International Journal of Finance & Economics

WILEY

RESEARCH ARTICLE OPEN ACCESS

An Ensemble Model Minimising Misjudgment Cost: Empirical Evidence From Chinese Listed Companies

Kunpeng Yuan[1] | Mohammad Zoynul Abedin[2] | Petr Hajek[3]

[1]School of Economics and Management, Dalian University of Technology, Dalian, Liaoning, China | [2]Department of Accounting and Finance, School of Management, Swansea University, Swansea, Wales, UK | [3]Science and Research Centre, Faculty of Economics and Administration, University of Pardubice, Pardubice, Czech Republic

Correspondence: Mohammad Zoynul Abedin (m.z.abedin@swansea.ac.uk)

Received: 3 January 2023 | **Revised:** 10 November 2024 | **Accepted:** 27 November 2024

Funding: The authors received no specific funding for this work.

Keywords: dynamics | ensemble model | financial distress | misjudgment cost | prediction

ABSTRACT

Predicting corporate financial distress is critical for bank lending and corporate bond investment decisions. Incorrect identification of default status can mislead lenders and investors, leading to substantial losses. This paper proposes an ensemble model that minimises the overall cost of misjudgment by considering the imbalanced ratio weighted loss of the unbalanced ratio of Type I and Type II errors in the objective function. Unlike existing static financial distress prediction models, the proposed model integrates panel data by using time-shifting to account for credit risk dynamics. To validate the prediction model, data were collected for Chinese listed companies, considering geographic area, ownership structure and firm size. We demonstrate that by weighting predictions from different classification models, the overall misjudgment cost can be minimised. This study identifies earnings per share and the product price index as the most relevant indicators affecting the financial performance of Chinese-listed companies. Overall, the results indicate that the proposed model has a predictive capacity of up to 5 years, with 98.7% for 1-year forecasting horizons and 96.8% for 5-year-ahead forecasting horizons. Furthermore, the proposed model outperforms existing distress prediction models in overall prediction performance by correctly identifying defaulting companies while avoiding misjudging good companies.

1 | Introduction

The prediction of financial distress involves establishing a relationship between a company's current financial distress status and future outcomes based on explanatory variables such as financial ratios. This prediction aims to identify potential bankruptcy or default risks, providing crucial insights for managers to pre-emptively manage risks. Accurate forecasting of corporate financial distress prediction is essential for informing bank lending decisions and corporate bond investments, as failure to detect default promptly can lead to significant financial losses (Purnanandam 2008; Sun, Shang, and Li 2014; Geng, Bose, and Chen 2015; Yuan et al. 2022; Yang, Abedin, and Hajek 2023).

Timely forecasts of financial distress can offer investors early warnings about future defaults, thus mitigating losses for both companies and investors. This is especially vital in environments where all companies face the risk of financial distress. Financially distressed companies are more prone to fraudulent activities (Kolasinski and Yang 2018), which can have substantial detrimental effects on the global economy. Corporate fraud, accounting for 5% of total annual revenue loss across 125 countries, amounts to nearly $4 trillion due to corporate fraud (according to the ACFE 2018). Additionally, financial distress is confined to specific economic cycles; it has been steadily increasing in recent years, exacerbated by events like COVID-19 (Ding et al. 2023). The urgency to assess and predict financial distress risks is highlighted by reports indicating a 37% rise in critically distressed companies compared to the previous year (as of August 2022 according to the Red Flag Alert Report[1] by Begbies Traynor). As a result, there is an urgent need to assess and predict the risk (Elsayed and Elshandidy 2020).

Recent empirical evidence suggests that ensemble learning models combining diverse base prediction algorithms outperform individual algorithms (Papouskova and Hajek 2019; Abedin et al. 2023a; Hasan et al. 2024). However, existing ensemble models often assign the same weight to all base prediction algorithms, regardless of their performance (Jiang et al. 2024; Jiang and Jones 2018; Tang et al. 2020; Hajek and Munk 2023). Sun and Li (2012) proposed a combination mechanism using weighted majority voting based on base classifiers' cross-validation accuracy. However, accuracy is not reliable in financial distress prediction due to imbalanced datasets and the need to consider the costs associated with Type-I and Type-II errors. Type-II errors, failing to identify truly distressed companies, can lead to significant financial losses for unprepared stakeholders. Type-I errors, classifying healthy companies as distressed, can result in costly interventions, damage to the firm's reputation and loss of investor confidence. Balancing both types of errors is crucial for effective risk management. A balanced misjudgment ratio between Type-I and Type-II errors is a more relevant objective in corporate financial distress prediction.

Most current models are static and fail to fully use historical data on financial distress predictors (Altman et al. 2017). To address this, Sun et al. (2017) and Zhou, Si, and Fujita (2017) developed dynamic financial distress prediction approaches that consider the time-weighting of samples. The dynamic hazard model of Yousaf, Jebran, and Wang (2022), a logistic model using survival analysis data, outperformed its static counterparts, particularly for unbalanced datasets. However, these approaches overlook variations in financial data critical for prediction over different periods and the unique requirements of different classifiers regarding historical data volumes. They also fail to adequately address classifiers' tendency to become biased when overloaded with high-dimensional data. To overcome these limitations, this study proposes a multi-model weighted average ensemble model that incorporates a time-shifting component and overall misjudgment cost as the objective function during the learning process.

Predicting corporate financial distress with an ensemble weighted average model involves several key considerations. First, determining the optimal weighting of default probabilities predicted by different classifiers is crucial. Different classifiers may yield disparate results for the same company, leading to varied overall evaluations. Thus, exploring the reasonable ensemble weighting of classifiers is essential for improving financial distress prediction performance. Second, selecting the optimal prediction time window is critical. Static discriminant models, which only use data from a single year to predict default status in the future, have shown limited accuracy (Sun et al. 2020; Shen et al. 2020). Therefore, dynamic financial default prediction, which adapts prediction models to evolving financial data streams, is preferable. Additionally, dealing with imbalanced datasets, where non-defaulted enterprises outnumber defaulted ones, presents another challenge. Constructing datasets with different prediction time windows, such as a 5-year window, allows for consideration of varying financial performance and default status over time.

The misjudgment of default (Type-II error) can lead to significant losses for investors when high-risk companies are incorrectly perceived as viable investments (Geng, Bose, and Chen 2015). Conversely, misjudgment non-default (Type-I error) companies exclude low-risk firms from investment opportunities, potentially damaging their reputation and eroding investor confidence (Geng, Bose, and Chen 2015; Liang et al. 2018). Balancing the misjudgment ratio between Type-II error and Type-I error is crucial, with the former typically posing higher potential losses. Adjusting the misjudgment ratio can also help address imbalanced data problems, particularly in the context of a larger number of non-risky companies compared to risky ones.

Our paper is closely related to previous research on ensemble methods, such as Perols, Chari, and Agrawal (2009) and Ala'raj and Abbod (2016), and the optimal use of historical financial data, as explored by Du Jardin (2015). However, our approach differs in its focus on minimising misjudgment costs and exploring prediction performance over varying time windows, up to 5 years.

In summary, our paper contributes by:

- Developing a novel financial distress prediction model integrating dynamic time shifting and heterogeneous ensemble learning based on misjudgment cost.
- Demonstrating the efficiency of our approach compared to existing methods, particularly in terms of misjudgment cost and overall performance.

The structure of our paper is as follows. Section 2 reviews relevant literature on ensemble learning and dynamic financial distress prediction. Section 3 outlines the methodology for selecting financial indicators and constructing prediction models. Section 4 details the construction of our financial distress prediction model based on minimising misjudgment cost. Section 5 presents empirical findings from predicting financial distress in Chinese listed companies. Finally, Section 6 concludes our study.

2 | Related Literature

This section delves into two critical aspects of financial distress prediction, ensemble learning and dynamic financial distress prediction. Recent empirical evidence has demonstrated that ensemble learning models, which integrate various base prediction algorithms, often surpass their counterparts in financial distress prediction (Abedin et al. 2023a; Huang et al. 2023; Bouteska et al. 2024). This superiority stems from the diversity among the base models used in ensemble learning, leading to reduced model variance and enhanced prediction accuracy. However, traditional ensemble models typically assign equal weight to all base prediction algorithms, regardless of their performance. This approach can lead to suboptimal performance, particularly in financial distress prediction scenarios characterised by unbalanced datasets and varying costs associated with Type-I and Type-II errors (Liang et al. 2018). Moreover, financial distress is a dynamic phenomenon, and a model's ability to incorporate time dynamics can significantly enhance its predictive capabilities (Sun et al. 2017).

2.1 | Ensemble Learning in Financial Distress Prediction

Ensemble learning, a technique where multiple individual classifiers are combined to make a final classification decision, has gained prominence in credit risk prediction (Abelian and Castellano 2017; Garcia, Marques, and Sanchez 2019; Lu et al. 2022). This approach improves the accuracy and robustness of the classification systems by aggregating information from diverse classifiers (Lessmann et al. 2015; Traczynski 2017; Ghosh, Jana, and Abedin 2023; Hajek, Abedin, and Sivarajah 2023).

One commonly used fusion method is majority voting, where each base model predicts a class, and the class with the most votes is chosen as the final prediction (Ertimur, Ferri, and Oesch 2015). Simple averaging, another widely employed method, computes the final output by equally averaging the outputs of all ensemble members (Yang and Browne 2004). Additionally, alternative averaging methods such as those proposed by Perols, Chari, and Agrawal (2009), Hsieh and Hung (2010), Leow and Crook (2016), Figini, Savona, and Vezzoli (2016), Ala'raj and Abbod (2016) and Beque et al. (2017) have been introduced to address specific challenges in ensemble weighting and classification improvement. Specifically, Perols, Chari, and Agrawal (2009) introduced a novel averaging method based on Information Market-based Fusion (IMF) to address the issue of ensemble weighting. Leow and Crook (2016) developed a weighted average model in discrete time for the survival of repeated events, aimed at predicting the probability of default. This model assigns weights based on the likelihood of a borrower having a balance exceeding a specified limit. Figini, Savona, and Vezzoli (2016) suggested an ensemble weighting technique that relies on the area under the curve (AUC) values from various parametric and nonparametric models. Ala'raj and Abbod (2016) introduced a consensus-based method that merges the base classifier rankings within a decision profile, considering the classifier's uncertainty regarding the decision. Beque et al. (2017) initially created scorecards using diverse classifiers and then applied calibration methods to these predictions to measure the improvement in calibration.

Homogeneous and heterogeneous ensemble learning methods have both been effectively applied in predicting corporate financial distress, showing enhanced performance compared to single classifiers (Tsai 2014). Heterogeneous ensemble learning, which combines different learning algorithms, offers advantages such as capturing diverse patterns in financial data, reducing overfitting and handling imbalanced datasets more effectively. These advantages are particularly relevant in financial distress prediction, where rapid changes in economic conditions require models to generalise well to unseen data.

This study extends previous research on ensemble learning in financial distress prediction by introducing a novel ensemble-weighted averaging method aimed at minimising misjudgment costs while optimising the weights of base classifiers (see Table 1). We propose using a misjudgment cost performance measure in weighted averaging due to the data imbalance and differing financial costs associated with misclassifying defaulted and non-defaulted companies. Combining diverse base classifiers helps address overfitting issues and enhances prediction performance. Our approach also incorporates weighted majority voting based on the cost of misjudgment, providing a perspective not addressed in previous studies.

2.2 | Dynamic Financial Distress Prediction

Beneish (1999) observed that, on average, it takes 2.33 years to detect financial distress, with a median of 2.21 years. Abbasi et al. (2012) designed an ensemble learning framework for improved financial fraud detection, finding that shorter time windows between training and testing data yielded better performance. Tinoco and Wilson (2013) used data from the previous 2 years (t-1 and t-2) to predict company default status using logistic regression, achieving reliable identification of financially distressed companies approximately 1.17 years before failure.

Zhou, Tam, and Fujita (2016) addressed data imbalance by under-sampling and predicted bankruptcy by using data up until companies were delisted. Sun et al. (2017) weighted support vector machine predictions based on re-substitution error rates, assigning larger weights to newer company instances and using data from t-2 for prediction. Zhou, Si, and Fujita (2017) predicted the 'special treatment' status of Chinese listed companies using a rolling-time window method with financial data from the past 5 years (t-5 to t-1), outperforming static models.

Overall, previous research on dynamic financial distress prediction has focused on determining relevant prediction horizons or using fixed-panel data. However, this approach overlooks the dynamics of financial data needed for prediction in different periods and the varying requirements of classifiers for historical data, including their susceptibility to bias with excessively high-dimensional data. To address these limitations, this study constructs five different time window datasets from the original panel dataset.

3 | Theoretical Framework

3.1 | Feature Selection

Given the correlations and redundancies among financial indicators, the prediction performance of classification models may suffer (Kohavi and John 1997; Hajek and Michalak 2013; Zhang et al. 2022; Abedin et al. 2023b). Therefore, determining the most relevant combination of financial indicators becomes crucial for the establishment of a default prediction indicator system.

