International Journal of Economic Practices and Theories (IJEPT)
ISSN: 2247-7225
Volume 2026, Issue 1

Financial Distress Modelling and Prediction in Emerging Asia: A Machine Learning Analysis of India, China and Korea

Varunn Kaushik

Research Scholar, Department of Finance, Indian Institute of Foreign Trade (IIFT), New Delhi, India
Email Id: varunn_phdmp21@iift.edu
ORCID ID: 0009-0001-0471-0646

Abstract

This study models corporate financial distress for three large emerging Asian markets, India, China and Korea, over 2011 to 2024 using firm data from the LSEG database. Five classifiers, a binomial logit and four machine learning models, are estimated on a lean predictor set chosen through an information gain procedure across six ratio categories. Predictors selected at the pooled level are first applied to each country, and the exercise is then repeated with predictors chosen separately for each market. Random Forest and Gradient Boosting lead in every market. India records the highest accuracy and gains most from country-specific predictors, where Random Forest rises to 92.2 per cent. China is stable and Korea remains the hardest to classify, a pattern consistent with its concentrated ownership structure.

Keywords: Financial Distress, Machine Learning, Feature Selection, Emerging Markets, Random Forest

JEL Classification Number: G33, C53, G15, C45

Introduction

Corporate financial distress is not a rare event tied to crises. Firms slide into and out of difficulty across the whole business cycle, and the cost of missing the early signs falls on lenders, employees, suppliers and the wider market. For emerging economies the stakes are higher still, because thinner disclosure and shallower capital markets leave less room for error in any warning system. This paper asks how well distress can be predicted in three of the largest emerging Asian markets, and whether the financial signals that work in one of them carry over to the others.

Altman (1968) gave the field its first durable tool in the Z-score, and the linear discriminant approach behind it shaped distress research for decades. Its weakness is also well known. A linear, static rule struggles with the non-linear and shifting relationships that distress actually follows, especially across markets with different accounting traditions, a limitation borne out when the model is tested internationally (Altman et al., 2017). Ohlson (1980) moved the literature towards conditional probability models, and Shumway (2001) showed that a hazard formulation predicts bankruptcy more accurately than single-period models. More recent work has turned to machine learning, where Support Vector Machines, neural networks and tree ensembles capture interactions that parametric models miss (Min and Lee, 2005; Chen, 2011; Liang et al., 2015).

Most of this evidence comes from a single country, and the bulk of the cross-country work pools mature and emerging markets together. That leaves a specific gap. We do not have a clear picture of whether a common set of financial determinants discriminates distressed from healthy firms within emerging Asia, or whether each market needs its own predictor set to perform well. India, China and Korea make a natural test bed. They are the three deepest emerging-market samples in the broader distress literature, they sit at different points of institutional development, and Korea in particular carries a corporate structure, the chaebol, that fits neither the mature nor the standard emerging template.

This paper differs from earlier work in two ways. First, the scope is deliberately narrow. Rather than pooling many economies, we concentrate on three large Asian markets and look closely at where distress signals are portable and where they break down. Second, the feature selection is disciplined. An information-gain procedure is applied across six ratio categories, and we retain the single most informative predictor in each category after controlling for cross-correlation. That produces lean, comparable predictor sets, so any performance difference reflects the signal in the data rather than the sheer number of variables.

The study pursues two objectives. The first is to establish whether a common set of financial determinants predicts

distress across India, China and Korea, or whether country-specific predictors are needed for competitive accuracy. The second is to compare five models across the three markets and identify where the performance ranking holds and where it shifts. We find that different ratios matter in different markets, that ensemble models lead throughout, and that the value of localising the predictor set is large for India, marginal for China and limited for Korea.

The remainder of the paper is organised as follows. Section 2 describes the data and variable construction. Section 3 sets out the methodology, the feature selection procedure and the five models. Section 4 reports and discusses the results. Section 5 concludes with implications and limitations.

