The Role of Feature Importance in Predicting Corporate Financial Distress in Pre and Post COVID Periods: Evidence from China

Abstract

The prediction of firm financial distress during the COVID-19 crisis episode attracted massive academic attention since economic uncertainty was exacerbated. In this paper, we propose a firm financial distress prediction model based on the Extreme Gradient Boosting-Genetic Programming (XGB-GP) framework by investigating subsamples of pre-COVID and post-COVID periods. The key contribution of our paper is that we explore time-varying prediction features for pre-COVID and post-COVID periods. We illuminate that the earning financial indicator is the dominant feature for financial distress prediction during the pre-COVID period, whereas total financial leverage is the most important factor during the post-COVID period. On this basis, our XGB-GP financial distress prediction model exhibits higher prediction accuracy than the traditional models. As a result, managers can modify the financial leverage level to improve the financial situation of the firm by reducing the debt burden and increasing profitability during the post-COVID period.

Keywords: Financial distress prediction, Time-varying feature selection, Extreme Gradient Boosting, Genetic Programming, COVID-19 crisis

JEL classification: G33

1. Introduction

The outbreak of COVID-19 has created adverse effects on economic situations worldwide, such as supply chain disruptions (Hosseini and Ivanov, 2021; Orlando et al., 2022) and economic crises (Vo et al., 2021; Brennan et al., 2022; Khan, 2022), resulting in an increase in the probability of financial distress and bankruptcy risk (Crespi-Cladera et al., 2021; Bozkurt and Kaya, 2022; Dorr et al., 2022; Zhao et al., 2023).

Therefore, the prediction of financial distress with the COVID-19 effect becomes pivotal under such an economic crisis circumstance. More importantly, factors that influence the probability of financial distress would be extremely helpful for the management board to raise the survival rate of the company. Accordingly, the time-varying circumstance motivates us to determine the superiority of the factors that can affect the probability of financial distress and then to formulate a prediction model for firm financial distress based on highly influential factors.

Our research can be strongly underpinned by the pecking order theory. The pecking order theory states that firms have a preferring order for financing sources from their internal financing resources to external financing resources. For external financing, firms usually prefer debt financing than equity financing (Chen et al., 2013). Firms that have high leverage ratio with debt financing are prone to enter the financial distress condition. Consequently, our XGB-GP model with feature selection can help firms to identify the most relevant debt-related factors to predict financial distress and thus enable managers to better manage the capital structure, which can exempt firms from financial distress. Both practical and theoretical sides provide us strong motivation to undertake this study to formulate the financial distress prediction model with feature selection.

Our study attempts to select features by using the Extreme Gradient Boosting-Genetic Programming (XGB-GP) method to construct financial distress prediction model. The preciousness of our contribution is that we explore time-varying prediction features for pre-COVID and post-COVID periods. By providing the comparative empirical evidence, we elucidate that the earning financial indicator is the dominating feature for financial distress prediction during the pre-COVID period. We further illuminate that total financial leverage is the most important factor in predicting firm financial distress during the post-COVID period. It is arguable that when the economy is in an expanding episode, firms are expected to expand earnings to cover the debt payment (White, 1983), and companies with substantial earning reduction could be likely to be financially distressed before the COVID period (Wu et al., 2010).

On the other hand, firms cut their debt payments to survive during economic downturn periods since the influence of financial leverage on the firm financial situation is threefold. First, financial leverage can increase the debt burden of a company with high interest payments and thus increase bankruptcy risk (Modigliani and Miller, 1963; Myers, 2001). Secondly, the increase in financial leverage can generate high cash flow volatility, which can create disruptions in the capital chain, especially during crisis periods (Mackay and Phillips, 2005; Elkhal, 2019). Finally, financial leverage creates a negative effect on firm profitability since financial leverage is positively associated with cost fixity and agency cost (Baker, 1973; Dalci, 2018). Thus, managers can modify the financial leverage level to improve the financial situation of the firm by reducing the debt burden and increasing profitability.

Besides, corporate financial distress is a highly concerning issue because it can be detrimental to various stakeholders of the company. The experience of financial distress can render a company unable to meet its financial obligations, resulting in a plummeting stock price. This can lead to a downward spiral, where the company's financial situation continues to deteriorate, making it increasingly difficult to recover. Given these potential negative impacts, understanding the crucial factors that can influence corporate financial distress is a pivotal aspect for the fundamental understanding of finance. Therefore, our study identifies the key features affecting corporate financial distress to strengthen the fundamental understanding of finance.

Therefore, the first contribution of our paper is to determine the superiority of the factors that can affect the probability of financial distress. This contribution would remarkably distinguish our study from extant research since existing studies mainly focus on the prediction accuracy of the model (Geng et al., 2015; Huang and Yen, 2019; Tsai et al., 2021; Jiang et al., 2022; Yang et al., 2023). Those studies have paid little attention to the factors behind the prediction accuracy. This would also shed new insights on corporate governance, as managers can revamp key financial indicators suggested by our model to alleviate the financial distress situation. Dirman (2020) reveals that different financial indicators respond to the financial crisis differently. Therefore, it is sensible to use different features for predicting financial distress conditions in different economic development episodes, especially before and after the COVID crisis. However, Qian et al. (2022) unveil that the consideration of feature selection importance is scarce in current financial distress prediction methods and thus the truly important features may not be assigned sufficient attention for different economic development phases.

Second, we use the listed companies sample from China, which is an emerging market. During the COVID-19 crisis, country characteristics play more crucial roles in predicting financial distress and bankruptcy risk than individual firm characteristics (Huang and Yen, 2019). Nevertheless, forecasting firm financial distress in developed countries such as the US and European countries is the focal point of existing studies (Chiaramonte and Casu, 2017; Mselmi et al., 2017; Jabeur and Fahmi, 2018; Charalambakis and Garrett, 2019; Gandhi et al., 2019). These results may not be as helpful in emerging markets such as China. Therefore, we complement the literature by using the emerging market sample with the COVID-19 crisis. Our study has unveiled that the financial leverage factor has the largest predicting power, which is quite different from the existing models since we have included the sample data of post-COVID-19 crisis periods. We unfold the fact that during crisis periods, high financial leverage is more likely to turn companies into financial distress.

Finally, we propose a novel hybridized framework that combines a supervised machine learning method (XGBoost) (Chen and Guestrin, 2016) and a metaheuristic algorithm Genetic Programming (GP) (Poli et al., 2008) to predict financial distress. Based on the tree-structured hierarchy, XGBoost is superior in mining the pattern in the data and is able to exploit a broad range of features that are relevant to the target variable. With the help of ensembles of decision trees, the feature importance can be estimated from the trained predictive model. GP has been very successful at evolving novel and unexpected ways of solving problems. One elegant characteristic of GP is that it can provide not only the predicted result but also the prediction method (symbolic expressions). These expressions can take the form of mathematical equations or logical rules, making it easier for humans to interpret and understand the relationships between input variables. This is highly desirable in financial institutions since it can provide a reliable human interpretable reference and valuable insights into the underlying mechanisms for managers compared with traditional machine learning methods, which are typically considered “black-box" methods. Since GP models often have the flexibility to generate complex solutions, post-processing techniques can be applied to simplify the models and improve interpretability. In this work, we use the top features selected by XGBoost to generate the non-linear financial distress prediction function, which can be easier to interpret and explain to stakeholders.