The challenge in selecting the optimal feature subset arises from the vast number of potential solutions. For

TABLE 1 | Summary of previous studies on predicting corporate financial distress using ensemble methods.

Study	Ensemble approach	Base algorithms	Fusion method
Sun and Li (2012)	Homogeneous bagging	SVM	Weighted majority vote based on accuracy and diversity
Tsai (2014)	Homogeneous and heterogeneous	*k*-Means, SOM, MLP, LR, DT	Majority vote/weighted majority vote based on accuracy
Figini, Savona, and Vezzoli (2016)	Homogeneous bagging and boosting	DT	Majority vote/weighted majority vote based on accuracy
Liang et al. (2018)	Heterogeneous	SVM, MLP, DT	Unanimous voting based on Type-II error
Chen, Chen, and Shi (2020)	Homogeneous bagging an boosting	SVM	Weighted majority vote based on accuracy
Du et al. (2020)	Heterogeneous	XGBoost, GBDT	Majority vote
Sun et al. (2020)	Homogeneous boosting	SVM	Weighted majority vote based on accuracy
Abedin et al. (2023a)	Homogeneous bagging	DT	Majority vote
Hajek and Munk (2023)	Homogeneous boosting	XGBoost	Weighted majority vote based on accuracy
This study	Heterogeneous	DT, LDA, LR, MLP, *k*-NN, SVM	Weighted majority vote based on misjudgment cost

Abbreviations: DT, decision tree; GBDT, gradient boosting decision tree; MLP, multilayer perceptron neural network; SOM, self-organising map.

instance, for n indicators, there are $2n-1$ feature combinations (Traczynski 2017; Yuan et al. 2022).

To address this challenge, this study adopts a two-stage feature selection approach recently proposed by Kou et al. (2021). Initially, correlation analysis and *F*-statistics are used for a preliminary screening of financial indicators. Subsequently, the SVM-based recursive feature elimination cross-validation (RFECV) algorithm (Yuan et al. 2022) is applied to explore all remaining indicator combinations. RFECV, using SVMs, a well-established method in financial distress prediction (Abedin et al. 2019), calculates discriminatory accuracy for different combinations of financial indicators. It selects the feature subset with the highest accuracy rate. Further details of the feature selection method can be found in Yuan et al. (2022) and Section 4.

3.2 | Optimal Weighting Scheme for the Ensemble Model

The classification outcome of the ensemble model can vary significantly based on the weights assigned to individual base classifiers (Sun and Li 2012). For instance, providing weights of (0.6, 0.3, 0.1) or (0.1, 0.3, 0.6) to three classifiers with predictions of (1, 1, 0) would yield different results, classifying respectively as default (0.9) or non-default (0.4). Thus, determining the optimal set of weights is essential to the ensemble's ability to discern the company's financial status.

Each model's weight, denoted as $0 \leq w_i \leq 1$, presents an infinite number of rational numbers between any two points on the numerical axis. Consequently, the combination of weights for multiple models is even more extensive. The challenge lies in identifying the optimal weight vector among these infinite possibilities (Shahhosseini, Hu, and Pham 2022).

The study proposes a solution to minimise the loss incurred for misclassifying defaults as non-defaults and vice versa. By optimising the set of model weight vectors, the overall misjudgment cost can be minimised.

3.3 | Time Windows

The classification model is trained to predict the default state (distressed/non-distressed) in year t using data from previous years t-1, t-2, ..., t-m. Here, in line with Geng, Bose, and Chen (2015), Sun et al. (2017) and Zhou, Si, and Fujita (2017), we consider $m=5$ years as the maximum time window. By predicting financial distress using different time windows ($m=\{1, 2, ..., 5\}$), we obtain different prediction models. However, the smaller number of defaulted companies compared to non-defaulted instances in the data for year t can lead to biased results if financial panel data are directly employed for prediction. This imbalance poses a challenge in effectively integrating unbalanced financial panel data and determining the optimal prediction time window m.

The challenge in determining the optimal prediction time window lies in constructing data for different prediction time windows so that the model has a high default discriminatory ability. As each enterprise exhibits different data and default statuses in various years, and defaulted and non-defaulted samples of different companies in the same year constitute generally unbalanced datasets, effectively combining data from different years is crucial (Sun et al. 2017).

To address this challenge, we first select companies that have defaulted historically from the panel data and use the 'earliest default year' of these companies as the base year t. This selection enables us to target the prediction of financial distress by analysing historical data where the outcome (default) is known, thereby enhancing model training and validation processes. Second, by taking the modal number of default years as the corresponding non-default year at the base year t, non-defaulted companies in the chosen non-default year form the non-defaulted sample at year t. Subsequently, the defaulted and non-defaulted samples together constitute the full sample at the base year t. Finally, data from the corresponding previous m years are extracted sequentially according to the names and years of the full sample in the benchmark year t, forming the time window t-m dataset ($m = 1, 2, ..., 5$), predicted m years in advance. This process is illustrated in Figure 1.

4 | Model Construction

The conceptual framework of this paper integrates the total cost of misjudgment in constructing an ensemble-based model for predicting corporate financial distress, as depicted in Figure 2. The framework consists of five components: (1) training and sample split; (2) feature selection; (3) prediction using individual base classifiers; (4) optimization of the ensemble construction and (5) prediction of customer default probability and default status using the weight-averaging ensemble model. These components are detailed below.

4.1 | Feature Selection

Optimising the combination of financial indicators serves two main purposes: first, it eliminates redundant indicators that lack significant predictive value while retaining those strongly associated with default status; second, it ensures the effectiveness of the selected indicator system by retaining only relevant indicators. However, balancing the removal of redundant indicators with the retention of informative ones is challenging, as informative indicators may overlap, while less informative ones may contribute to prediction accuracy when combined with others. To address this challenge, a two-step feature selection was employed in this study.

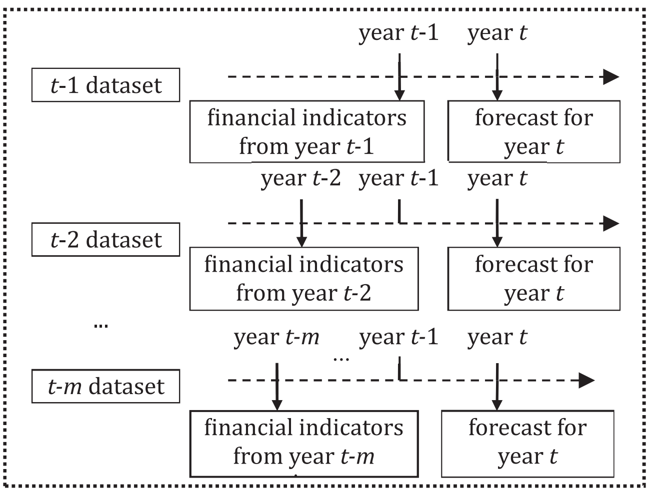

FIGURE 1 | Schematic diagram for constructing time window datasets. Forecasts for year t are generated using financial data from previous 1, 2, ..., m years. For instance, with $m = 5$, five different datasets are created. The first dataset employs only the data from the last year, while the fifth dataset incorporates data from the last 5 years. [Colour figure can be viewed at wileyonlinelibrary.com]

Initially, correlation analysis and F-statistics were used for preliminary feature selection. Indicators with correlation coefficients exceeding 0.7 were identified as highly correlated, indicating redundant information. Additionally, indicators with small F-values and weak discriminatory power regarding company default status were eliminated to prevent redundancy in the evaluation indicator system and inadvertent deletion of indicators crucial for identifying default status (Chi, Zhang, and Shi 2016).

Subsequently, the RFECV method was employed for further feature selection. This method explored all possible combinations of indicators (a total of $n(n+1)/2$), to identify the set with the highest accuracy, ensuring accurate discrimination of default status (Alkuhlani, Nassef, and Farag 2017; Farquad, Ravi, and Raju 2014). The specific steps of the RFECV method are outlined below.

Step 1: The training data is divided using a 10-fold cross-validation approach. The dataset is partitioned into 10 equal parts, with one part designated for validation and the remaining nine parts for training. This process is repeated to create 10 cross-validation datasets.

Step 2: The model prediction accuracy for the first cross-validation dataset with n indicators is calculated. This involves training a linear-SVM model using the n indicators of the training sample in the dataset and making predictions on the validation sample. The accuracy of this prediction is denoted as $Accuracy_{n,1}$.

Step 3: The model prediction accuracy for cross-validation datasets 2–10 with n indicators is calculated using the same procedure as in Step 2. The accuracy for each dataset is denoted as $Accuracy_{n,2}$, $Accuracy_{n,3}$ and so on.

Step 4: The 10-fold cross-validation accuracy for the combination of n indicators is calculated by averaging the accuracy results obtained in Steps 2 and 3. This average accuracy is represented $Accuracy_n$ as follows:

$$Accuracy_n = \sum_{j=1}^{10} Accuracy_{n,j} \qquad (1)$$

Step 5: The prediction accuracy of the combination of n-1 indicators is calculated by repeating Steps 2–4. The combination with the highest cross-validation accuracy is selected as the optimal combination of n-1 indicators.

Step 6: The prediction accuracy of the combination of n-2, n-3, ..., 1 indicator is calculated iteratively based on the optimal combination of n-1 indicators obtained in Step 5. The combination with the highest accuracy among these is chosen as the best combination of indicators for the RFECV method.

4.2 | Construction of the Ensemble Model

Construction of the ensemble model revolves around identifying an optimal set of model weights w_j ($j = 1, 2, ..., M$) to minimise

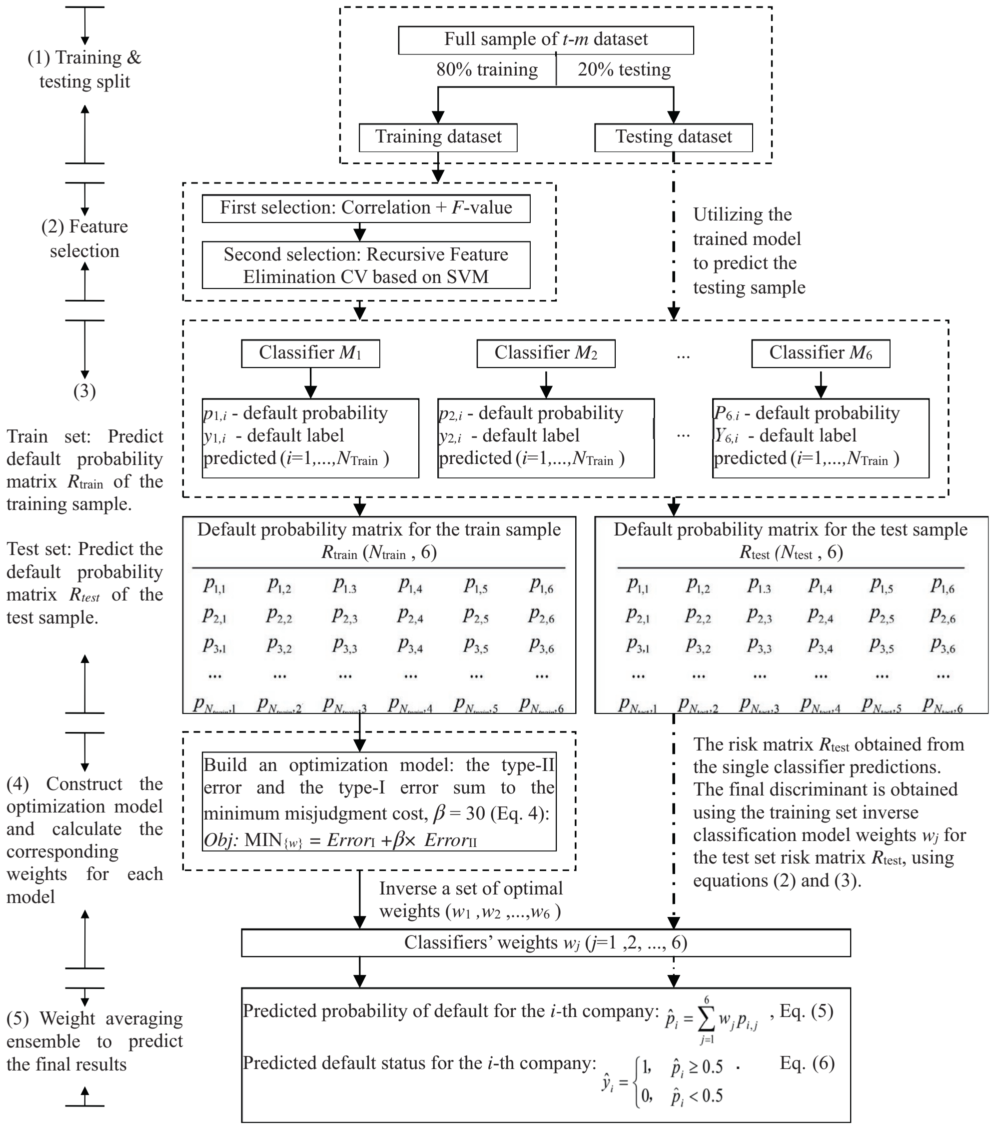

FIGURE 2 | Flowchart of the ensemble weighted averaging model based on minimizing the overall misjudgment cost.

the overall cost of misjudgment by considering the weighted sum of default probability prediction values from the base classification models.

