Prediction of Stock Pledge Financing Default Based on Machine Learning

Jialu Wu

{ zcahwul@ucl.ac.uk }

Department of Mathematics, University College London, London, WC1E 6AE, United Kingdom

Abstract. Stock pledge has become a widely used financing method among controlling listed companies, while its default causes significant challenges to financial stability and risk management. This study investigates default prediction in stock pledge financing by identifying key indicators for early warning signals using machine learning. The dataset contains information on the default status of pledge financing of controlling shareholders of Chinese listed companies from 2017-2020 are used to construct training and test sets for modelling. The performance of several models is compared, including Logistic Regression, Random Forests, and Gradient Boosting. The findings suggest that the Gradient Boosting model performs the best, achieving an AUC of 0.99 and an accuracy of 0.96. The feature importance analysis further points out key indicators related to default risk, providing valuable insights for risk monitoring and preventive interventions.

Keywords: Machine learning; Stock pledge; Default prediction

1 Introduction

Stock pledge (equity pledge) is a financing method where the shareholders grant the lender its security against its assets. The issue of pledging shares as security enables controlling shareholders to increase capital without loss of control. Nevertheless, the mentioned financing approach has high risk attached, because the pledged share is directly linked with the market prices. Rapid drops in share prices may have severe results, like margin calls and liquidation which may potentially exacerbate the stress.

Stock pledge financing is unique in its risk characteristics as compared to the traditional corporate financing practices. The default risks associated with corporate fundamentals and market conditions are extremely sensitive in the market valuation structure of pledged shares. Literature reports that with the economic crunch, pledges of stock increase the negative risks and market speculation, which imposes some uncertainty about corporate control and finance stability [1]. These attributes mean that stock promise default occurrences are often sudden and unexpected, hence, there is minimal opportunity to respond to it.

Although the traditional models in default prediction are a beneficial theoretical basis, they usually utilize few indicators and linear relations. In this regard, the latest trends in machine

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365111

learning have provided more adaptable models that are capable of capturing nonlinear relationships and other complicated interactions between different features. As such, the techniques are especially more useful in detecting the early warning signs in stock pledge financing, which is an issue marked by multi-dimensionalities and highly correlated risk processes.

Modelling presented in this study is based on a large structured dataset containing the information on the financing of pledges and other market indicators in addition to financial fundamentals of Chinese listed firms (the number of variables that were taken into account is 60). The samples included the years 2017 to 2020. The companies that lacked important variables were not included in the dataset. The dependent variable, which is referred to as "IsDefault" is a binary variable: 1 when the company encountered a judicial freeze event linked to stock pledge financing and 0 when not.

2 Literature review

2.1 Stock pledge and financial risk

Recent literature has increasingly focused on the financial risks associated with stock pledges, particularly in emerging markets, where firms often face difficulties financing, and therefore, controlling shareholders frequently pledge their shares to obtain external funds. A growing consensus suggests that stock pledges amplify corporate risks through multiple interrelated issues, including margin call pressures, agency conflicts and financing constraints. And that is why people care about the prediction of the default and early warning signals.

One research category has highlighted margin calls and forced liquidation arrangements. When share prices decline, pledged shares may cause margin calls, forcing controlling shareholders to provide additional collateral or suffer compulsory liquidation. Empirical studies indicate that equity-pledging-related margin pressures influence major corporate actions and increase financial vulnerability [2]. Similarly, there are results showing that controlling shareholders' equity pledges correlate with higher equity capital costs, indicating that capital markets have priced in the additional risks arising from pledge activities [3].

Another important category of studies examines stock pledges from the perspective of the agency problem and the risk of control rights. When a substantial proportion of a shareholder's shares is pledged, the extent of separation between control rights and cash flow rights intensifies, resulting in intensive risk-shifting behavior. Results from Wang et al. indicate that stock pledges can induce either an incentive or an adverse effect, depending on pledge intensity, but the higher the pledge ratio, the more likely it is to amplify the risk of dispossession and undermine corporate governance [4]. These findings show that stock pledges alter the management motivation mechanism and increase risk exposure at the corporate level.

Moreover, stock pledges have been linked to financial constraints and inefficient resource allocation, which can potentially increase default risk. In the previous work of Xie et al., it was found that shareholders' equity pledges heighten excessive financialization, leading to resource diversion from productive investments and weakening corporate long-term financial fundamentals [5]. Ren et al. reported that stock pledges are also related to lower innovation efficiency, potentially undermining the stability of future cash flow [6].

