KMANPUB | AI and Tech in Behavioral and Social Sciences Vol. 5 No. 1 (2027): 1-13

Article history:
Received 10 May 2026
Revised 13 August 2026
Accepted 20 August 2026
Initial Publication 10 September 2026
Final Publication 01 January 2027

An Explainable Hybrid Artificial Intelligence Model for Corporate Financial Distress Prediction Using Accounting Information

Elnaz Akbarlou[1*], Naser Mastchaman[2]

[1] Department of Management, Economics and Accounting, Payame Noor University, Tehran, Iran
[2] Department of Accounting, He.C., Islamic Azad University, Heris, Iran

*** Corresponding author email address**: Akbarlouelnaz@pnu.ac.ir

Article Info

Article type:
Original Research

How to cite this article:
Akbarlou, E. & Mastchaman, N. (2027). An Explainable Hybrid Artificial Intelligence Model for Corporate Financial Distress Prediction Using Accounting Information. *AI and Tech in Behavioral and Social Sciences, 5*(1), 1-13.
https://doi.org/10.61838/kman.aitech.5958

© 2027 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

This study aimed to develop an explainable hybrid artificial-intelligence model for predicting corporate financial distress using accounting information and to compare its predictive performance with conventional bankruptcy models and individual machine-learning algorithms. A longitudinal firm-year dataset comprising 2,946 observations from 285 non-financial companies listed on the Tehran Stock Exchange and Iran Fara Bourse was constructed. Financial distress was defined as first entry into the Article 141 condition in the subsequent year. Logistic regression, support vector machine, random forest, XGBoost, and artificial neural network models were developed and subsequently integrated through stacked generalization. Feature selection, class-imbalance management, time-aware validation, and SHapley Additive exPlanations (SHAP) were incorporated. The most recent 750 firm-year observations were retained as an untouched out-of-time test sample. Of the total observations, 247 (8.4%) subsequently became financially distressed. Machine-learning models outperformed traditional bankruptcy benchmarks, with XGBoost achieving the strongest performance among individual classifiers (ROC-AUC = .921; PR-AUC = .612). The hybrid stacking model produced the best overall performance (ROC-AUC = .938; PR-AUC = .684; sensitivity = .823; specificity = .927; balanced accuracy = .875). SHAP analysis identified accumulated losses, operating cash flow, leverage, profitability, working capital, retained earnings, and interest coverage as the principal predictors. After removing accumulated losses to registered capital, the model retained strong predictive ability (ROC-AUC = .919). Combining accounting information with stacked artificial-intelligence models provides more accurate, robust, and interpretable financial-distress forecasts than conventional statistical models or individual machine-learning algorithms.

Keywords: *accounting information; artificial intelligence; bankruptcy prediction; explainable AI; financial distress; machine learning; SHAP; stacking ensemble.*

1. Introduction

Corporate bankruptcy and financial distress have significant consequences for shareholders, creditors, employees, auditors, regulators, and capital markets. Because financial failure is usually preceded by deterioration in profitability, liquidity, solvency, and cash-flow capacity, accounting information has long been considered an important source for developing early-warning models. Seminal studies demonstrated that financial ratios can distinguish distressed from healthy firms and provide useful information before formal failure occurs (Altman, 1968; Beaver, 1966).

Traditional bankruptcy-prediction research initially relied on statistical techniques. Altman (1968) developed the well-known Z-score using multiple discriminant analysis (Altman, 1968), while Ohlson (1980) introduced a probabilistic model based on logistic regression (Ohlson, 1980). These approaches established the foundations of accounting-based bankruptcy prediction and remain important benchmarks. However, their predictive performance may decline when applied across different industries, economic periods, or institutional environments because they generally assume relatively stable and predefined relationships between financial indicators and corporate failure (Altman et al., 2017; Grice & Dugan, 2001).

In practice, financial distress is a complex phenomenon resulting from simultaneous changes in liquidity, profitability, leverage, operating efficiency, cash flows, and accumulated financial performance. Relationships among these variables may be nonlinear and interactive, limiting the ability of conventional linear models to capture the full structure of financial deterioration. Consequently, artificial-intelligence and machine-learning techniques have increasingly been applied to bankruptcy prediction. Methods such as support vector machines, random forests, neural networks, and gradient-boosting algorithms can identify complex patterns without requiring a predetermined functional relationship among predictors (Barboza et al., 2017; Hosaka, 2019).

Empirical evidence generally indicates that machine-learning methods can outperform traditional statistical models. Barboza et al. (2017), for example, reported substantial predictive improvements for several machine-learning algorithms compared with conventional approaches (Barboza et al., 2017). Nevertheless, no individual algorithm is consistently superior under all conditions. Neural networks may capture complex nonlinear relationships but can be difficult to interpret, tree-based algorithms efficiently identify interactions but may be sensitive to model specifications, and support vector machines depend strongly on parameter and kernel selection. These limitations have encouraged the development of hybrid and ensemble models, which combine complementary algorithms to achieve more stable and accurate predictions (Ainan et al., 2024; Cao et al., 2026).

Another important challenge is the substantial class imbalance typically observed in bankruptcy datasets. Failed or financially distressed firms usually constitute only a small proportion of the total sample. Consequently, a model may achieve high overall accuracy while failing to correctly identify distressed companies. For this reason, bankruptcy-prediction systems should not be evaluated solely through accuracy; measures such as sensitivity, precision, F1/F2 scores, ROC-AUC, and particularly precision-recall AUC are also required. Appropriate class-weighting or resampling procedures should also be integrated into model development (Ainan et al., 2024).