Let N_1 represent the number of defaulted companies in the dataset, M denotes the number of classification models, w_j signify the weight corresponding to the jth classification model and $p^1{}_{i,j}$ denotes the predicted default probability obtained by the jth classification model for the ith company. The Type-II error can be expressed as follows:

$$Error_{II} = \frac{\sum_{i=1}^{N_1}\left[0.5 - \sum_{j=1}^{M} w_j p_{i,j}^1\right]_+}{N_1}. \tag{2}$$

Since the plus function is defined by $[\ x\ ]_+ = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$, when $\sum_{j=1}^{M} w_j p_{i,j}^1 \geq 0.5$, that is, the predicted value of the weighted default probability is greater than or equal to the critical value of 0.5, then $\left[0.5 - \sum_{j=1}^{M} w_j p_{i,j}^1\right]_+ = 0$, indicating that the ith default sample is correctly classified. On the contrary, $\left[0.5 - \sum_{j=1}^{M} w_j p_{i,j}^1\right]_+ = 1$ when $\sum_{j=1}^{M} w_j p_{i,j}^1 < 0.5$ indicates that the ith defaulted sample is incorrectly judged to be non-defaulted. Thus, $\sum_{i=1}^{N_1}\left[0.5 - \sum_{j=1}^{M} w_j p_{i,j}^1\right]_+$ denotes the total number of misjudged samples, and $0.5 - \sum_{i=1}^{N_1}\left[\sum_{j=1}^{M} w_j p_{i,j}^1\right]_+ / N_1$ is the ratio of misjudged samples, which denotes the Type-II error ($Error_{II}$).

Furthermore, let N_0 be the number of all non-defaulted companies and $p^0{}_{i,j}$ be the predicted value of the probability of default obtained by the jth classifier for the ith non-defaulted company. Then the Type-I error can be expressed as:

$$Error_I = \frac{\sum_{i=1}^{N_0}\left[\sum_{j=1}^{M} w_j p_{i,j}^0 - 0.5\right]_+}{N_0} \tag{3}$$

Likewise, since the plus function is defined by $[\ x\ ]_+ = \begin{cases} 1, & x \ge 0 \\ 0, & x < 0 \end{cases}$, when $\sum_{j=1}^{M} w_j p_{i,j}^0 \ge 0.5$, that is, the predicted probability of default is greater than or equal to the critical value 0.5, which means that the ith non-defaulted sample is incorrectly identified as a defaulted one ($\left[\sum_{j=1}^{M} w_j p_{i,j}^0 - 0.5\right]_+ = 1$). In contrast, when $\sum_{j=1}^{M} w_j p_{i,j}^0 < 0.5$, the ith non-defaulted sample is correctly classified ($\left[\sum_{j=1}^{M} w_j p_{i,j}^0 - 0.5\right]_+ = 0$).

Finally, let $Error_I$ represent the Type-I error of non-defaulted samples misjudged as defaulted, $Error_{II}$ denote the type-II error of defaulted samples misjudged as non-defaulted and β signify the relative cost induced by the Type II error compared with the Type-I error. By combining Equations (2) and (3), the problem of minimising the total misjudgment cost ($Error_I + \beta \times Error_{II}$) can be defined as follows:

$$\begin{aligned} \min_{\omega,k} \quad & Error_I + \beta Error_{II} \\ \text{s. t.} \quad & \sum_{j=1}^{M} w_j = 1 \\ & 0 \le w_j \le 1 \end{aligned} \tag{4}$$

The constraint $\sum_{j=1}^{M} w_j = 1$ indicates that the summation of the model weight is 1. Equation (4) carries an economic implication where the weighted sum of the predicted default probabilities from different base classification models is compared with the default discriminant threshold value of 0.5 to obtain the predicted status labels of the ensemble model. This comparison allows for the assessment of individual classification models against the threshold 0.5, effectively solving the problem of weighted fusion of the different classification models. Ultimately, the expression for calculating the probability of default for the ith company is given by:

$$\hat{p}_i = \sum_{j=1}^{M} w_j p_{i,j} \tag{5}$$

At the same time, the default status for the ith company is obtained as follows:

$$\hat{y}_i = \begin{cases} 1, & \hat{p}_i \ge 0.5 \\ 0, & \hat{p}_i < 0.5 \end{cases} \tag{6}$$

The relative cost of misjudgment, $\beta = 30$ for the Type-II error, as chosen in this paper, was adopted from previous studies (Perols et al. 2017; Shajalal, Hajek, and Abedin 2021). This indicates that misclassifying a defaulted company costs 30 times more than misclassifying a non-defaulted one.

4.3 | Performance Evaluation Criteria

When validating the classification performance of a financial distress prediction model, the choice of evaluation metrics plays a crucial role, particularly due to the class imbalance problem. Therefore, multiple different evaluation metrics should be used to assess model performance (Hilscher and Wilson 2017; Moula, Guotai, and Abedin 2017).

In this study, six evaluation criteria for the classification model were used, in addition to the previously defined cost of misjudgment. These criteria include Accuracy, Type I error, Type II error, Geometric mean (G-means), F-value ($F1$-measure) and AUC. The confusion matrix used to calculate the evaluation criteria is presented in Table 2 (He, Zhang, and Zhang 2018). Accuracy, defined as the ratio of correctly classified defaulted and non-defaulted companies, serves as a fundamental metric. G-means, a comprehensive evaluation criterion, is represented by the geometric mean of the true negative (TN) rate and the true positive (TP) rate. The F-value, another comprehensive evaluation, considers the precision and recall of the classification model. The AUC value (Ferri, Hernandez-Orallo, and Modroiu 2009) is the area under the receiver operating characteristic (ROC) curve. This criterion is the most commonly used for evaluating the discriminative power of classification models in binary classification scenarios. Importantly, the AUC considers class imbalance by representing a balance between the TP rate and the false positive (FP) rate.

TABLE 2 | Confusion matrix for financial distress prediction.

	Predicted value		
Actual value	**Default (marked as 1)**	**Non-default (marked as 0)**	**Total**
Default (marked as 1)	Number of correctly classified defaulted companies (True Positive, TP)	Number of misclassified default companies (False Negative, FN)	Number of actual defaulted companies ($TP+FN$)
Non-default (marked as 0)	Number of misclassified non-defaulted companies (False Positive, FP)	Number of correctly classified non-defaulted companies (True Negative, TN)	Number of true non-defaulted companies ($FP+TN$)
Total	Number of companies predicted by the model ($TP+FP$)	Number of non-defaulted companies predicted by the model ($FN+TN$)	Total number of companies ($TP+FP+FN+TN$)

5 | Empirical Validation

5.1 | Data

5.1.1 | Data Source

The data used in this paper comprise information on Chinese listed companies spanning the period from 2000 to 2017. It encompasses financial, non-financial and macroeconomic data. A total of 610 indicators were included in the analysis. Among these, 342 financial indicators data were sourced from the Wind Financial Database, while 119 non-financial indicators data were obtained from the CSMAR Economic and Financial Database. Additionally, 147 macroeconomic indicators (Tinoco, Holmes, and Wilson 2018; Khoja, Chipulu, and Jayasekera 2019) data were gathered from the official website of the National Bureau of Statistics of the People's Republic of China and the CNKI database of China's economic and social development statistics. Two KMV (Kealhofer, McQuown and Vasicek) model indicators (Doumpos et al. 2015), namely Default Distance (DD) and Expected Default Frequency (EDF), were calculated based on data from the Regression Specification Error Test (RESET) database.

5.1.2 | Rationale for Data Selection

The choice of Chinese listed companies as the focus of the study is justified by China's prominent position in the global market. As of May 2018, China boasted the world's second-largest stock market capitalization,[2] with its capital market size ranking among the largest globally. Moreover, investigating financial distress forecasting for Chinese listed companies holds significant value beyond the confines of the United States or European emerging markets, as highlighted in prior research (Altman et al. 2017; Li, Lou, et al. 2021).

5.1.3 | Definition of Financial Distress

Listed companies marked as 'ST (Special Treatment)' were defined as companies in financial distress, denoted as '1', following Chinese bankruptcy law. This designation aligns with previous research, which indicates that companies receive the ST status due to prolonged negative net income, insufficient equity levels or cessation of operations with no prospect of resumption (Geng, Bose, and Chen 2015). Ding, Song, and Zen (2008) further demonstrated that Chinese listed companies labelled ST exhibit a high likelihood of financial distress, predicting ST status as a valuable early warning signal for bankruptcy or default. Therefore, predicting ST status serves as an effective means of anticipating bankruptcy or default.

5.1.4 | Sample Processing

During the data preprocessing step, the raw unprocessed panel data was first normalised. Subsequently, the original standardised samples underwent a 'moment shift', as detailed in Section 3.3, resulting in time window samples projected m years in advance. The time window samples are illustrated in Table 3. Table 3 displays five sets of samples constructed for each time window, encompassing $N = 1573$ companies. Among these, there were $N_0 = 1130$ non-default samples and $N_1 = 443$ default samples. For the training sample, 80% of non-defaulted firms and 80% of defaulted companies were randomly selected from a sample of t-m time windows using random stratified sampling, yielding 1258 companies, of which 354 were defaulted. The remaining 20% of companies constituted the test sample. It is noteworthy that stratified sampling was employed concerning the target class, firm size (micro and small/medium-sized/large) and industry sector (manufacturing/IT/retail, etc.). This approach offers several advantages in predicting financial distress, enhancing the relevance of predictive models. By using stratified sampling, models can capture distinct financial characteristics and risk profiles across various industries and firm sizes, thereby potentially improving the reliability and generalizability of predictions and reducing noise arising from comparisons among fundamentally different firms.

5.2 | Results of Feature Selection

In the initial stage of feature selection, indicators with weak discriminatory power or high correlation-based redundancy were eliminated, aiming to avoid inadvertent deletion of indicators with a significant impact on default status while reducing information redundancy (Chi, Zhang, and Shi 2016). The process yielded a portfolio of 253 indicators from the original set of 610.

Subsequently, the RFECV method was employed to enhance the prediction accuracy of the feature subset. The second stage of feature selection was carried out for each of the time window samples, leading to 27, 33, 42, 44 and 70 features for the t-1, t-2, ..., t-5 time windows, respectively. Detailed results of the feature selection process are shown in Table A1.

The analysis revealed that 'basic earnings per share' and 'total agricultural production price index' are pivotal indicators influencing the financial distress of Chinese listed companies. Basic earnings per share, serving as a measure of profitability for ordinary shares, is an important financial metric indicating the profitability and growth potential of companies.

The total agricultural production price index serves as a macroeconomic indicator reflecting cost of production and the level of production prices of agricultural influencing nationwide, influencing the liquidity and financing activities of listed companies. Table B1 presents the lists of indicators selected for the remaining time windows, with the respective feature selection curves depicted in Table C1.

Figure 3 illustrates that the RFECV method for the time window identified a subset of 27 indicators with the strongest default discrimination ability from the pool of 253 indicators, achieving a prediction accuracy of 0.9905. Table 4 enumerates these 27 indicators while considering the 5 Cs credit evaluation criteria. Notably, the indicator system in Table 4 meets four out of five criteria.

5.3 | Illustration of Financial Distress Prediction

Here is an illustration of financial distress prediction using the t-1 time window data sample as an example.

TABLE 3 | Illustration of sample data for time windows t-m ($m = 1, 2, ..., 5$).