2.2 Development and application of default prediction models

A distinct body of literature on stock pledges focuses on default prediction and early warning signals. Recent studies increasingly apply advanced econometric and machine learning methods to improve predictive accuracy and capture nonlinear relationships among risk factors.

Shaheen et al. developed five machine learning models for stock pledge default prediction, based on a dataset containing pledge ratios, stock volatility and return features [7]. Their results indicate that KNN, SVC, and Decision Tree models achieved perfect accuracy on the training data, and Decision Tree and SVC continued to achieve high accuracy. It highlighted the value of machine learning in detecting complex relations between features of stock pledge datasets.

Beyond stock-pledge-specific studies, there are many studies using machine learning models to predict other corporate defaults. These studies suggested that ensemble methods, such as Random Forests and Boosting, frequently perform better than traditional methods like KNN and logistic regression, especially when dealing with nonlinear relations and high-dimensional feature spaces. For instance, Guo compared several classification models to predict corporate credit default, including logistic regression, Decision Trees, AdaBoost and Random Forests [8]. The results show that the Random Forest model achieves the highest accuracy, recall, and F1-score, outperforming the others.

In terms of feature selection, recent work also explored alternative predictors beyond financial ratios. Chang et al. demonstrated that non-traditional information like ESG scores (Environment, Social and Governance) can improve corporate default prediction models under unbalanced sample conditions [9]. Although this study focuses more on corporate general default analysis, its findings support the idea that expanding feature dimensions, adding market-based and governance indicators, can improve the effectiveness of the prediction models.

2.3 Research gap

Although the existing literature has provided substantial evidence on the risk implications of stock pledges and has emphasised the growing prevalence of machine learning models in default prediction, certain gaps remain unfilled. Firstly, research about stock pledges mainly focused on the risk outcomes after default, such as the risk of share price collapse and financing constraints. However, little research attention is given to the issue of predicting default in the future. Second, the research on default prediction studies has largely focused on general corporate default highlighting little attention to any specific risk attributes of stock pledges.

In order to address these gaps, the paper is dedicated to stock-pledge-specific prediction of default using machine learning. By combining the data about the attributes of pledging structure, financial liquidity aspect, market based features and financial fundamentals, and comparing the performance of a set of classification machine learning models, the study will recognize the key early warning indicators of the stock pledging defaults.

3 Methodology

3.1 Feature engineering

The features are classified into four categories and they include as pledge structure features, financial liquidity features, market-based features and financial fundamentals. This combination assists in the modelling of interpretation and the identification of early warning of default.

The pledge structure characteristics represent how the pledging of the company fund depends on the financing by stock pledging. These variables are the portion of pledge ratio by the controlling shareholders, pledge ratio of the limited and unlimited sale shares. An increased pledge ratio means the availability of forced liquidation in case of a decrease in the value of the share and the request of the margin, which increases the default risk. It is this combination of characteristics that can be called systemic weakness, which has direct roots in arrangements of stock pledge.

The characteristics in the market are a reflection of the rating by the investors in terms of the financial risk of the company and market conditions. Variables such as stock volatility, stock price changes over the last year and valuation ratios (P/E and P/B) are included in order to determine when pledged shares put pressure. Since stock fluctuations and volatility often indicate potential corporate issues, these characteristics could be instrumental in identifying early warning signs.

The liquidity characteristics, including the annual turnover rate, are adopted to evaluate the investment strategies and the fund management activity and efficiency of the business. The scarcity of liquidity can amplify volatility of prices in forced sales due to margin calls, then a greater likelihood of companies being taken over to financial distress is possible [10].

Numerous indicators of financial stability, financing costs, leverage and profitability are included in the group of financial fundamentals. These features are used to measure an enterprise's solvency and risk-bearing capacity. Companies with weaker fundamentals often struggle to withstand the capital fluctuations from equity pledge financing, thus being at greater risk of default.

3.2 Models for comparison

To explore the most effective models in predicting stock pledge financing default, three classification models with different structures and interpretations are considered. First, the baseline model - Logistic regression, is built as a benchmark; it predicts the probability of default by fitting a linear combination of input features.

Random Forest is then included in the comparison, using the bagging strategy to construct a tree-based ensemble model. By integrating the predictive outputs of multiple decision trees, a random forest can reduce variance and detect non-linear relations between features and default risk.

The third model employed is Gradient Boosting (including implementations such as XGBoost and LightGBM), which combines weak models together to create a robust predictive system. Each model is trained sequentially, improving each model by correcting the errors of the previous ones. Gradient Boosting is regarded as a particularly powerful model for classification tasks, as it adjusts predictions by leveraging log-likelihood ratios and residual calculation, improving accuracy at each stage [11].