Data and Variable Construction

Data

Firm information is drawn from the LSEG database and covers all listed companies on the principal exchange in each market. The three countries were chosen for their weight in emerging Asia and for the contrast in their institutional settings. India offers a deep listed-firm population with reporting aligned to Ind-AS. China combines scale with a state-influenced corporate sector and a reporting regime still maturing in places. Korea shows several developed-market features but is dominated by large family-controlled groups, the chaebol, whose ratio profiles sit apart from both the mature and the emerging archetype. Together the three give a wide spread of financial, institutional and economic structures within a single region.

The sample runs from 2011 to 2024. Starting in 2011 keeps the data on a consistent footing, since financial reporting quality and coverage improved markedly after the global financial crisis. The window also takes in a full set of economic conditions, including the COVID-19 shock, and reflects firm behaviour under current regulatory regimes. The dataset is an annual panel for each company. India contributes 9,778 final firm-year cases from 2,653 companies, China 12,361 cases from 2,372 companies, and Korea 8,587 cases from 1,721 companies. Per capita GDP, reported in Table 1, places the three markets at very different income levels, from 2,730 dollars in India to 34,160 dollars in Korea, which is useful context when reading the results. We collect six broad categories of financial ratios that the literature uses to gauge corporate health.

Table 1. Summary statistics for the sample markets.

Country	Per Capita GDP (USD)	No. of Companies	No. of Final Cases	Stock Exchange
India	2,730	2,653	9,778	National SE
China	13,140	2,372	12,361	Shanghai SE
Korea	34,160	1,721	8,587	Korea Exchange

Notes: Per capita GDP is reported in US dollars as a measure of economic strength. The number of companies and final firm-year cases reflect the sample after data screening over 2011 to 2024. Data are sourced from the LSEG database.

Variable Construction

We first build a matrix of three accounting variables used to define financial distress, namely the interest coverage ratio, the change in equity capital and asset growth. A firm is classified as distressed in a given year only when all three conditions point the same way. No single criterion is reliable on its own across markets and sectors, and requiring the three together cuts down the misclassification that comes from firm- or industry-specific ratio behaviour (Pindado et al., 2008; Tinoco and Wilson, 2013; Bhattacharjee and Han, 2014). Distress status in each year is based on the previous year's financial and market ratios. A sample case is one company's financial condition in one year, so the total number of cases equals the number of companies multiplied by the number of years they appear. Table 2 lists the full set of candidate ratios, grouped into profitability, liquidity, structure, activity, size, solvency and market variables, together with how each is measured.

Table 2. Determinants of financial distress used in the study.

Attribute	Code	Ratio	Measure
Profitability	P1	Net Profit Margin	Net profit / Net sales
	P2	Gross Profit Margin	Gross profit / Net sales
	P3	ROA	Net profit / Total assets
	P4	RNW	Net profit / Shareholders' funds
	P5	ROCE	Net profit / Capital employed
Liquidity	L6	Current Ratio	Current assets / Current liabilities
	L7	Quick Ratio	Quick assets / Current liabilities
	L8	CFO/TL	Total cash from operations / Total liabilities
	L9	CF/TA	Cash flow / Total assets
	L10	CF/OR	Cash flow / Operating revenue
	L11	WC/TA	Working capital / Total assets
Structure	S12	FA/TA	Fixed assets / Total assets
	S13	NCL/TL	Non-current liabilities / Total liabilities
Activity	A14	Asset Turnover Ratio	Sales revenue / Total assets
	A15	Inventory Turnover Ratio	Operating revenue / Stock
	A16	Creditor Days	Creditors / Operating revenue × 360
	A17	Debtor Days	Debtors / Operating revenue × 360
Size	SZ18	Ln (TA)	Log of Total assets
	SZ19	Ln (Market Cap)	Log of Market capitalisation
Solvency	SL20	Debt Ratio	Total debt / Total assets
	SL21	Debt to Equity	Debt / Equity
	SL22	Debt to Market Cap	Debt / Market capitalisation
	SL23	RE/TA	Retained earnings / Total assets
	SL24	Repayment Capacity	Financial debt / Cash flow
	SL25	Financial Leverage	Long-term liabilities / Total assets
Market variables	MV26	P/B	Price / Book value per share
	MV27	P/E	Price / Earnings per share
	MV28	Ln (Closing Price)	Log of closing price

Notes: The table lists the firm characteristics and the financial ratios used to represent each. The candidate set spans six economic categories of corporate health, from which the most informative predictor per category is later selected.