			Raw data 610 indicators				Normalised data for 610 indicators			
Sample no.	**Time window**	**Company no.**	**Gearing ratio**	**...**	**Expected frequency of default**	**Default status in year *t***	**Gearing ratio**	**...**	**Expected frequency of default**	**Default status in year *t***
1	*t*-1	Company 1	96.5177	...	0.1435	0	0.6018	...	0.0000	0
2		Company 2	67.0017	...	0.0071	0	0.7250	...	0.9506	0
...		...	...	...	...	...	...	...	...	...
1573		Company 1573	74.7637	...	0.1435	1	0.6922	...	0.0000	1
1574	*t*-2	Company 1	96.5431	...	0.1435	0	0.6017	...	0.0000	0
1575		Company 2	67.4441	...	0.0563	0	0.7227	...	0.6079	0
...		...	...	...	...	...	...	...	...	...
3146		Company 1573	68.1329	...	0.1435	1	0.7198	...	0.0000	1
3147	*t*-3	Company 1	96.3107	...	0.1435	0	0.6027	...	0.0000	0
3148		Company 2	66.1125	...	0.0506	0	0.7282	...	0.6476	0
...		...	...	...	...	...	...	...	...	...
4719		Company 1573	58.7588	...	0.1435	1	0.7588	...	0.0000	1
4720	*t*-4	Company 1	97.5153	...	0.1435	0	0.5977	...	0.0000	0
4721		Company 2	65.0369	...	0.0103	0	0.7327	...	0.9284	0
...		...	...	...	...	...	...	...	...	...
6292		Company 1573	61.4826	...	0.1435	1	0.7474	...	0.0000	1
6293	*t*-5	Company 1	97.7266	...	0.1435	0	0.5968	...	0.0000	0
6294		Company 2	60.9809	...	0.0007	0	0.7495	...	0.9948	0
...		...	...	...	...	...	...	...	...	...
7865		Company 1573	47.6506	...	0.1435	1	0.8049	...	0.0000	1

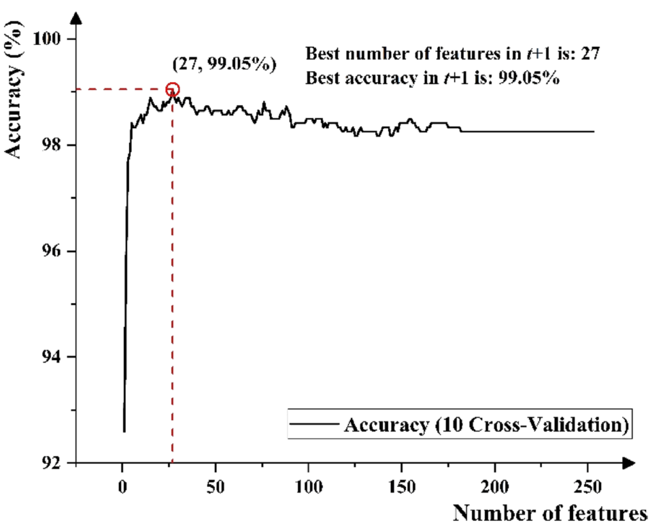

FIGURE 3 | Feature selection curve for the *t*-1 time window sample, representing a 1-year-ahead forecasting horizon. The horizontal axis displays the number of indicators in the evaluated subset, and the vertical axis indicates the accuracy results of the 10-fold cross-validation of the linear SVM model. [Colour figure can be viewed at wileyonlinelibrary.com]

Step 1: Divide the data into training and test samples in a 4:1 ratio (Ala'raj and Abbod 2016).

Step 2: Calculate the predicted default probabilities and labels for the six single classification models. The base classification models used were logistic regression (LR), linear support vector machine (SVM), linear discriminant analysis (LDA), *k*-nearest neighbour (*k*-NN), decision tree (DT) and neural network (NN). Table 5 presents the predicted default probabilities.

Step 3: Determine the optimal weight vector for the combination by solving the nonlinear programming model of Equations (2–4) for $\beta = 30$ (Perols et al. 2017). Use the genetic algorithm to determine the optimal weights w_j ($j = 1,2, ...,6$) of the model (Chen et al. 2011). The optimal model weights for the time window are listed in Table 6.

Step 4: Calculate the default probability values and labels in the proposed ensemble model. For instance, for company 1, the predicted default probability values were $(p_{11}, p_{12}, ..., p_{16}) = (0.0079, 0.0004, 0, 0, 0, 0)$ (Table 5). By combining these values with the model weights of Table 6, the predicted default value of the ensemble model was $\hat{p}_1 = 0.0020$ (Table 5). Comparing this value

TABLE 4 | List of indicators selected for the *t*-1 time window sample. The indicators are categorised in terms of 5 Cs credit criteria.

Indicator type		Indicator name	Character	Capacity	Capital	Collateral	Condition
Internal financial factors	Solvency	X_1 Gearing ratio excluding pre-receipts		√			
		X_2 Capital fixation ratio			√		
		X_3 Cash flow interest cover multiplier		√			
	Profitability	X_4 Net assets per share BPS			√		
		X_5 Retained earnings per share			√		
		X_6 Return on equity ROE		√			
		X_7 Return on assets ROA		√			
		X_8 EBITDA (TTM)/Total Assets		√			
		X_9 Net gain on change in value/total profit		√			
		X_{10} Net profit after non-recurring gains and losses		√			
		X_{11} Earnings per share EPS—net of/diluted		√			
		X_{12} Income tax		√			
		X_{13} Basic earnings per share		√			
	Operating capacity	X_{14} All assets cash recovery rate		√			
	Growth capacity	X_{15} Growth rate of net flows from operating activities		√			
Non-financial factors		X_{16} Number of earnings forecasts	√				
		X_{17} Ten largest shareholders outstanding Z index			√		
		X_{18} Type of audit opinion	√				
		X_{19} Number of executives			√		
		X_{20} Total remuneration of top three directors	√				
Senior management profile		X_{21} Chairman gender	√				
Corporate business reputation		X_{22} Are there litigation and arbitration cases	√				
External macro conditions		X_{23} GDP index: tertiary sector					√
		X_{24} Consumer price index					√
		X_{25} Commodity retail price index					√
		X_{26} Aggregate production price index					√
		X_{27} Industrial producer ex-factory price index					√
Total			5	12	5	0	5

Note: The *t*-1 time window indicator system satisfies four of the 5 Cs credit evaluation criteria, namely Character, Capacity, Capital and Condition. The reason why the zero indicator meets the 'Collateral' criterion is that the financial distress dataset did not include any collateral characteristics.

with the default discriminant threshold value of 0.5 indicates that the default label is $y_1 = 0$ (non-distressed company). The average weights in Table 6 indicate that the NN model had the highest average weight (0.20414), suggesting it was most effective in terms of classification performance, consistent with the findings of Sustersic, Mramor, and Zupan (2009).

TABLE 5 | Default prediction results of the six classification models and the proposed ensemble weighted averaging model at time window *t*-1.

Output	Classification model	Training data: Company 1	Company 2	...	Company 1258	Test data: Company 1259	Company 1260	...	Company 1573
Default probability	LR	0.0079	0.0088	...	0.0028	0.9989	0.0060	...	0.0005
	SVM	0.0004	0.0000	...	0.0000	1.0000	0.0005	...	0.0000
	LDA	0.0000	0.0000	...	0.0000	1.0000	0.0000	...	0.0000
	k-NN	0.0000	0.0000	...	0.0000	1.0000	0.0000	...	0.0000
	DT	0.0000	0.0000	...	0.0000	1.0000	0.0000	...	0.0000
	NN	0.0000	0.0000	...	0.0000	1.0000	0.0000	...	0.0000
Default probability	Ensemble weighted averaging model	0.0020	0.0022	...	0.0007	0.9997	0.0016	...	0.0001
Default label	Ensemble weighted averaging model	0	0	...	0	1	0	...	0
Actual default label		0	0	...	0	1	0	...	0

TABLE 6 | Optimal weights of the ensemble model for different time windows.

Classifier model	Model weight (*t*-1)	Model weight (*t*-2)	Model weight (*t*-3)	Model weight (*t*-4)	Model weight (*t*-5)	Average model weight
LR	**0.2486**	0.0539	0.0601	0.0951	0.0425	0.10004
SVM	0.1270	0.1678	0.1831	0.1624	0.1833	0.16472
LDA	0.1241	0.1833	0.1838	0.1859	0.1861	0.17264
k-NN	0.1567	0.1957	**0.1927**	0.1841	0.1860	0.18304
DT	0.1238	0.1941	0.1895	0.1845	0.1853	0.17544
NN	0.2198	**0.2053**	0.1907	**0.1881**	**0.2168**	**0.20414**

Note: The highest ensemble model weights are in bold.

5.4 | Comparative Analysis

A total of 11 comparative models were selected to validate the proposed ensemble weighted averaging model, including six single classification models commonly used in financial distress prediction: LR (Lessmann et al. 2015), SVM (Huang, Chen, and Wang 2007; Mselmi, Lahiani, and Hamza 2017), LDA (Desai, Crook, and Overstreet 1996), *k*-NN (Hand and Henley 1997), DT (Wang et al. 2012) and NN (Zhao et al. 2015; Chi, Abedin, and Moula 2017). Furthermore, five comparative ensemble models with the same base classifiers as the proposed ensemble weighted averaging model were used. These methods included majority voting (MV) (Ertimur, Ferri, and Oesch 2015), simple average (SA) weighting (Yang and Browne 2004), AUC weighting (Figini, Savona, and Vezzoli 2016), the IMF approach (Perols, Chari, and Agrawal 2009) and the Consensus theory-based Model (CM) (Ala'raj and Abbod 2016).

The empirical results of the compared models for different time windows are given in Tables 7–11. The analysis reveals the following insights:

1. The ensemble weighted averaging model proposed demonstrated the best overall financial distress prediction performance in terms of Accuracy, Type-II error, *G*-means, *F*-measure and AUC value across all time windows. This indicates the model's effectiveness in predicting the financial distress status of Chinese listed companies, with particularly low misclassification rates for distressed companies.
2. While the performance regarding Type I error was not consistently the best for the proposed model, the differences compared to other models were marginal, within 0.0179. However, the proposed model consistently showed a notable improvement of more than 0.02 in Type-II error

TABLE 7 | Results of financial distress prediction for time window *t*-1.

Prediction model	Accuracy [%]	Type-I error [%]	Type-II error [%]	*G*-means [%]	*F*-measure [%]	AUC [%]
LR	**98.73**	0.45	3.19	98.17	97.85	99.87
SVM	98.41	**0.00**	5.32	97.30	97.27	99.88
LDA	**98.73**	0.90	2.13	98.48	97.87	99.86
k-NN	98.10	0.90	4.26	97.41	96.77	99.37
DT	97.14	0.90	7.45	95.77	95.08	99.23
NN	98.41	1.36	2.13	98.26	97.35	99.63
Majority voting	98.41	0.90	3.19	97.95	97.33	97.95
Simple averaging	98.41	0.90	3.19	97.95	97.33	**99.93**
AUC averaging	98.41	0.90	3.19	97.95	97.33	**99.93**
IMF	96.83	4.07	**1.06**	97.42	94.90	98.17
Consensus model	**98.73**	1.36	**1.06**	**98.79**	**97.89**	98.59
Weighted averaging	**98.73**	1.36	**1.06**	**98.79**	**97.89**	**99.93**

Note: The best prediction performance is bold.

TABLE 8 | Results of financial distress prediction for time window *t*-2.

Prediction model	Accuracy [%]	Type-I error [%]	Type-II error [%]	*G*-means [%]	*F*-measure [%]	AUC [%]
LR	97.14	**0.00**	10.23	94.75	94.61	99.72
SVM	97.78	**0.00**	7.95	95.94	95.86	99.54
LDA	96.83	**0.00**	11.36	94.15	93.98	99.76
k-NN	96.19	0.44	12.50	93.34	92.77	98.08
DT	**98.10**	**0.00**	6.82	96.53	96.47	97.63
NN	97.78	**0.00**	7.95	95.94	95.86	99.78
Majority voting	97.46	**0.00**	9.09	95.35	95.24	95.45
Simple averaging	97.46	**0.00**	9.09	95.35	95.24	99.84
AUC averaging	97.46	**0.00**	9.09	95.35	95.24	99.84
IMF	96.83	3.52	**2.27**	97.10	94.51	99.23
Consensus model	**98.10**	**0.00**	6.82	96.53	96.47	99.71
Weighted averaging	**98.10**	1.76	**2.27**	**97.98**	**96.63**	**99.84**

Note: The best prediction performance is bold.

compared to other models, ensuring more balanced prediction outcomes for various forecast periods.

3. The IMF-based ensemble model (Perols, Chari, and Agrawal 2009) performed well in terms of Type-II error but exhibited relatively high values of Type-I error, ranging from 3.52% to 14.8%. Meanwhile, the consensus theory-based model (Ala'raj and Abbod 2016) ranked second for different forecasting periods, with slightly inferior performance compared to the proposed weighted averaging approach.

To determine the superiority of the proposed ensemble weighted averaging model over other models in terms of AUC prediction performance, we conducted a Chi-square test (Cleves 2002) for various time windows, with the original hypothesis assuming that the AUC values of the two models are equal. Table 12 illustrates that the proposed ensemble weighted averaging model significantly outperformed most single-classifier methods and two of the four alternative ensemble methods. Although the SA and AUC averaging methods were also surpassed, the difference was not deemed significant. In summary, it can be concluded that the proposed model, based on minimising the overall loss

TABLE 9 | Results of financial distress prediction for time window *t*-3.

