sustainability

MDPI

Review

Predicting Construction Company Insolvent Failure: A Scientometric Analysis and Qualitative Review of Research Trends

Jun Wang [1,*], Mao Li [1,2], Martin Skitmore [3] and Jianli Chen [4]

1 School of Management Engineering, Qingdao University of Technology, Qingdao 266520, China
2 PetroChina Huabei Oilfield Company, Cangzhou 062552, China
3 Faculty of Society and Design, Bond University, Robina, QLD 4226, Australia
4 Department of Civil Engineering, University of Utah, Salt Lake City, UT 84112, USA
* Correspondence: wangjun.gt@gmail.com

Abstract: The construction industry is infamous for its high insolvent failure rate because construction projects require complex processes, heavy investment, and long durations. However, there is a lack of a comprehensive framework and a requirement for such a framework in predicting the financial distress of construction firms. This paper reviews relevant literature to summarize the existing knowledge, identify current problems, and point out future research directions needed in this area using a scientometric analysis approach. Based on a total of 93 journal articles relating to predicting construction company failure extracted from multiple databases, this study conducts a holistic review in terms of chronological trends, journal sources, active researchers, frequent keywords, and most cited documents. Qualitative analysis is also provided to explore the data collection and processing procedures, model selection and development process, and detailed performance evaluation metrics. Four research gaps and future directions for predicting construction company failure are presented: selecting a broader data sample, incorporating more heterogeneous variables, balancing model predictability and interpretability, and quantifying the causality and intercorrelation of variables. This study provides a big picture of existing research on predicting construction company insolvent failure and presents outcomes that can help researchers to comprehend relevant literature, directing research policy-makers and editorial boards to adopt the promising themes for further research and development.

Keywords: prediction; construction company; business failure; scientometric analysis; qualitative review

Citation: Wang, J.; Li, M.; Skitmore, M.; Chen, J. Predicting Construction Company Insolvent Failure: A Scientometric Analysis and Qualitative Review of Research Trends. *Sustainability* **2024**, *16*, 2290. https://doi.org/10.3390/su16062290

Academic Editors: Małgorzata Renigier-Biłozor, Artur Janowski and Alina Źróbek-Rózańska

Received: 22 January 2024
Revised: 27 February 2024
Accepted: 7 March 2024
Published: 9 March 2024

1. Introduction

In 2020, the global construction industry reached a staggering market size of USD 10.7 trillion, which is expected to exceed USD 15.2 trillion by 2030 [1]. Compared to other industries, construction is particularly vulnerable to financial crises [2] and sensitive to economic cycles [3] due to its various specificities, including the uniqueness and long duration of construction projects, the complexity of the construction process, the involvement of multiple specific teams, and uncertainties surrounding construction activities [4]. Unsurprisingly, therefore, despite its large size, growing momentum, and notable economic contribution, the industry is infamous for its high business insolvent failure rate [5], making the accurate prediction of company failure important for both the companies themselves and such other stakeholders as investors, creditors, shareholders, and employees [6–9].

Business failure prediction, also known as bankruptcy prediction or default prediction, requires the quantitative analysis of a corporate enterprise to forecast the likelihood of its default, and much research has been conducted in different regions using various predictive approaches to achieve this goal. Earlier research, for instance, primarily focused on adopting statistical techniques to build linear models such as the multivariate discriminant analysis [10], multiple regression analysis [11], and logistic regression models [12]. Thanks to the recent development of artificial intelligence, emerging techniques such as machine learning [3], deep learning [13], and ensemble learning [2] have been adopted for prediction purposes and have enabled the use of more variables, larger sample sizes, and higher accuracy.

However, despite some fruitful academic results, limited reviews have been conducted that are specifically concerned with the construction industry. Of these, Edum-Fotwe et al. summarized how to utilize weighted financial ratios to construct a single index (known as a Z-score) that classified construction companies as failing, at risk, or non-failing [14]. Wong and Ng integrated the common causes of construction company failure and listed pertinent prediction techniques as ratio analysis, multiple discriminant analysis, conditional probability models, and subjective assessment [15]. Alaka et al. investigated 70 relevant journal articles and doctoral theses, summarizing their country, variables selected, techniques used, sample size, performance, etc., but incorporated articles from other industries like banking, IT, and manufacturing, and did not exclusively focus on the construction industry [16]. Alaka et al. summarized the critical factors for insolvency prediction regarding factor frequency and model accuracy and conducted a questionnaire survey of construction industry professionals to collect their feedback on those factors [17]. More recently, Assaad and El-adaway reviewed relevant research over the past 30 years to identify the failure factors impacting the business operations of construction firms using both simplified analysis and social network analysis [18].

Nonetheless, although these authors made a valuable contribution to the current body of knowledge, they primarily focused on evaluating and selecting variables involved and failed to consider other important technical aspects such as data processing methods, model development process, and performance evaluation criteria. In addition, they are essentially manual and ad hoc qualitative reviews and do not adopt any scientometric approach to conduct a systematic review. At the same time, recent research has found that humans are better at discovering and comprehending domain knowledge presented in graphical forms [19,20]. To update the research trends, this paper adopts the science mapping approach, which reveals the inherent relationships among existing research work using graphical representation and conducts a quantitative review of construction company failure prediction studies to complement existing qualitative work. The specific research objectives of this review include: (1) applying a science mapping approach to analyze the journals, keywords, researchers, and articles in the domain of predicting construction company insolvent failure; (2) analyzing the existing key research works related to predicting construction company insolvent failure; (3) revealing the recent research gaps and pointing out some possible future research directions of predicting construction company insolvent failure.

The paper is structured as follows. Section 2 lists the research methods used in the study with all the inclusion and exclusion criteria. Section 3 contains the results of the scientometric analysis. Section 4 includes further qualitative discussions by reviewing data collection and processing procedures, model selection and development process, and detailed performance evaluation metrics and identifying research gaps and future directions. Finally, Section 5 concludes the study.

2. Research Methods

This section describes the three-step review methods, a commonly adopted method for the science mapping-based systematic review [21–24] that comprises the literature search, scientometric analysis, and qualitative discussion. Figure 1 illustrates the detailed workflow of the research methods.

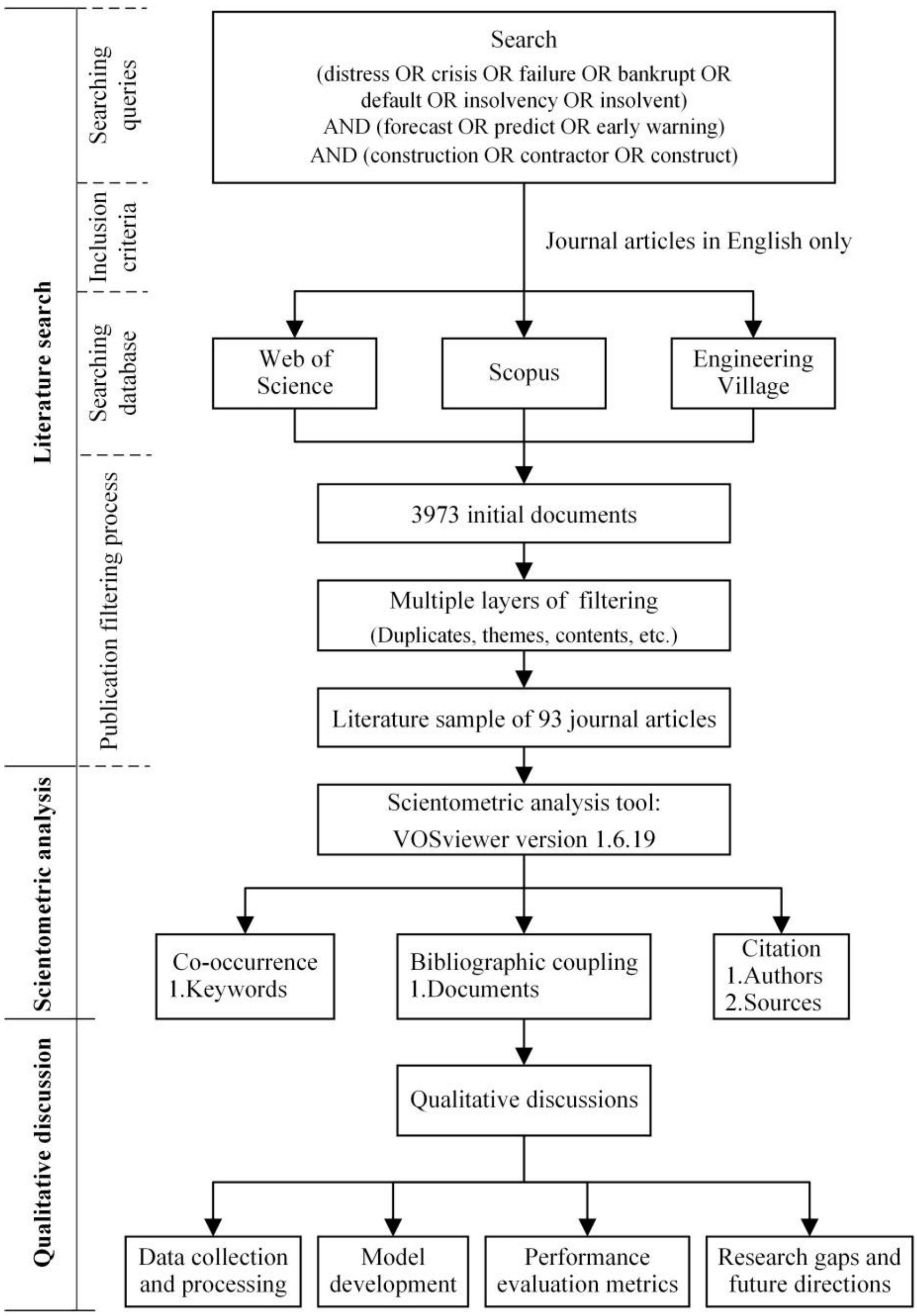

Figure 1. Detailed workflow of the research methods.

2.1. Literature Search

Following PRISMA protocols, a literature search was carried out using the query "(distress OR crisis OR failure OR bankrupt OR default OR insolvency OR insolvent) AND (forecast OR predict OR early warning) AND (construction OR contractor OR construct)" to confine the topic to predicting construction company insolvent failure. Three commonly used databases in the construction company failure domain [16,18,25–27]—Web of Science, Scopus, and Engineering Village—were selected to perform the search and only journal articles published in English were retained. In total, 1088, 1764, and 1121 articles were retrieved from each database, yielding 3973 initial documents. Next, 2085 duplicated articles were removed, and the remaining 1888 articles' titles and abstracts were manually reviewed for further filtering. This resulted in the further exclusion of articles concerning industries other than construction (e.g., banking, manufacturing, high-tech, fashion, and hospitality); other topics (e.g., forecasting construction cost or price [28], predicting project performance [29], and estimating technological capabilities [30]); and only analyzing business insolvent failure factors without using predictive models [17]—leaving 93 journal articles eventually selected to constitute the literature sample for analysis.