The remainder of the paper is structured as follows. Section 2 provides a relevant literature review, and section 3 introduces the data and methodology used in this paper. This section also presents three traditional financial distress prediction models with Extreme Gradient Boosting (XGBoost) and Genetic Programming (GP) techniques. Section 4 presents the empirical results for the model performance evaluation, and section 5 concludes the paper.

2. Literature review

2.1. Traditional financial distress prediction models

The prediction of financial distress is one of the corporate finance research topics that attracts a large number of scholars (Lian, 2017; Andrikopoulos and Khorasgani, 2018; Gao et al., 2018; Andreou et al., 2021; Aslan, 2022). Prior to the landmark study of Altman (1968), the early prediction models of financial distress were mainly the single factor prediction model, where multivariate and comprehensive index analyses were scarce. Altman (1968) introduces multiple discriminant analysis (MDA) into financial distress predictions to create the well-known Z-score model. He developed the first multivariate financial distress prediction model in the financial distress research area. Subsequently, the multivariable method was widely used in the fields of credit risk and financial distress predictions. Zmijewski (1984) conducts a follow-up study, taking 40 bankrupt enterprises and 800 non-bankrupt enterprises listed on the New York Stock Exchange as the firm sample and applying the unweighted probability and the weighted exogenous sample maximum likelihood test to develop the financial distress model.

Likewise, West (1985) creates compound variables through factor analysis to describe the financial and operational characteristics of banks and uses the data of the United States to illustrate the advantages of the combination of factor analysis and logit estimation in assessing the status of banks. Jones and Hensher (2004) proposed a mixed logit model for predicting the plight of enterprises and demonstrated examples that the mixed logit model obtained better prediction accuracy than the polynomial logit model. Altman et al. (2017) examine the performance of the Z-score Model of enterprises from 31 European countries and three non-European countries. They prove that the original Z-score model exhibits outstanding performance within the international context. The aforementioned models only focus on the prediction accuracy of financial distress, and a deep investigation of the underlying factors inside the models is scarce. Recently, scholars have paid attention to the effectiveness of tone features in predicting corporate financial distress.

2.2. Newly-developed financial distress prediction models

Unlike the traditional models, with the rapid development of big data and intelligent technology, advanced technology has become the main ingredient in terms of financial distress prediction (Kumar and Ravi, 2007). Cao (2012) proposed a combined classifier for financial crisis early warning based on the Choquet integral, which uses an adaptive fuzzy measure to identify the dynamic information in the results. It is found that the combined classifier based on the Choquet integral has higher average accuracy and stability than a single classifier. Huang and Yen (2019) apply four supervised machine learning models, including the traditional support vector machine (SVM), recently developed hybrid associative memory with translation (HACT), hybrid GA-fuzzy clustering and extreme gradient boosting (XGBoost), to predict financial distress. Through the sample analysis, it is concluded that XGBoost provides the most accurate financial bankruptcy prediction.

Similarly, Li et al. (2021) employ a word vector model and classifier based on deep learning; they constructed a framework based on deep learning to construct a domain emotion dictionary and predicted the financial distress of enterprises by generating the China Financial Domain Emotion Dictionary (CFDSL). The empirical results show that CFDSL can effectively capture information on financial distress. More recently, Zhu et al. (2022) also maintain that the application of artificial intelligence algorithms in monitoring enterprise financial risks would be favored in building a financial early warning model. In addition, Geng, Bose & Chen (2015) have adopted financial indicators, including net profit margin of total assets, return on total assets, earnings per share, and cash flow per share, have strong predictive power in forecasting company profitability deterioration. Mselmi, Lahiani & Hamza (2017) demonstrated that Support Vector Machine is verified to be the best classifier of financial distressed firms. Financial distressed firms are featured with high leverage, lower repayment capacity and low liquidity and profitability. Li et al. (2020) employed the survival model to reveal that corporate governance measure with financial ratios and macroeconomic factors can formulate high-accuracy financial distress prediction models. Qian et al. (2022) also used the gradient boosted decision tree with corrected feature selection measure, suggesting that the algorithm based on permutation importance has a higher accuracy in forecasting financial distress than traditional models.

2.3. Financial distress prediction with feature selection

The development of computer technology brings massive machine learning methods to the construction of financial early warning models, yet some of the research gap has not been addressed, especially for feature selection. The data sample constraint would be one of the most important factors that frustrates future research on the financial distress model (Huang and Yen, 2019). It is arguable that feature selection plays an important role in financial distress prediction, especially for modern AI analysis methods (Lin et al., 2014; Liang et al., 2015). Zhao et al. (2022) promote a novel model based on the CatBoost model that contains not only financial features but also sentiment tone features. They collect data on China's stock market from 2016 to 2020 as the research object and find that sentiment tone features can help them effectively identify enterprises in financial distress. Therefore, factors that influence firm financial distress the most would deliver helpful implications for managers to alleviate insolvency risk (Khoja et al., 2019) and, more importantly, to formulate a turnaround strategy that can aid firms' recovery (Sudarsanam and Lai, 2001). Nevertheless, as argued in Qian et al. (2022), the consideration of feature selection importance is scarce in current prediction financial distress methods, and thus, the truly important features may not be assigned sufficient attention in different economic development episodes. Our study attempts to fill this research gap by using the XGB-GP method.

3. Data and methodology

3.1. Data

The sample we obtain is from the WIND database. We have employed a sample of 1150 listed companies from the Chinese A share in the Shanghai Stock Exchange, as we remove all companies listed after 2016 since they have no financial data for 2016. The sample period is from 1 January 2016 to 31 December 2021, after the market crash in 2015. We use the criteria in Table 3 to classify the sample firms as stable firms and financially distressed firms. We presume that there are no reversion cases for firm financial status, which indicates that once the firm becomes financially distressed, it cannot return to a stable status during our sample period. It is arguable since the reversion of financial distress condition is difficult for firms in a short period, especially during the economic downturn period.

We have taken all financial ratio data from 2016 to 2021 as feature sample data, which constructs a firm-year panel sample. Within the sample, all industry sectors are included. The firms that have taken special consideration, such as ST and *ST, are all considered distressed firms. The criteria in Table 3 are commonly used in the literature (see Gilson, 1989; Keasey and Watson, 1991; Ashraf et al., 2019; Qian et al., 2022). According to those criteria, we classify firms as stable firms and financially distressed firms.

