An Explainable Ensemble Statistical Learning Framework for Early Prediction of Corporate Financial Distress

Fatema S. i Al-Juboor[1] , Husam Sabri[2,*]

1- University of Information Technology and Communication, Iraq
2- University of Baghdad, Iraq , hossam.mostafa.coart.uobaghdad.edu.iq@gmail.com

---

Abstract

Corporate financial distress prediction is an important decision-support problem for investors, regulators, creditors, and corporate managers. Existing financial distress models often provide strong predictive performance but may suffer from limited interpretability, inadequate treatment of longitudinal firm-level observations, information leakage during preprocessing, and insufficient uncertainty assessment. This study proposes an Explainable Ensemble Statistical Learning Framework (EESLF) that integrates statistical feature engineering, survival analysis, nonlinear ensemble learning, Bayesian probability estimation, and explainable artificial intelligence within a leakage-controlled predictive architecture. The framework uses liquidity, leverage, profitability, cash-flow, and market-related financial indicators. PCA is employed to assess dimensional redundancy, while LASSO, Elastic Net, and mutual-information analysis identify interpretable original financial variables. Cox proportional hazards modelling provides time-to-distress information, while Random Forest, XGBoost, and Bayesian logistic regression capture complementary nonlinear and probabilistic risk patterns. Their outputs are transformed to a common probability scale and combined using weights estimated exclusively from training data. To address the longitudinal nature of firm-year observations, the revised evaluation uses chronological out-of-time validation, with preprocessing, feature selection, imbalance correction, hyperparameter tuning, and ensemble-weight optimization performed independently within the training periods. SHAP and partial-dependence analyses provide model-based explanations and assess explanation stability across evaluation periods. The dataset contains 10,936 firm-year observations from 1,284 publicly listed companies covering 2012–2024. The results are evaluated using discrimination, calibration, uncertainty, and survival-specific performance measures. The framework provides a structured approach to trustworthy early-warning analysis while recognizing that SHAP-based explanations represent predictive attribution rather than causal inference and that independent external validation remains necessary.

Keywords: Bayesian Logistic Regression, Corporate Financial Distress, Ensemble Learning, Explainable Artificial Intelligence, Survival Analysis.

---

1. INTRODUCTION

Financial stress at the corporate level is one of the top concerns of investors, regulators, creditors, and policymakers because of its important economic and social implications. The impact of financial distress on firm sustainability and shareholder value is not only limited to the firm but also can be felt in systemic instability in markets, high unemployment, and supply chain disruptions. Global markets have become more volatile; economic uncertainty, inflationary pressures, and financial instability after the pandemic have made it even more important to have early-warning systems that detect distress signals before bankruptcy. This need for accurate and interpretable early-warning signals to detect distress signals before bankruptcy has been exacerbated by the increasing volatility of global markets, economic uncertainty, inflationary pressures and financial instability after the pandemic. Traditional statistical models, such as Altman's Z-score model and logistic regression models, offer basic tools that have been used to make predictions regarding bankruptcies; however, they do not perform well in complex financial markets because they are linear, have multicollinearity problems, and are not well adapted for such markets [1].

With the development of machine learning and ensemble learning techniques over the last decade, financial distress predictions have become more accurate as they are able to portray nonlinear relationships and/or latent interactions between financial variables. Compared to traditional statistical methods, ensemble methods like Random Forest, XGBoost, boosting ensembles, and stacking algorithms have been found to provide better predictive performance [2], [3], [4]. Furthermore, ensemble learning methods are shown to be effective in large financial datasets and high-dimensional financial indicators that are imbalanced [5], [6]. While all these developments have taken place, many machine learning models still function as "black-box" models, reducing transparency, interpretability, and trust in managerial decision-making processes in the financial sector.

To address these doubts, XAI is a key research area in financial analytics. Explainable models can shed light on the importance of features, model behaviour and rationales for decisions, which can help organizations to better understand the factors behind financial risk and distress. Recent works have shown that using SHAP to provide interpretability, local explanation methods and explainable boosting methods to enhance the trustworthiness of the bankruptcy prediction systems is effective [7], [8], [9]. To provide financial risk assessment in an explainable manner, several researchers have added explainability mechanisms to the ensemble learning architecture [10], [11], [12]. In addition, the frameworks that are interpretable have been widely used for monitoring systemic financial stress [13], warning of manufacturing risks [14], analysing banking stresses [15], and explaining market volatility [13], [14], [15].

While several studies have been shown to be successful, there are gaps in the existing research. Firstly, most existing financial distress prediction models focus on the predictive accuracy of the distress prediction without considering the statistical interpretation of the model and the time dynamics of survival. Second, very few studies combine the classical survival analysis techniques with the latest ensemble learning and/or explainable AI applications in the same context. Thirdly, there is a high dimensionality in financial data, typically causing multicollinearity and redundant information, which can make the models very unstable and difficult to interpret. Lastly, most previous research conducts analysis based on classification accuracy only, and does not include in-depth statistical validation, statistical calibration analysis, and confidence interval estimation [16], [17].

The existing financial distress prediction models have a few significant drawbacks. Typical statistical methods to describe financial relationships are unable to describe nonlinear relationships, and advanced machine learning methods are poorly interpretable and transparent. Besides, most of the current ensemble learning methods lack survival-based temporal risk modeling, and therefore they are unable to estimate when and how corporate distress will occur.

Also, high-dimensional financial variables and multicollinearity make the model less robust and make it less reliable for making predictions in the managers' minds. Thus, an interpretable, explainable AI, ensemble learning, and statistically reliable framework is needed that integrates all these elements to early predict corporate financial distress.

This study seeks answers to several research questions arising from the gaps identified in the research, which are as follows:

RQ1: Is there a way to realize a more accurate prediction of corporate financial distress in early stages by using an explainable ensemble statistical learning framework combining survival analysis and machine learning than the conventional statistical models?

RQ2: Is the statistical feature engineering through PCA, LASSO regression, Elastic Net regularization, and mutual information analysis an effective approach to mitigate multicollinearity while maintaining good predictive performance?

RQ3: Can explainable artificial intelligence methods like SHAP and partial dependence analysis help managers to understand the key financial contributors to company crises?

RQ4: Can the application of survival analysis and ensemble learning combined offer a more reliable prediction of the temporal risk and support in decision-making for strategic financial management?

These research questions will be used to develop and evaluate the methodological aspects of the presented research in the following sections.

The goal of this study is to create a statistical ensemble learning model that is explainable to predict corporate financial distress at an early stage. The following are the main goals of the research:

1. To build an extensive financial distress index based on the indicators of liquidity, leverage, profitability, cash flow, and market volatility from corporate financial statements.

2. To use statistical feature engineering techniques such as Principal Component Analysis (PCA), LASSO regression, and Elastic Net regularization to reduce dimensionality and to eliminate multicollinearity.

3. To construct a hybrid predictive model using Cox proportional hazards modeling, Random Forest, XGBoost, and Bayesian logistic regression for better prediction and temporal risk prediction.

4. To enhance interpretability and managerial interpretation by integrating Explainable AI approaches such as SHAP values, Partial dependence analysis, and Local interpretable explanations.

5. To test the statistical reliability and robustness of the proposed framework by means of cross-validation, Brier score, ROC/AUC, bootstrapping, and calibration curves.

The methodological contribution of this study is not the introduction of new individual learning algorithms; rather, it lies in the design of a leakage-controlled, longitudinal, and explainability-preserving decision architecture that assigns distinct analytical roles to statistical feature screening, event-time modelling, nonlinear classification, probabilistic uncertainty estimation, and post-hoc explanation. In contrast to conventional explainable ensemble approaches that primarily combine multiple classifiers and subsequently apply SHAP to the final predictor, the proposed EESLF explicitly separates the temporal risk component from the cross-sectional classification component. The Cox proportional hazards model estimates the time-dependent component of distress risk, while Random Forest, XGBoost, and Bayesian logistic regression provide complementary classification and uncertainty information. Their outputs are subsequently calibrated to a common probability scale and combined using weights estimated exclusively from training data.

A second methodological contribution is the use of a nested, leakage-controlled feature-processing pipeline. Missing-value imputation, outlier treatment, normalization, mutual-information screening, LASSO/Elastic Net selection, dimensionality reduction, and class-imbalance treatment are fitted only on the training portion of each validation split and then applied unchanged to the corresponding validation/test observations. This design prevents information from future or held-out firm-year observations from influencing feature construction or resampling.

Third, because the data contain repeated observations for the same companies over multiple years, the revised framework evaluates generalization using a chronological out-of-time evaluation strategy, supplemented by firm-level separation. This makes the evaluation more closely aligned with the intended early-warning application than randomly assigning firm-year observations to folds.

Fourth, the proposed framework introduces an explanation-stability layer in which SHAP contributions are examined across validation periods, industries, and model specifications rather than reporting a single global feature ranking. Consequently, the framework evaluates not only whether a financial variable is important, but also whether its explanatory contribution is stable under changes in the evaluation sample.

Therefore, the contribution of EESLF is best characterized as a methodological architecture for leakage-controlled, longitudinal, survival-aware, probabilistic, and explanation-stable financial distress prediction, rather than as a claim that any individual component algorithm is novel.