The performance of the three models is evaluated using multiple metrics. The area under the ROC curve (AUC) is reported to assess overall predictive capability, as well as Accuracy, Precision, Recall (sensitivity) and F1-Score. Furthermore, 10-fold cross-validation is applied within the training set to prevent overfitting and provide more reliable estimates than a single train-test split. In each fold, the data are divided into 10 subsets: 9 for estimation and 1 for validation. This is repeated until each fold has served as the validation fold.

4 Results

4.1 Model performance comparison

This study uses multiple classification evaluation metrics - AUC, Accuracy, Precision, Recall and F1-score to evaluate the predictive performance of the three models (Logistic Regression, Random Forest and Gradient Boosting).

Table 1. Model Performance Comparison on the Training Set.

	AUC	**Accuracy**	**Precision**	**Recall**	**F1-score**
Gradient Boosting	0.997681	0.982699	1.000000	0.905660	0.950495
Random Forest	0.968180	0.833910	1.000000	0.094340	0.172414
Logistic Regression	0.948033	0.916955	0.871795	0.641509	0.739130

As shown in Table 1, the Gradient Boosting model outperforms the other two in most of the metrics. It reaches the highest AUC of 0.9977, which is close to 1 (perfect predictive performance) , indicating it is a well-performing model with a high rate of true positives and a low rate of false positives. The Gradient Boosting model also shows strong classification performance, with the highest accuracy of 98.27, a high recall rate of 0.906, a perfect 100% precision and an F1-score of 0.9505.

In contrast, the Logistic Regression model obtains a relatively high AUC of 0.9480 and an accuracy of 0.9170, while the precision (0.8718) and recall (0.6415) are a more balanced trade-off, and the final F1-score is 0.7391. These results suggest that although the logistic regression can provide interpretable and stable results, its effectiveness in detecting complex nonlinear relations between features about stock pledge default is quite low.

The Random Forest model shows notably distinct performance characteristics. Although the AUC (0.9682), Accuracy (0.8339), and Precision (1.0000) are high, the extremely low Recall value indicated that the majority of default events were not detected. Consequently, the F1-score falls to just 0.1724.

4.2 10-fold cross-validation results

To further assess the models and prevent overfitting, 10-fold cross-validation was used on each of the three models. Mean values of metrics of all folds were computed for each model.

Table 2. 10-fold Cross-validation Mean Performance Comparison.

	AUC (CVmean)	Accuracy (CVmean)	Precision (CVmean)	Recall (CVmean)	F1-score (CVmean)
Gradient Boosting	0.991100	0.969488	0.972222	0.861111	0.912320
Random Forest	0.939374	0.839784	0.950000	0.139174	0.235246
Logistic Regression	0.918441	0.892490	0.782271	0.605983	0.674852

As summarised in Table 2, the cross-validation results further proved the model comparison results. The Gradient Boosting model is still listed first across the three models, with the highest values of all evaluation indicators. It achieved a mean AUC of 0.9911, an Accuracy of 0.9695, a Precision of 0.9722, a Recall of 0.8611 and an F1-score of 0.9123.The Logistic Regression model remains stable in performance, achieving a mean AUC of 0.9394 and an F1-score of 0.6749. The mean values of metrics for the Random Forest are also close to the single comparison results above. The small difference between Table 1 and Table 2 proves a strong generalization of models, showing that the performances are not dependent on any particular partitioning strategy.

Based on both validation and 10-fold cross-validation results, the Gradient Boosting model is considered to be the best predictive model and is selected for follow-up analysis.

4.3 Identification of early warning signals

To identify the early warning signals of stock pledge financing default using the chose Gradient Boosting Model, feature importance analysis is conducted.

Table 3. Top 10 Early Warning Signals.

Feature Name	Feature Importance
Share pledge ratio of controlling shareholders	0.360954
Current ratio	0.343117
Financial cycle m2/gdp	0.096041
Ratio of prepayments to operating income	0.051082
Downgrade or negative	0.030565
Z-SCORE	0.012917
Pledge ratio of unlimited shares	0.012861
Stock price rise and fall in the last year	0.007086
EBITDA/interest bearing debt	0.005753
Number of research institutions concerned	0.004084

As shown in Table 3, the share pledge ratio of controlling shareholders is most influential, that is to say, when shareholders pledge a large proportion of their shares, the risk of facing forced liquidation and financial distress significantly increases. The current ratio is ranked second, which is a liquidity indicator, indicating that weak liquidity may also significantly amplify the default risk.