Given the COVID crisis effect, we divide our sample into the pre-COVID period, which is 2018-2019, and the post-COVID period, which is 2020-2021. In order to enrich the dataset we can obtain, we not only include the financial data from annual reports of the listed companies but also the financial data from quarterly reports. Sample in each period is further split into 75% training data and 25% testing data respectively, where the training dataset is from quarterly reports and the testing dataset is from annual reports. In addition, we use three-year rolling financial data to predict firm financial distress conditions. In summary, there were 4142 observations in total sample, with 3735 stable observations and 407 distressed observations during the pre-COVID period. There were 3697 stable observations and 445 distressed observations during the post-COVID period.

3.2. Three traditional financial distress prediction models

We have used three main traditional financial distress prediction models including Altman (1968)'s Z-Score Model, the Probit model proposed by Zmijewski (1984) and the D-Score Model proposed by Blums (2003).

The Z-Score Model is defined in equation (1). The Probit Model is defined in equation (2). The D-Score Model is defined in equation (3).

$$Z = \alpha_1 WCTA + \alpha_2 RETA + \alpha_3 EBITTA + \alpha_4 MVBL + \alpha_5 STA, \quad (1)$$

where WCTA is Working capital/Total assets, RETA is Retained earnings/Total assets, EBITTA is Earnings before interest & taxes/Total assets, MVBL is Market value of equity/Book value of total liabilities and STA is Sales/Total assets. For the original Z-model, $\alpha_1 = 1.2$, $\alpha_2 = 1.4$, $\alpha_3 = 3.3$, $\alpha_4 = 0.6$, $\alpha_5 = 1.0$.

$$P = \Phi(\beta_0 + \beta_1 NITA + \beta_2 TLTA + \beta_3 CACL), \quad (2)$$

where NITA is the Net income/Total assets, TLTA is Total liabilities/Total assets, and CACL is Current assets/Current liabilities and Φ is the cumulative density function for a standard normal random variable. For the original Probit-model, $\beta_0 = -4.336$, $\beta_1 = -4.513$, $\beta_2 = 5.679$, $\beta_3 = 0.004$.

$$D = \gamma_0 + \gamma_1 NITA + \gamma_2 TDME + \gamma_3 META + \gamma_4 \Delta SP + \gamma_5 \Delta SG + \gamma_6 CLTA, \tag{3}$$

where NITA is the Net income/Total assets, TDME is Total Debt/Market equity, and META is Market Equity/Total assets, ΔSP is 6-month Stock Price change, ΔSG is the 3-year Sales Growth rate, and CLTA is Current liabilities/Total assets.

3.3. Model performance evaluation

To evaluate the model performance, we have two types of errors to consider, namely, Type I error and Type II error. Type I error indicates the model's misclassification by putting distressed companies into the stable group. On the other hand, Type II error suggests that the model's misclassification by putting stable companies into the distressed group (see Table 2).

To calculate both Type I error and Type II error for each model, we use five criteria to distinguish distressed firms from stable firms (see Table 3). If one firm satisfies any of the five criteria, it is categorized as a financially distressed firm. Otherwise, the firm would be categorized into the stable group. Then, we compare the model prediction result with the actual firm financial situation according to the categorization results. Model prediction accuracy with Type I error and Type II error can be obtained correspondingly.

Based on the data from listed Chinese companies, we undertake the model performance according to Type I and Type II error evaluation, where

Table 1: Descriptive statistics of sampled firms.

Notes: CACL is Current assets/Current liabilities, CLTA is Current liabilities/Total assets, EBITTA is Earnings before interest & taxes/Total assets, META is Market Equity/Total assets, MVBL is Market value of equity/Book value of total liabilities, NITA is the Net income/Total assets, RETA is Retained earnings/Total assets, STA is Sales/Total assets, TDME is Total Debt/Market equity, TLTA is Total liabilities/Total assets, WCTA is Working capital/Total assets, ΔSG is the 3-year Sales Growth rate, ΔSP is 6-month Stock price change.

	Mean	Std Dev	Max	Min
CACL	2.0579	3.31	11.5266	0
CLTA	0.2492	0.2042	2.787	0
EBITTA	0.0329	0.1502	28.2866	-4.6021
META	4.8469	69.0132	40.8544	0
MVBL	1.4572	97.5339	66.1398	0
NITA	0.0279	0.1291	0.716	-18.5889
RETA	0.1631	0.1902	0.8591	-10.0471
STA	0.3174	0.3921	9.7214	-0.0634
TDME	3.9969	40.1153	25.4896	0
TLTA	4.8469	69.0132	40.8544	0
WCTA	0.1953	0.2467	0.9807	-1.279
ΔSG	10.6502	64.8744	90.3791	-17.6885
ΔSP	0.9084	15.8498	70.3990	-55.407

the descriptive statistics of the listed company sample are presented in Table 1.

Table 2: Model prediction accuracy regarding Type I and Type II errors.

Actual Position	Model Prediction	
	Distressed	Stable
Distressed	Correctly predicted	Type I Error
Stable	Type II Error	Correctly predicted

Table 3: Categorization criteria for distressed and stable firms.

Description	Status of Firm Financial Position
Stable Financial Status	Stable
(1) Company share price is less than 50% book value for consecutive 3 years	
(2) Fail to announce dividend/bonus declaration for continuous 5 years	
(3) Failed to conduct AGM for 3 consecutive years	Distressed
(4) Failed to pay the yearly listing fee for 2 years	
(5) Face Delisted/Suspended/Liquidation/Winding up/Bankruptcy	

3.4. Extreme Gradient Boosting (XGBoost)

The gradient boost decision tree algorithm (GBDT) has many successful applications in different finance areas, such as stock price prediction and loan application scoring. XGBoost is an efficient and scalable variant of the Gradient Boosting Machine (GBM) proposed by Chen and Guestrin (2016). Essentially, XGBoost is an ensemble of K Classification and Regression Trees (CART), where each tree is assigned a weighted score and each leaf of the tree represents a target outcome. The final predication of the CART is calculated by summing up the K functions representing k trees of the CART. Mathematically, the general model can be written as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(X_i), f_k \in \mathcal{F}, \tag{4}$$

where K is the number of trees and $\mathcal{F}$ is the set of all possible CARTs. f is a function in the functional space $\mathcal{F}$, and X_i is the vector of the features adopted in the model. The objective function to be optimized is given by:

$$\text{obj}(\theta) = \sum_{i}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{5}$$

where L obtains the training loss between the predicted value $\hat{y}_i$ and the actual observation $\hat{y}_i$, and Ω is the regularization term.

XGBoost is widely used by data scientists to achieve state-of-the-art results in many machine learning competitions due to its advantages, such as ease of use, use of regularization to prevent overfitting, support of self-defined loss functions, support of sparse representations, and so on (Volkovs et al., 2017). Furthermore, since XGBoost has the capacity to perform parallel computation on a single machine and is over 10 times faster than the classical GBM, it possesses the intrinsic ability to handle a large amount of data and can be efficiently used for big data analytics. These advantages fit our requirement for our empirical experiments.