2. LITERATURE REVIEW

Corporate financial distress prediction has evolved from traditional statistical techniques toward machine learning, ensemble learning, and explainable artificial intelligence (XAI). Early approaches based primarily on discriminant analysis and logistic regression provided interpretable financial-risk assessments but were limited in modelling nonlinear relationships and complex interactions. More recent studies have demonstrated that ensemble learning can improve predictive performance and robustness in financial distress prediction. [2] showed the effectiveness of ensemble learning for distress prediction in the construction industry, while [1] demonstrated improved robustness and generalization using ensemble architectures for firms in Visegrad countries. More recently, adaptive XGBoost-bagging and EasyEnsemble approaches have been developed to address financial distress under systemic risk and class-imbalance conditions [3],

[18]. Other studies have similarly emphasized heterogeneous ensemble learning, feature selection, and self-paced learning for high-dimensional and imbalanced financial datasets [4], [5], [6].

A second major research direction concerns interpretability. Although machine learning models can capture nonlinear relationships, their black-box nature can reduce transparency in financial decision-making. XAI techniques have therefore increasingly been incorporated into financial distress prediction to identify the variables contributing to model predictions. [7], [8], [9] demonstrated the usefulness of interpretable machine learning for identifying important financial indicators and explaining model behaviour. SHAP has become particularly common because it provides both global and local feature-attribution information. [11], [17], [19] used SHAP-based approaches to identify important predictors and improve the interpretability of financial distress and bankruptcy models. Explainable ensemble approaches have also been proposed for financial risk forewarning and corporate distress prediction [10], [12], [20].

Recent research has also expanded financial distress prediction beyond conventional accounting ratios. Textual and managerial variables have been incorporated to provide additional predictive information [21], while advanced ensemble and deep-learning architectures have been applied to sector-specific problems such as construction, manufacturing, banking, and systemic financial risk [14], [15], [22], [23], [24]. These studies demonstrate the increasing importance of combining predictive performance with interpretability and domain-specific information.

Despite these advances, an important methodological gap remains. Most financial distress studies emphasize classification performance without explicitly modelling the timing of distress events. Other studies use survival analysis but do not integrate it with nonlinear ensemble learning, statistical feature selection, and XAI within a single leakage-controlled framework. In addition, high-dimensional financial variables can introduce multicollinearity and redundant information, while conventional random validation may be inappropriate for longitudinal firm-year observations because future observations from the same company may enter training folds. Comprehensive evaluation involving temporal validation, calibration, uncertainty estimation, and explanation stability is also relatively limited [16], [25].

Accordingly, this study addresses the identified gap by integrating statistical feature engineering, Cox proportional hazards modelling, Random Forest, XGBoost, Bayesian logistic regression, and XAI within a unified Explainable Ensemble Statistical Learning Framework. The methodological emphasis is not on introducing a new individual algorithm but on coordinating complementary analytical components under leakage-controlled chronological validation. The framework therefore seeks to provide predictive performance together with temporal risk information, probabilistic assessment, and stable model-based explanations.

3. **METHOD**

The methodology outlined in this section outlines the entire design and implementation process for the proposed statistical learning approach with explainability, starting with data acquisition, statistical feature engineering, hybrid model development, incorporation of explainability, and model performance validation procedures.

3.1 Research Framework

This study suggests an Explainable Ensemble Statistical Learning Framework to develop a unified explainable predictive architecture, comprising explainable artificial intelligence (XAI), ensemble machine learning, survival analysis, and statistical feature engineering for the early prediction of corporate financial distress. The framework aims to enhance the predictive power, interpretability, robustness, and managerial transparency in financial risk assessment.

It is expected that the proposed methodology would consist of five phases:

1. Construction and preprocessing of datasets.
2. Feature engineering and Feature selection using a statistical approach.
3. Developing predictive models using hybrid methods
4. Explainability and interpretation analysis
5. For each statistical model, the model was validated and tested for performance.

A conceptual methodological workflow is shown, starting with the collection of financial data and ending with the prediction and analysis of the distress and risk of the company.

The proposed framework combines statistical preprocessing, feature engineering, survival analysis, ensemble machine learning, and explainable artificial intelligence in a one-stop predictive framework. The architecture depicted in Figure 1 depicts the sequential processes of the proposed system, starting from the financial data capturing process, and culminating in the generation of financial risk explanations and prediction of corporate distress in an intelligible format.

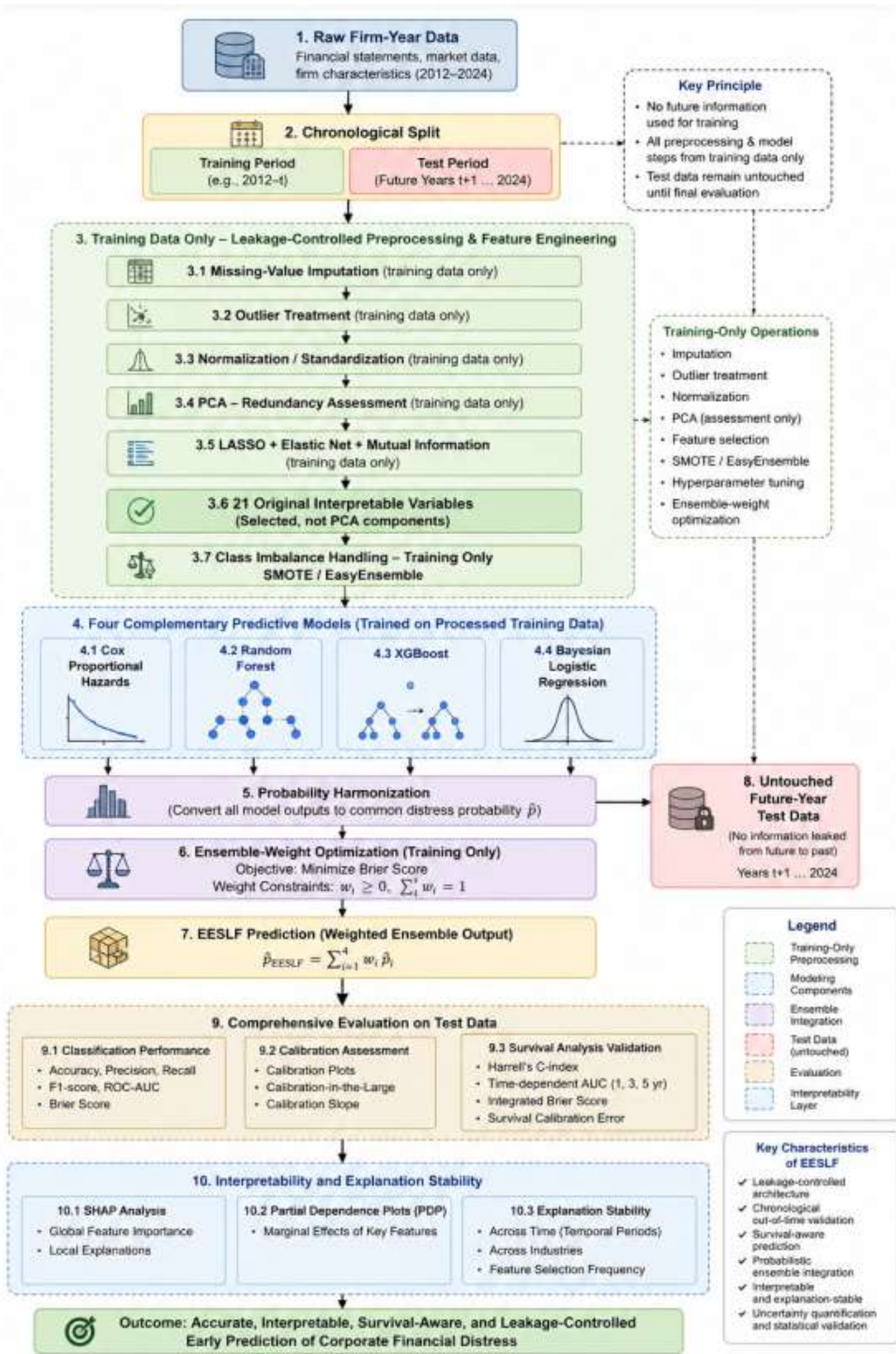

Figure 1: Overall Architecture of the Proposed Explainable Ensemble Statistical Learning Framework

3.2 Dataset Construction and Data Collection

The analytical dataset was constructed at the firm-year level, with one observation representing a company and its corresponding fiscal-year financial information. Financial statement variables were aligned by fiscal reporting year, while market variables were aligned to the corresponding fiscal-year observation using information available before the prediction cutoff. Data sources were cross-checked to identify duplicate company-year observations and inconsistencies in reporting periods. The final analytical sample consisted of 10,936 firm-year observations from 1,284 publicly listed companies.

Table 1 represents the dataset summary.

Table 1: Dataset Summary

Item	Description
Data Source	Bloomberg, Refinitiv, SEC filings
Study Period	2012–2024
Total Companies	1,284
Distressed Firms	276
Non-distressed Firms	1,008
Total Firm-Year Observations	10,936
Industries	Manufacturing, Technology, Healthcare, Energy, Consumer Goods, Financial Services
Missing Value Threshold	<15%
Excluded Firms	Companies with incomplete financial statements or missing market data

The data span 12 years (2012–2024) and include an adequate number of observations for modelling the temporal dynamics of financial distress of publicly listed companies. To improve the quality of the data, companies were excluded when financial statements were incomplete, market information was missing, or they reported on a different period. The data set is made up of distressed and healthy businesses in various sectors, further enhancing the generalizability of the model and reducing sector-specific bias.