2.2. Scientometric Analysis

The second step of the review involved the use of scientometric analysis, which is a generic process of domain analysis and visualization [31] that was broadly adopted to facilitate the systematic literature review of different building construction-related topics [21–24]. A text-mining tool VOSViewer (v1.6.19) [32,33] was used, which created distance-based visualizations of networks. With this, each node in the network represented information such as the source journal, author, organization, country, and keyword, and the distance between nodes reflected the closeness of nodes measured in such different metrics as co-authorship, shared references, and co-occurrence [34]. The 93 articles were transported into VOSViewer for scientometric analysis to generate results related to the influences of journals, keywords, researchers, and articles in predicting construction company insolvent failure.

2.3. Qualitative Discussion

Finally, an in-depth qualitative discussion was carried out to explore the results of the scientometric analysis from multiple perspectives, including the data collection and processing procedures used, comparing different predictive models, summarizing different performance evaluation metrics, identifying existing research gaps, and suggesting future research directions.

3. Results

This section presents the science mapping results of the 93 articles. It starts with a chronological trend and journal source analysis and concludes with a researcher, keyword, and document analysis using VOSViewer.

3.1. Chronological Trend and Journal Source Analysis

Figure 2 displays the annual publication count of articles, highlighting the growing focus of researchers on creating models to predict the insolvent failure of construction companies. This period can be segmented into two phases: the first phase (1977–2008) and the second phase (2009–2022). During the first phase, there was a lower output of articles, averaging 0.72 articles per year. In contrast, the second phase saw a marked increase in publication frequency, with an average of 5.00 articles per year. Notably, the articles published in the second phase comprise 75% of the total sample, underscoring the escalated academic interest in recent years towards precise prediction construction company insolvent failure. Considering the exponentially growing scientific outputs across different domains, the relatively slow growth speed shown in Figure 2 indicates that the research domain of predicting construction companies' bankruptcy needs further exploration.

Figure 3 shows the journal sources involved, as represented by the node, the size of which is proportionate to the number of publications. The internode distance approximately reflects their cross-reference times [34], and the node color demonstrates the clustering results, which were automatically determined by VOSViewer using a smart local moving average algorithm [33–35]. This indicates the *Journal of Construction Engineering and Management* is ranked first due to the large number of publications in the corresponding area, followed by *Construction Management and Economics*, *Engineering Construction and Architectural Management*, and *Expert Systems with Applications*.

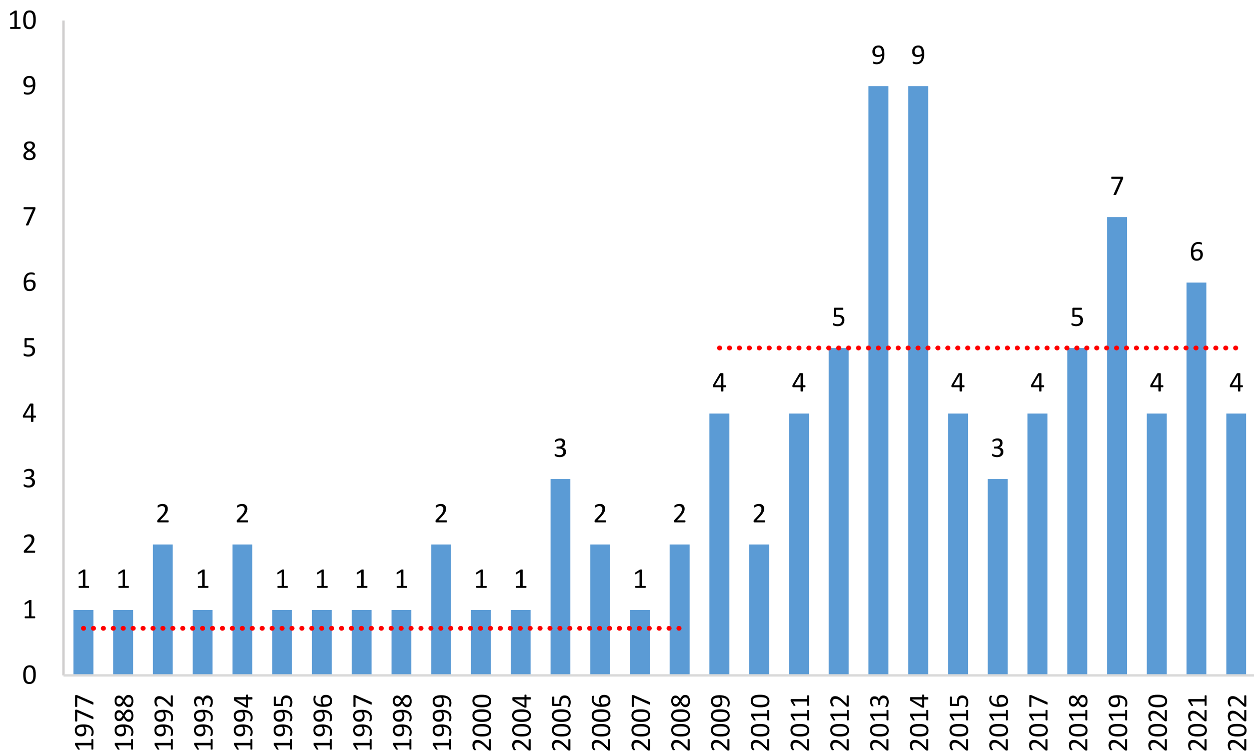

Figure 2. Chronological distribution of journal articles (dotted lines representing phase average).

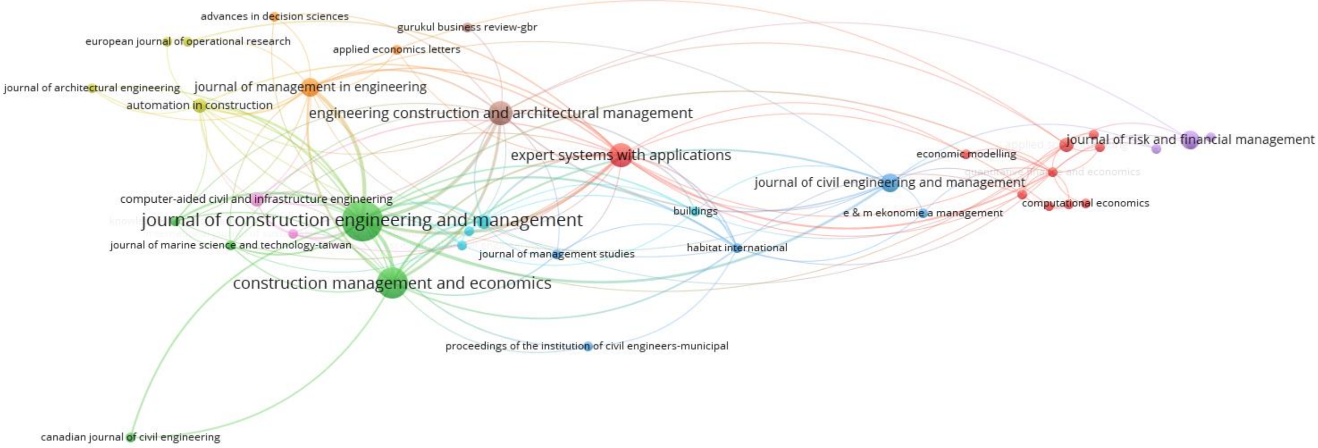

Figure 3. Visualization of journal sources.

Table 1 summarizes the associated quantitative metrics, with the *Journal of Construction Engineering and Management*, *Construction Management and Economics*, *Engineering Construction and Architectural Management*, and *Expert Systems with Applications* having received the highest number of publications and the publications from *Journal of Construction Engineering and Management*, *Expert Systems with Applications*, and *Journal of Civil Engineering and Management* having received the most total citations. To adjust for bias, given that earlier documents are more likely to receive more citations than recent documents, a normalized citation that equals the citation number of a document divided by the average citation number of all documents was introduced and calculated by VOSViewer [32]. On average, the documents in the *Journal of Risk and Financial Management* and *Journal of Management in Engineering* are more recent.

Table 1. Quantitative metrics of top journal sources.

Journal Source	Number of Publications	Total Citations	Average Publication Year	Average Citations	Normalized Citations	Average Normalized Citations
Journal of Construction Engineering and Management	13	420	2004	32.31	11.67	0.90
Construction Management and Economics	8	70	2003	8.75	7.13	0.89
Engineering Construction and Architectural Management	5	68	2010	13.60	5.32	1.06
Expert Systems with Applications	5	240	2014	48.00	14.18	2.84
Journal of Civil Engineering and Management	3	89	2013	29.67	4.27	1.42
Journal of Management in Engineering	3	40	2017	13.33	3.97	1.32
Journal of Risk and Financial Management	3	13	2021	4.33	2.52	0.84

3.2. Researchers, Keyword, and Document Analysis

Figure 4 shows the number of articles (≥2) published by each researcher in VOSViewer. This time, a node represents a researcher, with the node size representing the number of citations, and the internode distance reflecting the number of times two researchers cited each other. This indicates that H.P. Tserng, D. Arditi, P.C. Chen, R. Kangari, and J.S. Russell have received the most citations. Table 2 provides the associated quantitative metrics of the researchers involved.

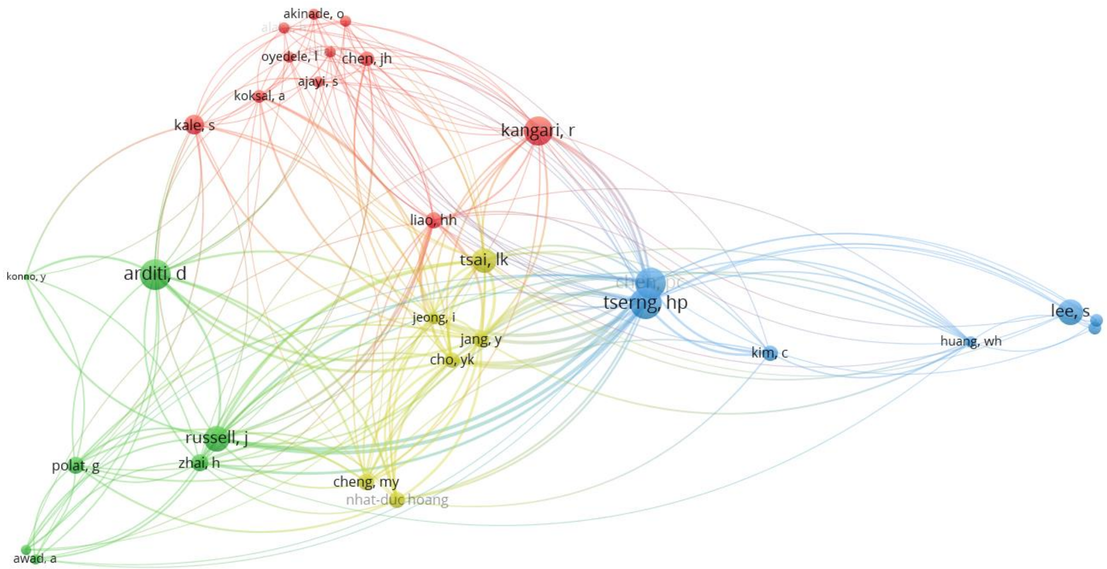

Figure 4. Visualization of researchers.