3.5. Genetic Programming (GP)

Genetic Programming (GP) is a specialized form of Genetic Algorithm (GA) where each evolved individual is a function. The field of GP took off in the early 1990s, developed by the pioneering application in various complex optimization and search problems (Koza, 1994). A typical GP system consists of the following components (Poli et al., 2008):

- *Terminal Set*: The inputs of the function (i.e., variables, constants, functions with no arguments, etc.)
- *Function Set*: The primitive operations used by the function (i.e. arithmetic, mathematical, Boolean, conditional, looping, etc.)

- *Fitness measure*: The quality of an individual function (i.e. the error between the output of the function and the desired output.)
- *GP parameters*: The control of the whole GP system (i.e. population size, maximum program size, crossover rate, mutation rate, etc.)
- *Termination criterion*: The stopping condition (i.e. a desired solution is found, the maximum number of generations is reached, etc.)

When the number of generations reaches a certain level, GP can transform populations of functions into new and potentially better ones. A distinct feature of GP compared to other evolutionary methods is the tree structure, which gives not only an optimized solution but also the solution method. The drawback is that such a tree is sometimes difficult to interpret semantically. As stated in Poli et al. (2008), GP has been very successful at evolving novel and unexpected ways of solving problems. To evaluate the model developed from XGBoost and GP, we use the same evaluation process as we adopted for the three traditional financial distress prediction models mentioned in section 3.2.

4. Empirical result

4.1. Baseline results

In this section, we first introduce the results from three traditional benchmark models for predicting company financial distress. We have used the three models shown in equations (1), (2) and (3) to estimate the model prediction accuracy regarding firm financial distress based on the original model coefficients for both pandemic and the post-pandemic periods.

The prediction results for the three models are presented in Table 4 for pre-COVID period. In Table 4, it can be observed that the Probit model exhibits the highest prediction accuracy with an overall accuracy of 82.1% for the pre-COVID period, followed by the D-score model with an overall accuracy of 77.9%. The Z-score model has the lowest prediction accuracy, namely, 61.2%.

The prediction results for the three models are presented in Table 5 for post-COVID period. From Table 5, it can be observed that the Probit model exhibits the highest prediction accuracy with an overall accuracy of 80.9%, followed by the D-score model with an overall accuracy of 77.2%. The Z-score model has the lowest prediction accuracy, namely, 57.9%.

For this period, the Probit model showcases the lowest prediction error in both type I and type II errors among those three models, which are 13.3% and 15.9%, respectively. On the other hand, the Z-score model displays the largest prediction error in both type I and type II errors among those three models, which are 43.8% and 41.3% respectively.

Table 4: Prediction accuracy of traditional financial distress prediction models for pre-COVID period.

Applied Model	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	Accuracy
Z-Score	60.2%	39.8%	61.4%	38.6%	61.2%
Probit	81.2%	18.8%	82.8%	17.2%	82.1%
D-Score	77.3%	22.7%	78.6%	21.4%	77.9%

4.2. Classification of distressed firms based on XGBoost

Similar to the previous traditional prediction models, we evaluate our XGBoost model by employing two types of errors, namely, Type I error and

Table 5: Prediction accuracy of traditional financial distress prediction models for post-COVID period.

Applied Model	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	Accuracy
Z-Score	56.2%	43.8%	58.7%	41.3%	57.9%
Probit	80.7%	19.3%	81.1%	18.9%	80.9%
D-Score	75.4%	24.6%	79.4%	20.6%	77.2%

Type II error. Type I error indicates the model's misclassification by putting distressed companies into the *Stable* group. On the other hand, Type II error suggests that the model's misclassification by putting stable companies into the *Distressed* group (see Table 2). 100 estimators with the maximum depth of 10 using binary logistic regression and $\gamma = 0.1$ are applied for our XGBoost model. The adopted features in the XGBoost model are those used in the traditional financial distress prediction models (Table 1). Table 7 shows the prediction accuracy using XGBoost.

Table 6: XGBoost prediction accuracy regarding Type I and Type II errors for pre-COVID period.

Model Prediction	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	
XGBoost	366/407 (89.9%)	10.1%	3320/3735 (88.8%)	11.1%	3686/4142 (89.0%)

Table 7: XGBoost prediction accuracy regarding Type I and Type II errors for post-COVID period.

Model Prediction	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	
XGBoost	377/445 (84.7%)	15.3%	3475 /3697 (93.9%)	6.1%	3852/4142 (93.0%)

Furthermore, the XGBoost model assists us in illuminating the feature importance according to their predictability of firm financial distress for both pre-COVID and post-COVID periods. Generally, the total information gain and the total split number are the common measures used to describe feature importance. The total information gain is the greatest amount of information gained about the targeted variable. The total split number is the split times of a feature when generating the decision tree. In this work, we use total information gain as our feature importance measure. The total information gain across all splits of a feature used in the tree for pre-COVID period prediction model and post-COVID period prediction model can be found in Figures 1 and 2, respectively.

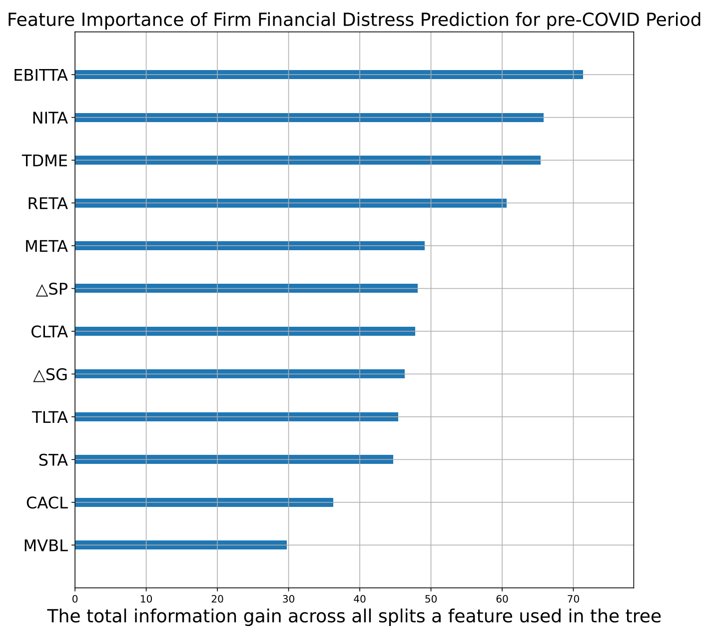

Figure 1: Feature importance of Firm Financial Distress Prediction for pre-COVID Period measured by total information gain.