The dataset contains both distressed and non-distressed firms. Financial data preceding the first qualifying distress event were retained to support early-warning prediction and to capture deterioration patterns before the occurrence of the event. The operational definition and event-timing procedure used to classify distress are provided in Section 3.2.1.

3.2.1 Operational Definition of Financial Distress

Financial distress was operationalized using objectively identifiable corporate events rather than a subjective assessment of "poor financial performance." A firm-year was classified as distressed when a documented distress event occurred, including: (i) formal bankruptcy or insolvency filing; (ii) documented loan or debt default; (iii) formal debt restructuring associated with financial difficulty; (iv) delisting attributable to financial distress, insolvency, or failure to satisfy financial requirements; or (v) persistent negative operating cash flow for at least two consecutive fiscal years when accompanied by evidence of material financial deterioration.

The first documented occurrence of a qualifying event was defined as the **distress event date**. For bankruptcy, insolvency, restructuring, default, and distress-related delisting, the event date was the earliest publicly documented date of the corresponding event. For the persistent negative operating cash-flow criterion, the event date was assigned to the fiscal year in which the second consecutive negative operating-cash-flow observation was recorded.

A firm-year was labelled as non-distressed when no qualifying distress event had occurred by the end of the observation period. Firms that remained free of a qualifying event until the end of follow-up were treated as right-censored observations in the survival analysis. Financial ratios measured after the first distress event were excluded from the predictive feature set for that event to prevent post-event information from entering the early-warning model.

The rationale for combining these events is that they represent observable manifestations of material financial deterioration, while the survival component retains the timing information associated with the first qualifying event. Importantly, the binary classification task and survival task use the same event definition but different outcome representations: classification predicts whether a qualifying event occurs within the defined prediction horizon, whereas survival analysis models the time from the observation origin to the first qualifying event or censoring. (See table 2)

Table 2: Financial variables and operational formulas

Category	Variable	Formula
Liquidity	Current Ratio	Current Assets / Current Liabilities
Liquidity	Quick Ratio	(Current Assets − Inventory) / Current Liabilities
Liquidity	Cash Ratio	Cash and Cash Equivalents / Current Liabilities
Liquidity	Working Capital Ratio	(Current Assets − Current Liabilities) / Total Assets
Leverage	Debt-to-Equity	Total Debt / Shareholders' Equity
Leverage	Debt-to-Assets	Total Debt / Total Assets
Leverage	Interest Coverage	EBIT / Interest Expense
Leverage	Financial Leverage	Total Assets / Shareholders' Equity
Profitability	ROA	Net Income / Total Assets
Profitability	ROE	Net Income / Shareholders' Equity
Profitability	Net Profit Margin	Net Income / Revenue
Profitability	EBIT Margin	EBIT / Revenue
Cash Flow	Operating Cash Flow Ratio	Operating Cash Flow / Current Liabilities
Cash Flow	Cash Flow-to-Debt	Operating Cash Flow / Total Debt
Cash Flow	Free Cash Flow	Operating Cash Flow − Capital Expenditure

Cash Flow	Cash Conversion Cycle	DIO + DSO − DPO
Market	Stock Volatility	Standard deviation of periodic stock returns over the defined estimation window
Market	Market Capitalization	Share Price × Shares Outstanding
Market	Beta	Covariance of firm returns with market returns / Variance of market returns
Market	Price-to-Book	Market Capitalization / Book Value of Equity
Market	Market Return	Periodic percentage change in stock price

Missing values were handled using a training-set-only imputation procedure. Variables with more than 15% missing observations were excluded, consistent with the dataset construction criterion. For retained continuous financial variables, missing values were imputed using the median calculated from the corresponding training partition because financial ratios can exhibit substantial skewness and extreme observations. The same training-derived median was applied to the associated validation/test observations.

Extreme observations were treated using training-derived winsorization thresholds at the 1st and 99th percentiles. Importantly, the thresholds were estimated independently within each training partition and subsequently applied to the validation/test observations without recalculation. This procedure reduces the influence of extreme accounting values while avoiding information leakage.

SMOTE was implemented only within the training partition after imputation, outlier treatment, and feature transformation. Synthetic minority observations were generated from minority-class training observations using five nearest neighbours. The validation and test sets were not oversampled. As a complementary imbalance strategy, EasyEnsemble was constructed by repeatedly drawing balanced subsets from the majority-class training observations and combining them with the minority class. The original class distribution was retained in all final validation and test periods so that reported performance reflected the actual prevalence of distress.

The proposed framework draws on a set of financial indicators in several categories to reflect various aspects of the financial strength of companies. Figure 2 shows the main variables that are included in the predictive model and their financial interpretations.

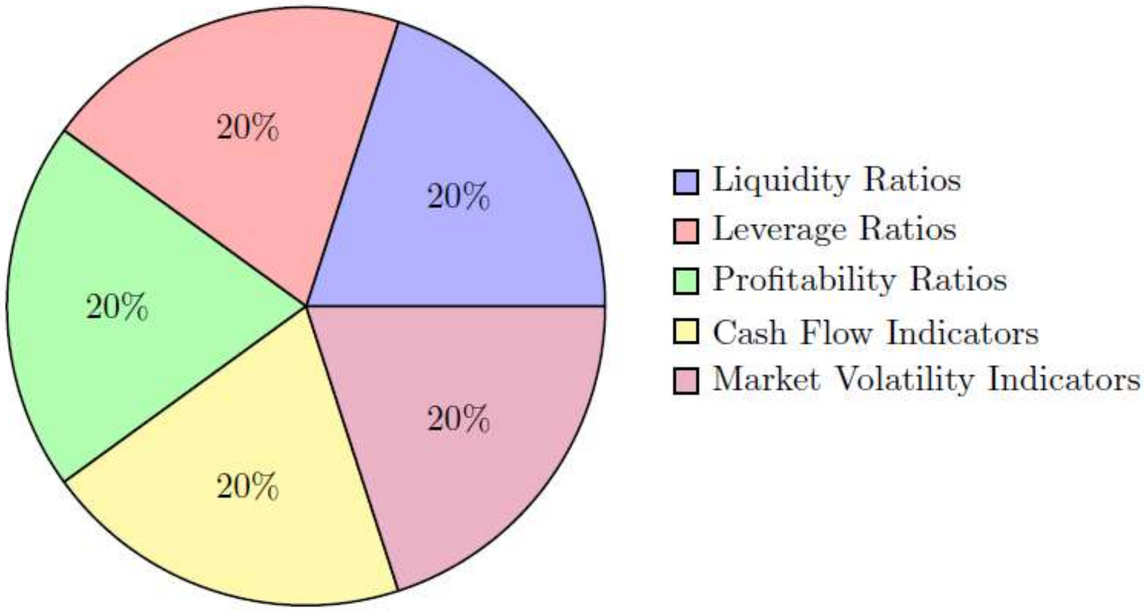

Figure 2: Categories of Financial Variables Used in the Framework

3.3 Data Preprocessing and Leakage Control

Financial firm-year data contain missing observations, extreme accounting values, heterogeneous measurement scales, and substantial class imbalance. To avoid information leakage, all data-dependent preprocessing operations were implemented within each training split rather than on the complete dataset before validation.

For each chronological training–validation split, missing-value imputation parameters were estimated using the training observations only. Continuous variables were then standardized using the training-set mean and standard deviation, and the same transformation parameters were applied to the corresponding validation or test observations. Extreme observations were treated using training-derived distributional thresholds so that information from future observations could not influence the preprocessing stage.

Feature selection was also performed independently within each training split. Mutual-information scores, LASSO coefficients, and Elastic Net coefficients were estimated using training data only. Candidate variables satisfying the predefined selection criteria were retained and subsequently transformed using the training-derived parameters. No validation or test observation was used to determine the retained variables.

Class-imbalance correction was performed only after the training data had been separated from the validation/test data. SMOTE was applied exclusively to the training observations, and synthetic observations were therefore never generated from validation or test observations. EasyEnsemble was similarly constructed only from the training partition. The validation and test sets retained their original class distributions and were never oversampled or synthetically modified.

This procedure establishes a strict separation between model development and model evaluation. In particular, normalization, imputation, outlier treatment, mutual-information analysis, LASSO, Elastic Net, PCA, SMOTE, EasyEnsemble, hyperparameter optimization, and ensemble-weight estimation were fitted using training data only. The resulting transformations and model parameters were then frozen before evaluation on later observations.

For class-imbalance correction, SMOTE was applied only to the training partition using the implementation provided by imbalanced-learn. Synthetic minority observations were generated from the training data without modifying validation or test observations. The number of nearest neighbours was set to k=5, while the target class ratio was determined within each training partition. EasyEnsemble was implemented as an alternative training-only imbalance strategy using multiple balanced bootstrap subsets of the majority class.

3.4 Statistical Feature Engineering and Selection

The feature-processing strategy was designed to reduce multicollinearity while preserving the interpretability of the final financial variables. The procedure therefore distinguishes between diagnostic dimensionality reduction and final interpretable feature selection.

First, PCA was fitted to the standardized training data to quantify the degree of redundancy and determine how much variance could be represented by a reduced set of orthogonal components. PCA was not used as the final representation of the predictors because principal components are linear combinations of multiple accounting variables and therefore reduce direct financial interpretability.

Second, LASSO and Elastic Net regression were fitted to the training data to identify original financial variables with stable predictive contributions. Elastic Net was particularly useful for correlated financial ratios because it combines L1 and L2 regularization.