A further limitation of advanced artificial-intelligence models concerns interpretability. Although complex algorithms may provide superior predictions, their “black-box” nature can limit their usefulness in accounting and financial decision-making. Auditors, creditors, investors, managers, and regulators generally require an explanation of why a firm has been classified as financially distressed. Explainable artificial intelligence (XAI), particularly SHapley Additive exPlanations (SHAP), can address this problem by identifying the financial variables that increase or decrease predicted bankruptcy risk and by providing both global and firm-specific explanations (Lundberg & Lee, 2017; Zhang et al., 2022).

The relevance of artificial-intelligence methods has also been demonstrated in the Iranian capital market. Etemadi et al. (2009) showed that genetic programming could successfully predict bankruptcy among Tehran Stock Exchange companies and outperform conventional discriminant analysis (Etemadi et al., 2009). However, contemporary predictive methods now permit the integration of feature selection, class-imbalance management, advanced machine-learning algorithms, ensemble learning, temporal validation, and explainability within a unified framework. Therefore, further research is required to develop models that are not only more accurate but also robust and economically interpretable.

Accordingly, the present study aims to design a hybrid artificial-intelligence model for predicting corporate financial distress using accounting information. The proposed framework integrates logistic regression, support vector machines, random forests, XGBoost, and artificial neural networks through stacked generalization. It also incorporates feature selection, class-imbalance management, out-of-time validation, and SHAP-based explainability. The model is compared with traditional bankruptcy-prediction approaches and individual machine-learning algorithms using multiple performance criteria. By integrating predictive accuracy, temporal robustness, and interpretability, the study seeks to provide a more effective

E-ISSN: 3041-9433

early-warning framework for investors, creditors, auditors, managers, and regulatory authorities.

2. Methods and Materials

- **Research Design**

This study employed a longitudinal predictive design to develop and evaluate a hybrid artificial-intelligence model for one-year-ahead prediction of corporate bankruptcy risk. The unit of analysis was the firm-year observation. Accounting and firm-specific information available at the end of fiscal year *t* was used to predict whether the firm would enter financial distress during fiscal year *t + 1*.

The empirical design emphasized genuine out-of-sample prediction rather than retrospective classification. Accordingly, all preprocessing, feature selection, class-balancing procedures, hyperparameter optimization, and probability-threshold selection were performed exclusively within the development sample. The most recent observations were retained as an untouched temporal test set to approximate the conditions under which the model would be applied to future companies. The proposed hybrid framework consisted of five stages: (1) construction of accounting and financial predictors; (2) data preprocessing and feature selection; (3) management of class imbalance; (4) estimation of individual statistical and artificial-intelligence classifiers; and (5) combination of the individual classifiers through stacked generalization. Explainable artificial intelligence was subsequently applied to identify the accounting variables contributing most strongly to predicted distress.

- **Population and Data Sources**

The statistical population consisted of non-financial companies listed on the Tehran Stock Exchange and Iran Fara Bourse during the 2011–2024 period. Rather than drawing a random sample, all firm-year observations satisfying the study criteria were included. Audited annual financial statements, explanatory notes, capital information, and other accounting disclosures were obtained from the publicly available corporate disclosure system. Market-related information used in supplementary analyses was obtained from historical trading records.

Financial institutions, including banks, insurance companies, investment companies, leasing companies, and other financial intermediaries, were excluded because their financial statement structure, leverage, regulatory capital requirements, and operating models differ substantially from those of non-financial firms. A firm-year observation was eligible when: (a) audited annual financial statements were available; (b) the fiscal year covered approximately 12 months; (c) the information necessary to determine distress status was available; (d) sufficient information existed to calculate the principal financial ratios; and (e) the firm had not already entered the distress condition before the beginning of the prediction period. Firm-years following the first distress event were excluded from the primary analysis because the purpose was to predict entry into distress, rather than repeatedly classify a company that was already known to be distressed.

- **Operational Definition of Bankruptcy/Financial Distress**

Because formal judicial bankruptcy events are relatively rare among publicly traded Iranian companies, financial distress was operationalized using Article 141 of the amendment to the Iranian Commercial Code. Under Article 141, when accumulated losses eliminate at least one-half of a company's registered capital, the board of directors is required to convene an extraordinary general meeting to consider continuation or dissolution of the company. Accordingly, this study treated entry into the Article 141 condition as an observable accounting-based proxy for severe financial distress and bankruptcy risk. For each firm-year, the binary outcome variable was defined as:

Distress = 1 when the company entered the Article 141 condition for the first time in fiscal year *t + 1*.

Distress = 0 when the company remained outside the Article 141 condition in fiscal year *t + 1*.

Thus, all predictors were measured using information available at fiscal year-end *t*, while the outcome referred to the subsequent year. This temporal separation was imposed to prevent contemporaneous information leakage. A company that had already been subject to Article 141 at year *t* was not included as a healthy observation for predicting year *t + 1*. This restriction ensured that the model represented an early-warning prediction rather than recognition of already-observed financial distress. Because the outcome represents a legally defined financial-distress condition rather than a judicial declaration of bankruptcy, the terms *bankruptcy prediction* and *financial-distress prediction* are used in the study in a predictive accounting sense, with this operational distinction explicitly acknowledged.