Based on our empirical observations, the common top four features are "EBITTA", "NITA", "TDME", and "RETA" for the pre-COVID period, where EBITTA is the Earnings before interest & taxes/Total assets, NITA is the Net income/Total assets, TDME is Total Debt/Market equity, and RETA is Retained earnings/Total assets. On the other hand, for post-COVID period, the common top four features are "TDME", "META", "RETA" and "EBITTA". Correspondingly, TDME is Total Debt/Market equity, META is Market Equity/Total assets, RETA is Retained earnings/Total assets, and EBITTA is the Earnings before interest & taxes/Total assets.

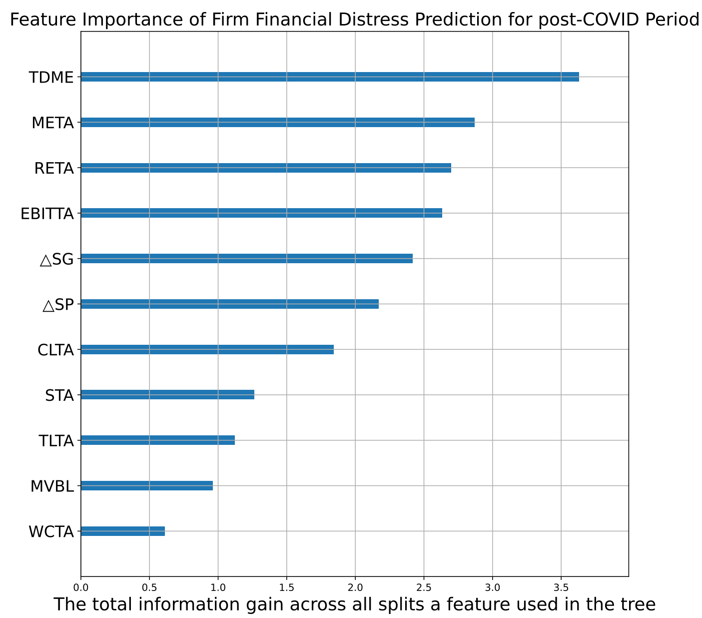

Figure 2: Feature importance of Firm Financial Distress Prediction for post-COVID Period measured by total information gain.

4.3. Model evolution using Genetic Programming

To further develop the prediction model, we adopt the GP method for our model construction. The reason why we further employ the GP method is twofold. First, we can significantly reduce the number of features that are used in the model construction. We only selected the top four features for the model for the pre-COVID and post-COVID periods, contrasting with the 13 features utilized in the XGB process. More importantly, GP can generate a specified model format that we can apply in an explicit manner.

Accordingly, we use GP to evolve the following function: $f(x_1, x_2, x_3, x_4) = I$, where x_1, x_2, x_3, x_4 are the top four features selected by XGBoost for both the pre-COVID period and post-COVID period. I denotes the indicator of whether a company belongs to the *Distressed* group or the *Stable* group according to the predetermined critical value acquired from the sample data distribution. Our goal is to develop a composite function that has effects on classifying financially distressed firms. The GP approach adopted in this work consists of the following parts:

- *Terminal Set*: x_1, x_2, x_3, x_4.
- *Function Set*: $+, -, \times, /$.
- *Fitness measure*: the error between the value of the individual function and the corresponding desired output (i.e. I).
- *GP parameters*: population = 10000, the maximum length of the program = 1000, probability of crossover operation = 0.9, and probability of mutation operation = 0.1.
- *Termination criterion*: Up to 1000 generations.

Parameter tuning is a common practice in evolutionary computation. Since the algorithmic parameters are not independent and often interact in a complex way, it is impractical to test all the possible combinations systematically. We do not claim the algorithmic parameters used for this work is optimal. Our primary aim is rather to test the effectiveness of GP for financial distress prediction problem. All the tests were run on the same Intel(R) Core(TM) i7-11800H 2.30GHz processor with 32.00 GB RAM PC and Windows 10 operating system. It takes approximately 5 minutes for 1000 iterations.

4.4. Genetic Programming based financial distress prediction for pre-COVID period

Table 8: GP-developed function prediction accuracy for pre-COVID period regarding Type I and Type II errors.

Model Prediction	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	
GP	336/407 (82.5%)	17.5%	3254/3735 (87.1%)	12.9%	3590/4142 (86.6%)

This section presents the XGB-GP based financial distress prediction results for the pre-COVID period. We first construct the financial distress prediction model by the GP method and the model is depicted in equation (6):

$$I = \frac{x_4}{x_1^2 + x_3^2} + \frac{x_2^2}{x_1^2 + 1}, \tag{6}$$

where x_1, x_2, x_3, x_4 are the top 4 features, namely, "EBITTA", "NITA", "TDME", and "RETA", respectively.

We then forecast firm financial distress conditions using the model in equation (6), for pre-COVID period, which are exhibited in Table 8. We demonstrate that our model prediction errors are lower than those of the traditional models. The overall accuracy predicted by our GP method is approximately 86.6%, which is superior than the traditional models displayed in Table 4. The Probit model has the highest prediction accuracy, with an overall accuracy of 82.1%, which is still lower than the GP model's accuracy.

Additionally, our evolved GP model performance is close to the performance of XGB model (see Table 6 and Table 8).

4.5. Genetic Programming based financial distress prediction for post-COVID period

In order to provide comparative evidence across both pre-COVID and post-COVID periods. This section delivers the XGB-GP based financial distress prediction results for the post-COVID period. With the settings stated in the section 4.3, the best function obtained by GP is:

$$I = \frac{1}{x_1 * x_2} + \frac{x_1 + x_3}{x_4^2 + 1}, \tag{7}$$

where x_1, x_2, x_3, x_4 are the top 4 features, namely, "TDME", "META", "RETA" and "EBITTA", respectively.

The model that our XGB-GP method created is a non-linear model. This nonlinearity feature of our model could be tremendously helpful since a non-linear model can be more flexible than a linear model in capturing data patterns (Bae, 2012). Furthermore, from previous empirical studies, non-linear models generally outperform linear models such as discriminant analysis, in predicting firm financial distress (Desai et al., 1996; Van Gestel et al., 2006).

According to our model in equation (7), the first term of the function is the inverse of "TD/ME*ME/TA", which yields TD/TA (total debt/total asset), suggesting the total financial leverage of the company. It has been documented in the literature that the impact of financial leverage on firm financial distress is threefold. First, it is obvious that the increase in financial leverage would raise the proportion of total debt against total assets. As revealed in tradeoff theory, this would increase the debt burden of a company with high interest payments and thus increase bankruptcy risk (Modigliani and Miller, 1963; Myers, 2001; Dudley and Yin, 2018). During the post-COVID period, the economic recession could exacerbate this debt burden and bankruptcy risk, which makes this factor more considerable.