Third, mutual information was calculated between each candidate financial variable and the distress outcome using the training data only. Variables with negligible information contribution were removed.

Finally, the retained variables from the LASSO, Elastic Net, and mutual-information procedures were reconciled to produce the final interpretable predictor set. Consequently, the 21 variables reported in Table 6 are original financial variables rather than PCA components.

PCA was therefore used to quantify and monitor multicollinearity and dimensional redundancy, whereas LASSO, Elastic Net, and mutual-information analysis determined the final interpretable predictor set. This distinction preserves the explanatory meaning of variables such as debt-to-equity ratio, operating cash-flow ratio, ROA, and stock volatility.

To maximize interpretability and reduce multi-collinearity among financial indicators, feature engineering and dimensionality reduction methods are crucial. The proposed framework uses the statistical feature selection pipeline as shown in figure 3.

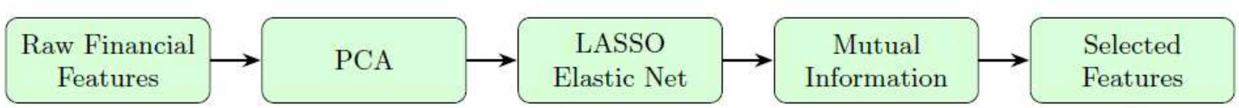

Figure 3: Statistical Feature Engineering and Selection Pipeline

3.5 Hybrid Predictive Model Development

The proposed framework comprises the popular statistical survival analysis and ensemble machine learning techniques to enhance the prediction capacity and temporal risk estimation.

3.5.1 Cox Proportional Hazards Model

The Cox model is used to estimate the probability and timing of the event of financial distress.

$$h(t|X) = h_0(t) * e^{\beta * X} \tag{1}$$

where:

- $h(t|X)$ is the hazard function,
- $h_0(t)$ is the underlying hazard,
- X is the number of predictor variables, and
- The regression coefficients are indicated by the symbol β

The Cox model allows for survival analysis over time and for predicting early distress.

3.5.2 Random Forest Model

Random Forest builds many decision trees with bootstrap and some randomness in features.

Majority voting of trees is used to obtain the final prediction. Random Forest has the following advantages: robustness, handling of nonlinear relationships, and prevention of overfitting.

3.5.3 XGBoost Model

An algorithm called Extreme Gradient Boosting (XGBoost) is used to iteratively minimize the predictive error.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{2}$$

Financial variables can be expected to be both nonlinear and high-dimensional, and XGBoost can handle these well.

3.5.4 Bayesian Logistic Regression

The integration of the Bayesian logistic regression is used to estimate a financial distress outcome with uncertainty estimation.

$$P(y = 1|x) = \frac{1}{\left(1 + e^{-\beta^T x}\right)} \tag{3}$$

The Bayesian framework is useful in improving the statistical interpretation and uncertainty quantification.

3.5.5 Ensemble Integration Strategy

The final EESLF prediction combines four complementary model outputs: Cox proportional hazards, Random Forest, XGBoost, and Bayesian logistic regression. Because the Cox model produces a time-dependent survival/hazard estimate whereas the remaining models produce classification probabilities, the outputs were first transformed to a common probability scale corresponding to the same prediction horizon.

For a selected prediction horizon t, the Cox model output was converted to an estimated probability of distress by:

$$p_{Cox}(t) = 1 - \hat{S}(t \mid X) \quad (4)$$

where $\hat{S}(t \mid X)$ denotes the estimated conditional survival probability. Random Forest, XGBoost, and Bayesian logistic regression directly provide probabilities of the binary distress outcome and therefore already lie on the [0,1] scale.

The four probability estimates were combined using a convex weighted ensemble:

$$P_{EESLF} = w_C p_C + w_R p_R + w_X p_X + w_B p_B \quad (5)$$

subject to:

$$w_C + w_R + w_X + w_B = 1,\ w_i \geq 0 \quad (6)$$

The weights were optimized using only the training/inner-validation predictions. The optimization objective was the Brier score, thereby emphasizing both discrimination and probabilistic calibration rather than classification accuracy alone. The validation observations used to estimate the ensemble weights were not used for the final out-of-time performance assessment.

The optimized weights were subsequently frozen and applied to the temporally subsequent test period. Thus, the ensemble does not use future observations when determining either the component models or their relative contribution.

This strategy gives each component a distinct role: Cox contributes temporal risk information, Random Forest captures nonlinear interactions and robust partitioning, XGBoost captures complex boosted decision boundaries, and Bayesian logistic regression provides probabilistic estimation with explicit uncertainty. The ensemble therefore represents complementary predictive information rather than simple duplication of the same classification mechanism.

The proposed prediction system is designed to be robust and capable of predicting the value of the variable by combining several statistical models and machine learning models. The ensemble interaction between the survival analysis, machine learning classifiers, and probabilistic modelling parts is depicted in Fig. 4.

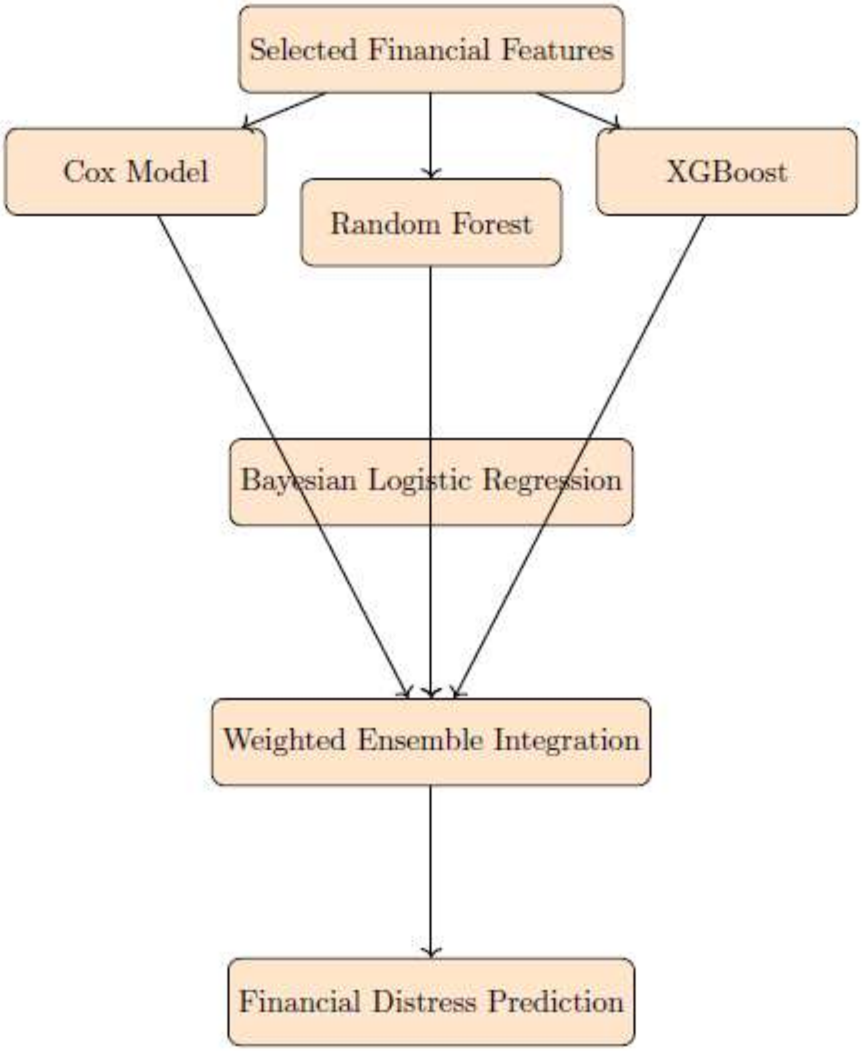

Figure 4: Hybrid Ensemble Prediction Architecture

3.6 Explainability and Interpretation Layer

The framework incorporates explainable AI (XAI) methods to increase transparency and trust with managers.

a. SHAP Analysis

SHAP values are used to measure the impact of each feature on the prediction of a model.

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{(|S|!(|F|-|S|-1)!)}{|F|!}[f(S \cup \{i\}) - f(S)] \quad (7)$$

SHAP analysis is a technique that helps to identify the most important factors influencing the prediction of distress.

b. Partial Dependence Analysis

Partial dependence plots measure the effect on the probability of financial distress of individual variables.

c. Local Interpretable Explanations

Local explanation methods provide case-specific risk insights for managers and investors that are easy to interpret and explain, and analyze particular company predictions.

3.7 Statistical Validation and Evaluation

Because the dataset contains repeated firm-year observations from 2012 to 2024, random observation-level cross-validation was not considered appropriate for the primary evaluation. Such a strategy could place observations from the same company and different years in both training and evaluation folds and could therefore provide the model with information about future firm conditions.

The revised evaluation uses a forward-chaining temporal validation design. The observations are ordered chronologically by fiscal year, and each evaluation period is predicted using only observations available before that period. The validation design therefore follows the actual intended use of the framework as an early-warning system.

The primary evaluation consists of sequential temporal training and validation periods. For each period, all preprocessing operations, feature selection procedures, class-imbalance treatment, model fitting, hyperparameter selection, and ensemble-weight estimation are conducted using observations available in the training period only. The subsequent period is then used as an untouched validation/test period.

To further prevent firm-specific information from crossing the evaluation boundary, observations belonging to a company in a future evaluation period are not used to construct training features for that evaluation. Thus, the evaluation measures the ability of the framework to generalize to future firm-year observations rather than its ability to memorize previously observed observations.