Table 2. Quantitative metrics of most-cited researchers.

Researcher Name	Number of Publications	Total Citations	Average Publication Year	Average Citations	Normalized Citations	Average Normalized Citations
H.P. Tserng	8	141	2013	17.63	7.11	0.89
D. Arditi	6	138	2002	23.00	6.18	1.03
P.C. Chen	7	137	2013	19.57	6.65	0.95
R. Kangari	2	123	1990	61.50	2.03	1.02
J. Russell	3	100	1996	33.33	2.97	0.99
S. Lee	2	96	2015	48.00	6.13	3.07
L.K. Tsai	4	90	2012	22.50	3.62	0.91
S. Kale	3	62	1999	20.67	3	1.00

Keywords, including both Author Keywords that authors believe best represent the context of their paper and Keywords Plus that appear frequently in the titles of an article's references and not necessarily in the title of the article or as Author Keywords, are critical textual information for scientific publications. Both Author Keywords (manually input by the document authors) and Keywords Plus (automatically suggested by Web of Science programs based on cross-referenced articles) were utilized here. Textual data pre-processing was also manually conducted to unify the writing formats of different semantically similar expressions. For example, in the context of this research, "artificial neural network model (ANN)", "artificial neural networks", and "artificial neural networks" all equal "artificial neural networks". The number of keywords was reduced from 416 to 148 after the data pre-processing process. After setting the minimum number of occurrences of a keyword at three in VOSViewer, 54 out of 148 keywords were selected and visualized in Figure 5. Each note represents a keyword, and the node size indicates its number of occurrences. The internode distance approximately reflects the number of times two keywords co-occur, and the node color demonstrates the average occurrence year of each node, as indicated by the timeline legend. The ten most frequent keywords are "bankruptcy prediction" (42), "financial ratios" (32), "construction industry" (29), "performance metric" (23), "financial distress prediction" (22), "prediction model" (21), "neural networks" (19), "risk management" (18), "discriminant analysis model" (16), "contractor" (16), "business failure prediction" (14), and "support vector machine" (13), where the values in parentheses denote the number of occurrences.

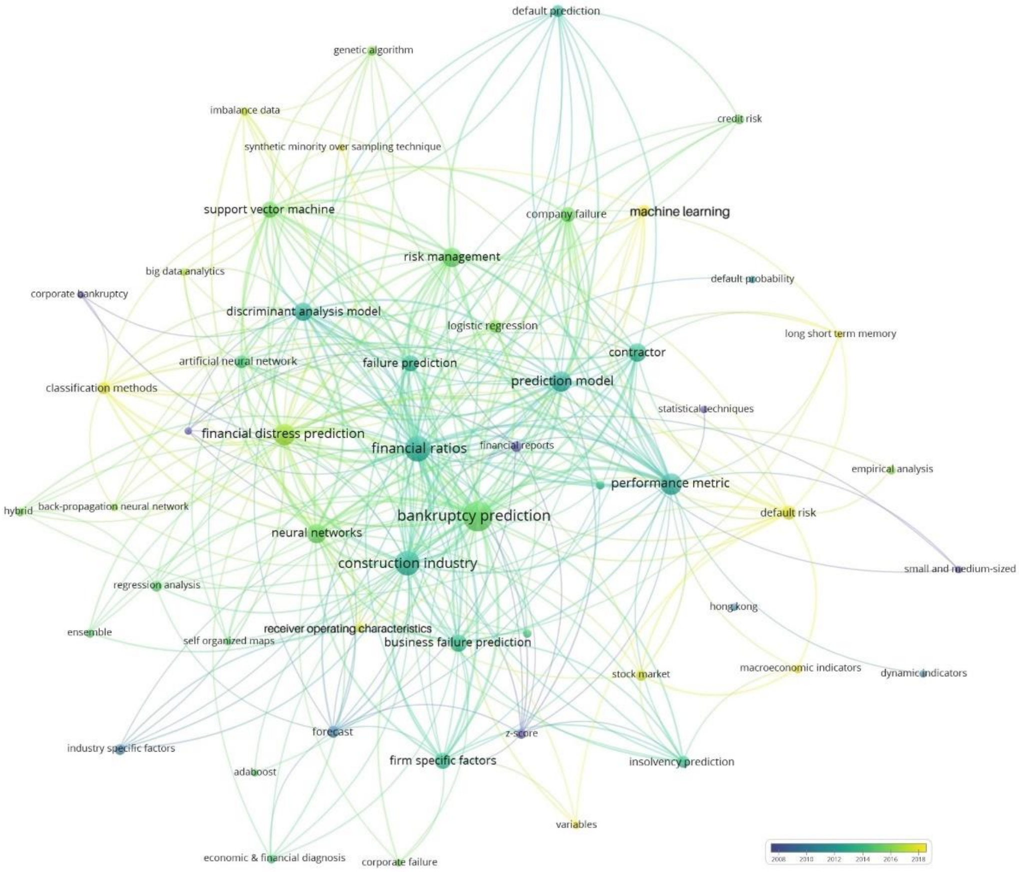

Figure 5. Visualization of keywords.

Figure 5 shows that predicting construction company insolvent failure can also identify "bankruptcy prediction", "financial distress prediction", "insolvency prediction", "corporate failure", or "company failure" problems within the "construction industry". In terms of "prediction model", "statistical techniques" such as "z-score", "discriminant analysis model", and "regression analysis" were utilized in the early stage. More recently, more advanced techniques such as "machine learning", "neural networks", "support vector machine", and "long short-term memory" have emerged. In addition to problem definition and predictive methods, input "variables" such as "macroeconomic indicators" and "stock market" and evaluation metrics like "receiver operating characteristics" have also recently occurred in the keyword visualization map.

Finally, the literature samples are analyzed based on the documents. Figure 6 shows the 43 selected connected documents obtained in VOSViewer after setting the minimum citations of a document to 10. In this case, each node represents a document, and the node size indicates the total number of citations the document had received. The closer documents are more related, and their internode distance approximately reflects the number of references shared.

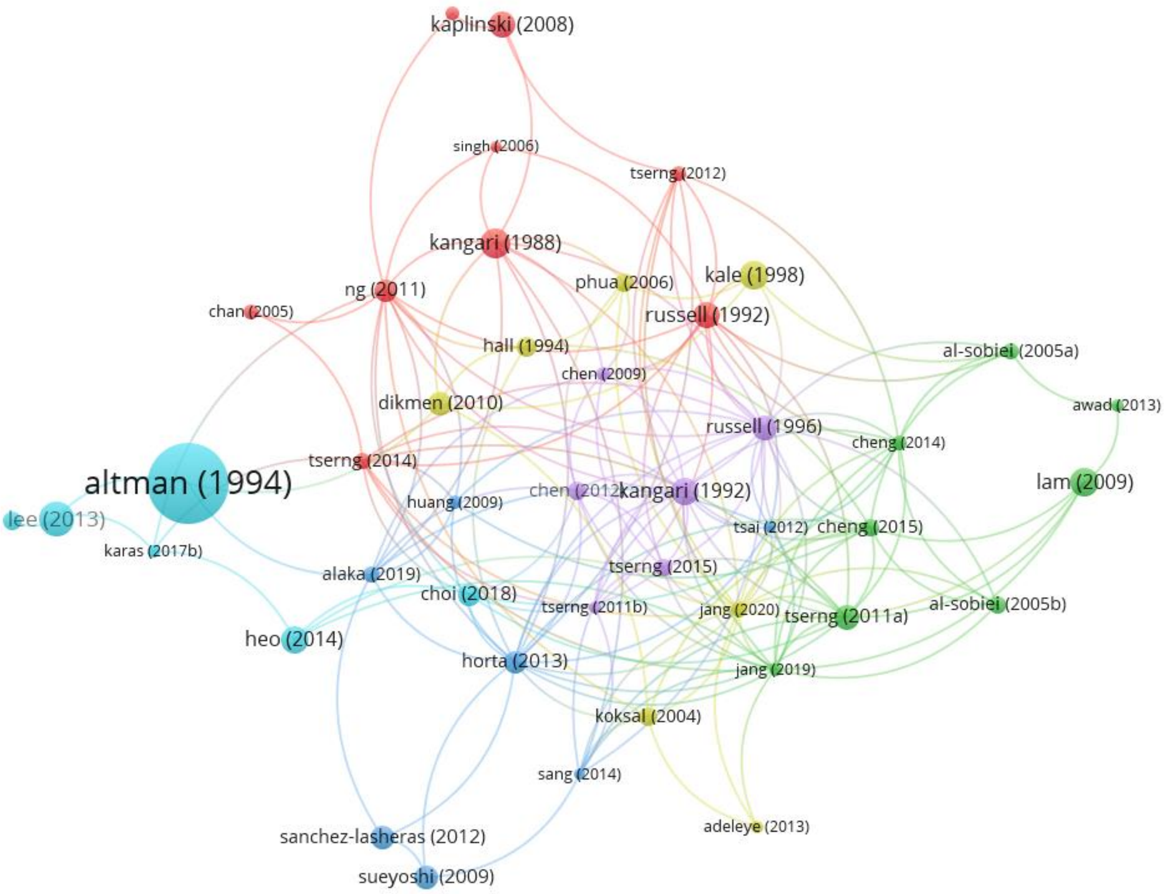

Figure 6. Visualization of documents.

Table 3 summarizes the titles and citations of the 20 most-cited documents sorted by their citations. These highly cited articles demonstrate that their common goal is to identify the factors that significantly impact construction company insolvent failure and make corresponding predictions as accurately as possible. Predictive methods that have worked well for other industries may not be suitable for the construction sector due to its unique nature [36–38], especially when considering the relatively long duration of construction projects [2]. Early-stage highly cited research focused on exploring and evaluating the impact of different factors involved [10,11,39–43] and, with the notable exception of Altman's research (which also adopted neural networks), primarily utilized linear methods [39]. More recently, highly cited research has increasingly incorporated more non-linear predictive methods and has generally achieved higher prediction accuracy [2,3,7,36,37,44,45].

Table 3. Top 20 highly cited articles in the literature sample.

