PREDICTING FINANCIAL DISTRESS IN SELECT INDIAN LISTED COMPANIES THROUGH MACHINE LEARNING ALGORITHMS

[1]Ramesh Kannan K and Dr. Firdaus Bashir

ABSTRACT

In the era of digital revolution, identifying and predicting the signs of financial distress in advance have become a strategic consideration. It enables in preserving the confidence of investors, while making informed decision to avoid distress. The financial distress concerns of listed companies become a matter of debate in academic and professional realms, as early detection enables companies, investors and regulators to implement corrective measures even before it strikes. In developing Indian markets, the vulnerabilities of businesses and structural imbalance have added to the pressure of wanting to have a sound predictive framework. This study aims to examine the effectiveness of machine learning algorithms in forecasting financial distress in select Indian listed companies. This research is based on 7,705 observations derived from 1,938 Indian listed companies across 11 sectors. The financial distress was analysed through three well-established traditional models namely Altman Z-score, Fich and Slezak, as well as Zmijewski index. In addition, key financial ratios representing profitability, leverage, liquidity, and operational efficiency are incorporated into it. The outcomes from these conventional Models were examined with the machine Learning algorithms encompassing of logistic regression, decision tree, K nearest neighbours and random forest. The predictive performance of these models was then evaluated on metrics such as precision, recall, F1 score and area under curve measure. The results indicate that the random algorithm shows strong predictive performance and delivered the highest accuracy of 93% for Model 1- Altman Z-score, 99% for Model 2- Fich and Slezak and Model 3 - Zmijewski. The results evidenced that asset turnover, return on asset, interest coverage ratio, and debt ratio have the significant impact on financial distress under different models.

Keywords— Altman Z-score, area under curve, Fich and Slezak, financial distress, F1 score, Indian listed companies, Machine learning, precision, random forest, Zmijewski index.

1. INTRODUCTION

Financial distress prediction has emerged as a vital area of investigation for securing the stability and continuity of companies and stakeholders like creditors, regulators. (Ha, Dang, & Tran, 2023). The competency to identify and forecast this concern accurately permits management to take remedial measures to avoid operational failure and the risk of bankruptcy, which disrupts financial markets and wider economy (Ha et al., 2023). Besides, investors and creditors can take proactive measures to equip themselves from such risks, thereby minimising losses (Tinoco, M., & Wilson, N2013). Financial distress can be described as a situation where a business is facing problems in generating adequate cash flows to meet their maturing financial obligations. (Ghazali, M. F., Mansor, N., & Saleh, N. 2020). This concept is different from bankruptcy, where the latter is a formal legal proceeding that arises when a company becomes insolvent. Companies address distress by restructuring or employing other avenues prior to bankruptcy to avail themselves from out-of-court remedies like insolvency (Felman et al., 2024; Kumar, A., & Sharma, R. (2022). Financial distress was earlier measured using univariate financial ratio developed by Beaver (1966) and has now progressed to be refined by multivariate models. Discriminant analysis techniques, as demonstrated through Altman's Z-score model, integrates multiple financial ratios into a single composite index which enables to enhance the estimation of financial distress (Altman, 1968).

[1]Department of Management, PSG Institute of Management, Coimbatore, Tamil Nadu, India.

E-mail: rameshkannanrk480@gmail.com and firdaus@psgim.ac.in
Far along, Ohlson (1980) presented a logistic regression framework to evaluate the possibility of distress within a definite period, which was further extended via Zmijewski's (1984) probit model (Ha et al., 2023). Although these standard models remain to serve as foundational pillars in distress prediction, their dependance on restrictive assumptions like the normal distribution and constant variance diminishes their applicability in varied and fast changing business scenario, predominantly in developing nations (Tinoco, M., & Wilson, N 2013).

Recent advancements in Machine Learning Algorithms (MLAs) have considerably enhanced the accuracy in prediction. As it is very good at detecting complex, nonlinear relationships without the strict limits of parametric constraints (Chen et al., 2021). Traditional statistical methods are often outperformed by MLAs, suitable for Indian diverse and emerging marketplaces. It provides flexible and scalable solutions after identifying patterns of risk (Mahesh, Singh & Kumar, 2025; Singh, M., Mishra, A., & Goel, A. 2023) and generate outputs highlighting the importance of features, key drivers of distress useful for managers to make informed decisions (Zhang, Z., Wu, C., Qu, S., & Chen, X. 2022). In spite of significant developments, most existing research has been concentrated in developed countries namely United States and Europe, where there are robust systems of data to underpin advanced model building and testing (Ha et al., 2023). On the contrary this domain is relatively under-researched in countries like India which is progressing to be a developing capital market. The present research uses predictive models based on MLAs to forecast distress in select Indian listed companies. It also tries to identify prominent financial ratios that are significantly consistent in predicting financial distress.