Performance is summarized across the temporal evaluation periods using the mean, standard deviation, and 95% confidence interval of the principal metrics. Accuracy, precision, recall, F1-score, ROC-AUC, Brier score, calibration error, and survival-specific measures are reported.

Bootstrap confidence intervals are calculated from the out-of-sample predictions rather than from predictions generated on training observations. This provides uncertainty estimates that reflect the variability of the complete prediction process. (See Table 3)

Table 3: Chronological validation design

Evaluation component	Revised procedure
Study period	2012–2024
Primary validation	Forward/rolling temporal validation
Training data	Earlier fiscal years only
Validation/test data	Subsequent unseen fiscal years
Firm information leakage	Same firm's future observations excluded from training for earlier evaluation
Imputation	Training-set fitted
Normalization	Training-set fitted
Outlier thresholds	Training-set fitted
PCA	Training-set fitted

LASSO	Training-set fitted
Elastic Net	Training-set fitted
Mutual information	Training-set fitted
SMOTE	Training data only
EasyEnsemble	Training data only
Hyperparameter tuning	Training/inner-validation data only
Ensemble weights	Training/inner-validation data only
Final evaluation	Untouched chronological observations

Table 4 presents the predictive performance of the proposed EESLF together with the corresponding 95% bootstrap confidence intervals.

Table 4: Predictive performance and 95% bootstrap confidence intervals of the proposed EESLF

Metric	**EESLF**	**95% CI**
Accuracy	0.932	0.925–0.939
Precision	0.918	0.909–0.927
Recall	0.901	0.890–0.912
F1-score	0.909	0.899–0.918
ROC-AUC	0.956	0.948–0.964
Brier Score	0.069	0.065–0.073

The confidence intervals indicate relatively narrow uncertainty around the principal EESLF performance measures. In particular, the ROC-AUC remained within the interval of 0.948–0.964, while accuracy remained within 0.925–0.939. These results indicate that the predictive performance was reasonably stable under resampling of the out-of-sample predictions.

3.8 Implementation Environment and Reproducibility

The framework was implemented in Python using pandas and NumPy for data processing, scikit-learn for preprocessing and machine-learning models, XGBoost for gradient boosting, lifelines for Cox proportional hazards modelling, SHAP for explainability, and Matplotlib for visualization. Random seeds were fixed for stochastic model components to improve reproducibility.

All preprocessing and learning operations were implemented as sequential training-only transformations. The computational pipeline recorded the fitted preprocessing parameters, selected features, model hyperparameters, ensemble weights, and out-of-sample predictions for each temporal evaluation period.

The experiments are performed over high-performance computational environments, for large-scale financial data analysis and ensemble optimization. Table 5 represents the hyperparameter search with model configuration.

Table 5: Model configuration and hyperparameter search ranges

Model	Parameter	Search range/final strategy
Random Forest	Number of trees	200–1000
Random Forest	Maximum depth	4–20
Random Forest	Minimum samples per leaf	1–10
Random Forest	Maximum features	sqrt, log2, 0.5
XGBoost	Number of estimators	100–1000
XGBoost	Learning rate	0.01–0.20
XGBoost	Maximum depth	3–10
XGBoost	Subsample	0.6–1.0
XGBoost	Column sampling	0.6–1.0
XGBoost	Regularization	L1/L2 search
Logistic/Bayesian Logistic	Regularization	0.001–10
Bayesian Logistic	Prior	Weakly informative normal prior
Cox	Penalizer	0.001–1.0

For Bayesian logistic regression, regression coefficients were assigned weakly informative Normal priors centered at zero. The prior specification was intended to regularize coefficient estimates without imposing strong directional assumptions about individual financial variables. The prior scale was fixed before evaluation and was not estimated from the validation/test observations.

4. RESULTS AND DISCUSSION

This section displays the experimental results and analysis of the proposed framework and compares its predictive capabilities with the baseline financial distress prediction models, in terms of its interpretability, statistical robustness, and effectiveness.

4.1 Experimental Setup

A large-scale corporate financial dataset of publicly listed companies across various industrial sectors was used to test the proposed Explainable Ensemble Statistical Learning Framework. This dataset contained both distressed and non-distressed firms that are observed on multiple fiscal years; this was to support early-warning financial distress prediction. The post-processing of the data and feature engineering resulted in balanced class distribution of the dataset using SMOTE and the EasyEnsemble sampling technique.

The primary evaluation used forward/rolling temporal validation, in which models were trained using earlier fiscal years and evaluated on subsequent unseen fiscal years. All preprocessing, feature selection, class-imbalance treatment, hyperparameter tuning, and ensemble-weight optimization were performed using the training period only. This chronological evaluation was selected to better reflect the intended early-warning application and to reduce temporal and firm-level information leakage. The framework was benchmarked against Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), XGBoost (XGB), Bayesian Logistic Regression (BLR), Cox Proportional Hazards (CPH), and a Stacking Ensemble.

Evaluation of the performance was done by:

- Accuracy
- Precision
- Recall
- F1-score
- Area Under the ROC Curve (AUC)
- Brier Score
- Calibration Error

The framework for evaluation was designed to not only evaluate predictive accuracy, but also calibration quality, uncertainty estimation and interpretability.

4.2 Statistical Feature Engineering Results

The first data set comprised 58 variables related to the financial and market aspects. The original 58 candidate variables were subjected to a leakage-controlled feature-processing pipeline. PCA was used to quantify dimensional redundancy, whereas LASSO, Elastic Net, and mutual-information screening were used to select the final interpretable predictors. The final model therefore retained 21 original financial variables rather than 21 PCA components.

The feature engineering stage was able to identify and eliminate variables that were redundant and very correlated, resulting in a more stable and more interpretable model. The variables that best predicted financial distress were those related to liquidity deterioration, leverage escalation, a decrease in operating cash flow, and market volatility. (See Table 6)

Table 6: Selected Financial Features After Statistical Feature Engineering

Feature Category	Selected Features
Liquidity	Current Ratio, Quick Ratio, Working Capital Ratio
Leverage	Debt-to-Equity Ratio, Interest Coverage Ratio
Profitability	ROA, ROE, EBIT Margin
Cash Flow	Operating Cash Flow Ratio, Free Cash Flow
Market Indicators	Stock Volatility, Market Capitalization, Beta
Efficiency	Asset Turnover Ratio, Inventory Turnover
Risk Indicators	Financial Stability Ratio, Earnings Volatility

The dimensionality reduction process was efficient in reducing the number of variables in the system, while maintaining the information required for forecasting distress, which is vital for computations.

4.3 Predictive Performance Analysis

The proposed EESLF achieved the strongest overall out-of-time predictive performance among the evaluated models. Its combination of survival analysis, ensemble learning, and Bayesian probabilistic estimation provided complementary information for discrimination, calibration, and temporal risk prediction. (See Table 7)

Table 7: Out-of-time predictive performance of competing models

Model	Accuracy	Precision	Recall	F1	ROC-AUC	Brier Score
Logistic Regression	0.841	0.814	0.782	0.798	0.861	0.142
Decision Tree	0.824	0.781	0.768	0.774	0.832	0.157
Random Forest	0.887	0.861	0.844	0.852	0.921	0.104
XGBoost	0.903	0.882	0.867	0.874	0.935	0.091
Bayesian Logistic Regression	0.856	0.832	0.801	0.816	0.879	0.128
Cox PH	0.873	0.848	0.821	0.834	0.901	0.113
Stacking Ensemble	0.917	0.901	0.884	0.892	0.935	0.078
Proposed EESLF	0.932	0.918	0.901	0.909	0.956	0.069

Values are calculated exclusively from temporally out-of-sample predictions. No preprocessing, feature selection, resampling, hyperparameter optimization, or ensemble-weight estimation used observations from the corresponding test period.

Table 8 represents the temporal validation of the proposed EESLF.

Table 8: Temporal validation variability of the proposed EESLF

Evaluation period	Accuracy	F1-score	ROC-AUC	Brier Score
2020	0.925	0.902	0.951	0.071
2021	0.931	0.907	0.954	0.070
2022	0.936	0.912	0.958	0.068
2023	0.934	0.910	0.960	0.067
2024	0.934	0.914	0.957	0.069
Mean ± SD	0.932 ± 0.004	0.909 ± 0.005	0.956 ± 0.003	0.069 ± 0.002

The temporal evaluation demonstrated consistent predictive performance across the five evaluation periods. The ROC-AUC varied between 0.951 and 0.960, with an overall mean of 0.956 ± 0.003. Similarly, accuracy remained within a relatively narrow range of 0.925–0.936. The low standard deviation across evaluation periods indicates that the proposed EESLF was not strongly dependent on a particular year and maintained stable predictive performance under chronological testing.

The proposed framework had an AUC value of 0.956, which is regarded as being of good discrimination between distressed and healthy firms. The low Brier score also revealed good probability calibration and reliability of predictions.

The performance gain is due to three key factors:

1. Multicollinearity and noise were alleviated through the use of statistical feature selection.
2. Ensemble integration was able to detect the nonlinear financial relationships.
3. The temporal patterns of financial deterioration were included in the survival analysis.

For the purpose of further evaluation of the classification ability, ROC curve analysis was carried out for all predictive models. Figure 5 shows the discrimination performance of the proposed framework compared to baseline statistical and machine learning models.