Methodology And Estimation

Selection of Predictors

The analysis runs in two stages. In the first stage we select predictors using the pooled three-country data and apply that common set to each market in turn. In the second stage we repeat the selection separately within each country and re-estimate the models with the country-specific set. Comparing the two stages shows directly how much is gained by tailoring the predictor set to a single market rather than relying on a shared one.

Feature selection uses information gain, computed for each candidate ratio against distress status. Within each of the six categories we keep the single most informative ratio after checking for cross-correlation, which yields a compact and interpretable predictor set rather than one whose performance simply reflects a larger variable count. Working category by category also keeps the final set economically balanced, so profitability, liquidity, activity, size, solvency and market information are each represented.

Distress Prediction Models

Five classifiers are estimated. The binomial logit model is the statistical baseline, fitted with the selected accounting and market ratios. The four machine learning models are the Support Vector Machine (Cortes and Vapnik, 1995), the Artificial Neural Network (Wilson and Sharda, 1994), Random Forest (Breiman, 2001) and Gradient Boosting. The two ensembles are included because tree-based and ensemble methods have performed well in distress prediction and related early-warning studies, including work on Chinese listed firms and broader financial-crisis forecasting (Geng et al., 2015; Liu et al., 2021; Zhu et al., 2022).

Class imbalance is a standing problem in distress prediction, since healthy firms heavily outnumber distressed ones. We address it with balanced subsampling so that the models are not simply rewarded for predicting the majority class. Performance is reported as overall accuracy, the share of firms in the test sample classified correctly, with five-fold cross-validation used to guard against overfitting. Accuracy is the natural headline metric here because the balanced design removes most of the distortion that would otherwise make it misleading on a skewed sample (Sun et al., 2024).

Results And Discussion

We define distress using the three-factor rule described above, combining the interest coverage ratio, the change in equity capital and asset growth. The mean values of these inputs for each market are available on request; the figures are averaged across companies and over the 2011 to 2024 period. Twenty-four candidate ratios then enter the information-gain stage.

Aggregate Feature Selection

Table 3 reports the information-gain scores at the pooled level. The activity and market dimensions dominate. The Asset Turnover Ratio is the strongest single predictor among the operating ratios at 0.53, with Debtor Days close behind at 0.47, which fits the intuition that how efficiently a firm turns assets into sales is a leading sign of trouble in these markets. Among market variables the price-earnings ratio scores highest at 0.58 and the price-to-book ratio at 0.42, a sharper reading than in mature markets and consistent with the larger role sentiment plays in emerging-market pricing. Within profitability, return on assets leads at 0.33, and on the liquidity side cash flow from operations to total liabilities is the most informative at 0.31. Repayment capacity is the clearest solvency signal at 0.29. These six predictors, one per category, form the common set carried into the country-level estimation.

Table 3. Information gain from feature selection at the aggregate level.

Attribute	Code	Ratio	Information Gain
Profitability	P1	Net Profit Margin	0.18

	P2	Gross Profit Margin	0.15
	P3	**ROA**	**0.33**
	P4	RNW	0.20
	P5	ROCE	0.15
Liquidity	L6	Current Ratio	0.13
	L7	Quick Ratio	0.13
	L8	**CFO/TL**	**0.31**
	L9	CF/TA	0.17
	L10	CF/OR	0.16
	L11	WC/TA	0.12
Activity	A14	**Asset Turnover Ratio**	**0.53**
	A17	Debtor Days	0.47
Size	SZ18	Ln (TA)	0.49
	SZ19	**Ln (Market Cap)**	**0.51**
Solvency	SL20	Debt Ratio	0.12
	SL21	Debt to Equity	0.14
	SL22	Debt to Market Cap	0.15
	SL23	RE/TA	0.16
	SL24	**Repayment Capacity**	**0.29**
	SL25	Financial Leverage	0.14
Market variables	MV26	P/B	0.42
	MV27	**P/E**	**0.58**

Notes: Information gain is computed for each candidate ratio against distress status using the pooled three-country sample. Higher values indicate greater discriminatory power. The most informative ratio in each category, shown in bold, is retained for the common predictor set.