1. REVIEW OF LITERATURE

When companies are in financial distress and struggling with bill payments, loan repayments, or critical operations due to poor financial management, there are certain warning signs indicating deficit in cash flows, high leverage, and declining of liquidity ratios, which leads organisational crises (Rai et al., 2025). It could be caused via short-term liquidity problems or an extreme, recurrent losses demanding a legal insolvency procedure.

Various Techniques in forecasting Distress of select Companies

Initially statistical discriminant analysis were employed widely before the application of machine learning (Altman, 1968, Beaver, 1966). It represents cash-to-total liabilities ratio as the key financial measurement for predicting oncoming distress. Companies with lower cash-to-total liabilities ratios have higher difficulties in making their debt payments, and therefore the risk of default and insolvency is more likely to occur. This ratio ha been empirical validated in further researches, underscoring its importance in representing early warning signal of financial distress (Al Najjar & Anfimiadou, 2020; Nguyen & Ahmed, 2023). Beside liquidity, Return on Assets (ROA) and debt ratio are the significant indicators for detecting as it denotes that when a company is utilising its assets to earn profits, its ROA declines leading to the period of financial stress. Using diverse financial ratios Altman proposed a Multivariate Discriminant Analysis (MDA) to develop a composite score, named as Z-score. It is widely used as a benchmark, based on which recent MLMs are assessed (Nguyen & Ahmed, 2023).

The Emergence of MLAs

The use of MLAs based analysis allows efficient processing for large datasets while generating more accurate predictions and thereby helps in reducing false positives and generating accurate decisions (Muhammad Atif Khan Achakzai & Peng Juan 2022). In recent era, they are widely used to predict distress enabling their capacity for modelling nonlinearities and multifaceted interactions. Logistic regression is also used for binary prediction because of its interpretability and strong performance in predicting financial distress(Hosmer & Lemeshow, 2000) and is a popular standard with high predictive power (Sultana et al., 2023). Decision trees predict accurately therefore, widely used in predictive modelling purpose (Sahin et al., 2013). Random forests have exhibited strong performance when used in Indian and Southeast Asian companies (Lokanan & Ramzan, 2024, Mahesh et al., 2025). Resolving class imbalance through methods like Synthetic Minority Oversampling Technique (SMOTE) is pertinent to identifying distressed companies, which typically belong to the minority category (Kuizinė et al., 2022).

I. RESEARCH METHODOLOGY

Variable Measurement

The present research concentrates on three traditional models of financial distress namely: 1. Altman Z-score (1968), 2. Fich and Slezak's (2008) Dummy variable and 3. Zmijewski (1984) Index.

Model 1- Altman Z-score: The Z-score is calculated from five ratios: X_1(current assets minus current liabilities / total assets), X_2(Retained earnings / total assets), X_3(profit before tax and interest / total assets), X_4(book value of equity / total liabilities), and X_5(revenue / total assets) using the formula;

$Z = 0.717X_1 + 0.847X_2 + 0.107X_3 + 0.420X_4 + 0.998X_5$.

Model 2 - Fich and Slezak describe financial distress by means of a dummy variable that takes the value of 1 when the ratio of return to interest expense is less than 1, indicating that the firm’s returns are insufficient to cover its interest obligations, and 0 otherwise, indicating that the firm’s returns are sufficient to meet its interest expenses.

Model 3 - Zmijewski Index: It combines profitability, leverage, and liquidity ratios into a single score. The formula is given by:

$Z=-4.3-4.5\times ROA+5.7\times FINL+0.004\times LIQ$. Where ROA = Net profit divided by total asset, LIQ (current ratio) = Current assets divided by current liabilities and FINL (Financial Leverage) = Total debt divided by total assets.