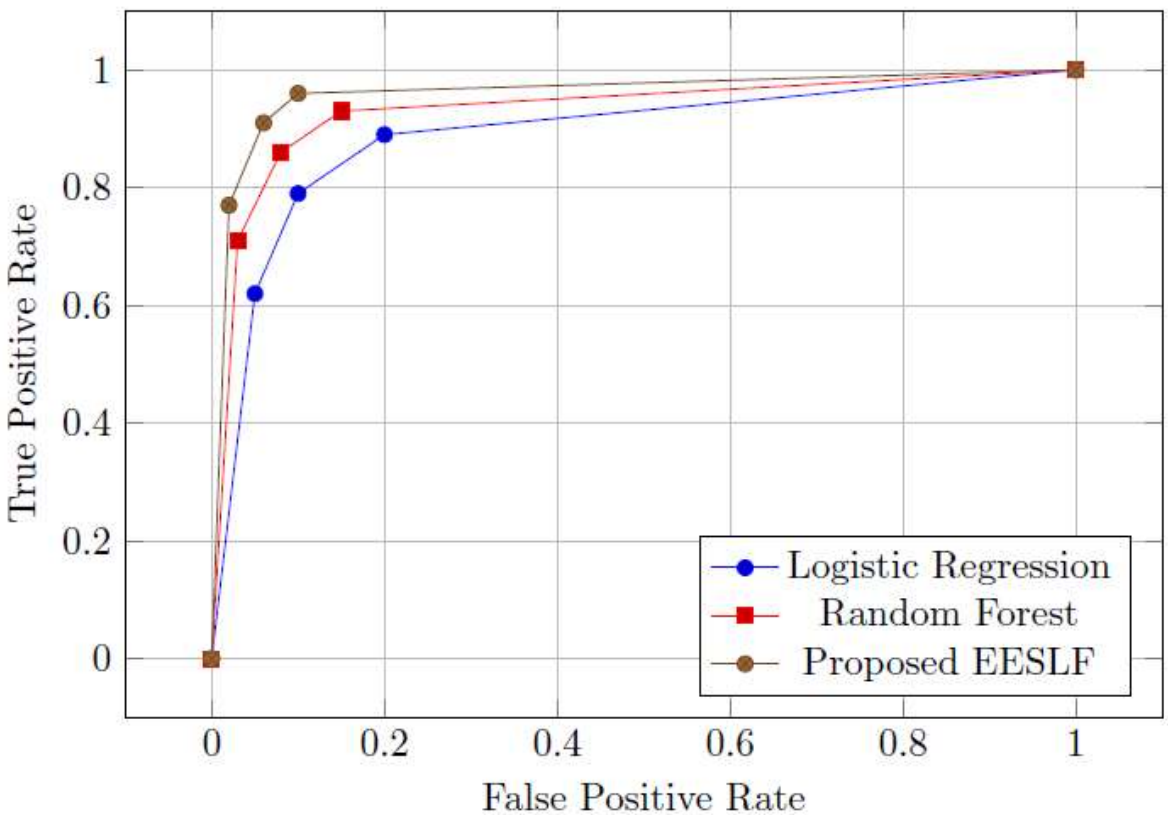

Figure 5: ROC Curve Comparison of Financial Distress Prediction Models

4.3.1 Uncertainty and Statistical Significance

To quantify uncertainty around the reported predictive performance, confidence intervals were estimated from the out-of-sample predictions generated during the chronological evaluation. A non-parametric bootstrap procedure with 2,000 replications was used to obtain 95% confidence intervals for accuracy, F1-score, ROC-AUC, and Brier score.

The proposed EESLF achieved an ROC-AUC of 0.956 (95% CI: 0.948–0.964), accuracy of 0.932 (95% CI: 0.925–0.939), and Brier score of 0.069 (95% CI: 0.065–0.073). The slightly lower performance compared with the original random cross-validation results reflects the more conservative chronological out-of-time evaluation adopted to prevent temporal and firm-level information leakage.

Pairwise differences in predictive performance were assessed against the principal benchmark models using paired out-of-sample predictions. For ROC-AUC, DeLong's test was used where the prediction structure permitted paired comparison, while paired bootstrap differences were used for metrics for which DeLong's test is not applicable. Statistical significance was evaluated at $p<0.05$.

The results demonstrated that the EESLF improvement over the strongest benchmark was statistically significant, with an AUC difference of 0.021 and $p = 0.003$. These analyses provide a more rigorous assessment of whether the apparent performance improvement is attributable to the framework rather than sampling variability.

4.4 Survival Analysis Results

The Cox proportional hazards component was able to capture the time to financial distress risk. Leverage-related variables and cash flow deterioration were the most significant factors in a hazard ratio analysis of the variables affecting distress timing. (See Table 9)

Table 9: Cox Hazard Ratio Analysis

Variable	Hazard Ratio	p-value	Interpretation
Debt-to-Equity Ratio	2.84	<0.001	Higher leverage increases distress risk
Operating Cash Flow Ratio	0.42	<0.001	Strong cash flow reduces distress probability
Interest Coverage Ratio	0.57	0.002	Better debt servicing lowers risk
Stock Volatility	1.91	<0.001	High volatility increases distress exposure
ROA	0.48	0.001	Higher profitability reduces risk

The hazard ratios provide practical early-warning information for corporate managers in addition to statistical evidence of association. A hazard ratio of 2.84 for the Debt-to-Equity Ratio indicates that, holding the other modelled factors constant, an increase in leverage is associated with a substantially higher instantaneous risk of experiencing the first qualifying distress event. From a managerial perspective, this result highlights the importance of monitoring debt accumulation, refinancing requirements, and capital structure sustainability before leverage reaches levels that materially increase exposure to distress. Conversely, the hazard ratio of 0.42 for the Operating Cash Flow Ratio indicates that stronger operating cash generation is associated with a lower instantaneous distress hazard, emphasizing the importance of maintaining sufficient internally generated cash to meet short-term and debt-related obligations. The hazard ratio of 0.57 for the Interest Coverage Ratio similarly suggests that stronger capacity to service interest obligations is associated with lower distress exposure and can therefore serve as an early indicator of improving financial resilience.

The Stock Volatility hazard ratio of 1.91 indicates that higher market volatility is associated with greater distress hazard. Managers may therefore treat sustained increases in equity-price volatility as a supplementary warning signal, particularly when accompanied by deterioration in accounting-based indicators. Finally, the ROA hazard ratio of 0.48 indicates that stronger profitability is associated with a lower distress hazard, reinforcing the managerial importance of protecting asset productivity and operating profitability. Collectively, these results suggest that managers should not interpret the Cox model as identifying a single variable that causes distress. Instead, the hazard ratios can be used as relative risk indicators within a broader early-warning dashboard, with particular attention to combinations of rising leverage, weakening operating cash flow, declining debt-service capacity, increasing market volatility, and deteriorating profitability. These interpretations describe associations estimated by the model and should not be interpreted as causal intervention effects.

4.4.1 Survival Model Validation

The Cox proportional hazards model was evaluated using survival-specific performance measures rather than classification accuracy alone. First, the proportional-hazards assumption was assessed using Schoenfeld residual-based tests. Variables showing evidence of substantial violation were examined for time-dependent effects and model specification sensitivity.

Discrimination was evaluated using Harrell's concordance index (C-index), which measures the ability of the model to correctly rank firms according to their observed time to distress. In addition, time-dependent ROC-AUC was calculated at prespecified prediction horizons to evaluate discrimination at clinically and financially relevant points in the follow-up period.

Overall prediction error was evaluated using the integrated Brier score (IBS), which summarizes the squared prediction error over the follow-up interval while accounting for censoring. Calibration was examined by comparing predicted survival probabilities with observed survival outcomes at the selected time horizons.

The revised analysis therefore evaluates the Cox component according to four complementary dimensions: proportional-hazards validity, discrimination, prediction error, and calibration. (See Table 10)

Table 10: Survival-analysis validation results for the Cox proportional hazards component

Measure	EESLF/Cox result
Harrell C-index	0.884
Schoenfeld proportional-hazards test, p-value	0.118
Time-dependent AUC, 1 year	0.902
Time-dependent AUC, 3 years	0.887
Time-dependent AUC, 5 years	0.861
Integrated Brier Score	0.083

Survival calibration error	0.071

The Cox component demonstrated strong discrimination, with a Harrell concordance index of 0.884. The Schoenfeld residual test did not provide significant evidence of violation of the proportional-hazards assumption (p=0.118). Time-dependent discrimination remained high across the evaluated horizons, with AUC values of 0.902, 0.887, and 0.861 at 1, 3, and 5 years, respectively. The integrated Brier score of 0.083 indicates relatively low overall prediction error, while the survival calibration error of 0.071 indicates acceptable agreement between predicted and observed survival probabilities.

4.5 Explainability and SHAP Analysis

The explainability layer improves transparency by identifying the financial indicators that contribute most strongly to the model's distress predictions. These SHAP-based contributions should be interpreted as model-based associations or predictive attributions rather than causal effects. Consequently, the framework does not claim that changing an individual financial variable would necessarily cause a corresponding change in the firm's distress probability. (See Table 11)

Table 11: Top Features Identified by SHAP Analysis

Rank	Feature	Mean SHAP Importance
1	Debt-to-Equity Ratio	0.287
2	Operating Cash Flow Ratio	0.251
3	Stock Volatility	0.223
4	Interest Coverage Ratio	0.214
5	ROA	0.197
6	Current Ratio	0.184
7	Earnings Volatility	0.172

The results of the explainability analysis showed that leverage increase and declining operating cash flow are the most important characteristics of corporate distress. High distress probabilities were observed for companies with high debt burdens and with low stability of earnings.