- **Prediction Horizon and Construction of the Panel Dataset**

The primary prediction horizon was one year. Financial information from year *t* was used to predict distress in year

E-ISSN: 3041-9433

t + 1. Consequently, accounting predictors from 2011 through 2023 were linked to distress outcomes observed from 2012 through 2024. A secondary two-year prediction horizon was examined in sensitivity analyses. In this analysis, information at year *t* was used to predict whether the firm entered financial distress during either of the following two fiscal years. The panel was organized chronologically by company and fiscal year. Firm identifiers were retained solely for data management and were never entered as predictive features.

- **Predictor Variables**

Predictor selection was guided by bankruptcy-prediction theory and previous accounting research originating from the financial-ratio approaches of Beaver (1966), Altman (1968), and Ohlson (1980) (Altman, 1968; Beaver, 1966; Ohlson, 1980). The initial candidate set included indicators of liquidity, profitability, leverage and solvency, operating efficiency, cash flow, growth, firm size, and accumulated financial performance.

- ***Accounting Predictors***

Table 1

Candidate Accounting Predictors and Operational Definitions

Dimension	Predictor	Calculation
Liquidity	Current ratio	Current assets / Current liabilities
Liquidity	Quick ratio	(Current assets − Inventories) / Current liabilities
Liquidity	Cash ratio	Cash and cash equivalents / Current liabilities
Liquidity	Working capital to assets	Working capital / Total assets
Liquidity	Operating cash flow coverage	Operating cash flow / Current liabilities
Profitability	Return on assets	Net income / Average total assets
Profitability	Return on equity	Net income / Average equity
Profitability	EBIT to assets	EBIT / Total assets
Profitability	Gross profit margin	Gross profit / Sales
Profitability	Net profit margin	Net income / Sales
Profitability	Retained earnings to assets	Retained earnings / Total assets
Leverage	Total liabilities to assets	Total liabilities / Total assets
Leverage	Debt-to-equity ratio	Total liabilities / Equity
Solvency	Equity ratio	Equity / Total assets
Solvency	Interest coverage	EBIT / Finance expense
Activity	Total asset turnover	Sales / Average total assets
Activity	Receivables turnover	Sales / Average receivables
Activity	Inventory turnover	Cost of goods sold / Average inventory
Cash flow	Operating cash flow to assets	Operating cash flow / Total assets
Cash flow	Operating cash flow to liabilities	Operating cash flow / Total liabilities
Cash flow	Operating cash flow to sales	Operating cash flow / Sales
Cash flow	Free cash flow to assets	Free cash flow / Total assets
Growth	Sales growth	Annual percentage change in sales
Growth	Asset growth	Annual percentage change in total assets
Growth	Earnings growth	Annual percentage change in net income
Size	Firm size	Natural logarithm of total assets
Capital structure	Accumulated losses to capital	Accumulated losses / Registered capital

Lagged changes in selected profitability, leverage, liquidity, and cash-flow indicators were additionally calculated to allow the algorithms to capture deterioration trends rather than relying exclusively on financial statement levels. The accumulated-loss-to-capital variable was retained in the primary model because it represents information genuinely available to investors before the subsequent-year distress event. However, because Article 141 itself is defined partly through accumulated losses and registered capital, the complete analysis was repeated without this variable as an important robustness test. This analysis was designed to determine whether the artificial-intelligence model could identify future financial distress from broader accounting information rather than relying mechanically on proximity to the legal threshold.

- ***Supplementary Market Variables***

A supplementary specification incorporated market-based variables when reliable data were available, including annual stock return, market-to-book ratio, market capitalization, trading liquidity, and annualized stock-return volatility. The accounting-only specification was considered the primary model because the central purpose of the study was to assess the predictive content of accounting information. The extended specification was used to determine whether market information provided incremental predictive value.

- **Data Preprocessing**

All preprocessing steps were embedded within the training procedure to prevent information from the future test observations from influencing model construction.

- ***Missing Data***

Variables with excessive missingness were first identified. Predictors with more than 30% missing observations across the eligible development sample were excluded unless their theoretical importance justified retention. Firm-year observations missing the outcome variable or the majority of core accounting variables were excluded. For the remaining predictors, missing continuous values were imputed using the median calculated from the corresponding training data. When adequate observations were available, industry-specific medians were used. Importantly, imputation values were estimated separately inside each training fold and subsequently applied to the corresponding validation fold. No information from the final test period was used to calculate imputation values.

- ***Outliers***

Financial ratios frequently exhibit extreme distributions because denominators can approach zero or because distressed firms may have unusually large accounting values. Continuous predictors were therefore winsorized at the 1st and 99th percentiles. Winsorization thresholds were estimated using the training sample only and then applied unchanged to validation and test observations. Extreme observations were not automatically deleted because unusually poor accounting values may contain legitimate information about approaching financial distress.

- ***Scaling***

Continuous variables supplied to logistic regression, support vector machines, and artificial neural networks were standardized to zero mean and unit variance. Means and standard deviations were calculated using training observations only. Tree-based models, including random forest and XGBoost, were trained on the processed but unstandardized feature values because tree-splitting procedures do not require variables to share the same scale.

- **Preliminary Feature Screening**

Predictor screening was conducted exclusively within the development data. First, variables with near-zero variance were removed. Second, pairwise correlations among continuous predictors were examined. When two variables showed an absolute correlation greater than .90, the variable with weaker theoretical relevance, greater missingness, or poorer univariate predictive performance was removed. To reduce dimensionality further, penalized logistic regression using the least absolute shrinkage and selection operator (LASSO) was applied. The LASSO penalty can shrink less informative regression coefficients toward zero and thereby provide simultaneous regularization and variable selection (Tibshirani, 1996). The regularization parameter was selected within the internal cross-validation procedure rather than from the full dataset. Feature selection was repeated independently within each training fold to prevent feature-selection leakage. The final predictor set was therefore determined exclusively from historical information available before evaluation of the held-out period.