Machine learning Approaches

MLAs like Logistic Regression, Random Forests, Decision Trees, and K-Nearest Neighbours (KNN) are employed in the study to analysis financial distress. Logistic Regression, as introduced by Berkson (1944), examines binary outcome variables, whereas the Decision Tree, associated with Belson (1959), divides the entire dataset into a large number of subsets indicating whether a company is financially healthy or financially distressed. Random Forest, developed by Breiman (2001), enables a deeper understanding of the financial indicators that influence decision-making in the overall evaluation of a company’s financial risk. This study evaluates the accuracy of different MLAs in comparison with the results generated by traditional financial distress models. In certain cases, financially distressed companies are few in number when compared to healthier firms, and this class imbalance affects the predictive accuracy of the minority class belonging to the distressed category. Therefore, to properly evaluate model performance, alternative metrics derived from the confusion matrix are used. The confusion matrix is a performance measurement tool for classification models and provides a clearer assessment of model effectiveness in distinguishing distressed firms from non-distressed ones. It is represented as a table which visualises the accuracy of prediction against actual and predicted values which is represented in Table I

TABLE I. THE CONFUSION MATRIX

		Actual	
		Positive	***Negative***
Predicted	***Positive***	True positives (TP)	False positives (FP) Type I error
	Negative	False negatives (FN) Type II error	True negative (TN)

The confusion matrix serves as the fundamental source for computing important classification evaluation metrics such as Precision, Recall, F1-score, and Accuracy. In addition, the Area Under the Receiver Operating Characteristic Curve (AUC–ROC) evaluates model performance across different classification thresholds. The ROC curve represents the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR), illustrating the model's ability to distinguish between positive and negative classes. The AUC value ranges from 0 to 1, where values closer to 1 indicate stronger discriminative capability. Generally, AUC values below 0.6 indicate weak predictive power, values between 0.8 and 0.9 suggest good predictive performance, and values above 0.9 represent excellent predictive ability. Precision measures the proportion of predicted positive cases that are actually positive, calculated as Precision = TP/(TP + FP). Recall, also referred to as sensitivity, indicates the proportion of actual positive cases that are correctly identified by the model and is calculated as Recall = TP/(TP + FN). The F1-score represents the harmonic mean of precision and recall, providing a balanced measure of model performance, particularly when class imbalance exists. Accuracy measures the overall proportion of correct predictions made by the model and is calculated as (TP + TN)/(TP + TN + FP + FN). In this study, models that achieve higher values across these evaluation metrics are considered to demonstrate superior predictive performance and stronger capability in identifying financially distressed firms.

Research data

The present study considers companies listed on both the National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE). Secondary financial data were collected from Yahoo Finance for the years 2022 to 2025. The final sample comprised 1,938 companies, generating 7,705 firm-year observations over the four-year period. Table II presents the sector-wise distribution of observations, calculated by multiplying the number of companies selected in each sector by four, as financial data for each company were examined for all four years. To ensure uniformity in classification, the sampled firms were categorized into 11 broad sectors in accordance with the standardized Global Industry Classification Standard (MOSPI, 2023).

TABLE II. OVERVIEW OF RESEARCH SAMPLES IN VARIOUS SECTORS

Sectors Classifications	Sector wise selected Companies	Sector wise Observations	Percentage
Industrials	406	1614	20.90%
Consumer	376	1494	19.40%
Basic Materials	322	1275	16.50%
Financial Services	214	854	11.10%
Technology	143	570	7.40%
Healthcare	138	550	7.10%
Consumer Defensive	134	531	6.90%

Real Estate	69	275	3.60%
Communication Services	65	260	3.40%
Utilities	35	140	1.80%
Energy	34	134	1.70%
Others	2	8	0.10%
Total	1938	7705	100%

Source: Compiled by Author

Variables and measurements

Table III shows the key financial ratios used to predict financial distress. This research analysis shows how these ratios when applied to machine learning acts as key input features for predictive models.

TABLE III. SUMMARY OF THE RATIOS USED IN THIS STUDY

Index Group	Ratio	Formula
Solvency	Debt to Equity	Total Liabilities / Shareholders' Equity
	Interest Coverage	Earnings Before Interest and Taxes (EBIT) / Interest Expense
	Debt Ratio	Total Liabilities / Total Assets
Profitability	Return on Assets (ROA)	Net Income / Total Assets
	Net Profit Margin	Net Income / Net Sales
	Return on Equity (ROE)	Net Income / Shareholders' Equity
	Operating Margin	Operating Income / Net Sales
Efficiency	Asset Turnover	Net Sales / Total Assets
Liquidity	Current Ratio	Current Assets / Current Liabilities
Market Value	Price to Earnings (PE)	Share Price / Earnings Per Share (EPS)