Besides these two features, the feature importances of all other indicators are relatively low (all below 0.1). This may be due to the large number of indicators – a total of 60. However, collectively, these supplementary indicators enhance the effectiveness of early warning systems by integrating corporate financial fundamentals, market conditions and external perception.

4.4 Default prediction for the test set

The Gradient Boosting model is then applied on the independent test set to conduct out-of-sample prediction, outputting the probabilities of stock pledge default occurring. Part of the results are shown in Table 4.

Table 4. Predictive Default Probabilities (partial examples).

Stock code	pre_default_prob
X01443	0.628719
X01444	0.540202
X01445	0.935718
X01446	0.587596
X01447	0.103173

These predictive probabilities provide an important quantitative basis for constructing an early warning system. Companies with high default risk can enhance supervision and implement preventive interventions on the earlier stage.

5 Conclusion

This study focuses on stock pledge financing default prediction, constructs and compares three classification machine learning models to identify key early warning signals. The results indicate that machine learning approaches - especially the Gradient Boosting- can effectively detect complex nonlinear relations between risk features, and compute accurate default probability prediction. A small set of core indicators is found, the share pledge ratio of controlling shareholders and the current ratio are the most influential two, playing important roles in default risk. Additionally, this study applied the Gradient Boosting model to the independent test set to generate predictive default probabilities of companies, providing a practical early warning tool to support risk monitoring and advanced intervention.

While this study provided useful insights, there are several limitations. Firstly, the features included in the models lack qualitative information, such as corporate disclosures, media sentiment and changes in management personnel, which would require much more work on data collection. Secondly, although the model shows strong predictive performance, its effectiveness may vary across different economic environments and market conditions. Therefore, periodic modification is essential to maintain the model's high scores in each metric. Future studies may deepen the analysis of stock-pledge financing defaults and address these limitations, further improving the performance of prediction by integrating dynamic data, alternative modelling approaches or wider market conditions.

References

[1] Pang, C.& Wang, Y.: Stock pledge, risk of losing control and corporate innovation, Journal of Corporate Finance, Volume 60, 101534, ISSN 0929-1199, (2020)
[2] Li, X. & Li, Q.: Does Margin Call Pressure Affect the M&A Decision of Controlling Shareholders With Equity Pledges: Evidence From Chinese Listed Company. Front. Psychol. 13:898118. (2022)
[3] Liu, W. & Tian, G.G.: Controlling shareholder share pledging and the cost of equity capital: Evidence from China, British Accounting Review, 54(6), 1-31. Article 101057. (2022)
[4] Wang, L., & Cao, W.: Separation of control and cash flow rights, ultimate controlling shareholders' equity pledge ratio and risk of control transfer--evidence from Chinese A-share listed companies. Applied Economics Letters, 29(16), 1533–1540. (2022)
[5] Xie, Y., Wang, T., Zhang, J., & Wang, N.: Does controlling shareholders' share pledge exacerbate excessive financialization of enterprises?-Evidence from performance pressure perspective. PloS one, 18(7), e0288705. (2023)
[6] Ren, G., Mo, Y., Liu, L., Zheng, M., & Shen, L.: Equity pledge of controlling shareholders, property right structure and enterprise innovation efficiency: evidence from Chinese firms. Economic Research-Ekonomska Istraživanja, 35(1), 6558–6578. (2022)
[7] Shaheen, H., Irziqat, S., & Al-Dardour, E.: Stock Pledge Defaults Prediction Using Machine Learning. Ahliya Journal of Business Technology and MEAN Economies, 1(2), 22–32. (2024)
[8] Guo, K.: Corporate Credit Default Prediction Based on Machine Learning. In Proceedings of the 2025 International Conference on Economic Management and Big Data Application (ICEMBDA '25). Association for Computing Machinery, New York, NY, USA, 340–348. (2025)
[9] Chang, R., Liu, X.& Deng, W.: Prediction of corporate default risk considering ESG performance and unbalanced samples, Applied Soft Computing, Volume 171, 112864, ISSN 1568-4946, (2025)
[10] The Investopedia Team: Understanding Annual Turnover: Definition, Formula, and Examples. Available at: https://www.investopedia.com/terms/a/annual-turnover.asp (2025)
[11] Pourmahdavi, A.: Master Gradient Boosting for Classification: Enhance Accuracy with Machine Learning. Available at: https://caasify.com/master-gradient-boosting-for-classification-enhance -accuracy-with-machine-learning/ (2025)