- **Class Imbalance**

Financial-distress datasets are inherently imbalanced because distressed companies usually represent a smaller group than financially healthy firms. Therefore, overall classification accuracy was not used as the principal basis for model selection. Three imbalance-handling approaches were compared within the training data: class-weighted learning; Synthetic Minority Over-sampling Technique (SMOTE); and the original class distribution without resampling. SMOTE creates synthetic minority observations between neighboring minority-class examples rather than simply duplicating existing observations (Chawla et al., 2002). SMOTE was applied only within each training fold after the training-validation split. Synthetic observations were never created before data splitting and were never added to validation or test samples. The preferred imbalance strategy for each classifier was selected according to validation-set PR-AUC, sensitivity to distressed firms, and F1/F2 performance.

- **Benchmark Models**

The proposed hybrid model was compared with established bankruptcy-prediction benchmarks.

- ***Traditional Accounting Benchmarks***

Altman's Z-score and Ohlson's O-score were calculated where the required accounting variables were available. These models were used as historically important reference models rather than being re-estimated to optimize their coefficients.

- ***Logistic Regression***

AITBSS E-ISSN: 3041-9433

Binary logistic regression was estimated as the primary conventional statistical benchmark. Both standard logistic regression and regularized logistic regression were evaluated. The probability of distress was modeled as a function of the selected accounting predictors. Logistic regression also provided a transparent benchmark against which the incremental predictive value of more complex machine-learning methods could be assessed.

- **Artificial-Intelligence Base Learners**

Four complementary artificial-intelligence algorithms were selected as base learners because they represent substantially different learning mechanisms.

- ***Support Vector Machine***

A support vector machine with a radial basis function kernel was estimated. SVMs identify a decision boundary that maximizes separation between classes and can use kernel functions to model nonlinear classification boundaries (Cortes & Vapnik, 1995). The principal hyperparameters included the penalty parameter C and the radial-kernel coefficient γ.

- ***Random Forest***

Random forest was used as a bagging-based tree ensemble. It constructs multiple randomized decision trees and aggregates their predictions, thereby allowing nonlinear relationships and interactions among accounting variables to be learned without requiring a predetermined functional form (Breiman, 2001). The tuned parameters included the number of trees, maximum tree depth, number of variables considered at each split, and minimum observations required in terminal nodes.

- ***Extreme Gradient Boosting***

XGBoost was selected as the principal boosting algorithm. XGBoost sequentially constructs regularized decision trees to reduce prediction error and is capable of handling complex nonlinear and interaction effects in structured tabular data (Chen & Guestrin, 2016). Hyperparameters included the number of boosting rounds, learning rate, maximum tree depth, minimum child weight, subsampling fraction, feature-subsampling fraction, and L1/L2 regularization terms.

- ***Artificial Neural Network***

A feed-forward multilayer perceptron was used as the neural-network component. The architecture contained an input layer corresponding to the selected predictors, one to three hidden layers, nonlinear activation functions, dropout and/or L2 regularization, and a sigmoid output unit representing the predicted probability of distress. The number of hidden layers, number of neurons, learning rate, batch size, regularization strength, and dropout rate were treated as tunable parameters. Early stopping based on validation loss was used to reduce overfitting.

- **Hyperparameter Optimization**

Hyperparameters for SVM, random forest, XGBoost, regularized logistic regression, and the neural network were optimized exclusively within the development data. A nested time-aware validation procedure was employed. The outer development folds preserved chronological order, such that each validation period occurred after the observations used to train the corresponding model. Hyperparameter selection occurred in internal folds using only historical information. The optimization criterion was primarily PR-AUC because the outcome was imbalanced. ROC-AUC, balanced accuracy, sensitivity, and F1-score were considered secondary criteria. This procedure prevented hyperparameters from being selected on the final test data.

- **Development of the Hybrid Stacking Model**

The principal contribution of the study was the development of a stacked ensemble combining the complementary predictive signals generated by logistic regression, SVM, random forest, XGBoost, and the artificial neural network. Stacked generalization uses predictions from first-level learners as inputs to a second-level model rather than selecting a single algorithm as the winner (Wolpert, 1992). For every observation in the development sample, out-of-fold predicted probabilities were generated separately by each base learner. These probabilities formed a new meta-level dataset: probability predicted by logistic regression; probability predicted by SVM; probability predicted by random forest; probability predicted by XGBoost; and probability predicted by the artificial neural network.

A regularized logistic regression model was then trained as the meta-learner using these out-of-fold probabilities. Only out-of-fold predictions were used to train the meta-learner. Predictions generated by a base model for observations on which that same base model had been trained were never supplied to the stacking algorithm. After model specification had been finalized, each base learner was retrained on the complete development sample. Their predictions for the untouched temporal test set were then entered into the trained meta-learner to obtain the final hybrid probability of bankruptcy risk.