Article	Title	Total Citations	Normalized Citations
Altman et al. [39]	Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)	440	1.87
Lee and Choi [36]	A multi-industry bankruptcy prediction model using a back-propagation neural network and multivariate discriminant analysis	87	4.33
Kangari [40]	Business failure in the construction industry	67	1
Kale and Arditi [41]	Business failures: Liabilities of newness, adolescence, and smallness	62	1
Lam et al. [44]	A support vector machine model for contractor prequalification	60	1.74
Heo and Yang [37]	AdaBoost-based bankruptcy forecasting of Korean construction companies	56	3.89
Kangari et al. [11]	Financial performance analysis for the construction industry	56	1
Russell and Jaselskis [42]	Predicting construction contractor failure prior to contract award	53	1
Kapliński [38]	Usefulness and credibility of scoring methods in the construction industry	52	1.65
Tserng et al. [3]	An enforced support vector machine model for construction contractor default prediction	47	1.76
Russell and Zhai [43]	Predicting contractor failure using stochastic dynamics of economic and financial variables	47	1
Sueyoshi and Goto [46]	DEA–DA for bankruptcy-based performance assessment: Misclassification analysis of Japanese construction industry	45	1.30
Sánchez-Lasheras et al. [47]	A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy	44	2.04
Dikmen et al. [48]	Using analytic network process to assess business failure risks of construction firms	42	1.58
Horta and Camanho [7]	Company failure prediction in the construction industry	41	2.04
Ng et al. [49]	Applying the Z-score model to distinguish insolvent construction companies in China	39	1.46
Choi et al. [2]	Predicting financial distress of contractors in the construction industry using ensemble learning	39	4.43
Hall [12]	Factors distinguishing survivors from failures amongst small firms in the UK construction sector	30	0.13
Chen [45]	Developing SFNN models to predict the financial distress of construction companies	29	1.34
Zoričák et al. [50]	Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets	29	2.04

4. Qualitative Discussion

Following previous scientometric analysis and results, this section provides an in-depth qualitative discussion of research related to predicting construction company insolvent failure. This summarizes up-to-date research into the perspectives of data collection and processing, predictive models, performance evaluation, and research gaps and future directions.

4.1. Data Collection and Processing

4.1.1. Sample Determination

Sample determination is central to any prediction model, as it informs the initial investigation of the data and facilitates effective model construction [51]. Of the 93 investigated articles, 22 focus on construction companies in the United States, followed by Taiwan (8), the United Kingdom (6), South Korea (6), the Czech Republic (5), Japan (4), Spain (4), China (4), Russia (4), Hong Kong (3), Poland (3), Portugal (3), and the Slovak Republic (2) (the value in parentheses following each region represents the number of associated articles). This points to a research gap for future research to be conducted into the less studied regions because predictive models that work well for a particular region may not necessarily be effective for other regions due to their different construction market dynamics, accounting systems, and definitions of corporate insolvent failure. For this reason, the articles tend to focus on a single region.

Sample size and sample type are often critical factors that influence predictive model construction and accuracy. A total of 86 out of the 93 investigated articles disclose the sample sizes of their studies, of which 65 have a sample size of less than 1000, 15 have a sample size between 1000 and 10,000, and the remaining 6 have a sample size greater than 10,000. Due to the advent and utilization of big data analytical techniques, the average sample size has grown from 142 for articles published in and before 2010, to 4005 subsequently. In terms of type, many studies focus on publicly trading companies, as they are required by different Securities and Exchange Commissions to periodically disclose their financial reports [50], making their financial data more accessible than their small- and medium-sized counterparts. In addition, the insolvent failure of large companies tends to have a greater impact on the construction market [52].

4.1.2. Variable Selection

Once an appropriate sample has been determined, relevant variables (features, factors, or indicators) are selected for model construction. However, there is no extant theory or academic consensus to indicate which variables are better than others for distinguishing between failing and non-failing companies [53–55], and hence this comprehensive review of the variables adopted in previous studies will be helpful for future variable selection.

One type of commonly utilized variable, for instance, is the financial ratios of companies [56–58], which have been considered objective measurements of companies based on publicly available information [59] and can generally be classified into five different categories that demonstrate the liquidity, leverage, activity, profitability, and cash flow situations of those companies when they are selected as explanatory variables to construct predictive models. It is not surprising that financial ratios are the most frequently used variables for corporate insolvent failure prediction due to their irrefutable relationship discovered as early as 1966 [60]. Several studies rely solely on financial ratios as explanatory variables, with the assumption that these ratios contain all relevant information for corporate insolvent failure prediction [54,57,58]. Table 4 summarizes the financial ratio variables of each category based on their occurrence in the sample publications.

Table 4. Frequently selected financial ratio variables for each category.

Type	Variable Name	Variable Description	Occurrence
Liquidity	Current ratio	Current assets/current liabilities	40
	Working capital to total assets	(Current assets − current liabilities)/total assets	37
	Quick ratio	(Current assets − inventory)/current liabilities	22
	Debt to net worth	Total liabilities/(total assets + total liabilities)	17
	Current assets to net assets	Current assets/(total assets − current liabilities)	14
	Fixed assets to total assets	Fixed assets/total assets	9
	Current liabilities to total assets	Current liabilities/total assets	9
	Current assets to total assets	Current assets/total assets	6

Table 4. *Cont.*

Type	Variable Name	Variable Description	Occurrence
Leverage	Debt ratio	Total liabilities/total assets	26
	Debt to equity ratio	Total liabilities/total shareholders' equity	22
	Times interest earned	Earnings before interest and taxes/interest expense	20
	Retained earnings to sales	(Beginning period retained earnings + net income − cash dividends − stock dividends)/sales	14
	Retained earnings to total assets	(Beginning period retained earnings + net income − cash dividends − stock dividends)/total assets	12
Activity	Total assets turnover	Net sales/average total assets	40
	Accounts receivable turnover	Net sales/average account receivables	22
	Working capital turnover	Net sales/(current assets − current liabilities)	20
	Equity turnover	Net sales/average shareholders' equity	13
	Inventory turnover	Net sales/average inventory	13
	Fixed asset turnover	Net sales/average fixed assets	13
	Accounts payable turnover	Net sales/average account payables	12
	Current assets turnover	Net sales/average current assets	7
Profitability	Return on assets	Net income/average total assets	42
	Return on equity	Net income/shareholders' equity	31
	Return on sales	Net profit/net sales	25
	Earnings before interest and taxes to total assets	Earnings before interest and taxes/total assets	12
	Operating profit margin	Operating profit/total income	11
	EBIT margin	Earnings before interest and taxes/total sales	8
	Gross margin	Gross profit/total income	8
	Return on invested capital	(Net income − dividends)/(debt + equity)	7
	Profits to net working capital	Net profit/(Current assets − current liabilities)	6
	Sales to net income	Sales/net income	5
Cash flow	Cash and cash equivalents to total debts	Cash and cash equivalents/total debts	7
	Cash and cash equivalents to current liabilities	Cash and cash equivalents/current liabilities	6
	Cash and cash equivalents to total assets	Cash and cash equivalents/total assets	5

However, other studies have questioned such assumptions and have pointed out limitations of the financial ratio variables, in that they may reflect a company's recorded book value rather than its true value [61] and do not contain all the information related to financial distress [62]. Consequently, researchers have started incorporating other complementary variables to improve the predictive power of the models. Other commonly used variables include those that represent the characteristics of the companies, the characteristics of construction projects undertaken by companies, and the economic conditions of the construction industry, and the broader stock market. Table 5 summarizes the other frequently selected variables based on their occurrence in publications.

The importance of meticulously cataloging the frequency of selected variables in prior studies cannot be overstated, as this quantification underscores the significance of each variable in the research domain [63]. Of those non-financial variables, the primary focus should be placed on the strategic variables that encapsulate the unique characteristics of construction companies. These variables are designed to mirror the strategic positioning of these companies within the industry, which is pivotal in determining their operational performance. This assertion is substantiated by the findings of Horta and Camanho, who highlighted the profound impact of industry positioning on company performance [7]. Furthermore, the incorporation of project characteristic indicators is imperative. Given that the construction sector is inherently project-centric, with projects dictating the majority of a company's operations, these indicators significantly influence the financial health of a company [64]. Khanzadi's research further emphasizes the pervasive nature of risks within construction projects, noting that failing to account for these risks inevitably leads to challenges in meeting project objectives on time, within budget, and to the required quality standards [65]. In addition, variables pertaining to economic conditions are critical for the prediction of bankruptcy among construction companies. The susceptibility of these companies to macroeconomic fluctuations is particularly pronounced, given the construction industry's substantial contribution to the global economy. Variations in economic conditions have a direct impact on the financial stability of construction firms and the financial capacity of consumers [13,66]. Lastly, stock market variables, which serve as proxies for macroeconomic factors, warrant consideration for their role in enhancing the precision of predictive models. These variables encapsulate timely information that is crucial for assessing the likelihood of success or failure among contractors. Tserng's research corroborates the significance of these variables in providing insights into contractors' survivability or failure [4].

Table 5. Other selected non-financial variables.

Type	Variable Name	Occurrence
Company characteristics	Company age	12
	Number of employees	9
	Academic and professional qualifications of senior management	7
	Number of senior management personnel	6
	Credit granted	5
	Technical competency	5
	Environmental considerations	4
	Headquarter geographic location	3
	Claims history	3
Project characteristics	Project value	6
	Project contract value in progress	4
	Number of projects in progress	2
	Duration and complexity of project	2
	Number of past similar projects	1
Economic conditions	Interest rates	11
	Gross domestic product	10
	Consumer price index	5
	Total number of companies	5
	Number of employees in the construction industry	4
	Inflation rate	4
	Unemployment rate	3
	Housing starts	3
	Housing prices	2
	Construction consumption price index	2
	Volume of performed construction work	1
Stock market	Earnings per share	7
	Index of stock price	3
	Price to book ratio	3
	Market to book ratio	3

4.1.3. Data Processing

After selecting a set of initial variables, feature selection and extraction are common steps in the data processing phase. The main advantage of this is in dimensionality reduction, which reduces the computational effort of training predictive models and the risk of overfitting, making classification problems easier to solve [67]. Studies have also shown that feature selection enables predictive models to provide better results by allowing models to focus on the more important features involved [68].

Data filter [66,69] or wrapper [13,50] methods are also utilized to conduct feature selection and extraction that create a subset of refined variables for model construction. Filter methods carry out the feature selection before the model construction stage and are therefore independent of the predictive algorithms. Wrapper methods, on the other hand, use predictive algorithms as subroutines that measure the effectiveness of the variables with the prediction accuracy over a validation set [70,71]. Of these, filter methods tend to be used more in general because of their computational efficiency and lower risk of overfitting [72].