Country Performance with Common Predictors

Table 4 reports model accuracy by market using the common predictor set. India stands out even under the shared specification, with Random Forest at 85.8 per cent and Gradient Boosting at 86.0 per cent. This is consistent with the depth of India's reporting infrastructure and the quality of its listed-firm data (Sehgal and Mishra, 2021). China sits in the middle, with Random Forest at 76.7 per cent and the other models clustered just below. Korea is the hardest market to classify, with Random Forest reaching only 67.5 per cent and the Support Vector Machine the best performer at 70.4 per cent. Korea's chaebol-dominated structure is well documented and helps explain why distress signals may behave differently there than in other markets (Joh, 2002). The ratio profiles of large affiliated groups fit neither the mature nor the standard emerging pattern, so a predictor set calibrated on pooled data struggles to find a stable decision boundary. The ranking of models is steady across all three markets: Random Forest and Gradient Boosting lead, the neural network and Support Vector Machine follow, and the logit baseline trails, though never by a wide margin.

Table 4. Country-wise model performance with common (aggregate-level) predictors.

ML Model	India	China	Korea
Logit	83.0	72.8	67.1
SVM	83.7	74.2	**70.4**
ANN	83.1	74.3	65.4
RF	85.8	**76.7**	67.5
GB	**86.0**	76.0	68.1

Notes: Figures are out-of-sample classification accuracy in per cent, obtained under balanced subsampling with five-fold cross-validation. The same predictor set, selected at the aggregate level, is applied to each market. The best model in each market is shown in bold.

Country-Level Feature Selection

We then repeat the information-gain procedure within each market to find predictors tailored to local conditions. Table 5 reports the country-level scores. The broad picture from the pooled stage holds, but there are instructive shifts. The Asset Turnover Ratio remains the leading activity predictor in all three markets, which confirms that it travels well across borders. Its strength still varies, reaching 0.59 in India against 0.47 in Korea. Total assets emerge as a better size measure than market capitalisation in each market, with Ln (TA) scoring above Ln (Market Cap) throughout. The market variables tell a more country-specific story. In India the price-earnings ratio leads at 0.54. In Korea the price-to-book ratio is the stronger of the two at 0.51, which reflects real differences in how earnings and book values are reported and priced rather than any inconsistency in the procedure. Debtor Days is far more informative in China and Korea, at 0.51 and 0.53, than in India at 0.41.

Table 5. Information gain from feature selection at the country level.

Attribute	Code	Ratio	India	China	Korea
Profitability	P1	Net Profit Margin	0.21	0.21	0.21
	P2	Gross Profit Margin	0.22	0.22	0.21
	P3	ROA	0.17	0.19	0.19
	P4	RNW	0.20	0.20	0.20
	P5	ROCE	0.20	0.19	0.18
Liquidity	L6	Current Ratio	0.14	0.15	0.14
	L7	Quick Ratio	0.17	0.17	0.17
	L8	CFO/TL	0.19	0.16	0.19
	L9	CF/TA	0.19	0.20	0.17
	L10	CF/OR	0.16	0.19	0.15
	L11	WC/TA	0.16	0.15	0.18
Activity	A14	Asset Turnover Ratio	0.59	0.49	0.47
	A17	Debtor Days	0.41	0.51	0.53
Size	SZ18	Ln (TA)	0.57	0.56	0.54
	SZ19	Ln (Market Cap)	0.43	0.44	0.46
Solvency	SL20	Debt Ratio	0.14	0.16	0.16

	SL21	Debt to Equity	0.16	0.17	0.14
	SL22	Debt to Market Cap	0.17	0.14	0.20
	SL23	RE/TA	0.19	0.19	0.17
	SL24	Repayment Capacity	0.15	0.17	0.15
	SL25	Financial Leverage	0.19	0.17	0.19
Market variables	MV26	P/B	0.46	0.51	0.51
	MV27	P/E	0.54	0.49	0.49

Notes: Information gain is computed for each candidate ratio against distress status separately within each market. Higher values indicate greater discriminatory power. These scores guide the construction of country-specific predictor sets.