- **Temporal Validation and Prevention of Data Leakage**

AITBSS AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

Random train-test splitting was deliberately avoided because random splitting could allow observations from later economic periods to influence models evaluated on earlier periods. The development and testing strategy was chronological. Predictor observations from 2011–2020 were used for model development and hyperparameter selection through rolling or expanding-window validation. Predictor observations from 2021–2023, corresponding to distress outcomes in 2022–2024, were reserved as the final out-of-time test sample. The test set remained completely untouched until preprocessing rules, feature selection, imbalance treatment, algorithms, hyperparameters, stacking structure, and classification threshold had been finalized. All procedures capable of learning information from the data—including imputation, winsorization, scaling, feature selection, SMOTE, and hyperparameter tuning—were fitted using training observations only.

- **Classification Threshold**

The models generated continuous distress probabilities rather than only binary labels. Performance was first assessed using threshold-independent metrics such as ROC-AUC and PR-AUC. For binary classification measures, the probability threshold was selected in the development sample before examination of test outcomes. Because failure to identify a truly distressed company may be more consequential than issuing an additional false warning, the primary decision threshold was selected by maximizing the F2-score, which assigns greater weight to recall than precision. Performance at the conventional .50 threshold was also reported for comparison.

- **Model Evaluation**

The final models were evaluated exclusively on the out-of-time test sample. Given the imbalance of the bankruptcy outcome, PR-AUC was designated as a primary performance measure, because precision-recall analysis can be more informative than ROC analysis when the positive class is rare (Saito & Rehmsmeier, 2015). The following metrics were reported: area under the receiver operating characteristic curve (ROC-AUC); area under the precision-recall curve (PR-AUC); sensitivity/recall; specificity; precision/positive predictive value; F1-score; F2-score; balanced accuracy; Matthews correlation coefficient; and Brier score for probability accuracy. Confusion matrices were also reported for every model. A model was not considered superior merely because it obtained greater overall accuracy. Particular emphasis was placed on correctly identifying distressed firms while maintaining an acceptable false-positive rate.

- **Comparison of Predictive Performance**

Ninety-five percent confidence intervals for performance measures were estimated using bootstrap resampling of the independent test sample. Differences in ROC-AUC between the hybrid model and competing algorithms were tested using the DeLong procedure for correlated ROC curves (DeLong et al., 1988). Paired bootstrap confidence intervals were used to compare PR-AUC, F1-score, sensitivity, and balanced accuracy. When multiple model comparisons were conducted, adjusted significance levels were used to reduce the risk of Type I error. The hybrid model was considered to provide meaningful incremental predictive value when it showed consistently superior out-of-sample discrimination across several complementary metrics rather than only a marginal improvement in a single statistic.

- **Probability Calibration**

Because the intended output was an early-warning probability rather than merely a binary classification, probability calibration was evaluated in addition to discrimination. Calibration plots compared predicted distress probabilities with observed event frequencies. Calibration intercept, calibration slope, and Brier score were calculated. When necessary, probability calibration was performed using development data only. The calibration transformation was subsequently applied unchanged to the test sample.

- **Explainable Artificial Intelligence**

Explainability analysis was performed after final evaluation to determine which financial variables drove the AI predictions. SHapley Additive exPlanations (SHAP) were used because the approach attributes a model prediction to individual features and allows both global and observation-specific explanations to be generated (Lundberg & Lee, 2017). Global SHAP importance was calculated as the mean absolute SHAP value of each predictor. This analysis identified the accounting variables contributing most strongly to predicted financial distress across the sample.

SHAP dependence plots were used to investigate nonlinear relationships between important financial ratios and predicted distress probabilities. For selected distressed and healthy companies, local SHAP explanations were generated to show how individual accounting indicators increased or decreased the predicted probability of distress. For example, the analysis could identify whether a firm's elevated predicted bankruptcy risk was principally associated with negative operating cash flow, high

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

leverage, declining profitability, poor liquidity, accumulated losses, or interactions among these factors. Interpretations were based on the final fitted model and were not used to alter the test results retrospectively.

- **Robustness and Sensitivity Analyses**

Several additional analyses were performed to assess robustness. First, the complete modeling process was repeated after removing the accumulated-loss-to-capital variable. This test examined whether model performance depended mechanically on a predictor closely related to the Article 141 outcome definition. Second, a two-year-ahead prediction specification was estimated. Third, performance was examined separately across major industries when subgroup sample sizes were sufficient. Fourth, an accounting-only feature set was compared with the extended accounting-plus-market-information model. Fifth, the hybrid model was compared with individual classifiers under both the original class distribution and alternative class-balancing procedures. Sixth, models were re-estimated using only firm-years with complete audited financial information to determine whether imputation affected the conclusions. Finally, sensitivity and false-negative rates were examined specifically because failure to identify future distressed firms represents a critical limitation of any bankruptcy early-warning system.

- **Statistical Analysis and Software**

Descriptive statistical analyses were used to characterize the distributions of the predictor variables and the prevalence of financial distress. Differences between distressed and non-distressed firm-years were initially evaluated using independent-samples tests appropriate to the distributions of continuous variables and chi-square tests for categorical variables. These inferential comparisons were descriptive and were not used to select predictors outside the cross-validation procedure. All machine-learning analyses were implemented in Python using established libraries for data manipulation, statistical modeling, machine learning, imbalanced-data processing, gradient boosting, neural networks, and SHAP-based explanation. To ensure reproducibility, random seeds were fixed for stochastic procedures. The final analysis code, model hyperparameters, predictor definitions, preprocessing rules, and software package versions were archived with the research materials. All hypothesis tests were two-sided with $\alpha = .05$ unless otherwise specified.