Source: Compiled by Author

Profitability ratios have been identified as strong predictors of financial distress as they capture company's financial performance and its efficiency in generating revenue (Demetriades and Owusu-Agyei, 2022). Companies' probability of default has been assessed using solvency ratios (Horobet et al., 2021). Earlier studies have incorporated liquidity ratios namely the current ratio and cash-to-total assets, as key explanatory financial indicators, while analysing distress (Song et al., 2014). Incorporating liquidity ratios in predictive models, enables the application of MLAs to uncover patterns and recognise liquidity driven indicators of financial distress. Price-to-earnings ratio are commonly used in distress because they incorporate forward-looking market information that complements accounting-based ratios. Apart from these financial ratios several financial data are extracted from financial statements including Total Assets, Current Assets, Total Liabilities, Current Liabilities, Retained Earnings, Total Revenue, EBIT for calculating financial distress by using three traditional models. All these ratios and results given by three models can be incorporated into ML to identify patterns associated with operational efficiency and to improve the predictive model's ability to identify the signs of distress embedded within operational metrics.

V. MACHINE LEARNING WORKFLOW

The analytical procedure commences with data collection, wherein the relevant dataset is obtained and examined to confirm its appropriateness for the study. The next stage is data preprocessing, which involves treating missing values and correcting data types to ensure consistency and accuracy. Thereafter, the scores of the three traditional models are calculated to derive the distress labels. Following data preparation, the dataset is divided into training and testing subsets, with the former used for model development and the latter for performance evaluation on unseen observations. The machine learning algorithms are then trained, while class imbalance is addressed through the application of the Synthetic Minority Oversampling Technique (SMOTE). The trained models are subsequently evaluated using reliability and validation techniques. In the final stage, visual outputs, including ROC curves, confusion matrices, feature importance plots, and decision tree diagrams, are generated and preserved along with the final results.

VI. RESULTS AND DISCUSSION

The Table IV represents the distressed and normal observations of three standard financial distress prediction models

TABLE IV. DISTRESSED AND NORMAL OBSERVATIONS OF THREE MODELS

Model	Total Observations	Healthy observations	Distressed observations
Altman Z-Score	7707	4831	2876
Fich & Slezak	7707	6775	932
Zmijewski	7707	4157	3550

Source: Compiled by Author

Table V shows that the Altman Z-Score model classifies 27.45% of the company-year observations as financially distressed and 72.55% as financially normal. The relatively higher proportion of distress identified by this model reflects its stronger sensitivity, which is in line with prior evidence from emerging economy settings (Altman, 1968). The Fich and Slezak model, by comparison, identifies 12.09% of the observations as distressed and 87.91% as non-distressed. The Zmijewski model provides an intermediate result, classifying 13.58% of the observations as financially distressed and 86.42% as financially healthy. This outcome may be attributed to its reliance on liquidity, profitability, and leverage measures in predicting financial condition (Zmijewski, 1984).

Thereafter, the dataset is divided into training and testing subsets containing both the financial predictor variables and the binary target variable, where 1 represents distress and 0 represents non-distress. Around 70% of the observations are used for training the model, enabling it to learn the relationships between the financial indicators and the distress status of firms. The remaining 30% are used for testing in order to evaluate the predictive performance of the model on unseen data. Such a split improves the assessment of the model's capacity to generalize beyond the training sample.

Further, feature selection is undertaken to retain only the most relevant financial indicators for prediction. Including all available variables may unnecessarily increase model complexity and lead to overfitting. To address this, the Random Forest algorithm is employed to rank the predictor variables according to their importance in improving predictive accuracy. Based on this ranking, the ten most influential indicators are selected. These variables are subsequently standardized and transformed using scikit-learn preprocessing techniques in Python so that the data are presented in a form appropriate for machine learning models. This step is particularly useful for ensuring stability and consistency in logistic regression estimation. Table V depicts the results of the logistic regression coefficients for the select key financial data and ratios.