Prediction model	Accuracy [%]	Type-I error [%]	Type-II error [%]	*G*-means [%]	*F*-measure [%]	AUC [%]
LR	93.97	**0.00**	20.00	89.44	88.89	95.00
SVM	94.60	**0.00**	17.89	90.61	90.17	95.22
LDA	94.92	0.45	15.79	91.56	90.91	95.12
k-NN	95.87	**0.00**	13.68	92.91	92.66	93.22
DT	92.38	**0.00**	25.26	86.45	85.54	91.47
NN	93.65	**0.00**	21.05	88.85	88.24	93.24
Majority voting	94.29	**0.00**	18.95	90.03	89.53	90.53
Simple averaging	94.29	**0.00**	18.95	90.03	89.53	**96.55**
AUC averaging	94.29	**0.00**	18.95	90.03	89.53	**96.55**
IMF	92.38	6.36	**10.53**	91.53	87.63	96.34
Consensus model	94.92	**0.00**	16.84	91.19	90.80	94.63
Weighted averaging	**96.83**	**0.00**	**10.53**	**94.59**	**94.44**	96.54

Note: The best prediction performance is bold.

TABLE 10 | Results of financial distress prediction for time window *t*-4.

Prediction model	Accuracy [%]	Type-I error [%]	Type-II error [%]	*G*-means [%]	*F*-measure [%]	AUC [%]
LR	95.24	0.43	17.07	90.87	90.07	98.13
SVM	97.14	0.43	9.76	94.79	94.27	**98.22**
LDA	96.51	0.86	10.98	93.95	92.99	97.67
k-NN	97.14	**0.00**	10.98	94.35	94.19	95.46
DT	93.65	0.86	21.95	87.97	86.49	92.95
NN	94.60	3.00	12.20	92.29	89.44	97.30
Majority voting	97.14	0.43	9.76	94.79	94.27	94.91
Simple averaging	97.46	0.43	8.54	95.43	94.94	98.03
AUC averaging	97.46	0.43	8.54	95.43	94.94	98.02
IMF	90.16	12.02	3.66	92.07	83.60	96.69
Consensus model	96.83	1.29	8.54	95.02	93.75	97.57
Weighted averaging	**97.78**	0.43	**7.32**	**96.07**	**95.60**	97.96

Note: The best prediction performance is bold.

of misjudgment loss, is superior to the compared classification models and outperforms the existing ensemble learning models.

5.5 | Robustness Tests

To further validate the proposed model, we conducted a series of robustness tests to assess potential factors that could influence our results.

Our first examination addressed concerns raised by Sayari and Mugan (2017) regarding the influence of industry-specific distributional characteristics on financial indicators. Industries have distinct financial standards and benchmarks, and operational risks vary. For instance, high leverage in the technology sector may not align with norms in utilities. By considering industry-specific indicators, the robustness check ensures predictions are sensitive to variations, improving predictive power across diverse market conditions. Additionally, industries react differently to economic cycles, impacting financial stability. Some industries, perceived as stable, have better access to finance, affecting their resilience. Adjusting distress prediction models for economic sensitivities enhances their efficacy.

TABLE 11 | Results of financial distress prediction for time window *t*-5.

Prediction model	Accuracy [%]	Type-I error [%]	Type-II error [%]	*G*-means [%]	*F*-measure [%]	AUC [%]
LR	95.24	**0.00**	16.30	91.49	91.12	94.77
SVM	94.92	**0.00**	17.39	90.89	90.48	95.35
LDA	94.60	**0.00**	18.48	90.29	89.82	96.26
k-NN	96.51	1.35	8.70	94.91	93.85	95.56
DT	93.97	0.45	19.57	89.48	88.62	93.38
NN	94.92	1.35	14.13	92.04	90.80	94.67
Majority voting	95.56	**0.00**	15.22	92.08	91.76	92.39
Simple averaging	95.56	**0.00**	15.22	92.08	91.76	97.37
AUC averaging	95.56	**0.00**	15.22	92.08	91.76	97.37
IMF	87.62	14.80	6.52	89.24	81.52	96.62
Consensus model	96.51	0.45	10.87	94.20	93.71	96.23
Weighted averaging	**96.83**	1.79	**6.52**	**95.81**	**94.51**	**97.45**

Note: The best prediction performance is bold.

TABLE 12 | Chi-square test for differences in AUC values between the ensemble weighted averaging model and comparative models.

Prediction model	*t*-1	*t*-2	*t*-3	*t*-4	*t*-5
LR	2.30	0.58	2.44	0.23	4.61**
SVM	1.26	1.38	1.52	2.44	2.75*
LDA	1.61	0.86	1.58	0.71	1.37
k-NN	1.15	3.16**	6.47**	3.71*	2.96*
DT	1.79	4.13***	9.46***	7.84***	7.20***
NN	0.90	0.65	6.42**	0.91	4.45**
Majority voting	4.59*	8.68***	12.63***	5.00**	9.71***
Simple averaging	0.00	0.00	0.01	1.46	0.59
AUC averaging	0.00	0.00	0.01	1.08	0.59
IMF	8.29***	3.05**	0.64	6.15**	4.63**
Consensus model	1.63	1.75	1.35	0.16	1.43

Note: ***/**/* denotes significant differences at the confidence level of 99%/95%/90%, respectively.

Our second check, prompted by Gupta, Gregoriou, and Healy (2015), examined the impact of firm size. Large, medium and small firms exhibit varying risk profiles affecting vulnerability to financial distress. Understanding this aids policymakers in designing targeted interventions for different-size firms, bolstering economic resilience.

Figures 4 and 5 illustrate that the accuracy of the proposed ensemble weighted averaging model remains stable across industry sectors and firm size categories. There is a slight decrease in performance for the mining industry, particularly in the *t*-4 and *t*-5 forecasting periods. These results align with the industry's specific asset structure (Hájek 2012) and high cyclicality, sensitive to global economic conditions, and geopolitical events, making accurate prediction distress challenging.

5.6 | Analysis of Credit Score

In this section, we use the predictions from the model to derive default probabilities and credit scores, aiming to explore the credit characteristics of Chinese listed companies across geographical location, corporate ownership and firm size.

The credit score of each company (indexed as *i*) can be calculated from its default probability p_i using the formula: $Score_i = (1 - p_i) \times 100$. That transformation converts default probabilities into credit scores within the interval [0,100].

Geographical location, corporate ownership and firm size are chosen as key factors for analysing the credit score of Chinese listed companies due to their significant impact on financial health and risk profiles (McCann and McIndoe-Calder 2015; Fernandes and Artes 2016; Liu et al. 2018; Che et al. 2024). Geographical location influences market access, regulatory environments and resource availability. Firms in economically developed regions may benefit from better infrastructure, favourable policies and easier markets and capital access, affecting their financial stability and credit risk. Corporate ownership structure plays a crucial role in governance, strategic decisions and access to resources. State-owned enterprises may exhibit different risk profiles and financial health compared with privately owned counterparts due to government support and objectives. Additionally, ownership can shape a firm's financial policies and risk-taking behaviour, thereby impacting its credit score. Furthermore, the size of a company

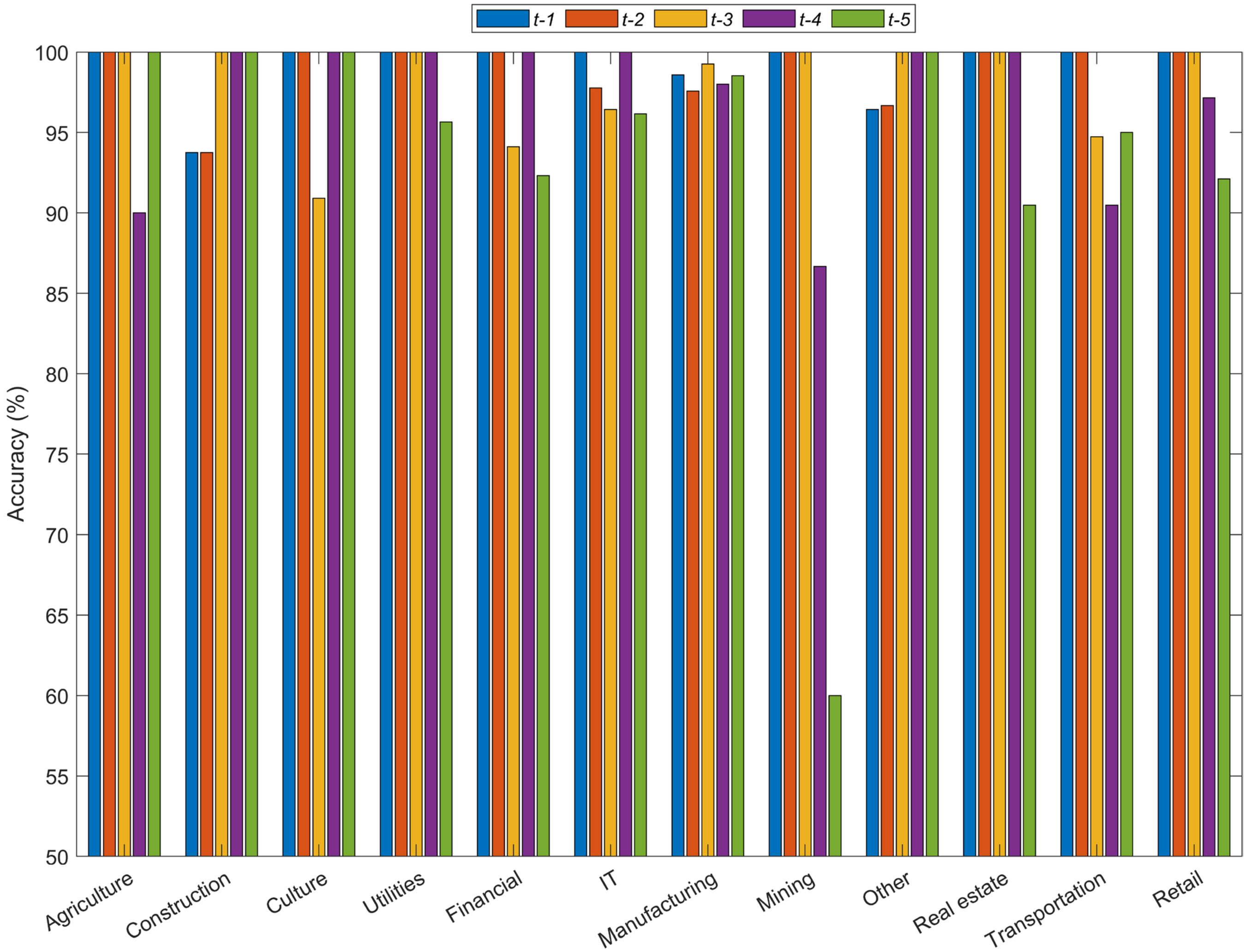

FIGURE 4 | Accuracy of the ensemble weighted averaging model across industries. [Colour figure can be viewed at wileyonlinelibrary.com]

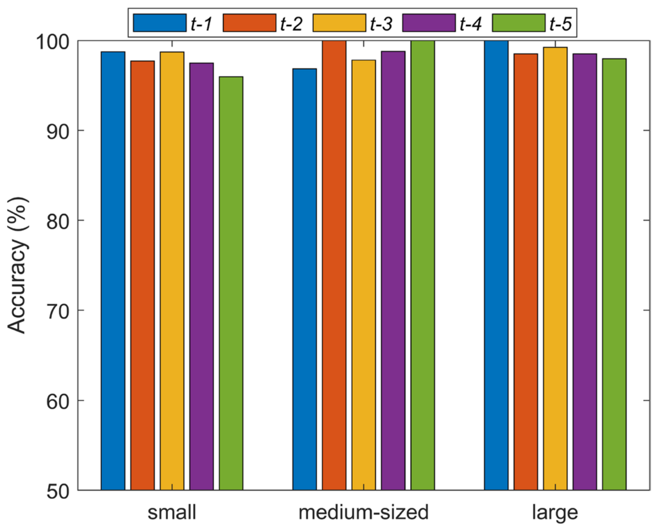

FIGURE 5 | Accuracy of the ensemble weighted averaging model for different firm sizes. [Colour figure can be viewed at wileyonlinelibrary.com]

was chosen for analysis larger firms often engage in more diversified business activities, enjoy better access to capital markets and wield greater bargaining power with suppliers and creditors. Consequently, larger companies typically face lower financial distress risk compared to smaller counterparts.

5.6.1 | Effect of Geographical Area

The spatial distribution of geographical regions significantly impacts corporate credit risk (Fernandes and Artes 2016). To assess this effect, Chinese companies were categorised into three regions: eastern, central and western. Table 13 presents the corresponding credit scores for each geographical category. To determine if there are significant differences in credit scores between these regions, a Student's t-test for independent samples was conducted. The highest credit scores were observed in eastern provinces, including Shanghai, Fujian and Zhejiang, followed by central regions such as Anhui, Hunan and Henan. Conversely, companies in western regions such as Xinjiang, Guangxi and Guizhou had the lowest credit scores. Significant differences ($p < 0.01$) in credit quality were noted among the three geographical regions—eastern, western and central China. This trend may be attributed to better access to financing and investment opportunities for companies in eastern coastal regions compared with those in central and western inland areas.

TABLE 13 | Descriptive statistics of credit scores for different geographical areas.