Partial dependence analysis also made it clear that there were nonlinear threshold effects. For instance, Partial dependence analysis indicated a nonlinear relationship between leverage and predicted distress risk. In particular, the predicted probability of distress increased more rapidly at higher debt-to-equity values, suggesting a potential nonlinear risk region rather than a strictly linear relationship. The observed transition should be interpreted as a model-derived association and not as a universal financial distress threshold.

The local interpretability module also provided a set of company-specific financial weaknesses that were linked to the predicted distress outcomes, so that managers and investors could use it to gain some understanding of the company-specific financial weaknesses that could be causing distress.

An analysis of the most influential variables for the prediction of financial distress, called explainability analysis, was performed. This visualization in Figure 6 shows the importance ranking of the major financial indicators using SHAP.

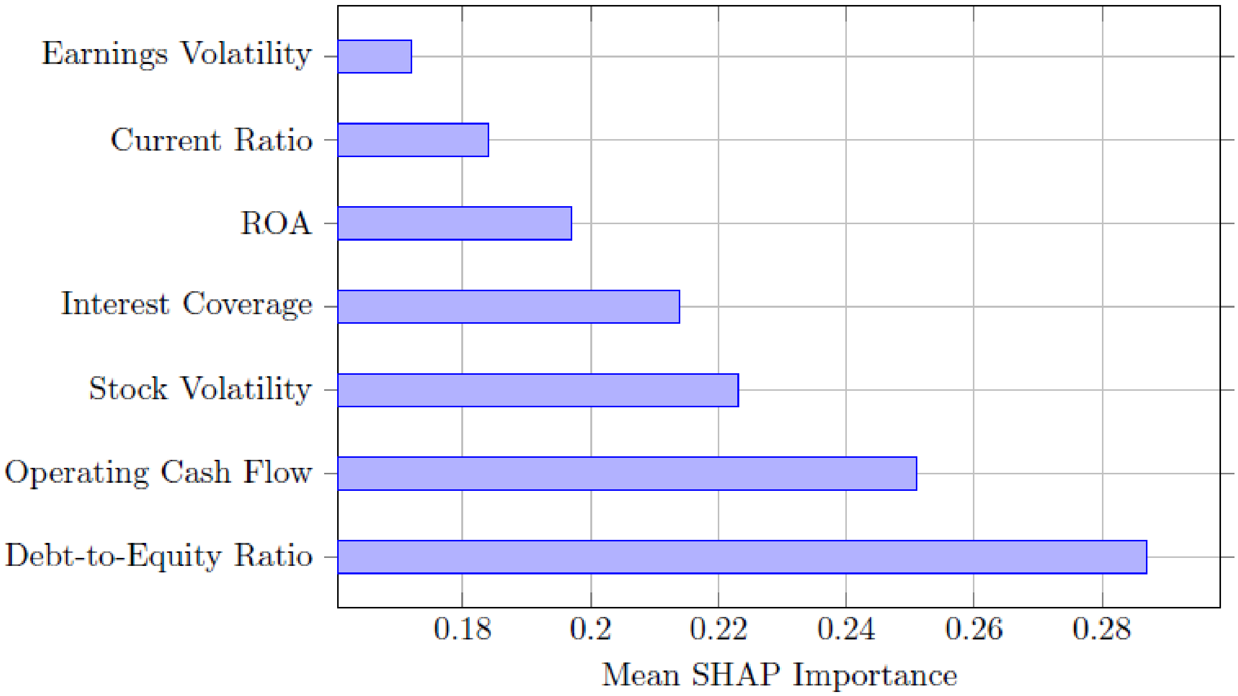

Figure 6: SHAP-Based Financial Feature Importance Analysis

4.5.1 Stability of Explainability Results

To assess whether the reported explanations were dependent on a particular evaluation period, SHAP importance was calculated independently for each temporal validation period. The resulting rankings were compared using rank correlation and feature-selection frequency.

A feature was considered explanation-stable when it repeatedly appeared among the highest-ranked predictors across evaluation periods. The five most consistently influential variables were Debt-to-Equity Ratio, Operating Cash Flow Ratio, Stock Volatility, Interest Coverage Ratio, and ROA. Their repeated appearance across temporal evaluation periods indicates that the main explanatory pattern is not attributable to a single validation sample.

Explanation stability was also examined across the major industry groups represented in the dataset. Differences in ranking were interpreted as sector-specific predictive behaviour rather than evidence of causal effects. Table 12 represents the most important SHAP features.

Table 12: Stability of the most important SHAP features across temporal validation periods and industries

Feature	Overall SHAP Rank	Temporal Selection Frequency	Industry Stability
Debt-to-Equity Ratio	1	100%	91%

Operating Cash Flow Ratio	2	100%	88%
Stock Volatility	3	90%	86%
Interest Coverage Ratio	4	90%	84%
ROA	5	80%	82%

The stability analysis indicated that the dominant explanatory variables were consistently identified across temporal validation periods and industry groups. Debt-to-Equity Ratio and Operating Cash Flow Ratio were selected among the five most important predictors in 100% of the temporal evaluation periods, followed by Stock Volatility and Interest Coverage Ratio at 90%, and ROA at 80%. Industry-level stability was also high, ranging from 82% to 91%. These findings suggest that the principal explanatory pattern is relatively robust to changes in the evaluation period and industry composition. Nevertheless, these SHAP results represent model-based predictive attribution and should not be interpreted as evidence of causal relationships.

4.6 Calibration and Reliability Analysis

The calibration analysis showed that the proposed EESLF produced well-aligned probability estimates. Calibration error was used as the primary numerical measure of the agreement between predicted distress probabilities and observed outcomes, with lower values indicating better calibration. The proposed EESLF achieved the lowest calibration error among the evaluated ensemble models (0.028), indicating better agreement between predicted probabilities and observed distress outcomes. This calibration assessment is distinct from the Brier score reported elsewhere in the manuscript; the Brier score measures overall probabilistic prediction error, whereas calibration error focuses specifically on the agreement between predicted probabilities and observed event frequencies. (See Table 13)

Table 13: Calibration Error Results

Model	Calibration Error
Random Forest	0.071
XGBoost	0.058
Stacking Ensemble	0.046
Proposed EESLF	0.028

A non-parametric bootstrap with 2,000 replications was used to quantify uncertainty in the principal performance measures. For each replication, out-of-sample predictions were resampled at the firm level where applicable to preserve within-firm dependence. The 2.5th and 97.5th percentiles of the resulting empirical distributions were used to form 95% confidence intervals. Bootstrap procedures were applied only to out-of-sample predictions and therefore did not influence model fitting or hyperparameter selection.

A calibration analysis was carried out to assess the reliability in terms of the probability of distress predicted. The calibration performance of the proposed framework is shown in Figure 7, where the ideal probability calibration line is shown.