Because the number of non-failing companies is much higher than that of failing companies, the datasets collected for predicting company insolvent failure in the construction industry or many other industries are often imbalanced [73]. Learning from an imbalanced dataset can result in predictive models that do not accurately represent the data characteristics and may lead to suboptimal classification models that provide poor prediction results across data classes [74]. An empirical study has shown that an imbalanced dataset, in which the minority class represents 20%, significantly disturbed prediction accuracy [75]. This introduces another important but often overlooked data processing step for constructing reliable models to predict construction company insolvent failure (of the 93 investigated articles, only 14 disclosed that they adopted relevant methods to deal with the data imbalance issue).

Two different techniques are often suggested to deal with the data imbalance issue: under-sampling (e.g., random under-sampling [76]) and oversampling (e.g., synthetic minority oversampling technique (SMOTE) [77]). Under-sampling techniques seek to decrease the number of majority class members (non-failing companies), while oversampling seeks to increase the number of minority class members (failing companies) in the training set. The advantage of oversampling is that no information from the original training set is lost, since all original members remain, but oversampling would increase the size of the training set and thus increase training time and amount of memory accordingly [78]. Some recent studies have adopted the SMOTE + Tomek links approach, a hybrid method that combines oversampling and under-sampling techniques [13,79,80]. Extant studies disagree on which technique is better, and conflicting results are likely due to the combination of different datasets and classification algorithms [78,81–83].

4.2. Predictive Models

4.2.1. Statistical Models

Various statistical models were adopted to predict construction company insolvent failure, especially in early studies. Those most commonly utilized included the Z-score [38,49,84,85], discriminant analysis—both linear discriminant analysis (LDA) [39,86–89] and multivariate discriminate analysis (MDA) [90–93]—multiple linear regression [11,40,43,94], and logistic regression [12,95–97].

The Z-Score model was proposed by Altman [98] and has been widely utilized to predict corporate bankruptcy. It calculates the Z-Score by linearly combining different financial ratios and generally sets two thresholds to classify companies into three zones of safe, grey, and distress. Discriminant analysis is a parametric technique that determines the weightings of predictors that best discriminates between two or more categories by constructing a suitable statistical decision function—the discriminant function [99,100]. Depending on the number of categories and the method of constructing the discriminant function, discriminant analysis can be categorized into LDA, MDA, and quadratic discriminant analysis (QDA). Multiple linear regression constructs a linear equation to determine the response variable as a function of multiple explanatory variables. Logistic regression, introduced by Ohlson [101] in his study of default prediction, is a conditional probability model that estimates the probability of an event taking place, and can be used to solve both binary and multi-category classification problems [102].

Although these statistical models are popular in financial distress research and have been widely used to predict construction company insolvent failure, their assumptions often limit them quite restrictively. For example, discriminant analysis is based on the assumption that participants are independent of each other (randomly sampled) and the explanatory variables are normally distributed and not linearly interrelated with other explanatory variables (i.e., lack of multicollinearity) [103]. Multiple linear regression assumes that the response variable is a linear combination of explanatory variables, the error term variance does not depend on the values of explanatory variables, errors of explanatory variables are not intercorrelated, and explanatory variables cannot be linearly predicted from other explanatory variables [104]. Although logistic regression does not make assumptions about the distributions of explanatory variables and does not assume a linear relationship between explanatory and response variables, it does assume that the explanatory variables are linearly related to the logarithm of the response variable [105]. The model's predictive power decreases should these assumptions be violated.

4.2.2. Artificial Intelligence Models

Artificial intelligence models supported by advanced computational power and enhanced data supply have become increasingly popular. These models learn from the data to capture the relationship between the explanatory and response variables, and with fewer constraints than statistical models [106]. Commonly adopted artificial intelligence models are neural networks [36,45,107,108], support vector machines [7,44,50,109], and decision trees [86,110–114]. Lee and Choi's predictive comparison found that artificial intelligence models usually outperform statistical methods due to their ability to handle non-linear relationships and lack of restrictive model assumptions [36]. Nevertheless, unlike statistical methods, artificial intelligence methods require more sample data and have complex training processes. They are also exposed to overfitting risk, which reduces the stability of cross-sample prediction.

Ensemble models that combine multiple artificial intelligence models and statistical models (base learners) have recently gained popularity in predicting construction company insolvent failure [2,37,69,115–117] due to their lower variance, better accuracy, and higher model stability than base learners [118]. Commonly used ensemble methods include voting, bagging, boosting, and stacking. Voting combines the model outputs (hard voting) or predicted probabilities (soft voting) of base learners through weighted or unweighted averages to produce an aggregated predictive result [115,119]. Bagging, developed by Breiman [120], generates different training subsets using random sampling to train base learners and combine them using majority voting to obtain a strong classifier [121]. Boosting trains each base learner separately by iteratively selecting misclassified instances from the training set to convert the weak learners into strong learners and then combines their results using weighted mechanisms [122]. Stacking adopts a two-level structure that consists of level-0 (base level) learners and level-1 (meta-level) learners [123], and first trains level-0 learners whose predicted results will then be used as input to train the level-1 learners [124].

More recently, deep learning models have been used to predict construction company insolvent failure, and relevant research has primarily focused on the long short-term memory (LSTM) recurrent neural network (RNN) model [13,79,80,125]. Jang et al. first adopted the LSTM RNN model, and it was found that this deep learning model had outperformed both the neural networks model and the support vector machine model for both one-year ahead [79] and three-year ahead [80] predictions. Jing et al. later combined the structural default probability estimation model (ZPP) and time-series neural networks (LSTM) to estimate the corporate default probabilities [125]. A Shapley value-based approach was also proposed to quantify the impacts of input variables in deep learning models [13].

Artificial intelligence models can generally provide better predictive results than statistical models, and ensemble models and deep learning models can outperform single classifiers. Deep learning models also have an advantage over ensemble models when analyzing data with high dimensionality [126]. Despite their superior predictive ability, these models have their limitations. For example, the vital steps of building ensemble models determine the learning strategies for generating base learners and ensemble criteria for creating ensemble learners, but there are no stationary rules for deciding which learning strategies or ensemble criteria are better than others [127]. One has to construct the appropriate ensemble classifiers by referring to specific situations, which inevitably hinders the universality of using ensemble models in many artificial intelligence applications. Deep learning models are often limited because they are intractable when distinguishing the impact of input variables [125]. They are usually considered "black box" models whose internal structure and learned parameters are not interpretable [26].

4.3. Performance Evaluation Metrics

4.3.1. Threshold Metrics

Evaluation metrics play a critical role in model optimization and comparison during and after the model development. Threshold metrics set a predefined threshold and evaluate models by assessing how the predicted values fall relative to that threshold [128]. The threshold metrics in construction company insolvent failure research are summarized in Table 6. Accuracy (the ratio of correctly classified samples to the total number of samples) and error rate (1 − accuracy) are the threshold metrics most widely used by researchers to discriminate and select optimal classifiers. Moreover, the F measure (the harmonic mean of precision and recall) and the geometric mean (a metric that maximizes both true positive and true negative while keeping them relatively balanced) were also adopted to evaluate model performance.

Table 6. Threshold metrics in construction company insolvent failure research.

Metrics	Formula [1]	Relevant Articles
Accuracy	$\frac{tp+tn}{tp+fp+tn+fn}$	[7,13,36,44,45,47,79,87,88,96,97,108,111–113,116,117,129–133]
Error rate	$\frac{fp+fn}{tp+fp+tn+fn}$	[46,69,90,92,93,134]
F-measure	$\frac{2*P*R}{P+R}$	[56,74,79,135]
Geometric mean	$\sqrt{tp*tn}$	[50]

[1] Note: tp—true positive, tn—true negative, fp—false positive, fn—false negative, P—precision, R—recall.

These threshold metrics are easy for humans to compute and understand. However, their effectiveness is often limited when handling imbalanced datasets, which is often the case for predicting construction company insolvent failure. In such cases, these metrics do not provide adequate information about prediction ability, and their results are often biased towards the majority class (non-failing companies), leading to a higher misclassification rate for the minority class or a higher Type I error rate (identifying failing companies as non-failing companies) [26,52,136].

4.3.2. Area under the ROC Curve

The area under the receiver operating characteristic (ROC) curve is better than the accuracy metric [137] because it visually represents classifiers' performance and is insensitive to imbalanced datasets. It visualizes the trade-off between the tp and fp rates and includes misclassification costs regardless of the sample proportion. Studies have widely adopted this evaluation metric to evaluate predictive models for construction company insolvent failure with imbalanced datasets [50,64,66,89,90,114,115,125,138]. In addition, the precision–recall curve (PRC) has also been used to evaluate construction company insolvent failure models as an alternative to the area under ROC curve [125,129].

4.3.3. Other Metrics

In addition to the metrics already mentioned, several less frequently used metrics exist for construction company insolvent failure research. For example, entropy is a probability metric used for evaluating the utility of attributes of data in building optimized decision tree classifiers [138]. The cumulative accuracy profile (CAP) curve has been used to visualize the predictive capability of a developed model and compares it with a perfect classification model and a random classification model [139]. Both the root mean squared error (RMSE) [129] and the mean absolute percentage error (MAPE) [140] have been used to evaluate how the predicted output deviates from the desired output. It is worth mentioning, however, that different metrics involve different tradeoffs, often making it difficult to determine the best evaluation metric in practice [128]. For this reason, the ability of a given (or novel) predictive model should be evaluated by multiple metrics, resulting in a multi-dimensional metric space.

4.4. Research Gaps and Future Directions

Although an increasing number of articles have been published in the construction company insolvent failure domain, several limitations can be identified and addressed to recognize research gaps and outline future research directions.

4.4.1. Selecting a Broader Data Sample

Much research has selected listed (large) construction companies from developed regions for their research samples. There is, therefore, the potential for future studies to be carried out of construction companies in much less-studied developing regions because of the different accounting principles, construction market characteristics, and economic conditions involved. The high-performing predictive models for developed regions may not work well for those developing regions, and it is necessary to test the universality of those models and increase their prediction potentials accordingly. More involvement is also suggested for construction SMEs because of their growing proportion within the construction industry and increased vulnerability to economic volatilities, thus making them more likely to suffer from corporate insolvent failure. However, it is likely that future work of this kind would be subject to major data availability restrictions.

4.4.2. Incorporating More Heterogeneous Variables

Research to date has relied heavily on data from financial statements (financial ratios) for model development, but has neglected the utilization of non-financial indicators such as such as project characteristics, economic conditions, construction industry dynamics, and stock market performance that can complement the existing financial ratios and provide additional information for model development. It is suggested that future work incorporate variables from different data sources to constitute a more heterogeneous dataset. In addition to numerical data, textual data such as management discussions outlined in annual filings, actions described in corporate social responsibility (CSR) reports, and sentiment information disclosed in corporate news could also be utilized with the help of natural language processing (NLP) techniques, which would inevitably require more advanced classification techniques. In this case, it would be necessary to consider the interactions and synergies between different variables and choose an optimal set of input variables, and using feature selection and removing redundant variables could reduce storage requirements and computational time and enhance training and prediction efficiency.