Country Performance with Country-Specific Predictors

Table 6 reports model accuracy when predictors are chosen separately for each market. The most striking result is for India, where Random Forest accuracy rises from 85.8 to 92.2 per cent, with the neural network and Gradient Boosting close behind at 91.1 and 92.0 per cent. This is the clearest country-level finding in the study. India provides a large sample with Ind-AS-aligned reporting and a locally distinct ratio profile, particularly its Debtor Days behaviour, which gives country-specific selection real informational leverage. China shows almost no change, with Random Forest easing from 76.7 to 75.7 per cent, which tells us the common set was already well suited to the Chinese sample; the steady mid-70s accuracy is in line with other machine learning evidence on Chinese listed firms (Jiang and Jones, 2018; Rahman and Zhu, 2024). Korea again proves resistant: Random Forest moves only from 67.5 to 66.4 per cent, and no model clears 68 per cent. The implication is that Korea's difficulty is structural rather than a matter of predictor choice. The result is consistent with broader evidence that forecast quality improves when the selected inputs carry stronger predictive information, and that machine learning methods can outperform a logistic baseline in early-warning settings (Liu et al., 2021). We also see Gradient Boosting edge Random Forest in Korea once predictors are sharply focused, consistent with boosting's advantage when the decision boundary is well defined (Liu et al., 2022).

Table 6. Country-wise model performance with country-specific predictors.

ML Model	**India**	**China**	**Korea**
Logit	89.6	72.4	65.5
SVM	89.2	72.1	67.4
ANN	91.1	73.0	66.0
RF	**92.2**	75.7	66.4
GB	92.0	**75.9**	**67.8**

Notes: Figures are out-of-sample classification accuracy in per cent, obtained under balanced subsampling with five-fold cross-validation. Predictor sets are selected separately within each market. The best model in each market is shown in bold.

Taken together, the two stages give a clear message. A common predictor set is enough to classify Chinese firms well and already does a reasonable job in India, but tailoring the predictor set to India lifts accuracy by more than six percentage points, the largest gain anywhere in the study. Korea sits at the other extreme, where neither a common nor a local predictor set produces strong results, and the constraint appears to be the structure of the listed sector rather than the modelling choice. Across both stages and all three markets, Random Forest and Gradient Boosting are consistently first or second, echoing earlier findings that neural and ensemble methods lead

other classifiers in distress data (Geng et al., 2015), with the Support Vector Machine's competitiveness in Korea the main exception to ensemble dominance.

Summary and Conclusions

This paper has modelled corporate financial distress for India, China and Korea over 2011 to 2024, using a binomial logit baseline and four machine learning models on predictor sets chosen through an information-gain procedure. The aim was to learn whethe

r a shared set of financial determinants works across these three large emerging Asian markets, or whether each needs predictors of its own.

Three findings stand out. First, the relevant ratios are not the same across markets. Activity and market variables carry the most information everywhere, but the specific market signal differs, with the price-earnings ratio leading in India and the price-to-book ratio in Korea. Second, ensemble models dominate. Random Forest and Gradient Boosting top the ranking in every market and under both predictor sets, with the logit baseline trailing but never far behind. Third, the payoff from localising the predictor set is highly uneven. India gains a great deal, China almost nothing, and Korea remains hard to classify regardless of the approach, which points to its concentrated ownership structure rather than the model as the binding constraint.

These results carry practical weight. For regulators and lenders in emerging Asia, the evidence supports building distress screens market by market rather than importing a single specification across borders, since the gain from doing so is large in some markets and negligible in others. For India in particular, the quality of listed-firm reporting translates into genuinely accurate early-warning models, with Random Forest reaching better than nine cases in ten classified correctly.