TABLE V. COEFFICIENT OF FINANCIAL INDICATORS BY THREE MODELS

Altman Z Score		Fich and Slezak		Zmijewski	
Financial indicators	***Coefficient***	***Financial indicators***	***Coefficient***	***Financial indicators***	***Coefficient***
Asset Turnover	-2.76844	Interest Coverage	-1.54577	Debt Ratio	16.32553
Debt Ratio	6.742158	ROA	-6.78384	ROA	-3.82385
ROA	0.270877	ROE	-3.35964	Retained Earnings	-0.50317
Current Ratio	-0.07187	Net Profit Margin	0.076674	Current Ratio	3.423875
Interest Coverage	-0.81265	EBIT	-3.22003	Asset Turnover	-0.40604
P/E	-0.08673	Operating Margin	0.001778	Current Liabilities	-0.06549
Market Cap	-1.66739	P/E	-0.0902	Total Liabilities	1.957362
Total Liabilities	7.734157	Current Ratio	-0.29427	ROE	-0.42431
Retained Earnings	-0.07875	Retained Earnings	-0.83756	Debt/Equity	0.287849
Debt/Equity	0.13594	Debt Ratio	-0.0368	Net Profit Margin	-0.09829

Source: Compiled by Author

The regression coefficients reported in Table V reveal that, across all three models, leverage-related variables such as debt ratio, total liabilities, and debt-to-equity ratio serve as the most prominent positive predictors of financial distress. This suggests that firms with higher debt exposure are more vulnerable to financial failure. Conversely, profitability indicators, including ROA and ROE, exhibit substantial negative coefficients, indicating that stronger profitability reduces the probability of distress. The Fich and Slezak model appears to assign greater importance to profitability-based measures such as ROA and EBIT, whereas the Altman Z-Score and Zmijewski models place relatively stronger emphasis on leverage-related variables.

As shown in Table VI, the predictive accuracy and AUC values vary across the three distress models. Under the Altman Z-Score framework, accuracy ranges from 0.86 to 0.93, with Random Forest producing the best results at an accuracy of 0.93 and an AUC of 0.97. Under the Fich and Slezak model, the accuracy values range from 0.93 to 1.00, and both Decision Tree and Random Forest achieve perfect performance with accuracy and AUC values of 1.00. Similarly, under the Zmijewski model, the accuracy ranges between 0.96 and 0.99, where Random Forest again outperforms the other algorithms with an accuracy of 0.99 and an AUC of 1.00. These results clearly suggest that Random Forest demonstrates the strongest predictive potential for forecasting financial distress.

TABLE VI. RESULTS DEPCTING THE ACCURACY AND AUC OF SELECT MODELS

Target Model	Classifier	Accuracy	AUC
Model 1-Altman Z Score	Logistic Regression	0.86	0.93
	Decision Tree	0.87	0.84
	Random Forest	0.93	0.97
	K-nearest neighbours	0.89	0.93
Model 2-Fich and Slezak	Logistic Regression	0.93	0.98
	Decision Tree	1.00	1.00
	Random Forest	0.99	1.00

	K-nearest neighbours	0.94	0.94
Model 3 - Zmijewski	Logistic Regression	0.98	1.00
	Decision Tree	0.99	0.99
	Random Forest	0.99	1.00
	K-nearest neighbours	0.96	0.97

Source: Compiled by Author

Table VII reports the forecasting performance of all three models estimated using the Random Forest algorithm. As Random Forest demonstrated the highest predictive accuracy among the machine learning algorithms employed in the study, it was selected as the benchmark for evaluating model performance using precision, recall, and F1-score. The findings show that the algorithm consistently classifies both financially sound and financially distressed firms with a high degree of accuracy across all three models. The precision, recall, and F1-score values range from 0.90 to 0.99, indicating robust and reliable predictive performance. This result supports the findings of Zhang (2024), who identified Random Forest as a superior approach for financial distress prediction.

TABLE VII. FORECASTING RESULTS OF EACH MODELS BY RANDOM FOREST ALGORITHM

Models	**Normal**			**Financially Distressed**		
	Precision	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Model 1 – Altman Z-Score	**0.94**	**0.96**	**0.95**	**0.90**	**0.85**	**0.88**
Model 2 – Fich and Slezak	**0.99**	**1.00**	**0.99**	**1.00**	**0.99**	**0.99**
Model 3 – Zmijewski	**0.99**	**0.99**	**0.99**	**0.97**	**0.97**	**0.97**