Geographical area	Average credit score	Max	Min	Median	St. dev.	# Observations
Eastern regions	96.63	99.98	0.59	99.90	13.47	27,315
Central regions	94.36	99.98	0.61	99.84	17.57	6940
Western regions	93.05	99.98	0.63	99.78	19.30	6058

TABLE 14 | Descriptive statistics of credit scores for different types of corporate ownership.

Ownership	Average credit score	Max	Min	Median	St. dev.	# Observations
Collective-owned	96.61	99.98	5.85	99.87	12.28	319
Central state-owned	96.39	99.98	0.71	99.82	12.47	5269
Private-owned	96.07	99.98	0.62	99.91	15.16	20,885
Public-owned	95.44	99.98	0.62	99.89	15.48	2058
Local state-owned	94.92	99.98	0.61	99.80	16.26	10,037
Foreign-owned	94.55	99.98	0.59	99.89	17.46	1221
Other companies	92.12	99.98	0.87	99.78	20.86	524

TABLE 15 | Descriptive statistics of credit scores for different firm sizes.

Firm size	Average credit score	Max	Min	Median	St. dev.	# Observations
Large	97.37	99.98	0.62	99.88	10.95	25,585
Medium	94.83	99.98	0.61	99.90	17.16	4832
Small	92.27	99.98	0.59	99.84	21.05	9811
Micro	40.26	99.90	0.78	18.57	40.82	85

5.6.2 | Effect of Corporate Ownership

The analysis of credit scores for companies with different ownership structures is a significant area of study, with Liu et al. (2018) suggesting that Chinese state-owned enterprises (SOEs) may possess competitive advantages, including higher average rates of return attributed to political-level ties. This section examines seven ownership categories based on major shareholders and effective controllers: central SOEs, local SOEs, private companies, collective companies, public companies, foreign companies and others effectively controlled by associations.

Empirical findings in Table 14 indicate that corporate ownership significantly impacts the credit scores of Chinese listed companies. Collective companies, central state-owned companies and private companies exhibit the highest average scores, followed by public companies and local state-owned companies, while foreign and other companies have the lowest scores. Student's paired *t*-tests were conducted to statistically validate these results, revealing that collective, central state and private-owned companies have significantly higher credit scores than local state, foreign and other companies at $p < 0.05$.

5.6.3 | Effect of Firm Size

Firm size is recognised as another crucial determinant of a company's creditworthiness (McCann and McIndoe-Calder 2015). Companies were categorised into large, medium, small and micro enterprises based on their total assets and number of employees.

Results in Table 15 indicate that the creditworthiness of Chinese listed companies increases with company size. Significant differences were observed in the creditworthiness of companies of different sizes at $p < 0.01$, as determined by Student's paired *t*-test. The positive correlation between company creditworthiness and size may stem from larger companies being more profitable and resilient to risk.

6 | Conclusions

In this study, we proposed a weighted average ensemble model based on minimising the cost of misjudgments. Our empirical findings demonstrate that misjudgment cost in constructing a heterogeneous ensemble model can significantly enhance financial distress prediction performance. This improvement is attributed to the reduction in the number of attributes achieved through a two-stage feature selection approach, which aligns with previous research indicating the benefits of feature subset reduction in prediction models (Kou et al. 2021). This procedure enhanced prediction performance by preventing overfitting, indicating the need to identify and eliminate irrelevant or redundant features before employing the ensemble model. Moreover, this feature reduction can lead to significant financial benefits for banks and creditors by reducing data acquisition costs, a benefit also emphasised in prior similar studies (Papouskova and Hajek 2019).

Additionally, our study highlights the importance of achieving a balanced prediction performance in terms of Type-I and Type-II errors, allowing for high prediction accuracy up to 5 years before financial distress. This balanced performance is particularly valuable for stakeholders such as banks, regulators and investors, enabling early identification of financial difficulties and proactive risk mitigation measures. Indeed, previous research in this field has faced challenges in achieving a balance between Type-I and Type-II errors when predicting financial distress (Mselmi, Lahiani, and Hamza 2017; Liang et al. 2018; Sun et al. 2020).

Consistent with the findings of Altman et al. (2017), our model demonstrates stable prediction accuracy across extended forecasting horizons, outperforming traditional models like Altman's *Z*-score model which often exhibit declining performance over longer horizons (Altman et al. 2017). Notably, our model maintains strong performance for both financially distressed and healthy firms, although the Type-II error increases as the prediction horizon lengthens.

Furthermore, our empirical analysis of Chinese listed companies confirms the significance of geographical area, ownership and firm size in credit scoring, aligning with prior research on firm-level and institutional factors affecting financial distress among Chinese firms (Bhattacharjee and Han 2014; Li, Crook, et al. 2021). This suggests potential policy implications for targeted government support and interventions based on regional, ownership and company size characteristics.

Our study underscores several implications for shareholders, policymakers and other stakeholders involved in corporate financial distress prediction. The early forecasting capability of our model allows for timely risk management decisions, reducing the potential financial losses associated with incorrect predictions. Moreover, by incorporating a cost function that weights misjudgment costs, the model prioritises minimising overall mispricing costs, benefiting policymakers and investors alike. This offers significant benefits to policymakers and investors by minimising potential financial losses due to inaccurate predictions about a company's financial health. For stakeholders like shareholders, the model serves as an effective risk management tool. By precisely predicting financial distress, shareholders are empowered to make well-informed decisions regarding their investments, such as divesting from companies with high risk or advocating for strategic changes to reduce risk. Furthermore, the model enhances corporate governance by enabling management to be alerted to potential financial troubles in time for appropriate corrective measures. The proposed model is designed for customization to accommodate varying cost sensitivities related to Type-I and Type-II errors. Its adaptability ensures its applicability across different industries and economic environments, enabling policymakers and stakeholders to adjust the model to meet their specific risk management requirements.

However, our study has limitations. It uses a fixed parameter for misjudgment cost, which could be made more flexible in future research to reflect changes in company characteristics. Future studies could also explore alternative profit-driven performance metrics (Gunnarsson et al. 2021) and optimise the selection of base classifiers to further enhance prediction performance. The ensemble model could incorporate a mechanism to update predictions with the latest financial data and indicators, enabling real-time adjustments to the firm's assessed status as new information emerges. Additionally, the treatment of ST firms warrants consideration, as they may recover or fail, requiring dynamic monitoring and nuanced forecasting approaches. In particular, the model could be tailored to distinguish between ST firms using specific criteria, such as the nature of their financial distress, existing recovery plans and any external financial support received. Furthermore, while our model integrates a wide range of indicators, it may not explicitly capture the effects of individual governance decisions. The complexity involved in capturing the effects of governance and individual decisions often necessitates a qualitative analysis of corporate governance, as discussed in Yousaf, Jebran, and Wang (2021); Yousaf, Jebran, and Ullah (2024), or the incorporation of additional data sources. These could include textual analysis of corporate disclosures or executive communications, as explored in Elsayed and Elshandidy (2020), Tang et al. (2020), and Huang et al. (2023). However, such approaches may fall outside the purview of this study. Future research could explore qualitative analysis of governance factors or additional data sources to enhance model comprehensiveness. Additionally, investigating the effects of industry sector, management quality and competitive positioning could provide valuable insights into firms' resilience to financial difficulties in varying regulatory and market environments.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Endnotes

[1] The Red Flag Alert (https://www.londonstockexchange.com/news-article/BEG/latest-red-flag-alert-report-for-q2-2022/15572417).

[2] https://www.ft.com/content/3ea51148-632f-11e8-a39d-4df188287fff.

References

Abbasi, A., C. Albrecht, A. Vance, and J. Hansen. 2012. "MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud." *MIS Quarterly* 36, no. 4: 1293–1327. https://doi.org/10.1002/asi.22741.

Abedin, M. Z., C. Guotai, P. Hajek, and T. Zhang. 2023a. "Combining Weighted SMOTE With Ensemble Learning for the Class-Imbalanced Prediction of Small Business Credit Risk." *Complex & Intelligent Systems* 9, no. 4: 3559–3579. https://doi.org/10.1007/s40747-021-00614-4.

Abedin, M. Z., C. Guotai, F. E. Moula, A. Azad, and M. Khan. 2019. "Topological Applications of Multilayer Perceptrons and Support Vector Machines in Financial Decision Support Systems." *International Journal of Finance and Economics* 24, no. 1: 474–507. https://doi.org/10.1002/ijfe.1675.

Abedin, M. Z., P. Hajek, T. Sharif, M. S. Satu, and M. I. Khan. 2023b. "Modelling Bank Customer Behaviour Using Feature Engineering and Classification Techniques." *Research in International Business and Finance* 65: 101913. https://doi.org/10.1016/j.ribaf.2023.101913.

Abelian, J., and J. G. Castellano. 2017. "A Comparative Study on Base Classifiers in Ensemble Methods for Credit Scoring." *Expert Systems With Applications* 73: 1–10. https://doi.org/10.1016/j.eswa.2016.12.020.

ACFE. 2018. "Report to the Nations 2018 Global Study on Occupational Fraud and Abuse." *Association of Certified Fraud Examiners*: 7–8.

Ala'raj, M., and M. F. Abbod. 2016. "A New Hybrid Ensemble Credit Scoring Model Based on Classifiers Consensus System Approach." *Expert Systems With Applications* 64: 36–55. https://doi.org/10.1016/j.eswa.2016.07.017.

Alkuhlani, A., M. Nassef, and I. Farag. 2017. "Multistage Feature Selection Approach for High-Dimensional Cancer Data." *Soft Computing* 21, no. 22: 6895–6906. https://doi.org/10.1007/s00500-016-2439-9.

Altman, E. I., M. Iwanicz-Drozdowska, E. K. Laitinen, and A. Suvas. 2017. "Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model." *Journal of International Financial Management & Accounting* 28, no. 2: 131–171. https://doi.org/10.1111/jifm.12053.

Beneish, M. D. 1999. "Incentives and Penalties Related to Earnings Overstatements That Violate GAAP." *Accounting Review* 74, no. 4: 425–457. https://doi.org/10.2308/accr.1999.74.4.425.

Beque, A., K. Coussement, R. Gayler, and S. Lessmann. 2017. "Approaches for Credit Scorecard Calibration: An Empirical Analysis." *Knowledge-Based Systems* 134: 213–227. https://doi.org/10.1016/j.knosys.2017.07.034.

Bhattacharjee, A., and J. Han. 2014. "Financial Distress of Chinese Firms: Microeconomic, Macroeconomic and Institutional Influences." *China Economic Review* 30: 244–262. https://doi.org/10.1016/j.chieco.2014.07.007.

Bouteska, A., M. Z. Abedin, P. Hajek, and K. Yuan. 2024. "Cryptocurrency Price Forecasting—A Comparative Analysis of Ensemble Learning and Deep Learning Methods." *International Review of Financial Analysis* 92: 103055. https://doi.org/10.1016/j.irfa.2023.103055.

Che, W., Z. Wang, C. Jiang, and M. Z. Abedin. 2024. "Predicting Financial Distress Using Multimodal Data: An Attentive and Regularized Deep Learning Method." *Information Processing & Management* 61, no. 4: 103703. https://doi.org/10.1016/j.ipm.2024.103703.

Chen, N., B. Ribeiro, A. S. Vieira, J. Duarte, and J. C. Neves. 2011. "A Genetic Algorithm-Based Approach to Cost-Sensitive Bankruptcy Prediction." *Expert Systems With Applications* 38, no. 10: 12939–12945. https://doi.org/10.1016/j.eswa.2011.04.090.

Chen, Z., W. Chen, and Y. Shi. 2020. "Ensemble Learning With Label Proportions for Bankruptcy Prediction." *Expert Systems With Applications* 146: 113155. https://doi.org/10.1016/j.eswa.2019.113155.

Chi, G., M. Z. Abedin, and F. E. Moula. 2017. "Modeling Credit Approval Data With Neural Networks: An Experimental Investigation and Optimization." *Journal of Business Economics and Management* 18, no. 2: 224–240. https://doi.org/10.3846/16111699.2017.1280844.

Chi, G., Y. Zhang, and B. Shi. 2016. "The Debt Rating for Small Enterprises Based on Probit Regression." *Journal of Management Sciences in China* 19, no. 6: 136–156. (In Chinese).

Cleves, M. A. 2002. "From the Help Desk: Comparing Areas Under Receiver Operating Characteristic Curves From Two or More Probit or Logit Models." *Stata Journal* 2, no. 3: 301–313. https://doi.org/10.1177/1536867X020020030.

Desai, V. S., J. N. Crook, and G. A. Overstreet. 1996. "A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment." *European Journal of Operational Research* 95, no. 1: 24–37. https://doi.org/10.1016/0377-2217(95)00246-4.

Ding, S., T. Cui, A. G. Bellotti, N. Z. Abedin, and B. Lucey. 2023. "The Role of Feature Importance in Predicting Corporate Financial Distress in Pre and Post COVID Periods: Evidence From China." *International Review of Financial Analysis* 90: 102851. https://doi.org/10.1016/j.irfa.2023.102851.