Second, an increase in financial leverage can generate a high level of future cash flow uncertainty, resulting in high cash flow volatility (Mackay and Phillips, 2005; Elkhal, 2019; Chowdhury et al., 2023). Firms with a high level of cash flow volatility are more likely to experience financial distress (Minton and Schrand, 1999). Finally, financial leverage creates a negative effect on firm profitability since financial leverage is positively associated with cost fixity and agency cost (Baker, 1973; Dalci, 2018; Moreno-Bromberg and Vo, 2018). The low level profitability can enlarge the probability of debt default because low level profit may not be able to cover the whole debt cost. This argument is also supported by the second term of the function. The second term of the function indicates that when the retrained earnings are high, the probability of firms involved in the financial distress situation is low. Furthermore, during the post-COVID period, it became more difficult to obtain loans from banks (Bao and Huang, 2021) and thus cash flow could be more valuable to keep firms surviving. Hence, the high cash flow volatility could jeopardize the company's stable operation and exacerbate the bankruptcy risk.

Therefore, as exhibited in Table 9, the model prediction errors are tremendously lower than those of the traditional models. The overall accuracy predicted by our GP method is approximately 90.9%, which is profoundly higher than the traditional models displayed in Table 5. The Probit model has the highest prediction accuracy, with an overall accuracy of 80.9%, which is still lower than the GP model's accuracy, which is 90.9%. It is thereby arguable that TDTA may have significant predictive power for the firm financial distress situation, which has not been included in the three traditional models. This sheds new insights for financial distress predictions with other new variable considerations.

Although the performance of GP is slightly below that of XGP (see Table 7 and Table 9), GP is preferable since it can provide an explainable model and the prediction results are still better than those of traditional linear models. On the other hand, the XGB method is mainly dedicated to identifying the key features to construct the model. Therefore, the GP method would be favored since it can provide an explainable model and has a close performance to XGB.

Table 9: GP-developed function prediction accuracy post-COVID period regarding Type I and Type II errors.

Model Prediction	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	
GP	368/445(82.6%)	17.4%	3401/3697 (91.9%)	8.1%	3769/4142 (90.9%)

4.6. Robustness check

Furthermore, the standard robustness check would not be applicable for our study in terms of changing measures. The factors of our model in equation (7), which has been extracted by XGB method, as the top relevant factors in predicting firms' financial distress probability. If we change the factors in our model (in equation 7), then the whole framework would be highly inconsistent as the importance ranking and the measure of the key factors will differ. As a result, it would be inappropriate for the data-mining based model to undertake the standard robustness check in terms of changing measures.

On the other hand, however, in order to verify our results' robustness, we further compare two machine learning based non-linear models used in the literature (see Geng et al., 2015; Elhoseny et al., 2022; Qian et al., 2022), namely, the Neural Network (NN) model (see Table 10) and the Support Vector Machines (SVM) model (see Table 11), for firms' financial distress prediction for the post-COVID period, which is the main focus of our study. For the NN model, a Multilayer perceptron (MLP) with 3 layers and the relu activation function are adopted. For the SVM model, the default parameters in sklearn are used ($C = 1.0$, γ is scale and kernel is rbf). From Tables 10 and 11, it is clear that our model has surpassed the two models in terms of predicting accuracy regarding both NN model and SVM model for the post-COVID period. Therefore, it is arguable that our results are robust.

Table 10: Neural network based financial distress prediction accuracy regarding Type I and Type II errors for post-COVID period.

Model Prediction	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	
NN	355/445 (79.7%)	20.3%	2834 /3697 (76.7%)	23.3%	3189/4142 (76.9%)

Table 11: Support vector machines based financial distress prediction accuracy regarding Type I and Type II errors for post-COVID period.

Model Prediction	Distressed		Stable		Overall
	Distressed	Type I Error	Stable	Type II Error	
SVM	334/445 (75.1%)	24.9%	2592/3697(70.1%)	29.9%	2926/4142 (70.6%)

5. Research Implications

In fact, our model is remarkably aligned with the current Chinese economic situation and economic policy, which yields significant policy implications. It is notable that the high leverage ratio has become a major risk factor faced by the Chinese economy, not only for listed companies in China. As a result, the Chinese government has successively launched a series of mandatory deleveraging policies, encouraging the removal of firms' redundant liabilities. Against this backdrop, this policy would help companies to concentrate their equity rights, releasing companies from the free rider behavior of small and medium-sized shareholders in corporate governance. Consequently, it can heavily empower major shareholders to effectively supervise the management team and reduce the first type of agency costs (Jensen, 1986).

On this basis, our model can thereby improve the operational efficiency of enterprises and thus achieve the purpose of reducing financial distress risk by deleveraging. Deleveraging could be extremely helpful during the post-COVID period to reduce both debt burden and cash flow volatility, which can assist firms in improving their financial distress conditions. Our model can be applied in optimizing firms' capital structure since our model can generate the threshold value of leverage ratio such that company will enter into the financial distress situation. Therefore, managers can actually use our model to preset the maximum firm's debt/equity level to prevent firms from entering into the financial distress situation.

The theoretical implications of our paper is twofold. The first theoretical implication of our XGB-GP model is the interpretability of our financial distress prediction model. Our XGB-GP model has created the explicit financial distress prediction model with explicit predicting factors, which can enhance the transparency of financial distress predicting. Our explicit financial distress prediction model thereby can provide stakeholders with a deeper understanding of the factors influencing financial distress, facilitating better decision-making and risk management. Additionally, accurate financial distress forecasting models require continuous updates for accommodating the evolving economic and financial circumstances. The theoretical implications of our XGB-GP model in terms of model adaptability and incremental learning can be inferred. Our XGB-GP model exhibits the adaptability and incremental learning characteristics. We have used our XGB-GP model for both pre-COVID and post-COVID periods, demonstrating how our XGB-GP model can incorporate time-varying information and embed such information into the financial distress predictions.

6. Conclusions

The prediction of financial distress becomes pivotal within the COVID-19 crisis circumstance. In this paper, we propose a firm financial distress prediction model based on the XGB-GP framework with both pre-COVID and post-COVID Chinese company samples. We first determine the superiority of the factors that can affect the probability of financial distress. Then, we develop our model based on the most relevant factors determined by the XGB technique adopting the GP method.

Our prevailling contribution is that we argue that different key prediction features shall be applied for pre-COVID and post-COVID periods by demonstrating a time-varying feature selection. We demonstrate that the earning financial indicator is the dominating feature for financial distress prediction during the pre-COVID period. We further unveil that total financial leverage is the most important factor in predicting firm financial distress during the post-COVID period. It is arguable that when the economy is in an expanding episode, firms are expected to expand earnings to cover the debt payment whereas firms shall cut the debt payment to survive during the economy downturn periods.