4.4.3. Balancing Model Predictability and Interpretability

An increasing number of studies have adopted complex models such as ensemble methods and deep learning methods to predict construction company insolvent failure, and many have demonstrated an empirically superior predictive ability compared to other single artificial intelligence models and statistical models. These complex models, however, are often criticized, as they function like "black boxes" whose internal structure and learned parameters are not fully interpretable. They cannot explore the relationships or causalities between the various input variables involved and business insolvent failure. Therefore, they face the limitation of being unable to provide corporate strategies to improve business conditions and reduce default risks. This yields a gap in balancing model predictability and interpretability when researchers choose the right models for their predictive and interpretable purposes in future research.

4.4.4. Quantifying the Causality and Intercorrelation of Variables

Finally, as most extant research has focused on building complex models for better predictive accuracy, it is also critical to quantify the causality of selected variables to the ultimate company insolvent failure as well as the intercorrelation between selected variables. When doing so, it is helpful to identify and incorporate variables that have a causal impact on company insolvent failure into the predictive model, thus reducing dataset dimensionality and enhancing model predictive power. This approach not only refines the model but also offers practical advice to stakeholders and management teams regarding the underlying causes of company insolvent failure, enabling them to focus on improving business operations. This analysis can be achieved by using data filter methods, data wrapper methods, and dimensionality reduction techniques. Additionally, employing approaches such as cross-lagged correlation analysis and longitudinal path analysis can be instrumental in examining the relationships among selected variables over time.

5. Conclusions

This study reviewed 93 key journal articles relating to predicting construction company insolvent failure using both scientometric analysis and qualitative discussion. The results of the scientometric analysis reveal the proliferation of relevant research over the last 12 years or so. The *Journal of Construction Engineering and Management*, *Construction Management and Economics*, *Engineering Construction and Architectural Management*, and *Expert Systems with Applications* have made the most contributions in terms of article numbers. H.P. Tserng, D. Arditi, and P.C. Chen are the most influential researchers who have both produced the most articles and received the most citations. A chronological keywords analysis found that the research focus—business insolvent failure prediction—can also be considered as "bankruptcy prediction", "financial distress prediction", or "insolvency prediction", but the "prediction models" have gradually shifted from "statistical techniques" such as "z-score", "discriminant analysis model", and "regression analysis" to more advanced "machine learning" techniques such as "neural networks" and "support vector machine". Finally, the articles receiving the most citations were identified, visualized, and discussed.

A detailed qualitative discussion was conducted to set out data collection and processing procedures, compare different predictive models, summarize performance evaluation metrics, point out research gaps, and suggest future research directions. The extant research focuses heavily on studying publicly traded construction companies in developed regions. With the help of big data analytical techniques, the average sample size has significantly increased from 142 for articles on and before 2010 to 4005 for articles after 2010. It is also found that financial ratio variables are commonly utilized for model development, followed by variables representing company characteristics and economic conditions. Due to the imbalanced nature of construction company insolvent failure datasets, under-sampling and oversampling techniques need to be utilized to yield more balanced datasets. Artificial intelligence models have recently gained more popularity than statistical models due to their generally superior prediction ability. However, they are also criticized because of their "black boxes" nature and, thus, lack of interpretability. Different model performance evaluation metrics are also discussed in terms of usage and limitations, and it is suggested that multiple metrics should be used to evaluate a given (or novel) predictive model, since different metrics show different tradeoffs. Finally, several research gaps and corresponding future directions are identified in areas: selecting a broader data sample, incorporating more heterogeneous variables, balancing model predictability and interpretability, and quantifying the causality and intercorrelation of variables.

This review-based study combines a novel scientometric analysis approach and traditional qualitative discussion to provide a holistic review of research related to predicting construction company insolvent failure. It provides an overview of relevant research works from both visual and textual perspectives and outlines research gaps and directions for future studies. One limitation of this study is that it focused on only English language journal articles from Web of Science, Scopus and Engineering Village databases when selecting the literature sample. Future studies will require extending the inclusion criteria by considering other publication outlets (e.g., conference articles, books) and articles published in other languages and other sorts of databases (e.g., PubMed, Google Scholar).

Author Contributions: Conceptualization, J.W. and M.L.; methodology, J.W. and M.L.; software, J.W.; validation, M.L., M.S. and J.C.; formal analysis, J.W., M.L. and J.C.; data curation, J.W. and M.L.; writing—original draft preparation, J.W. and M.L.; writing—review and editing, J.W., M.L., M.S. and J.C.; visualization, J.W. and M.L.; supervision, M.L., M.S. and J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Shandong Province Natural Science Foundation, grant number ZR2023QG168.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Author Mao Li was employed by PetroChina Huabei Oilfield Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Robinson, G.; Leonard, J.; Whittington, T. Future of construction. In *A Global Forecast for Construction to 2030*; Oxford Economics: London, UK, 2021.
2. Choi, H.; Son, H.; Kim, C. Predicting financial distress of contractors in the construction industry using ensemble learning. *Expert Syst. Appl.* **2018**, *110*, 1–10. [CrossRef]
3. Tserng, H.P.; Lin, G.-F.; Tsai, L.K.; Chen, P.-C. An enforced support vector machine model for construction contractor default prediction. *Autom. Constr.* **2011**, *20*, 1242–1249. [CrossRef]
4. Tserng, H.P.; Liao, H.-H.; Jaselskis, E.J.; Tsai, L.K.; Chen, P.-C. Predicting construction contractor default with barrier option model. *J. Constr. Eng. Manag.* **2012**, *138*, 621–630. [CrossRef]
5. Tsang, Y.K.; Abdelmageed, S.; Zayed, T. Development of a contractor failure prediction model using analytic network process. *J. Arch. Eng.* **2021**, *27*, 04021006. [CrossRef]
6. Cheng, M.-Y.; Hoang, N.-D. Evaluating contractor financial status using a hybrid fuzzy instance based classifier: Case study in the construction industry. *IEEE Trans. Eng. Manag.* **2015**, *62*, 184–192. [CrossRef]
7. Horta, I.; Camanho, A. Company failure prediction in the construction industry. *Expert Syst. Appl.* **2013**, *40*, 6253–6257. [CrossRef]
8. Tserng, H.P.; Liao, H.-H.; Tsai, L.K.; Chen, P.-C. Predicting construction contractor default with option-based credit models—models' performance and comparison with financial ratio models. *J. Constr. Eng. Manag.* **2011**, *137*, 412–420. [CrossRef]
9. Yeh, C.-C.; Chi, D.-J.; Hsu, M.-F. A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Syst. Appl.* **2010**, *37*, 1535–1541. [CrossRef]
10. Mason, R.J.; Harris, F.C. Predicting company failure in the construction industry. *Proc. Inst. Civ. Eng.* **1979**, *66*, 301–330.
11. Kangari, R.; Farid, F.; Elgharib, H.M. Financial performance analysis for construction industry. *J. Constr. Eng. Manag.* **1992**, *118*, 349–361. [CrossRef]
12. Hall, G. Factors distinguishing survivors from failures amongst small firms in the UK construction sector. *J. Manag. Stud.* **1994**, *31*, 737–760. [CrossRef]
13. Jang, Y.; Jeong, I.; Cho, Y.K. Identifying impact of variables in deep learning models on bankruptcy prediction of construction contractors. *Eng. Constr. Arch. Manag.* **2021**, *28*, 3282–3298. [CrossRef]
14. Edum-Fotwe, F.; Price, A.; Thorpe, A. A review of financial ratio tools for predicting contractor insolvency. *Constr. Manag. Econ.* **1996**, *14*, 189–198. [CrossRef]
15. Wong, J.M.; Ng, T.S.T. Company failure in the construction industry: A critical review and a future research agenda. In Proceedings of the 24th FIG International Congress: Facing the Challenges—Building the Capacity, Sydney, Australia, 11–16 April 2010.
16. Alaka, H.A.; Oyedele, L.O.; Owolabi, H.A.; Ajayi, S.O.; Bilal, M.; Akinade, O.O. Methodological approach of construction business failure prediction studies: A review. *Constr. Manag. Econ.* **2016**, *34*, 808–842. [CrossRef]
17. Alaka, H.A.; Oyedele, L.O.; Owolabi, H.A.; Bilal, M.; Ajayi, S.O.; Akinade, O.O. Insolvency of small civil engineering firms: Critical strategic factors. *J. Prof. Issues Eng. Educ. Pr.* **2017**, *143*, 04016026. [CrossRef]
18. Assaad, R.; El-Adaway, I.H. Enhancing the knowledge of construction business failure: A social network analysis approach. *J. Constr. Eng. Manag.* **2020**, *146*, 04020052. [CrossRef]