Ding, Y., X. Song, and Y. Zen. 2008. "Forecasting Financial Condition of Chinese Listed Companies Based on Support Vector Machine." *Expert Systems With Applications* 34, no. 4: 3081–3089. https://doi.org/10.1016/j.eswa.2007.06.037.

Doumpos, M., D. Niklis, C. Zopounidis, and K. Andriosopoulos. 2015. "Combining Accounting Data and a Structural Model for Predicting Credit Ratings: Empirical Evidence From European Listed Firms." *Journal of Banking and Finance* 50: 599–607. https://doi.org/10.1016/j.jbankfin.2014.01.010.

Du Jardin, P. 2015. "Bankruptcy Prediction Using Terminal Failure Processes." *European Journal of Operational Research* 242, no. 1: 286–303. https://doi.org/10.1016/j.ejor.2014.09.059.

Du, X., W. Li, S. Ruan, and L. Li. 2020. "CUS-Heterogeneous Ensemble-Based Financial Distress Prediction for Imbalanced Dataset With Ensemble Feature Selection." *Applied Soft Computing* 97: 106758. https://doi.org/10.1016/j.asoc.2020.106758.

Elsayed, M., and T. Elshandidy. 2020. "Do Narrative-Related Disclosures Predict Corporate Failure? Evidence From UK Non-financial Publicly Quoted Firms." *International Review of Financial Analysis* 71: 101555.

Ertimur, Y., F. Ferri, and D. Oesch. 2015. "Does the Director Election System Matter? Evidence From Majority Voting." *Review of Accounting Studies* 20, no. 1: 1–41. https://doi.org/10.1007/s11142-014-9284-9.

Farquad, M. A. H., V. Ravi, and S. B. Raju. 2014. "Churn Prediction Using Comprehensible Support Vector Machine: An Analytical CRM Application." *Applied Soft Computing* 19: 31–40. https://doi.org/10.1016/j.asoc.2014.01.031.

Fernandes, G. B., and R. Artes. 2016. "Spatial Dependence in Credit Risk and Its Improvement in Credit Scoring." *European Journal of Operational Research* 249, no. 2: 517–524. https://doi.org/10.1016/j.ejor.2015.07.013.

Ferri, C., J. Hernandez-Orallo, and R. Modroiu. 2009. "An Experimental Comparison of Performance Measures for Classification." *Pattern Recognition Letters* 30, no. 1: 27–38. https://doi.org/10.1016/j.patrec.2008.08.010.

Figini, S., R. Savona, and M. Vezzoli. 2016. "Corporate Default Prediction Model Averaging: A Normative Linear Pooling Approach." *Intelligent Systems in Accounting, Finance and Management* 23, no. 1–2: 6–20. https://doi.org/10.1002/isaf.1387.

Garcia, V., A. I. Marques, and J. S. Sanchez. 2019. "Exploring the Synergetic Effects of Sample Types on the Performance of Ensembles for Credit Risk and Corporate Bankruptcy Prediction." *Information Fusion* 47: 88–101. https://doi.org/10.1016/j.infus.2018.07.004.

Geng, R., I. Bose, and X. Chen. 2015. "Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining." *European Journal of Operational Research* 241, no. 1: 236–247. https://doi.org/10.1016/j.ejor.2014.08.016.

Ghosh, I., R. K. Jana, and M. Z. Abedin. 2023. "An Ensemble Machine Learning Framework for Airbnb Rental Price Modeling Without Using Amenity-Driven Features." *International Journal of Contemporary Hospitality Management* 35, no. 10: 3592–3611. https://doi.org/10.1108/IJCHM-05-2022-0562.

Gunnarsson, B. R., S. Vanden Broucke, B. Baesens, M. Óskarsdóttir, and W. Lemahieu. 2021. "Deep Learning for Credit Scoring: Do or Don't?" *European Journal of Operational Research* 295, no. 1: 292–305. https://doi.org/10.1016/j.ejor.2021.03.006.

Gupta, J., A. Gregoriou, and J. Healy. 2015. "Forecasting Bankruptcy for SMEs Using Hazard Function: To What Extent Does Size Matter?" *Review of Quantitative Finance and Accounting* 45: 845–869. https://doi.org/10.1007/s11156-014-0458-0.

Hájek, P. 2012. "Credit Rating Analysis Using Adaptive Fuzzy Rule-Based Systems: An Industry-Specific Approach." *Central European Journal of Operations Research* 20: 421–434. https://doi.org/10.1007/s10100-011-0229-0.

Hajek, P., M. Z. Abedin, and U. Sivarajah. 2023. "Fraud Detection in Mobile Payment Systems Using an XGBoost-Based Framework." *Information Systems Frontiers* 25: 1985–2003. https://doi.org/10.1007/s10796-022-10346-6.

Hajek, P., and K. Michalak. 2013. "Feature Selection in Corporate Credit Rating Prediction." *Knowledge-Based Systems* 51: 72–84. https://doi.org/10.1016/j.knosys.2013.07.008.

Hajek, P., and M. Munk. 2023. "Speech Emotion Recognition and Text Sentiment Analysis for Financial Distress Prediction." *Neural Computing and Applications* 35, no. 29: 21463–21477. https://doi.org/10.1007/s00521-023-08470-8.

Hand, D. J., and W. E. Henley. 1997. "Statistical Classification Methods in Consumer Credit Scoring: A Review." *Journal of the Royal Statistical Society Series A: Statistics in Society* 160: 523–541.

Hasan, M., M. Z. Abedin, P. Hajek, K. Coussement, M. N. Sultan, and B. Lucey. 2024. "A Blending Ensemble Learning Model for Crude Oil Price Forecasting." *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05810-8.

He, H., W. Zhang, and S. Zhang. 2018. "A Novel Ensemble Method for Credit Scoring: Adaption of Different Imbalance Ratios." *Expert Systems With Applications* 98: 105–117. https://doi.org/10.1016/j.eswa.2018.01.012.

Hilscher, J., and M. Wilson. 2017. "Credit Ratings and Credit Risk: Is One Measure Enough?" *Management Science* 63, no. 10: 3414–3437. https://doi.org/10.1287/mnsc.2016.2514.

Hsieh, N.-C., and L.-P. Hung. 2010. "A Data Driven Ensemble Classifier for Credit Scoring Analysis." *Expert Systems With Applications* 37, no. 1: 534–545. https://doi.org/10.1016/j.eswa.2009.05.059.

Huang, B., X. Yao, Y. Luo, and J. Li. 2023. "Improving Financial Distress Prediction Using Textual Sentiment of Annual Reports." *Annals of Operations Research* 330, no. 1: 457–484. https://doi.org/10.1007/s10479-022-04633-3.

Huang, C.-L., M.-C. Chen, and C.-J. Wang. 2007. "Credit Scoring With a Data Mining Approach Based on Support Vector Machines." *Expert Systems With Applications* 33, no. 4: 847–856. https://doi.org/10.1016/j.eswa.2006.07.007eswa.2006.07.007.

Jiang, P., Z. Liu, M. Z. Abedin, J. Wang, W. Yang, and Q. Dong. 2024. "Profit-Driven Weighted Classifier With Interpretable Ability for Customer Churn Prediction." *Omega* 125: 103034. https://doi.org/10.1016/j.omega.2024.103034.

Jiang, Y., and S. Jones. 2018. "Corporate Distress Prediction in China: A Machine Learning Approach." *Accounting and Finance* 58, no. 4: 1063–1109. https://doi.org/10.1111/acfi.12432.

Khoja, L., M. Chipulu, and R. Jayasekera. 2019. "Analysis of Financial Distress Cross Countries: Using Macroeconomic, Industrial Indicators and Accounting Data." *International Review of Financial Analysis* 66: 101379.

Kohavi, R., and G. H. John. 1997. "Wrappers for Feature Subset Selection." *Artificial Intelligence* 97, no. 1–2: 273–324. https://doi.org/10.1016/s0004-3702(97)00043-x.

Kolasinski, A. C., and N. Yang. 2018. "Managerial Myopia and the Mortgage Meltdown." *Journal of Financial Economics* 128, no. 3: 466–485. https://doi.org/10.1016/j.jfineco.2017.03.010.

Kou, G., Y. Xu, Y. Peng, et al. 2021. "Bankruptcy Prediction for SMEs Using Transactional Data and Two-Stage Multiobjective Feature Selection." *Decision Support Systems* 140: 113429. https://doi.org/10.1016/j.dss.2020.113429.

Leow, M., and J. Crook. 2016. "A New Mixture Model for the Estimation of Credit Card Exposure at Default." *European Journal of Operational Research* 249, no. 2: 487–497. https://doi.org/10.1016/j.ejor.2015.10.001.

Lessmann, S., B. Baesens, H.-V. Seow, and L. C. Thomas. 2015. "Benchmarking State-Of-The-Art Classification Algorithms for Credit Scoring: An Update of Research." *European Journal of Operational Research* 247, no. 1: 124–136. https://doi.org/10.1016/j.ejor.2015.05.030.

Li, C., C. Lou, D. Luo, and K. Xing. 2021. "Chinese Corporate Distress Prediction Using LASSO: The Role of Earnings Management." *International Review of Financial Analysis* 76: 101776.

Li, Z., J. Crook, G. Andreeva, and Y. Tang. 2021. "Predicting the Risk of Financial Distress Using Corporate Governance Measures." *Pacific-Basin Finance Journal* 68: 101334. https://doi.org/10.1016/j.pacfin.2020.101334.

Liang, D., C. F. Tsai, A. J. Dai, and W. Eberle. 2018. "A Novel Classifier Ensemble Approach for Financial Distress Prediction." *Knowledge and Information Systems* 54: 437–462. https://doi.org/10.1007/s10115-017-1061-1.

Liu, L., Q. G. Liu, G. Tian, and P. P. Wang. 2018. "Government Connections and the Persistence of Profitability: Evidence From Chinese Listed Firms." *Emerging Markets Review* 36: 110–129. https://doi.org/10.1016/j.ememar.2018.04.002.

Lu, Y., L. Yang, B. Shi, J. Li, and M. Z. Abedin. 2022. "A Novel Framework of Credit Risk Feature Selection for SMEs During Industry 4.0." *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-04849-3.

McCann, F., and T. McIndoe-Calder. 2015. "Firm Size, Credit Scoring Accuracy and banks' Production of Soft Information." *Applied Economics* 47, no. 33: 3594–3611. https://doi.org/10.1080/00036846.2015.1019034.

Moula, F. E., C. Guotai, and M. Z. Abedin. 2017. "Credit Default Prediction Modeling: An Application of Support Vector Machine." *Risk Management* 19: 158–187. https://doi.org/10.1057/s41283-017-0016-x.

Mselmi, N., A. Lahiani, and T. Hamza. 2017. "Financial Distress Prediction: The Case of French Small and Medium-Sized Firms." *International Review of Financial Analysis* 50: 67–80.

Papouskova, M., and P. Hajek. 2019. "Two-Stage Consumer Credit Risk Modelling Using Heterogeneous Ensemble Learning." *Decision Support Systems* 118: 33–45. https://doi.org/10.1016/j.dss.2019.01.002.

Perols, J., K. Chari, and M. Agrawal. 2009. "Information Market-Based Decision Fusion." *Management Science* 55, no. 5: 827–842. https://doi.org/10.1287/mnsc.1080.0977.

Perols, J. L., R. M. Bowen, C. Zimmermann, and B. Samba. 2017. "Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction." *Accounting Review* 92, no. 2: 221–245. https://doi.org/10.2308/accr-51562.

Purnanandam, A. 2008. "Financial Distress and Corporate Risk Management: Theory and Evidence." *Journal of Financial Economics* 87, no. 3: 706–739. https://doi.org/10.1016/j.jfineco.2007.04.003.

Sayari, N., and C. S. Mugan. 2017. "Industry Specific Financial Distress Modeling." *BRQ Business Research Quarterly* 20, no. 1: 45–62. https://doi.org/10.1016/j.brq.2016.03.003.

Shahhosseini, M., G. Hu, and H. Pham. 2022. "Optimizing Ensemble Weights and Hyperparameters of Machine Learning Models for

Regression Problems." *Machine Learning With Applications* 7: 100251. https://doi.org/10.1016/j.mlwa.2022.100251.

Shajalal, M., P. Hajek, and M. Z. Abedin. 2021. "Product Backorder Prediction Using Deep Neural Network on Imbalanced Data." *International Journal of Production Research* 12: 1–18. https://doi.org/10.1080/00207543.2021.1901153.

Shen, F., Y. Liu, R. Wang, and W. Zhou. 2020. "A Dynamic Financial Distress Forecast Model With Multiple Forecast Results Under Unbalanced Data Environment." *Knowledge-Based Systems* 192: 105365. https://doi.org/10.1016/j.knosys.2019.105365.

