, Cogger KO (1994) A comparative analysis of artificial neural networks using financial distress prediction. Intell Syst Account Financ Manag 3(4):241–252
21. Faris H, Mirjalili S, Aljarah I (2019) Automatic selection of hidden neurons and weights in neural networks using grey wolf optimizer based on a hybrid encoding scheme. Int J Mach Learn Cybern 10(10):2901–2920
22. Fawcett T (2004) Roc graphs: notes and practical considerations for researchers. Mach Learn 31(1):1–38
23. Franzen LA, Rodgers KJ, Simin TT (2007) Measuring distress risk: the effect of r&d intensity. J Financ 62(6):2931–2967
24. Friedman FH et al (1985) Introducing recursive partitioning for financial classification: the case of financial distress. J Financ 40(6):269–291
25. Garcia-Appendini E (2018) Financial distress and competitors’ investment. J Corp Financ 51:182–209
26. Geng R, Bose I, Chen X (2015) Prediction of financial distress: an empirical study of listed Chinese companies using data mining. Eur J Oper Res 241(1):236–247
27. Gentry JA, Newbold P, Whitford DT (1985) Classifying bankrupt firms with funds flow components. J Account Res 146–160
28. Gilson SC, John K, Lang LH (1990) Troubled debt restructurings: an empirical study of private reorganization of firms in default. J Financ Econ 27(2):315–353
29. Goel T, Murugan R, Mirjalili S, Chakrabortty DK (2021) Optimization of an optimized convolutional neural network for an automatic diagnosis of COVID-19. Appl Intell 51(3):1351–1366
30. Guo J, Liu Z, Chen C, Zhang T, Wang L, Fan K (2022) An efficient inspection system based on broad learning: nondestructively estimating cement compressive strength with internal factors. IEEE Trans Ind Inform 18(6):3787–3798
31. Guo J, Wang L, Fan K, Yang B (2020) An efficient model for predicting setting time of cement based on broad learning system. Appl Soft Comput 96:106698
32. Hillegeist SA, Keating EK, Cram DP, Lundstedt KG (2004) Assessing the probability of bankruptcy. Rev Acc Stud 9(1):5–34

3 Some variables have multiple references, but only the earliest ones are retained here.