3. Findings and Results

Sample Characteristics

The final dataset consisted of 2,946 firm-year observations from 285 non-financial companies. Of these, 247 observations (8.4%) subsequently entered the Article 141 financial-distress condition, confirming substantial class imbalance. A total of 2,196 firm-year observations were used for model development, whereas 750 observations from the most recent period were reserved as an untouched out-of-time test sample. The test sample included 62 distressed and 688 non-distressed observations.

Table 2

Distribution of the Development and Test Samples

Sample	Firm-Years	Distressed, n (%)	Non-Distressed, n (%)
Development	2,196	185 (8.4)	2,011 (91.6)
Out-of-time test	750	62 (8.3)	688 (91.7)
Total	2,946	247 (8.4)	2,699 (91.6)

Firms that subsequently became financially distressed showed significantly poorer liquidity, profitability, cash-flow generation, and solvency than non-distressed firms. In particular, distressed firms had lower working capital to assets, return on assets, operating cash flow, retained earnings, current ratios, and interest coverage, together with higher leverage and accumulated losses ($p < .001$ for the principal comparisons). Feature-selection procedures identified accumulated losses to registered capital, operating cash flow to total assets, total liabilities to total assets, return on assets, working capital to total assets, retained earnings to total assets, and interest coverage as the most stable predictors of subsequent distress. Class-weighted learning generally provided stronger validation performance than training on the original imbalanced

E-ISSN: 3041-9433

distribution, whereas the artificial neural network performed best with SMOTE-based training.

All machine-learning models outperformed the traditional bankruptcy benchmarks on the independent temporal test sample. Altman's Z-score and Ohlson's O-score achieved ROC-AUC values of .758 and .781, respectively, whereas logistic regression reached .845. Among the individual artificial-intelligence models, XGBoost showed the strongest performance, with ROC-AUC = .921 and PR-AUC = .612. However, the proposed hybrid stacking model achieved the best overall predictive performance, with ROC-AUC = .938, PR-AUC = .684, sensitivity = .823, specificity = .927, F2 = .731, balanced accuracy = .875, and MCC = .605.

Table 3

Out-of-Time Predictive Performance of the Main Models

Model	ROC-AUC	PR-AUC	Sensitivity	Specificity	F2	Balanced Accuracy
Altman Z-score	.758	.265	.613	.872	.508	.742
Ohlson O-score	.781	.304	.661	.876	.548	.769
Logistic regression	.845	.421	.710	.895	.604	.803
SVM	.873	.480	.758	.900	.646	.829
Random forest	.899	.537	.774	.908	.669	.841
ANN	.904	.558	.790	.907	.679	.849
XGBoost	.921	.612	.806	.917	.704	.862
Hybrid stacking	.938	.684	.823	.927	.731	.875

The hybrid model correctly identified 51 of the 62 distressed observations and 638 of the 688 non-distressed observations. It therefore produced only 11 false negatives, compared with 24 for Altman's Z-score and 18 for logistic regression. The improvement in ROC-AUC was statistically significant compared with logistic regression, $\Delta AUC = .093$, 95% CI [.062, .124], $p < .001$, and random forest, $\Delta AUC = .039$, 95% CI [.019, .059], $p < .001$. The hybrid model also modestly but significantly outperformed XGBoost, $\Delta AUC = .017$, 95% CI [.002, .032], $p = .027$. Its PR-AUC was .072 higher than that of XGBoost, 95% bootstrap CI [.020, .126].

The hybrid model also showed the best probability calibration, with a calibration intercept of .01, slope of .98, and Brier score of .053, indicating close agreement between predicted and observed distress probabilities. SHAP analysis showed that the most influential variables were accumulated losses relative to registered capital, operating cash flow to total assets, total liabilities to total assets, return on assets, working capital to total assets, retained earnings to total assets, and interest coverage. Higher leverage and accumulated losses increased predicted distress risk, whereas stronger operating cash flow, profitability, liquidity, and retained earnings reduced it. SHAP dependence analyses also showed nonlinear effects. Distress risk increased sharply when liabilities exceeded approximately 70% of total assets or when operating cash flow became negative. The combination of high leverage and negative operating cash flow produced particularly high predicted risk.

The principal findings remained stable across alternative model specifications. After removal of accumulated losses to registered capital—a variable closely related to the Article 141 definition—the hybrid model still achieved ROC-AUC = .919 and PR-AUC = .602. This finding indicates that predictive performance was not driven solely by proximity to the legal distress threshold. The model also retained useful predictive ability for a two-year forecasting horizon (ROC-AUC = .902; PR-AUC = .519). Adding market-based variables produced only a modest improvement over the accounting-only model, increasing ROC-AUC from .938 to .945.

Table 4

Robustness Analyses for the Hybrid Model

E-ISSN: 3041-9433

Specification	ROC-AUC	PR-AUC	Sensitivity	Balanced Accuracy
Primary accounting model	.938	.684	.823	.875
Excluding accumulated losses/capital	.919	.602	.774	.847
Two-year prediction horizon	.902	.519	.758	.821
Accounting + market variables	.945	.704	.823	.881

Overall, the results showed that deterioration in profitability, liquidity, operating cash flow, solvency, and accumulated financial performance preceded entry into financial distress. Machine-learning algorithms consistently outperformed traditional bankruptcy-prediction models, and XGBoost was the strongest individual classifier. Most importantly, integrating logistic regression, SVM, random forest, XGBoost, and an artificial neural network through stacked generalization produced the best out-of-time prediction. The hybrid model achieved an ROC-AUC of .938 and PR-AUC of .684 while correctly identifying more than 82% of subsequently distressed firms. The persistence of strong predictive performance across sensitivity analyses, together with the SHAP-based identification of economically meaningful accounting drivers, supports the robustness, interpretability, and practical usefulness of the proposed hybrid artificial-intelligence framework for corporate financial-distress prediction.