19. Felizardo, K.R.; Salleh, N.; Martins, R.M.; Mendes, E.; MacDonell, S.G.; Maldonado, J.C. Using visual text mining to support the study selection activity in systematic literature reviews. In Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement, Banff, AB, Canada, 22–23 September 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 77–86.
20. Keim, D.A. Information visualization and visual data mining. *IEEE Trans. Vis. Comput. Graph.* **2002**, *8*, 1–8. [CrossRef]
21. Wang, J.; Chen, J.; Hu, Y. A science mapping approach based review of model predictive control for smart building operation management. *J. Civ. Eng. Manag.* **2022**, *28*, 661–679. [CrossRef]
22. Jin, R.; Zou, P.X.; Piroozfar, P.; Wood, H.; Yang, Y.; Yan, L.; Han, Y. A science mapping approach based review of construction safety research. *Saf. Sci.* **2019**, *113*, 285–297. [CrossRef]
23. Sepasgozar, S.; Karimi, R.; Farahzadi, L.; Moezzi, F.; Shirowzhan, S.; Ebrahimzadeh, S.M.; Hui, F.; Aye, L. A Systematic content review of artificial intelligence and the internet of things applications in smart home. *Appl. Sci.* **2020**, *10*, 3074. [CrossRef]
24. Kim, H.; Choi, H.; Kang, H.; An, J.; Yeom, S.; Hong, T. A systematic review of the smart energy conservation system: From smart homes to sustainable smart cities. *Renew. Sustain. Energy Rev.* **2021**, *140*, 110755. [CrossRef]
25. Kuizinienė, D.; Krilavičius, T.; Damaševičius, R.; Maskeliūnas, R. Systematic review of financial distress identification using artificial intelligence methods. *Appl. Artif. Intell.* **2022**, *36*, 2138124. [CrossRef]
26. Veganzones, D.; Severin, E. Corporate failure prediction models in the twenty-first century: A review. *Eur. Bus. Rev.* **2021**, *33*, 204–226. [CrossRef]
27. Shi, Y.; Li, X. A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms. *Heliyon* **2019**, *5*, e02997. [CrossRef] [PubMed]
28. Wang, J.; Ashuri, B. Predicting ENR Construction cost index using machine-learning algorithms. *Int. J. Constr. Educ. Res.* **2017**, *13*, 47–63. [CrossRef]
29. Assaad, R.; El-Adaway, I.H.; Abotaleb, I.S. Predicting project performance in the construction industry. *J. Constr. Eng. Manag.* **2020**, *146*, 04020030. [CrossRef]
30. Li, S.; Zheng, W.; Xu, T.; Wang, Y. Artificial neural network model for predicting the local compression capacity of stirrups-confined concrete. *Structures* **2022**, *41*, 943–956. [CrossRef]
31. Chen, C. Science Mapping: A Systematic Review of the Literature. *J. Data Inf. Sci.* **2017**, *2*, 1–40. [CrossRef]
32. Van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [CrossRef]
33. Waltman, L.; van Eck, N.J.; Noyons, E.C.M. A unified approach to mapping and clustering of bibliometric networks. *J. Informetr.* **2010**, *4*, 629–635. [CrossRef]
34. Van Eck, N.J.; Waltman, L. Visualizing bibliometric networks. In *Measuring Scholarly Impact*; Springer: Cham, Switzerland, 2014; pp. 285–320.
35. Waltman, L.; Van Eck, N.J. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **2013**, *86*, 1–14. [CrossRef]
36. Lee, S.; Choi, W.S. A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Syst. Appl.* **2013**, *40*, 2941–2946. [CrossRef]
37. Heo, J.; Yang, J.Y. AdaBoost based bankruptcy forecasting of Korean construction companies. *Appl. Soft Comput.* **2014**, *24*, 494–499. [CrossRef]
38. Kapliński, O. Usefulness and credibility of scoring methods in construction industry. *J. Civ. Eng. Manag.* **2008**, *14*, 21–28. [CrossRef]
39. Altman, E.I.; Marco, G.; Varetto, F. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *J. Bank. Financ.* **1994**, *18*, 505–529. [CrossRef]
40. Kangari, R. Business failure in construction industry. *J. Constr. Eng. Manag.* **1988**, *114*, 172–190. [CrossRef]
41. Kale, S.; Arditi, D. Business failures: Liabilities of newness, adolescence, and smallness. *J. Constr. Eng. Manag.* **1998**, *124*, 458–464. [CrossRef]
42. Russell, J.S.; Jaselskis, E.J. Predicting construction contractor failure prior to contract award. *J. Constr. Eng. Manag.* **1992**, *118*, 791–811. [CrossRef]
43. Russell, J.S.; Zhai, H. Predicting Contractor failure using stochastic dynamics of economic and financial variables. *J. Constr. Eng. Manag.* **1996**, *122*, 183–191. [CrossRef]
44. Lam, K.C.; Palaneeswaran, E.; Yu, C.-Y. A support vector machine model for contractor prequalification. *Autom. Constr.* **2009**, *18*, 321–329. [CrossRef]
45. Chen, J.-H. Developing SFNN models to predict financial distress of construction companies. *Expert Syst. Appl.* **2012**, *39*, 823–827. [CrossRef]
46. Sueyoshi, T.; Goto, M. DEA–DA for bankruptcy-based performance assessment: Misclassification analysis of Japanese construction industry. *Eur. J. Oper. Res.* **2009**, *199*, 576–594. [CrossRef]
47. Sánchez-Lasheras, F.; de Andrés, J.; Lorca, P.; Juez, F.J.d.C. A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy. *Expert Syst. Appl.* **2012**, *39*, 7512–7523. [CrossRef]
48. Dikmen, I.; Birgonul, M.T.; Ozorhon, B.; Sapci, N.E. Using analytic network process to assess business failure risks of construction firms. *Eng. Constr. Arch. Manag.* **2010**, *17*, 369–386. [CrossRef]
49. Ng, S.T.; Wong, J.M.; Zhang, J. Applying Z-score model to distinguish insolvent construction companies in China. *Habitat Int.* **2011**, *35*, 599–607. [CrossRef]

50. Zoričák, M.; Gnip, P.; Drotár, P.; Gazda, V. Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets. *Econ. Model.* **2020**, *84*, 165–176. [CrossRef]
51. Tian, S.; Yu, Y.; Zhou, M. data sample selection issues for bankruptcy prediction. *Risk Hazards Crisis Public Policy* **2015**, *6*, 91–116. [CrossRef]
52. Chen, F.; Hope, O.-K.; Li, Q.; Wang, X. Financial reporting quality and investment efficiency of private firms in emerging markets. *Account. Rev.* **2011**, *86*, 1255–1288. [CrossRef]
53. Scott, J. The probability of bankruptcy: A comparison of empirical predictions and theoretical models. *J. Bank. Financ.* **1981**, *5*, 317–344. [CrossRef]
54. Becchetti, L.; Sierra, J. Bankruptcy risk and productive efficiency in manufacturing firms. *J. Bank. Financ.* **2003**, *27*, 2099–2120. [CrossRef]
55. I Altman, E.; Saunders, A. Credit risk measurement: Developments over the last 20 years. *J. Bank. Financ.* **1997**, *21*, 1721–1742. [CrossRef]
56. Appiah, K.O.; Chizema, A.; Arthur, J. Predicting corporate failure: A systematic literature review of methodological issues. *Int. J. Law Manag.* **2015**, *57*, 461–485. [CrossRef]
57. Kaski, S.; Sinkkonen, J.; Peltonen, J. Bankruptcy analysis with self-organizing maps in learning metrics. *IEEE Trans. Neural Netw.* **2001**, *12*, 936–947. [CrossRef] [PubMed]
58. Chauhan, N.; Ravi, V.; Chandra, D.K. Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks. *Expert Syst. Appl.* **2009**, *36*, 7659–7665. [CrossRef]
59. Micha, B. Analysis of business failures in France. *J. Bank. Financ.* **1984**, *8*, 281–291. [CrossRef]
60. Beaver, W.H. Financial ratios as predictors of failure. *J. Account. Res.* **1966**, *4*, 71–111. [CrossRef]
61. Agarwal, V.; Taffler, R. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *J. Bank. Financ.* **2008**, *32*, 1541–1551. [CrossRef]
62. Balcaen, S.; Ooghe, H. 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *Br. Account. Rev.* **2006**, *38*, 63–93. [CrossRef]
63. Murray, G.D. A cautionary note on selection of variables in discriminant analysis. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1977**, *26*, 246–250. [CrossRef]
64. Thanh-Long, N.; Hong-Chuong, L. A back propagation neural network model with the synthetic minority over-sampling technique for construction company bankruptcy prediction. *Int. J. Sustain. Constr. Eng. Technol.* **2022**, *13*, 68–79. [CrossRef]
65. Khanzadi, M.; Eshtehardian, E.; Esfahani, M.M. Cash flow forecasting with risk consideration using bayesian belief networks (BBNS). *J. Civ. Eng. Manag.* **2017**, *23*, 1045–1059. [CrossRef]
66. Sousa, A.; Braga, A.; Cunha, J. Impact of macroeconomic indicators on bankruptcy prediction models: Case of the Portuguese construction sector. *Math. Biosci. Eng.* **2022**, *6*, 405–432. [CrossRef]
67. Pestov, V. Is the k-NN classifier in high dimensions affected by the curse of dimensionality? *Comput. Math. Appl.* **2013**, *65*, 1427–1437. [CrossRef]
68. Tsai, C.-F.; Hsiao, Y.-C. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decis. Support Syst.* **2010**, *50*, 258–269. [CrossRef]
69. Costa, M.; Lisboa, I.; Gameiro, A. Is the Financial report quality important in the default prediction? SME Portuguese construction sector evidence. *Risks* **2022**, *10*, 98. [CrossRef]
70. Bolón-Canedo, V.; Sánchez-Maroño, N.; Alonso-Betanzos, A. *Feature Selection for High-Dimensional Data*; Springer International Publishing: Cham, Switzerland, 2015.
71. Li, Y.; Li, T.; Liu, H. Recent advances in feature selection and its applications. *Knowl. Inf. Syst.* **2017**, *53*, 551–577. [CrossRef]
72. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 1060–1073. [CrossRef]
73. Skogsvik, K.; Skogsvik, S. On the choice based sample bias in probabilistic bankruptcy prediction. *Investig. Manag. Financ. Innov.* **2013**, *10*, 29–37.
74. Fernández, A.; del Jesus, M.J.; Herrera, F. On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets. *Inf. Sci.* **2010**, *180*, 1268–1291. [CrossRef]
75. Veganzones, D.; Séverin, E. An investigation of bankruptcy prediction in imbalanced datasets. *Decis. Support Syst.* **2018**, *112*, 111–124. [CrossRef]
76. Seiffert, C.; Khoshgoftaar, T.M.; Van Hulse, J.; Napolitano, A. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2009**, *40*, 185–197. [CrossRef]
77. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
78. Liu, A.Y.C. The Effect of Oversampling and Undersampling on Classifying Imbalanced Text Datasets. Ph.D. Thesis, University of Texas at Austin, Austin, TX, USA, 2004.
79. Jang, Y.; Jeong, I.-B.; Cho, Y.K.; Ahn, Y. Predicting business failure of construction contractors using long short-term memory recurrent neural network. *J. Constr. Eng. Manag.* **2019**, *145*, 04019067. [CrossRef]
80. Jang, Y.; Jeong, I.; Cho, Y.K. Business failure prediction of construction contractors using a lstm rnn with accounting, construction market, and macroeconomic variables. *J. Manag. Eng.* **2020**, *36*, 04019039. [CrossRef]
81. Drummond, C.; Holte, R.C. C4. 5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. In Proceedings of the Workshop on Learning from Imbalanced Datasets II, Washington, DC, USA, 21 August 2003; Volume 11, pp. 1–8.