On the basis, our XGB-GP financial distress prediction model is more accurate than the traditional models. Our model has an overall prediction accuracy of 86.6% for the pre-COVID period and 90.9% for the post-COVID period, which are both superior than the Probit model of 82.1% and 80.9%, the D-score model of 77.9% and 77.2%, and the Z-score of 61.2% and 57.9%. In addition to the traditional financial distress prediction models, we also compared our XGB-GP model with artificial intelligence algorithms, including SVM and NN. Our model also surpasses both artificial intelligence algorithms in forecasting the financial distress.

Therefore, our study unravels the fact that the financial leverage factor retains the leading predictive power of financial distress, which is quite different from the existing prediction models we have used during the post-COVID period. The results also suggest that during the COVID-19 period, high financial leverage is more likely to turn companies into financial distress. The influence of financial leverage on the firm financial situation is threefold. First, financial leverage can increase the debt burden of a company with high interest payments and thus increase bankruptcy risk. Second, the increase in financial leverage can generate high cash flow volatility. Finally, financial leverage creates a negative effect on firm profitability since financial leverage is positively associated with cost fixity and agency cost. Thus, managers can modify the financial leverage level to improve the financial situation of the firm by reducing the debt burden and increasing profitability. Furthermore, our study has unveiled that the financial leverage factor has the largest predicting power, which is quite different from the existing models since we have included the sample data of post-COVID-19 crisis periods. We unfold the fact that during crisis periods, high financial leverage is more likely to turn companies into financial distress. We thereby shed new insights for both financial distress prediction and firm corporate governance during the COVID-19 period.

Since copious of companies and financial institutions are connected within the same network, the market disaster can arise from the chain reaction when several companies are involved into the financial distress situation simultaneously. This phenomena can be highly likely during the COVID period as the economy went into a recession. Consequently, our accurate financial distress prediction model with COVID data can prevent such market disaster from the initiation with the precautions beforehand. Thus, our financial distress prediction model is exceedingly valuable to policymakers because it can constitute an important part for the establishment of an early warning system regarding the company distress risk.

Aknowledgement

This Project is Supported by Ningbo Natural Science Foundation (Project ID 2023J194)

References

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

[2] Andreou, C. K., Lambertides, N., & Panayides, P. M. (2021). Distress risk anomaly and misvaluation. *The British Accounting Review*, 53(5), 100972.

[3] Andrikopoulos, P., & Khorasgani, A. (2018). Predicting unlisted SMEs' default: Incorporating market information on accounting-based models for improved accuracy. *The British Accounting Review*, 50(5), 559-573.

[4] Ashraf, S., Felix, E., & Serrasqueiro, Z. (2019). Do traditional financial distress prediction models predict the early warning signs of financial distress?. *Journal of Risk and Financial Management*, 12(2), 55.

[5] Aslan, H. (2022). Personal Financial Distress, Limited Attention. *Journal of Accounting Research*, 60(1), 97-128.

[6] Bae, J. K. (2012). Predicting financial distress of the South Korean manufacturing industries. *Expert Systems with Applications*, 39(10), 9159-9165.

[7] Bao, Z., & Huang, D. (2021). Shadow banking in a crisis: Evidence from FinTech during COVID-19. *Journal of Financial and Quantitative Analysis*, 56(7), 2320-2355.

[8] Baker, S. H. (1973). Risk, leverage and profitability: an industry analysis. *The Review of Economics and Statistics*, 55(4), 503-507.

[9] Blums, M. (2003). D-Score: Bankruptcy Prediction Model for Middle Market Public Firms. Available online: http://www.minneapolisfed.org/mea/contest/2004papers/blums.pdf.

[10] Bozkurt, I., & Kaya, M. V. (2022). Foremost features affecting financial distress and bankruptcy in the acute stage of COVID-19 crisis. *Applied Economics Letters*, 1-12.

[11] Brennan, N. M., Edgar, V. C., & Power, S. B. (2022). COVID-19 profit warnings: Delivering bad news in a time of crisis. *The British Accounting Review*, 54(2), 101054.

[12] Campa, D. (2015). The impact of SME's pre-bankruptcy financial distress on earnings management tools. *International Review of Financial Analysis*, 42, 222-234.

[13] Cao, Y. (2012). Aggregating multiple classification results using Choquet integral for financial distress early warning. *Expert Systems with Applications*, 39(2), 1830-1836.

[14] Charalambakis, E. C., & Garrett, I. (2019). On corporate financial distress prediction: What can we learn from private firms in a developing economy? Evidence from Greece. *Review of Quantitative Finance and Accounting*, 52(2), 467-491.

[15] Chiaramonte, L., & Casu, B. (2017). Capital and liquidity ratios and financial distress. Evidence from the European banking industry. *The British Accounting Review*, 49(2), 138-161.

[16] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *the 22nd ACM SIGKDD International Conference*, 785-794.

[17] Chen, X., Cheng, Q., & Lo, A. K. (2013). Accounting restatements and external financing choices.*Contemporary Accounting Research*, 30(2), 750-779.

[18] Chowdhury, MAF., Abdullah, M., Alam, M., Abedin, MZ., and Shi, B. (2023). NFTs, DeFi, and Other Assets Efficiency and Volatility Dynamics: An Asymmetric Multifractality Analysis. *International Review of Financial Analysis.*

[19] Crespi-Cladera, R., Martin-Oliver, A., & Pascual-Fuster, B. (2021). Financial distress in the hospitality industry during the COVID-19 disaster. *Tourism Management*, 85, 104301.

[20] Dalci, I. (2018). Impact of financial leverage on profitability of listed manufacturing firms in China. *Pacific Accounting Review*, 30(4), 410-432.

[21] Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95, 24-37.

[22] Dirman, A. (2020). Financial distress: the impacts of profitability, liquidity, leverage, firm size, and free cash flow. *International Journal of Business, Economics and Law*, 22(1), 17-25.

[23] Dorr, J. O., Licht, G., & Murmann, S. (2022). Small firms and the COVID-19 insolvency gap. *Small Business Economics*, 58(2), 887-917.

[24] Dudley, E., & Yin, Q. E. (2018). Financial distress, refinancing, and debt structure.*Journal of Banking & Finance*, 94, 185-207.

[25] Elhoseny, M., Metawa, N., Sztano, G., & El-Hasnony, I. M. (2022). Deep Learning-Based Model for Financial Distress Prediction. *Annals of Operations Research*, 1-23.

[26] Elkhal, K. (2019). Business uncertainty and financial leverage: should the firm double up on risk?. *Managerial Finance*, 45 (4), 536-544.

[27] Hosseini, S., & Ivanov, D. (2021). A multi-layer Bayesian network method for supply chain disruption modelling in the wake of the COVID-19 pandemic. *International Journal of Production Research*, 1-19.

[28] Huang, H., & Ye, Y. (2021). Rethinking capital structure decision and corporate social responsibility in response to COVID-19. *Accounting & Finance*, 61(3), 4757-4788.

[29] Huang, Y., & Yen, M. (2019). A new perspective of performance comparison among machine learning algorithms for financial distress prediction. *Applied Soft Computing*, 83, 105663.