Sun, J., H. Fujita, P. Chen, and H. Li. 2017. "Dynamic Financial Distress Prediction With Concept Drift Based on Time Weighting Combined With Adaboost Support Vector Machine Ensemble." *Knowledge-Based Systems* 120: 4–14. https://doi.org/10.1016/j.knosys.2016.12.019.

Sun, J., and H. Li. 2012. "Financial Distress Prediction Using Support Vector Machines: Ensemble vs. Individual." *Applied Soft Computing* 12, no. 8: 2254–2265. https://doi.org/10.1016/j.asoc.2012.03.028.

Sun, J., H. Li, H. Fujita, B. Fu, and W. Ai. 2020. "Class-Imbalanced Dynamic Financial Distress Prediction Based on Adaboost-SVM Ensemble Combined With SMOTE and Time Weighting." *Information Fusion* 54: 128–144. https://doi.org/10.1016/j.inffus.2019.07.006.

Sun, J., Z. M. Shang, and H. Li. 2014. "Imbalance-Oriented SVM Methods for Financial Distress Prediction: A Comparative Study Among the New SB-SVM-Ensemble Method and Traditional Methods." *Journal of the Operational Research Society* 65, no. 12: 1905–1919. https://doi.org/10.1057/jors.2013.117.

Sustersic, M., D. Mramor, and J. Zupan. 2009. "Consumer Credit Scoring Models With Limited Data." *Expert Systems With Applications* 36, no. 3: 4736–4744. https://doi.org/10.1016/j.eswa.2008.06.016.

Tang, X., S. Li, M. Tan, and W. Shi. 2020. "Incorporating Textual and Management Factors Into Financial Distress Prediction: A Comparative Study of Machine Learning Methods." *Journal of Forecasting* 39, no. 5: 769–787. https://doi.org/10.1002/for.2661.

Tinoco, M. H., P. Holmes, and N. Wilson. 2018. "Polytomous Response Financial Distress Models: The Role of Accounting, Market and Macroeconomic Variables." *International Review of Financial Analysis* 59: 276–289.

Tinoco, M. H., and N. Wilson. 2013. "Financial Distress and Bankruptcy Prediction Among Listed Companies Using Accounting, Market and Macroeconomic Variables." *International Review of Financial Analysis* 30: 394–419. https://doi.org/10.1016/j.irfa.2013.02.013.

Traczynski, J. 2017. "Firm Default Prediction: A Bayesian Model-Averaging Approach." *Journal of Financial and Quantitative Analysis* 52, no. 3: 1211–1245. https://doi.org/10.1017/s002210901700031x.

Tsai, C. F. 2014. "Combining Cluster Analysis With Classifier Ensembles to Predict Financial Distress." *Information Fusion* 16: 46–58. https://doi.org/10.1016/j.inffus.2011.12.001.

Wang, G., J. Ma, L. Huang, and K. Xu. 2012. "Two Credit Scoring Models Based on Dual Strategy Ensemble Trees." *Knowledge-Based Systems* 26: 61–68. https://doi.org/10.1016/j.knosys.2011.06.020.

Yang, F., M. Z. Abedin, and P. Hajek. 2023. "An Explainable Federated Learning and Blockchain-Based Secure Credit Modeling Method." *European Journal of Operational Research* 317, no. 2: 449–467. https://doi.org/10.1016/j.ejor.2023.08.040.

Yang, S., and A. Browne. 2004. "Neural Network Ensembles: Combining Multiple Models for Enhanced Performance Using a Multistage Approach." *Expert Systems* 21, no. 5: 279–288. https://doi.org/10.1111/j.1468-0394.2004.00285.x.

Yousaf, U., K. Jebran, and M. Wang. 2022. "A Comparison of Static Dynamic and Machine Learning Models in Predicting the Financial Distress of Chinese Firms." *Romanian Journal of Economic Forecasting* 25, no. 1: 122.

Yousaf, U. B., K. Jebran, and I. Ullah. 2024. "Corporate Governance and Financial Distress: A Review of the Theoretical and Empirical Literature." *International Journal of Finance and Economics* 29, no. 2: 1627–1679. https://doi.org/10.1002/ijfe.2752.

Yousaf, U. B., K. Jebran, and M. Wang. 2021. "Can Board Diversity Predict the Risk of Financial Distress? Corporate Governance." *International Journal of Business in Society* 21, no. 4: 663–684. https://doi.org/10.1108/CG-06-2020-0252.

Yuan, K. P., G. T. Chi, Y. Zhou, and H. L. Yin. 2022. "A Novel Two-Stage Hybrid Default Prediction Model With k-Means Clustering and Support Vector Domain Description." *Research in International Business and Finance* 59, no. 101536: 1–24. https://doi.org/10.1016/j.ribaf.2021.101536.

Zhang, Z. P., G. T. Chi, S. Colombage, and Y. Zhou. 2022. "Credit Scoring Model Based on a Novel Group Feature Selection Method: The Case of Chinese Small-Sized Manufacturing Enterprises." *Journal of the Operational Research Society* 73, no. 1: 122–138. https://doi.org/10.1080/01605682.2021.1880295.

Zhao, Z., S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu, and R. Wasinger. 2015. "Investigation and Improvement of Multi-Layer Perceptron Neural Networks for Credit Scoring." *Expert Systems With Applications* 42, no. 7: 3508–3516. https://doi.org/10.1016/j.eswa.2014.12.006.

Zhou, L., Y.-W. Si, and H. Fujita. 2017. "Predicting the Listing Statuses of Chinese-Listed Companies Using Decision Trees Combined With an Improved Filter Feature Selection Method." *Knowledge-Based Systems* 128: 93–101. https://doi.org/10.1016/j.knosys.2017.05.003.

Zhou, L., K. P. Tam, and H. Fujita. 2016. "Predicting the Listing Status of Chinese Listed Companies With Multi-Class Classification Models." *Information Sciences* 328: 222–236. https://doi.org/10.1016/j.ins.2015.08.036.

Appendix A

TABLE A1 | Summary of feature selection for different time windows.

Feature no.	Indicator type		Indicator name	Time window *t*-1	Time window *t*-2	Time window *t*-3	Time window *t*-4	Time window *t*-5	Total number of occurrences
1	Internal financial factors	Solvency	X_1 Gearing ratio excluding pre-receipts	√					1
2			X_2 Long-term asset suitability ratio				√		1
3			X_3 Current liabilities to equity ratio		√				1
…			…	…	…	…	…	…	…
21		Profitability	X_{21} Earnings per share EPS			√	√	√	3
22			X_{22} Net assets per share BPS	√				√	2
…			…	…	…	…	…	…	…
41			X_{41} Earnings per share EPS—net of/diluted	√		√	√	√	4
…			…	…	…	…	…	…	…
48			X_{48} Basic earnings per share	√	√	√	√	√	5
49			X_{49} Ordinary shares earning rate		√	√	√	√	4
50		Operational capabilities	X_{50} Net cash flow from operating activities			√		√	2
51			X_{51} Percentage of net cash flows from investing activities		√			√	2
…			…	…	…	…	…	…	…
66		Growth capacity	X_{66} Total assets growth			√			1
…			…	…	…	…	…	…	…
72	Non-financial factors		X_{72} Is a financial institution					√	1
73			X_{73} Pre-audit status			√			1
74			X_{74} Number of earnings forecasts'	√	√	√		√	4
…			…	…	…	…	…	…	…
84			X_{84} Annual salary disclosure method		√	√	√	√	4
…			…	…	…	…	…	…	…
91	Senior management profile		X_{91} Chairman gender	√					1
…			…	…	…	…	…	…	…
95	Corporate credit profile		X_{95} Defect type				√		1
96			X_{96} Change or no change in the first largest shareholder					√	1
…			…	…	…	…	…	…	…

(Continues)

Feature no.	Indicator type	Indicator name	Time window *t*-1	Time window *t*-2	Time window *t*-3	Time window *t*-4	Time window *t*-5	Total number of occurrences
100	Corporate business reputation	X_{100} Whether there are significant litigation and arbitration cases	√					1
…	External macro conditions	…	…	…	…	…	…	…
106		X_{106} GDP index: tertiary sector (previous year = 100)	√		√	√	√	4
…		…	…	…	…	…	…	…
113		X_{113} Average wage of employed persons in urban units (yuan)		√	√	√	√	4
…		…	…	…	…	…	…	…
117		X_{117} Consumer price index	√		√	√	√	4
118		X_{118} Retail price index	√	√	√		√	4
119		X_{119} Aggregate production price index for agricultural products	√	√	√	√	√	5
120		X_{120} Industrial producer ex-factory price index	√		√		√	3
…		…	…	…	…	…	…	…
123	KMV indicators	X_{123} Default distance DD				√		1
124		X_{124} Expected default frequency EDF		√		√		2
125	Total number of indicators		27	33	42	44	70	—

Note: This table counts the number of times the indicator appears in the 5-time window indicator system, pink is for indicators that appear five times and yellow is for indicators that appear four times.

Appendix B

TABLE B1 | Evaluation of selected features using 5 Cs of credit.

Feature no.	Indicator type		Indicator name	Character	Capacity	Capital	Collateral	Condition
A: Time window *t*-2								
1	Internal financial factors	Solvency	X_1 Current liabilities to equity ratio			√		
...			...	...	...	...	...	...
14		Operational capabilities	X_{14} Percentage of net cash flows from investing activities		√			
15			X_{15} Cash received for obtaining loans			√		
16	Non-financial factors		X_{16} Number of earnings forecasts	√				
...			...	...	...	...	...	...
24	External macro conditions		X_{24} Industry sentiment index					√
25			X_{25} GDP index (previous year = 100)					√
...			...	...	...	...	...	...
33	KMV indicators		X_{33} Expected default frequency EDF	√				
34	Total			7	13	4	0	9
B: Time window *t*-3								
Feature no.	**Indicator type**		**Indicator name**	**Character**	**Capacity**	**Capital**	**Collateral**	**Condition**
1	Internal financial factors	Solvency	X_1 Capital fixation ratio			√		
...			...	...	...	...	...	...
4		Profitability	X_4 Earnings per share EPS—basic		√			
...			...	...	...	...	...	...
13		Operational capabilities	X_{13} Net cash flow from operating activities as a percentage of		√			
...			...	...	...	...	...	...
17		Growth capacity	X_{17} Total assets (growth rate relative to beginning of year)			√		
...			...	...	...	...	...	...
19	Non-financial factors within the business		X_{19} Pre-audit status'	√				
...			...	...	...	...	...	...
24	Basic corporate credit profile		X_{24} Whether to disclose internal control evaluation reports	√				
25	Corporate Business Reputation		X_{25} Total amount of penalties'	√				
26	External macro conditions of the enterprise		X_{26} Industry Sentiment Index					√
...			...	...	...	...	...	...
43	Total			6	14	5	0	17

(Continues)

C: Time window *t*-4								
Feature no.	**Indicator type**		**Indicator name**	**Character**	**Capacity**	**Capital**	**Collateral**	**Condition**
1	Internal financial factors of the enterprise	Solvency	X_1 Long-term asset suitability ratio			√		
...			...	...	...	...	...	...
7		Profitability	X_7 Earnings per share EPS—basic		√			
...			...	...	...	...	...	...
20		Operational capabilities	X_{20} Current assets/total assets			√		
...			...	...	...	...	...	...
23		Growth capacity	X_{23} Total profit growth rate B′		√			
24	Non-financial factors within the business		X_{24} Ten largest shareholders outstanding Herfindahl_3 Index			√		
...			...	...	...	...	...	...
32	Corporate Business Reputation		X_{32} Total amount of penalties’	√				
...			...	...	...	...	...	...
34	External macro conditions of the enterprise		X_{34} GDP index (previous year = 100)					√
...			...	...	...	...	...	...
43	KMV Indicator		X_{43} Default Distance DD	√				
...			...	...	...	...	...	...
45		Total		9	15	11	0	9

D: Time window *t*-5								
Feature no.	**Indicator type**		**Indicator name**	**Character**	**Capacity**	**Capital**	**Collateral**	**Condition**
1	Internal financial factors	Solvency	X_1 Total current liabilities/liabilities			√		
...			...	...	...	...	...	...
12		Profitability	X_{12} Earnings per share EPS		√			
...			...	...	...	...	...	...
25		Operational capabilities	X_{25} Net cash flow from operating activities		√			
...			...	...	...	...	...	...
...		Growth capacity	...	...	...	...	...	...
36			X_{36} Total profit growth rate		√			
...			...	...	...	...	...	...
40	Non-financial factors within the business		X_{40} Is a financial institution	√				
...			...	...	...	...	...	...

(Continues)

D: Time window *t*-5							
Feature no.	**Indicator type**	**Indicator name**	**Character**	**Capacity**	**Capital**	**Collateral**	**Condition**
51	Senior management profile	X_{51} General manager gender	√				
...		...	...	...	...	...	...
56	Corporate credit profile	X_{56} Industry sentiment Index					√
...		...	...	...	...	...	...
71	Total		9	27	19	0	15

Appendix C

TABLE C1 | Feature selection curves for different forecasting horizons.