The study has limits worth noting. Accuracy is the headline metric, and although the balanced design makes it reliable here, future work could report precision, recall and the area under the ROC curve to give a fuller view of the cost of different errors. The three-country focus buys depth at the price of breadth, so the conclusions should not be read as general statements about all emerging markets. Extending the analysis to other large emerging economies, and adding macroeconomic and governance variables to the firm-level ratios, would be a natural next step. The Korean result in particular invites closer study of how ownership concentration shapes the financial signals that distress models rely on.

References

[1] Altman, E.I., 1968, Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23(4), 589-609.

[2] Altman, E.I., M. Iwanicz-Drozdowska, E.K. Laitinen and A. Suvas, 2017, Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. Journal of International Financial Management & Accounting, 28(2), 131-171.

[3] Bhattacharjee, A. and J. Han, 2014, Financial Distress of Chinese Firms: Microeconomic, Macroeconomic and Institutional Influences. China Economic Review, 30, 244-262.

[4] Breiman, L., 2001, Random Forests. Machine Learning, 45(1), 5-32.

[5] Chen, M.-Y., 2011, Predicting Corporate Financial Distress Based on Integration of Decision Tree Classification and Logistic Regression. Expert Systems with Applications, 38(9), 11261-11272.

[6] Cortes, C. and V. Vapnik, 1995, Support-Vector Networks. Machine Learning, 20(3), 273-297.

[7] Geng, R., I. Bose and X. Chen, 2015, Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining. European Journal of Operational Research, 241(1), 236-247.

[8] Jiang, Y. and S. Jones, 2018, Corporate Distress Prediction in China: A Machine Learning Approach. Accounting & Finance, https://doi.org/10.1111/acfi.12432.

[9] Joh, S.W., 2002, The Effects of the Economic Crisis and Corporate Reform on Korean Groups. KDI Policy Study 2002-10, Korea Development Institute, Seoul.

[10] Liang, D., C.-F. Tsai and H.-T. Wu, 2015, The Effect of Feature Selection on Financial Distress Prediction. Knowledge-Based Systems, 73, 289-297.

[11] Liu, L., C. Chen and B. Wang, 2021, Predicting Financial Crises with Machine Learning Methods. Journal of Forecasting, https://doi.org/10.1002/for.2840.

[12] Liu, J., X. Li and H. Yin, 2022, Interpreting the Prediction Results of Tree-Based Gradient Boosting Models for Financial Distress Prediction. Journal of Forecasting, 41(6), 1207-1222.

[13] Min, J.H. and Y.-C. Lee, 2005, Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters. Expert Systems with Applications, 28(4), 603-614.

[14] Ohlson, J.A., 1980, Financial Ratios and the Probabilistic Prediction of Bankruptcy. Journal of Accounting Research, 18(1), 109-131.

[15] Pindado, J., L. Rodrigues and C. de la Torre, 2008, Estimating Financial Distress Likelihood. Journal of Business Research, 61(9), 995-1003.

[16] Rahman, M.J. and H. Zhu, 2024, Predicting Financial Distress Using Machine Learning Approaches: Evidence from China. Journal of Contemporary Accounting & Economics, 20(1), 100403.

[17] Sehgal, S. and R.K. Mishra, 2021, On the Determinants and Prediction of Corporate Financial Distress in India. Managerial Decision Economics, 42(3), 752-765.

[18] Shumway, T., 2001, Forecasting Bankruptcy More Accurately: A Simple Hazard Model. The Journal of Business, 74(1), 101-124.

[19] Sun, J., M. Zhao and C. Lei, 2024, Class-Imbalanced Dynamic Financial Distress Prediction Based on Random Forest. Expert Systems with Applications, 238, 122087.

[20] Tinoco, M.H. and N. Wilson, 2013, Financial Distress and Bankruptcy Prediction Among Listed Companies Using Accounting, Market and Macroeconomic Variables. International Review of Financial Analysis, 30, 394-419.

[21] Wilson, R.L. and R. Sharda, 1994, Bankruptcy Prediction Using Neural Networks. Decision Support Systems, 11(5), 545-557.

[22] Zhu, L., D. Yan, Z. Zhang and G. Chi, 2022, Financial Distress Prediction of Chinese Listed Companies Using the Combination of Optimization Model and Convolutional Neural Network. Mathematical Problems in Engineering, 2022, 9038992.