4. Discussion

The present study developed and evaluated a hybrid artificial-intelligence model for predicting corporate financial distress using accounting information. The findings showed that firms entering subsequent financial distress exhibited clear deterioration in liquidity, profitability, operating cash flow, solvency, and accumulated financial performance. More importantly, machine-learning models consistently outperformed conventional bankruptcy-prediction approaches, and the proposed stacking model achieved the strongest out-of-time predictive performance, with ROC-AUC = .938, PR-AUC = .684, sensitivity = .823, and specificity = .927.

The observed financial patterns are consistent with the foundations of accounting-based bankruptcy prediction. Beaver (1966), Altman (1968), and Ohlson (1980) demonstrated that declining profitability, insufficient liquidity, excessive leverage, and weak financial performance provide important early signals of corporate failure (Altman, 1968; Beaver, 1966; Ohlson, 1980). The present results confirm the continuing relevance of these accounting indicators while showing that their predictive information can be extracted more effectively through nonlinear and ensemble-based methods. In particular, operating cash flow, leverage, profitability, working capital, retained earnings, and accumulated losses emerged as the most influential predictors of subsequent distress.

The superior performance of machine-learning algorithms is consistent with previous evidence showing that artificial-intelligence methods can capture nonlinear relationships and complex interactions that are difficult to represent through fixed statistical equations (Barboza et al., 2017; Hosaka, 2019). Among the individual classifiers, XGBoost achieved the strongest performance (ROC-AUC = .921; PR-AUC = .612), supporting the effectiveness of gradient-boosting methods for structured financial data. Nevertheless, the hybrid stacking model significantly outperformed XGBoost, indicating that combining heterogeneous learners provided additional predictive information beyond that available from the strongest individual algorithm.

This finding supports the theoretical rationale for ensemble learning. Logistic regression, SVM, random forest, XGBoost, and neural networks differ in how they represent decision boundaries, nonlinearities, interactions, and probability structures. Stacking allows these complementary signals to be integrated rather than relying on a single model. The findings are consistent with recent studies demonstrating that hybrid and ensemble approaches can improve bankruptcy forecasting by combining diverse machine-learning algorithms (Ainan et al., 2024; Cao et al., 2026).

An important strength of the findings is that performance was assessed using an untouched out-of-time test sample rather than a random train-test split. This design provides a more realistic assessment of whether a model trained on historical accounting information can predict financial distress in later periods. Previous research has shown that bankruptcy models may lose accuracy when applied across different periods or economic environments (Altman et al., 2017; Grice & Dugan, 2001). Therefore, the strong temporal performance of the hybrid model provides more convincing evidence of predictive usefulness than results obtained only from random cross-validation.

The results also highlight the importance of addressing class imbalance. Because only 8.4% of firm-year

AITBSS
AI and Tech in Behavioral and Social Sciences
E-ISSN: 3041-9433

observations represented subsequent financial distress, overall accuracy alone could have produced misleading conclusions. The emphasis on PR-AUC, sensitivity, F2-score, balanced accuracy, and MCC provided a more appropriate assessment of the model's ability to identify the minority distress class. This is particularly important in practical applications because failure to detect a genuinely distressed company may be considerably more costly than issuing an additional false warning.

The SHAP analysis strengthened the economic interpretability of the proposed model. Accumulated losses, operating cash flow, leverage, return on assets, working capital, retained earnings, and interest coverage were the most influential variables. These predictors are economically consistent with established theories of financial distress and demonstrate that the AI model did not rely on uninterpretable statistical artifacts. The nonlinear patterns identified by SHAP were particularly informative. For example, distress risk increased sharply when operating cash flow became negative and when liabilities exceeded approximately 70% of assets. High leverage combined with negative operating cash flow produced especially high predicted risk.

These results illustrate an important advantage of explainable AI in accounting research. Although complex machine-learning methods often outperform conventional models, their limited transparency can reduce their usefulness for auditors, creditors, investors, and regulators. SHAP-based explanations allow model predictions to be translated into financially meaningful information, thereby partially resolving the traditional trade-off between predictive accuracy and interpretability (Lundberg & Lee, 2017; Zhang et al., 2022).

The robustness analyses further supported the validity of the proposed framework. Because accumulated losses relative to registered capital is closely related to the Article 141 distress definition, there was a potential concern that the model might simply learn proximity to the legal threshold. However, after this variable was removed, the hybrid model still achieved ROC-AUC = .919 and PR-AUC = .602. This finding indicates that broader deterioration in cash flow, profitability, solvency, and liquidity independently contributed to future distress prediction.

The model also retained useful performance at a two-year prediction horizon, although predictive accuracy decreased compared with the one-year horizon. This decline is theoretically expected because uncertainty regarding future firm conditions increases as the forecasting horizon becomes longer. Nevertheless, the persistence of meaningful predictive ability indicates that accounting information contains early-warning signals before immediate financial distress becomes observable.

Adding market-based information resulted in only a modest increase in predictive performance, with ROC-AUC increasing from .938 to .945. This finding suggests that accounting information alone captured most of the economically relevant distress signal in the present setting. From an accounting perspective, this result is important because it confirms that audited financial statements continue to provide substantial predictive information even when advanced artificial-intelligence methods are used.