Source: Compiled by Author

The findings reveal that, in Model 3, the F1-score is 0.99 for financially healthy firms and 0.97 for financially distressed firms. This marginal difference may be explained by the class imbalance observed in the dataset. To address this issue, SMOTE was applied. In contrast, SMOTE was not used for Model 2, despite the distress proportion being 12%, because the model relies primarily on a single financial ratio, namely the interest coverage ratio. The results indicate that the application of SMOTE does not produce any notable variation in predictive performance, suggesting that the model is relatively insensitive to class imbalance and maintains reliable classification outcomes. Further, Random Forest demonstrates superior predictive performance compared to the other machine learning algorithms, recording AUC values of 0.97, 0.98, and 1.00 for Models 1, 2, and 3, respectively. These results, presented in Figures 6, 7, and 8 together with the confusion matrices, confirm the strong discriminative power of the algorithm. Model 1 attains an accuracy of 93.64%, while the Fich model achieves near-perfect accuracy of 99.96%, with minimal false classifications. The Zmijewski model also performs strongly, reporting an accuracy of 99.22% with only slight misclassification. Taken together, these results suggest that the integration of traditional financial distress models with machine learning techniques, particularly Random Forest, significantly improves predictive precision and strengthens early warning capability in financial distress assessment.

The decision tree structures of the three models also highlight the critical role of leverage and profitability indicators. In Model 1, firms with a debt ratio above 0.634 are classified as distressed, whereas lower values indicate financial soundness. An ROA below 0.033 similarly points to higher distress risk. In Model 2, the interest coverage ratio is the dominant predictor, with values below 0.998 indicating distress and higher values indicating stability. In Model 3, firms with a debt ratio above 0.774 are classified as distressed, while lower values suggest greater financial resilience. Overall, these findings confirm that higher profitability reduces distress risk, whereas greater leverage and weaker coverage capacity increase the likelihood of financial distress.

VIII. LIMITATIONS AND FUTURE DIRECTIONS

The present study is subject to certain limitations. First, it is confined to a four-year post-pandemic period, which may restrict the long-term generalizability of the findings. Future research may consider a longer time span in order to capture broader financial trends and improve the robustness of distress prediction. Second, the models employed in this study are based primarily on financial variables and do not account for non-financial factors such as corporate governance quality, managerial capability, or other organizational characteristics. Incorporating such dimensions in future research may enhance the explanatory strength of the models and support more informed investment, financing, and strategic decisions.

IX. CONCLUSION

This study establishes that machine learning algorithms, with particular emphasis on Random Forest, are highly effective in forecasting financial distress among selected Indian listed companies. Financial indicators such as asset turnover, ROA, interest coverage ratio, and debt ratio are found to be especially significant in predicting distress, thereby providing insight into the internal financial health of firms. The results indicate that higher leverage increases the probability of financial distress, whereas stronger profitability and operating efficiency reduce such risk. Among the traditional distress prediction approaches, the Zmijewski model and the Fich and Slezak model exhibit comparatively stronger predictive capability. The findings underscore the importance of data-driven prediction methods and suggest that machine learning algorithms can function as reliable early warning mechanisms. Their application can strengthen organizational resilience and support better financial decision-making in increasingly competitive and uncertain business environments.

REFERENCES

- Felman, J., et al. (2024). Navigating financial distress in emerging markets under new insolvency laws. *Journal of Economic Policy Reform*, 27(1), 45–6. https://doi.org/10.1080/17487870.2023.2189423
- Ghazali, M. F., Mansor, N., & Saleh, N. (2020). Financial distress predictors with governance and audit committee impact. *Procedia Economics and Finance*, 35, 77–98. https://doi.org/10.1016/S2212-5671(15)01445-0
- Kumar, A., & Sharma, R. (2022). Machine learning-based bankruptcy prediction models for Indian companies. *Expert Systems with Applications*, 187, 115835. https://doi.org/10.1016/j.eswa.2021.115835
- Muhammad Atif Khan Achakzai & Peng Juan (2022). *Using machine learning Meta-Classifiers to detect financial frauds. Finance Research Letters*, 48, 102915. https://doi.org/10.1016/j.frl.2022.102915
- Singh, M., Mishra, A., & Goel, A. (2023). Forecasting financial distress in Indian corporate sector: A hybrid machine learning approach. *Journal of Banking and Finance*, 148, 106615. https://doi.org/10.1016/j.jbankfin.2023.106615
- Zijiao Zhang, Chong Wu, Shiyou Qu, Xiaofang Chen (2022). Explainable AI approach for financial distress prediction. *Information Processing & Management*, 59(1), 102988. https://doi.org/10.1016/j.ipm.2021.102988