[30] Gandhi, P., Loughran, T., & McDonald, B. (2019). Using annual report sentiment as a proxy for financial distress in US banks. *Journal of Behavioral Finance*, 20(4), 424-436.

[31] Gao, P., Parsons, C. A., & Shen, J. (2018). Global relation between financial distress and equity returns. *The Review of Financial Studies*, 31(1), 239-277.

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

[33] Gilson, S. C. (1989). Management turnover and financial distress. *Journal of Financial Economics*, 25(2), 241-262.

[34] Habib, A., Bhuiyan, B. U., & Islam, A. (2013). Financial distress, earnings management and market pricing of accruals during the global financial crisis. *Managerial Finance*, 39 (2), 155-180.

[35] Jabeur, S. B., & Fahmi, Y. (2018). Forecasting financial distress for French firms: a comparative study. *Empirical Economics*, 54(3), 1173-1186.

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

[37] Jiang, C., Lyu, X., Yuan, Y., Wang, Z., & Ding, Y. (2022). Mining semantic features in current reports for financial distress prediction: Empirical evidence from unlisted public firms in China. *International Journal of Forecasting*, 38(3), 1086-1099.

[38] Jones S., & Hensher D. A. (2004). Predicting firm financial distress: A mixed logit model, *Accounting Review*, 79 (4),1011-1038.

[39] Keasey, K., & Watson, R. (1991). Financial distress prediction models: a review of their usefulness. *British Journal of Management*, 2(2), 89-102.

[40] Khan, S. U. (2022). Financing constraints and firm-level responses to the COVID-19 pandemic: International evidence. *Research in International Business and Finance*, 59, 101545.

[41] Khoja, L., Chipulu, M., & Jayasekera, R. (2019). Analysis of financial distress cross countries: Using macroeconomic, industrial indica-

tors and accounting data. *International Review of Financial Analysis*, 66, 101379.

[42] Koza, J. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, 4(2), 87-112.

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

[44] Li, S., Shi, W., Wang, J., & Zhou, H. (2021). A deep learning-based approach to constructing a domain sentiment lexicon: a case study in financial distress prediction. *Information Processing & Management*, 58(5), 102673.

[45] Li, Y., Li, X., Xiang, E., & Djajadikerta, H. G. (2020). Financial distress, internal control, and earnings management: Evidence from China. *Journal of Contemporary Accounting & Economics*, 16(3), 100210.

[46] Li, Z., Crook, J., Andreeva, G., & Tang, Y. (2021). Predicting the risk of financial distress using corporate governance measures. *Pacific-Basin Finance Journal*, 68, 101334.

[47] Lian, Y. (2017). Financial distress and customer-supplier relationships. *Journal of Corporate Finance*, 43, 397-406.

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

[49] Lin, F., Liang, D., Yeh, C. C., & Huang, J. C. (2014). Novel feature selection methods to financial distress prediction. *Expert Systems with Applications*, 41(5), 2472-2483.

[50] Mackay, P., & Phillips, G. M. (2005). How does industry affect firm financial structure?. *Review of Financial Studies*, 18 (4), 1433-1466.

[51] Mahtani, U. S., & Garg, C. P. (2018). An analysis of key factors of financial distress in airline companies in India using fuzzy AHP framework. *Transportation Research Part A: Policy and Practice*, 117, 87-102.

[52] Moreno-Bromberg, S., & Vo, Q. A. (2017). Resolution of financial distress under agency frictions.*Journal of Banking & Finance*, 82, 40-58.

[53] Minton, B. A., & Schrand, C. (1999). The impact of cash flow volatility on discretionary investment and the costs of debt and equity financing. *Journal of Financial Economics*, 54(3), 423-460.

[54] Mselmi, N., Lahiani, A., & Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis*, 50, 67-80.

[55] Orlando, B., Tortora, D., Pezzi, A., & Bitbol-Saba, N. (2022). The disruption of the international supply chain: Firm resilience and knowledge preparedness to tackle the COVID-19 outbreak. *Journal of International Management*, 28(1), 100876.

[56] Pindado, J., Rodrigues, L., & De La Torre, C. (2006). How does financial distress affect small firms' financial structure?. *Small Business Economics*, 26(4), 377-391.

[57] Poli, R., Langdon, B., & McPhee, N. (2008). A Field Guide to Genetic Programming. *Lulu Enterprises, UK Ltd.*

[58] Qian, H., Wang, B., Yuan, M., Gao, S., & Song, Y. (2022). Financial distress prediction using a corrected feature selection measure and

gradient boosted decision tree. *Expert Systems with Applications*, 190, 116202.

[59] Sudarsanam, S., & Lai, J. (2001). Corporate financial distress and turnaround strategies: An empirical analysis. *British Journal of Management*, 12(3), 183-199.

[60] Tsai, C. F., Sue, K. L., Hu, Y. H., & Chiu, A. (2021). Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *Journal of Business Research*, 130, 200-209.

[61] Van Gestel, T., Baesens, B., Suykens, J. A., Van den Poel, D., Baestaens, D. E., & Willekens, M. (2006). Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research*, 172(3), 979-1003.

[62] Vo, T. A., Mazur, M., & Thai, A. (2021). The impact of COVID-19 economic crisis on the speed of adjustment toward target leverage ratio: an international analysis. *Finance Research Letters*, 102157.

[63] Volkovs, M., Yu, G., & Poutanen, T. (2017). Content-based Neighbor Models for Cold Start in Recommender Systems. *Proceedings of the Recommender Systems Challenge 2017*, 7:1-7:6.

[64] West, R. C. (1985). A factor analytic approach to bank condition, *Journal of Banking & Finance*, 9, 253-266.

[65] White, M. J. (1983). Bankruptcy costs and the new bankruptcy code. *The Journal of Finance*, 38(2), 477-488.

[66] Wu, Y., Gaunt, C., & Gray, S. (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics*, 6(1), 34-45.

[67] Yang, C., Abedin, MZ., Zhang, H., Weng, F., & Petr, H. (2023). An Interpretable System for Predicting the Impact of COVID-19 Government Interventions on Stock Market Sectors. *Annals of Operations Research.*

[68] Zhao, S., Xu, K., Wang, Z., Liang, C., Lu, W., & Chen, B. (2022). Financial distress prediction by combining sentiment tone features. *Economic Modelling*, 106, 105709.

[69] Zhao, Y., Goodell, JW., Wang, Y. & Abedin, MZ. (2023). Fintech, Macroprudential Policies and Bank Risk: Evidence from China. *International Review of Financial Analysis.*

[70] Zhu, W., Zhang, T., Wu, Y., Li, S., & Li, Z. (2022). Research on optimization of an enterprise financial risk early warning method based on the DS-RF model. *International Review of Financial Analysis*, 81, 102140.

[71] Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82.