4.1. Practical Implications

The proposed hybrid model may serve as an early-warning decision-support tool for investors, lenders, auditors, managers, and regulatory institutions. Creditors may use predicted distress probabilities to support credit-risk assessment, investors may incorporate them into fundamental analysis, and auditors may use the identified financial-risk indicators as supplementary information when evaluating going-concern uncertainty. For managers, the SHAP explanations may be particularly useful because they identify the accounting dimensions contributing most strongly to elevated predicted risk. Rather than providing only a binary classification, the framework can indicate whether distress risk is primarily associated with deteriorating cash flow, excessive leverage, accumulated losses, weak profitability, or insufficient liquidity. However, the model should complement rather than replace professional judgment. Predictions represent statistical probabilities based on historical financial information and should not be interpreted as definitive legal or investment conclusions.

4.2. Limitations and Future Research

Several limitations should be considered. First, financial distress was operationalized using the Article 141 condition rather than formal judicial bankruptcy. Although this measure provides a meaningful and observable indicator of severe corporate financial distress in Iran, future studies should examine alternative failure definitions. Second, the analysis focused primarily on accounting variables. Although supplementary market indicators produced only modest incremental improvement, future research could incorporate corporate governance, audit characteristics, macroeconomic indicators, textual disclosures, managerial information, and industry-level variables. Third, predictive relationships may change as accounting standards, economic conditions, financing structures, and regulatory environments evolve. Future studies should therefore evaluate model stability across economic cycles and periodically recalibrate the model. Finally, although out-of-time validation provides stronger evidence than random splitting, external validation using companies from other capital markets would be necessary before claiming broader international generalizability.

5. Conclusion

This study demonstrated that a hybrid artificial-intelligence approach can substantially improve the prediction of corporate financial distress using accounting information. Traditional bankruptcy models retained useful predictive ability, but machine-learning algorithms provided superior out-of-time performance, with XGBoost emerging as the strongest individual classifier. The proposed stacking model produced the best overall results, achieving ROC-AUC = .938, PR-AUC = .684, sensitivity = .823, and specificity = .927. Its advantage remained evident after removal of the accounting variable most directly related to the Article 141 definition and under alternative prediction horizons and model specifications.

The SHAP analysis further showed that the model's predictions were driven by economically meaningful indicators, particularly accumulated losses, operating cash flow, leverage, profitability, working capital, retained earnings, and interest coverage. Therefore, the proposed framework combines three features that are essential for a useful financial-distress early-warning system: predictive accuracy, temporal robustness, and interpretability. Overall, the findings suggest that integrating accounting information with heterogeneous machine-learning algorithms through stacked generalization provides a more effective approach to financial-distress prediction than relying on conventional statistical models or any single artificial-intelligence algorithm. The proposed model may therefore provide a useful analytical framework for financial analysts, creditors, auditors, investors, managers, and regulators seeking timely identification and explanation of corporate financial distress.

Authors' Contributions

E.A. and N.M. contributed to the conception and design of the study, interpretation of the findings, manuscript preparation, and critical revision. Both authors approved the final manuscript and accept responsibility for the integrity of the work.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

The underlying corporate disclosures are publicly available. The derived dataset, analysis code, model hyperparameters, predictor definitions, preprocessing rules, and software-version information are available from the corresponding author upon reasonable request, subject to applicable data-provider and licensing restrictions.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethics Considerations

Not applicable. This study used publicly available company-level financial information and did not involve human participants, personal records, or interventions.

References

Ainan, U. H., Por, L. Y., Chen, Y. L., Yang, J., & Ku, C. S. (2024). Advancing bankruptcy forecasting with hybrid machine learning techniques: Insights from an unbalanced Polish dataset. *IEEE Access*, *12*, 9369-9381. https://doi.org/10.1109/ACCESS.2024.3354173

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

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management &*

AITBSS
E-ISSN: 3041-9433

Accounting, *28*(2), 131-171. https://doi.org/10.1111/jifm.12053

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405-417. https://doi.org/10.1016/j.eswa.2017.04.006

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, *4*, 71-111. https://doi.org/10.2307/2490171

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Cao, Y., Luo, Y. H., Wei, P., Zhai, J., & Shi, S. (2026). Bankruptcy forecasting-Market information with ensemble model. *The British Accounting Review*, *58*(3), 101530. https://doi.org/10.1016/j.bar.2024.101530

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321-357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273-297. https://doi.org/10.1007/BF00994018

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*(3), 837-845. https://doi.org/10.2307/2531595

Etemadi, H., Anvary Rostamy, A. A., & Dehkordi, H. F. (2009). A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert Systems with Applications*, *36*(2), 3199-3207. https://doi.org/10.1016/j.eswa.2008.01.012

Grice, J. S., & Dugan, M. T. (2001). The limitations of bankruptcy prediction models: Some cautions for the researcher. *Review of Quantitative Finance and Accounting*, *17*(2), 151-166. https://doi.org/10.1023/A:1017973604789

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, *117*, 287-299. https://doi.org/10.1016/j.eswa.2018.09.039

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 30,

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

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241-259. https://doi.org/10.1016/S0893-6080(05)80023-1

Zhang, Z., Wu, C., Qu, S., & Chen, X. (2022). An explainable artificial intelligence approach for financial distress prediction. *Information Processing & Management*, *59*(4), 102988. https://doi.org/10.1016/j.ipm.2022.102988

AITBSS
E-ISSN: 3041-9433