- Kuizinienė, S., Krilavičius, T., Damaševičius, R., & Maskeliūnas, R. (2022). Systematic Review of Financial Distress Identification using Artificial Intelligence Methods. *Applied Artificial Intelligence*, 36(1), 2138124. https://doi.org/10.1080/08839514.2022.2138124
- Rai, M., Sahu, S., Srivastava, S., & Basu, S. (2025). Predicting financial distress in emerging markets: The case of Indian listed companies. *Cogent Economics & Finance, 13*(1), 2489708. https://doi.org/10.1080/23322039.2025.2489708
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, *4*, 71–111. https://doi.org/10.2307/2490171
- Al Najjar, M., & Anfimiadou, A. (2020). Financial distress prediction using market and accounting variables: A review. *Scientometrics*, 124, 3459–3489. https://doi.org/10.1007/s11192-020-03672-w
- Nguyen, L. Q., & Ahmed, R. (2023). The role of financial ratios in predicting financial distress: A study of listed firms. *Cogent Economics & Finance*, *9*(1), 2165438. https://doi.org/10.1080/23322039.2023.2165438
- Sultana, N., Islam, M. A., & Rahman, M. M. (2023). Advances in logistic regression models for financial distress prediction: A comprehensive review. *Expert Systems with Applications*, 217, 117315. https://doi.org/10.1016/j.eswa.2022.117315
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109–131. https://doi.org/10.2307/2490395
- Chen, Y.-S., Lin, C.-K., Lo, C.-M., Chen, S.-F., & Liao, Q.-J. (2021). Comparable studies of financial bankruptcy prediction using advanced hybrid intelligent classification models to provide early warning in the electronics industry. *Mathematics*, 9(20), 2622. https://doi.org/10.3390/math9202622
- Şahin, Y., Bulkan, S., & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications, 40*(15), 5916-5923. https://doi.org/10.1016/j.eswa.2013.05.021
- Mahesh, G., Kumar, A., & Singh, R. (2025). Financial distress prediction in emerging markets: New perspectives and methodologies. *Cogent Business & Management*, 12(1), 2489708. https://doi.org/10.1080/23322039.2025.2489708
- Lokanan, M. E., & Ramzan, S. (2024). Predicting financial distress in TSX-listed firms using machine learning algorithms. *Frontiers in Artificial Intelligence*, 7, Article 1466321. https://doi.org/10.3389/frai.2024.1466321
- Ha, H. H., Dang, N. H., & Tran, M. D. (2023). Financial distress forecasting with a machine learning approach. *Corporate Governance and Organizational Behavior Review*, 7(3), 90–104. https://doi.org/10.22495/cgobrv7i3p8
- Chen, Y., Kuok, K., & Lee, S. (2021). Financial distress prediction using advanced machine learning models. *Journal of Financial Stability*, 58, 100839. https://doi.org/10.1016/j.jfs.2021.100839
- Tinoco, M., & Wilson, N. (2013). Financial distress and bankruptcy prediction in the presence of heteroscedasticity. *European Accounting Review*, 22(3), 523-550. https://doi.org/10.1080/09638180.2013.777596
- Ministry of Statistics and Programme Implementation (MOSPI). (2023). *Annual Survey of Industries*. https://mospi.gov.in/51-annual-survey-industries

- Zhang, J. (2024). Impact of an improved random forest-based financial distress model. *Finance Research Letters*. https://doi.org/10.1016/j.sasc.2024.200102
- Demetriades, P., & Owusu-Agyei, S. (2022). Fraudulent financial reporting: An application of the fraud diamond to Toshiba's accounting scandal. *Journal of Financial Crime, 29*(2), 729–763. https://doi.org/10.1108/JFC-05-2021-0108
- Horobet, A., Curea, S. C., Smedoiu Popoviciu, A., Botoroga, C.-A., Belascu, L., & Dumitrescu, D. G. (2021). *Journal of Risk and Financial Management*, 14(11), 536. https://doi.org/10.3390/jrfm14110536
- Song, X.-P., and Sheng, Z.-H. (2014). Application of machine learning methods to risk assessment of financial statement fraud: evidence from China: risk assessment of financial statement fraud. J. Forecast. 33, 611–626. https://doi.org/10.1002/for.2294
- Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression (2nd ed.). Wiley. https://doi.org/10.1002/0471722146