82. Kaur, P.; Gosain, A. Comparing the behavior of oversampling and undersampling approach of class imbalance learning by combining class imbalance problem with noise. In *ICT Based Innovations*; Springer: Singapore, 2018; pp. 23–30. [CrossRef]
83. Mohammed, R.; Rawashdeh, J.; Abdullah, M. Machine learning with oversampling and undersampling techniques: Overview study and experimental results. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 243–248.
84. Alexakis, P. 'Altman Z-score model' and prediction of business failures. *Int. J. Monet. Econ. Financ.* **2008**, *1*, 329. [CrossRef]
85. Chan, J.K.; Tam, C.M.; Cheung, R.K. Construction firms at the crossroads in Hong Kong: Going insolvency or seeking opportunity. *Eng. Constr. Archit. Manag.* **2005**, *12*, 111–124. [CrossRef]
86. Pompe, P.P.; Feelders, A.J. Using machine learning, neural networks, and statistics to predict corporate bankruptcy. *Comput.–Aided Civ. Infrastruct. Eng.* **1997**, *12*, 267–276. [CrossRef]
87. Marcinkevičius, R.; Kanapickienė, R. Bankruptcy prediction in the sector of construction in Lithuania. *Procedia Soc. Behav. Sci.* **2014**, *156*, 553–557. [CrossRef]
88. Kanapickiene, R.; Marcinkevičius, R. Possibilities to apply classical bankruptcy prediction models in the construction sector in Lithuania. *Econ. Manag.* **2014**, *19*, 317–332. [CrossRef]
89. Karas, M.; Režňáková, M. Building a bankruptcy prediction model: Could information about past development increase model accuracy? *Pol. J. Manag. Stud.* **2018**, *17*, 116–130. [CrossRef]
90. Alaka, H.; Oyedele, L.; Owolabi, H.; Akinade, O.; Bilal, M.; Ajayi, S. A big data analytics approach for construction firms failure prediction models. *IEEE Trans. Eng. Manag.* **2018**, *66*, 689–698. [CrossRef]
91. Abidali, A.F.; Harris, F. A methodology for predicting company failure in the construction industry. *Constr. Manag. Econ.* **1995**, *13*, 189–196. [CrossRef]
92. Langford, D.; Iyagba, R.; Komba, D.M. Prediction of solvency in construction companies. *Constr. Manag. Econ.* **1993**, *11*, 317–325. [CrossRef]
93. Makeeva, E.; Neretina, E. The prediction of bankruptcy in a construction industry of Russian Federation. *J. Mod. Account. Audit.* **2013**, *9*, 256.
94. Fadel, H. The predictive power of financial ratios in the british construction industrY. *J. Bus. Financ. Account.* **1977**, *4*, 339–352. [CrossRef]
95. Tserng, H.P.; Chen, P.-C.; Huang, W.-H.; Lei, M.C.; Tran, Q.H. Prediction of default probability for construction firms using the logit model. *J. Civ. Eng. Manag.* **2014**, *20*, 247–255. [CrossRef]
96. Adeleye, T.; Huang, M.; Huang, Z.; Sun, L. Predicting loss for large construction companies. *J. Constr. Eng. Manag.* **2013**, *139*, 1224–1236. [CrossRef]
97. Muscettola, M. Probability of default estimation for construction firms. *Int. Bus. Res.* **2014**, *7*, 153. [CrossRef]
98. Altman, E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Financ.* **1968**, *23*, 589–609. [CrossRef]
99. Dhamnetiya, D.; Goel, M.K.; Jha, R.P.; Shalini, S.; Bhattacharyya, K. How to perform discriminant analysis in medical research? Explained with Illustrations. *J. Lab. Physicians* **2022**, *14*, 511–520. [CrossRef]
100. Cramer, D. *Advanced Quantitative Data Analysis*; McGraw-Hill Education: London, UK, 2003.
101. Ohlson, J.A. Financial ratios and the probabilistic prediction of bankruptcy. *J. Account. Res.* **1980**, *18*, 109. [CrossRef]
102. Jackson, R.H.; Wood, A. The performance of insolvency prediction and credit risk models in the UK: A comparative study. *Br. Account. Rev.* **2013**, *45*, 183–202. [CrossRef]
103. Büyüköztürk, Ş.; Çokluk-Bökeoğlu, Ö. Discriminant function analysis: Concept and application. *Eurasian J. Educ. Res.* **2008**, *33*, 73–92.
104. Uyanık, G.K.; Güler, N. A study on multiple linear regression analysis. *Procedia Soc. Behav. Sci.* **2013**, *106*, 234–240. [CrossRef]
105. Keramati, A.; Yousefi, N. A proposed classification of data mining techniques in credit scoring. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Kuala Lumpur, Malaysia, 22–24 January 2011.
106. Jing, Z.; Fang, Y. Predicting US bank failures: A comparison of logit and data mining models. *J. Forecast.* **2018**, *37*, 235–256. [CrossRef]
107. Al-Sobiei, O.S.; Arditi, D.; Polat, G. Predicting the risk of contractor default in Saudi Arabia utilizing artificial neural network (ANN) and genetic algorithm (GA) techniques. *Constr. Manag. Econ.* **2005**, *23*, 423–430. [CrossRef]
108. Alaka, H.A.; Oyedele, L.O.; Owolabi, H.A.; Bilal, M.; Ajayi, S.O.; Akinade, O.O. A framework for big data analytics approach to failure prediction of construction firms. *Appl. Comput. Inform.* **2018**, *16*, 207–222. [CrossRef]
109. Tserng, H.P.; Ngo, T.L.; Chen, P.C.; Tran, L.Q. A grey system theory-based default prediction model for construction firms. *Comput. Civ. Infrastruct. Eng.* **2015**, *30*, 120–134. [CrossRef]
110. Karas, M.; Režňáková, M. The stability of bankruptcy predictors in the construction and manufacturing industries at various times before bankruptcy. *Econ. Manag.* **2017**, *20*, 116–133. [CrossRef]
111. Karas, M.; Režňáková, M. Predicting the bankruptcy of construction companies: A cart-based model. *Eng. Econ.* **2017**, *28*, 145–154. [CrossRef]
112. Pastor-Vega, D.; Fernández-Miguélez, S.M.; Diéguez-Soto, J.; Fernández-Gámez, M.A. A multi-class model to predict the result of the legal insolvency proceedings. *J. Sci. Ind. Res.* **2019**, *78*, 742–745.
113. Balina, R.; Idasz-Balina, M.; Achsani, N.A. Predicting insolvency of the construction companies in the creditworthiness assessment process—Empirical evidence from Poland. *J. Risk Financ. Manag.* **2021**, *14*, 453. [CrossRef]

114. Karminsky, A.; Burekhin, R. Comparative analysis of methods for forecasting bankruptcies of Russian construction companies. *Бизнес-информатика* **2019**, *13*, 52–66. [CrossRef]
115. Jeong, J.; Kim, C. Comparison of machine learning approaches for medium-to-long-term financial distress predictions in the construction industry. *Buildings* **2022**, *12*, 1759. [CrossRef]
116. Chen, J.-H.; Su, M.-C.; Badjie, B.A. Exploring and weighting features for financially distressed construction companies using swarm inspired projection algorithm. *Adv. Eng. Inform.* **2016**, *30*, 376–389. [CrossRef]
117. Sun, J.; Liao, B.; Li, H. AdaBoost and bagging ensemble approaches with neural network as base learner for financial distress prediction of chinese construction and real estate companies. *Recent Pat. Comput. Sci.* **2013**, *6*, 47–59. [CrossRef]
118. Sun, J.; Li, H.; Huang, Q.H.; He, K.Y. Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowl.-Based Syst.* **2014**, *57*, 41–56. [CrossRef]
119. Kim, H.; Kim, H.; Moon, H.; Ahn, H. A weight-adjusted voting algorithm for ensembles of classifiers. *J. Korean Stat. Soc.* **2011**, *40*, 437–449. [CrossRef]
120. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
121. Li, Q.-F.; Song, Z.-M. High-performance concrete strength prediction based on ensemble learning. *Constr. Build. Mater.* **2022**, *324*, 126694. [CrossRef]
122. Mienye, I.D.; Obaido, G.; Aruleba, K.; Dada, O.A. Enhanced prediction of chronic kidney disease using feature selection and boosted classifiers. In Proceedings of the Intelligent Systems Design and Applications: 21st International Conference on Intelligent Systems Design and Applications (ISDA 2021), Online, 13–15 December 2021; Springer International Publishing: Cham, Switzerland, 2021; pp. 527–537.
123. Rojarath, A.; Songpan, W. Probability-weighted voting ensemble learning for classification model. *J. Adv. Inf. Technol.* **2020**, *11*, 217–227. [CrossRef]
124. Liang, M.; Chang, T.; An, B.; Duan, X.; Du, L.; Wang, X.; Miao, J.; Xu, L.; Gao, X.; Zhang, L.; et al. A stacking ensemble learning framework for genomic prediction. *Front. Genet.* **2021**, *12*, 600040. [CrossRef]
125. Jing, J.; Yan, W.; Deng, X. A hybrid model to estimate corporate default probabilities in China based on zero-price probability model and long short-term memory. *Appl. Econ. Lett.* **2021**, *28*, 413–420. [CrossRef]
126. Tang, X.; Li, S.; Tan, M.; Shi, W. Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *J. Forecast.* **2020**, *39*, 769–787. [CrossRef]
127. Yang, Y.; Lv, H.; Chen, N. A Survey on ensemble learning under the era of deep learning. *Artif. Intell. Rev.* **2023**, *56*, 5545–5589. [CrossRef]
128. Caruana, R.; Niculescu-Mizil, A. Data mining in metric space: An empirical analysis of supervised learning performance criteria. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA, 24–27 August 2003; pp. 69–78.
129. Chen, N.; Chen, A.; Ribeiro, B. Towards tangible benefits of corporate failure prediction with business sector: A comparative study. *Intell. Decis. Technol.* **2016**, *10*, 431–442. [CrossRef]
130. Lucanera, J.P.; Fabregat-Aibar, L.; Scherger, V.; Vigier, H. Can the SOM analysis predict business failure using capital structure theory? Evidence from the subprime crisis in spain. *Axioms* **2020**, *9*, 46. [CrossRef]
131. Giriūniene, G.; Giriūnas, L.; Morkunas, M.; Brucaite, L. A comparison on leading methodologies for bankruptcy prediction: The case of the construction sector in Lithuania. *Economies* **2019**, *7*, 82. [CrossRef]
132. Awad, A.; Fayek, A.R. Contractor default prediction model for surety bonding. *Can. J. Civ. Eng.* **2012**, *39*, 1027–1042. [CrossRef]
133. Awad, A.; Fayek, A.R. Adaptive learning of contractor default prediction model for surety bonding. *J. Constr. Eng. Manag.* **2013**, *139*, 694–704. [CrossRef]
134. Rybárová, D.; Majdúchová, H.; Štetka, P.; Luščíková, D. Reliability and accuracy of alternative default prediction models: Evidence from slovakia. *Int. J. Financ. Stud.* **2021**, *9*, 65. [CrossRef]
135. Tinoco, M.H.; Wilson, N. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *Int. Rev. Financ. Anal.* **2013**, *30*, 394–419. [CrossRef]
136. Kumar, P.R.; Ravi, V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques—A review. *Eur. J. Oper. Res.* **2007**, *180*, 1–28. [CrossRef]
137. Huang, J.; Ling, C. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 299–310. [CrossRef]
138. Bal, J.; Cheung, Y.; Wu, H.-C. Entropy for business failure prediction: An improved prediction model for the construction industry. *Adv. Decis. Sci.* **2013**, *2013*, 1–14. [CrossRef]
139. Acosta-González, E.; Fernández-Rodríguez, F.; Ganga, H. Predicting corporate financial failure using macroeconomic variables and accounting data. *Comput. Econ.* **2019**, *53*, 227–257. [CrossRef]
140. Chen, H.L. Model for predicting financial performance of development and construction corporations. *J. Constr. Eng. Manag.* **2009**, *135*, 1190–1200. [CrossRef]