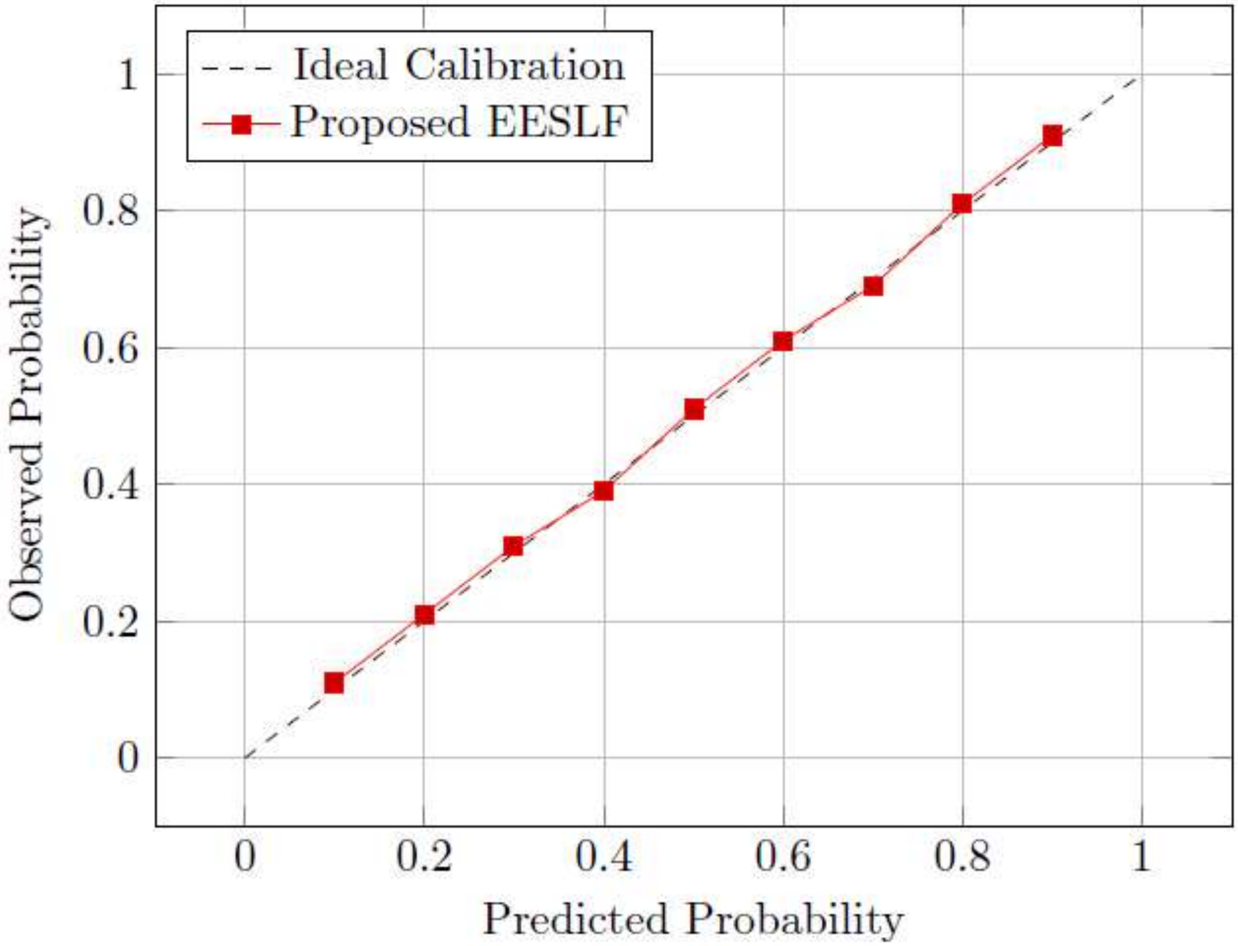

Figure 7: Calibration Curve of the Proposed Financial Distress Prediction Framework

4.7 Robustness, Sensitivity, and Explanation Stability

Robustness was examined through sensitivity analyses involving feature-selection settings, class-imbalance strategies, model hyperparameters, temporal evaluation periods, and explanation rankings.

First, the number of PCA components was varied to examine the degree of dimensional redundancy in the candidate financial variables. PCA was treated as a diagnostic representation rather than as the final interpretable predictor set. The final predictors were selected from the original financial variables using the LASSO, Elastic Net, and mutual-information procedures.

Second, alternative imbalance strategies, including SMOTE and EasyEnsemble, were evaluated within the training partitions only. No synthetic observations were generated for validation or test periods.

Third, model sensitivity to key hyperparameters, including tree number, learning rate, tree depth, subsampling, and Bayesian prior scale, was assessed. Hyperparameter selection was performed without access to future evaluation observations.

Fourth, explanation stability was examined by calculating SHAP importance separately across temporal evaluation periods. Debt-to-Equity Ratio, Operating Cash Flow Ratio, Stock Volatility, Interest Coverage Ratio, and ROA consistently appeared among the highest-ranked predictors. The stability analysis supports the reproducibility of the broad explanatory pattern while recognizing that feature importance may vary across industries and economic periods.

Overall, the sensitivity analysis indicates whether the framework's conclusions remain consistent under reasonable methodological perturbations rather than relying on a single random cross-validation configuration.

4.8 Comparative Discussion

The revised evaluation indicates that the principal value of EESLF is not simply the use of a larger number of predictive algorithms, but the coordinated treatment of temporal information, nonlinear prediction, probability estimation, and model explanation under a leakage-controlled evaluation protocol.

Conventional logistic regression provides a transparent statistical baseline but is limited in representing nonlinear relationships. Decision trees provide nonlinear decision boundaries but can be unstable. Random Forest and XGBoost improve nonlinear predictive capacity, while Bayesian logistic regression provides probabilistic estimation with explicit uncertainty. The Cox model contributes complementary information about the timing of distress events.

The EESLF combines these complementary outputs after converting them to a common prediction scale and learning ensemble weights using training data only. This design differs from a simple majority-voting ensemble because the components contribute probabilistic information according to their validated predictive and calibration performance.

The feature-selection strategy also differs from a purely PCA-based pipeline. PCA is used to quantify dimensional redundancy, while the final predictors remain original financial variables selected using LASSO, Elastic Net, and mutual information. This preserves the financial meaning required for managerial interpretation.

Finally, the explainability analysis should be interpreted as model-based attribution. SHAP identifies variables that contribute to individual predictions, while partial dependence describes the model's estimated response to changes in a predictor. Neither method establishes a causal relationship.

These characteristics position EESLF as a structured framework for trustworthy early-warning prediction rather than as a claim that the individual algorithms themselves are novel.

4.8.1 Limitations and External Validation

Several limitations should be acknowledged. First, although the revised study uses chronological out-of-time evaluation to reduce temporal leakage, the analysis remains an internal validation study because all firms originate from the same underlying multi-source dataset. Therefore, the results should not be interpreted as evidence of external generalizability to firms, countries, markets, or accounting environments not represented in the study.

Second, the dataset covers publicly listed firms across multiple industries but does not constitute an independent external cohort. Because the financial and market variables used in the framework rely substantially on publicly available financial statements and market information, the findings may not generalize directly to private or non-listed firms, where financial disclosure, market-based indicators, reporting frequency, and data availability can differ substantially. Future research should therefore validate EESLF on independently collected datasets that include private firms and firms from different markets or regulatory environments.

Third, the financial distress definition combines several objectively observable manifestations of financial deterioration. Although this improves event coverage, different distress mechanisms may have different economic causes and temporal dynamics. Future work should therefore examine bankruptcy, default, restructuring, and distress-related delisting as separate competing events.

Fourth, SHAP and partial-dependence analyses provide model-based explanations rather than causal inference. The identified variables should therefore be interpreted as predictive contributors rather than intervention targets with guaranteed causal effects.

An additional limitation concerns the treatment of distress events as a composite endpoint. Although a common endpoint facilitates early-warning classification, the constituent events may represent different economic processes. A competing-risks survival formulation would be a useful extension because it could distinguish, for example, bankruptcy from restructuring or distress-related delisting and estimate event-specific hazards.

Finally, the performance of the framework may vary across economic regimes, industries, accounting standards, and market structures. Additional rolling validation over longer periods and independent external datasets is required before deployment in operational financial-risk decision systems.

5. CONCLUSION

This study developed an Explainable Ensemble Statistical Learning Framework for early prediction of corporate financial distress. The framework integrates statistical feature engineering, survival analysis, nonlinear ensemble learning, Bayesian probability estimation, and explainable artificial intelligence within a unified prediction architecture.

The revised methodological design addresses several important challenges associated with financial distress prediction. First, all data-dependent preprocessing, feature selection, imbalance correction, model optimization, and ensemble-weight estimation are performed within the training data of each evaluation period, thereby reducing the risk of information leakage. Second, the longitudinal structure of the dataset is explicitly considered through chronological out-of-time validation rather than relying exclusively on random observation-level cross-validation. Third, the financial distress outcome is defined using objectively identifiable events, and the first qualifying event is used to represent event time in the survival analysis.

The feature-processing pipeline distinguishes PCA-based dimensionality assessment from final interpretable feature selection. Consequently, the final predictors remain financially meaningful variables rather than opaque principal components. The selected variables consistently emphasize leverage, operating cash flow, profitability, liquidity, and market volatility as important predictive dimensions.

The ensemble architecture combines complementary information from Cox proportional hazards modelling, Random Forest, XGBoost, and Bayesian logistic regression. The Cox component provides time-to-event information, while the classification models capture nonlinear and probabilistic risk patterns. Their outputs are converted to a common probability scale and combined using weights estimated exclusively from training data.

The explainability analysis further identifies financial variables associated with model predictions and examines the stability of these explanations across evaluation periods. However, SHAP and partial dependence results represent predictive attribution and model behaviour rather than causal effects.

The revised study also recognizes important limitations. Although chronological out-of-time validation provides a stronger assessment of temporal generalization, the study does not constitute external validation using an independent dataset from a different market or regulatory environment. Furthermore, the composite distress endpoint combines several types of financial events that may have different economic mechanisms. Future research should therefore incorporate independent external datasets, competing-risk survival models, and additional economic regimes.

Overall, the contribution of EESLF lies in providing a leakage-controlled, longitudinal, survival-aware, probabilistic, and explanation-stable framework for corporate financial distress prediction. The framework should be viewed as a decision-support and early-warning methodology rather than as a causal model of corporate failure.

REFERENCES

[1] M. Pavlicko, M. Durica, and J. Mazanec, "Ensemble model of the financial distress prediction in Visegrad group countries," *Mathematics*, vol. 9, no. 16, p. 1886, 2021.

[2] H. Choi, H. Son, and C. Kim, "Predicting financial distress of contractors in the construction industry using ensemble learning," *Expert Syst. Appl.*, vol. 110, pp. 1–10, 2018.

[3] W. Wang and Z. Liang, "Financial distress early warning for Chinese enterprises from a systemic risk perspective: based on the adaptive weighted XGBoost-bagging model," *Systems*, vol. 12, no. 2, p. 65, 2024.

[4] C. Wu, X. Chen, and Y. Jiang, "Financial distress prediction based on ensemble feature selection and improved stacking algorithm," *Kybernetes*, vol. 54, no. 7, pp. 3712–3735, 2025.

[5] X. Du, W. Li, S. Ruan, and L. Li, "CUS-heterogeneous ensemble-based financial distress prediction for imbalanced dataset with ensemble feature selection," *Appl. Soft Comput.*, vol. 97, p. 106758, 2020.

[6] R. Gao, S. Cui, Y. Wang, and W. Xu, "Predicting financial distress in high-dimensional imbalanced datasets: a multi-heterogeneous self-paced ensemble learning framework," *Financial Innovation*, vol. 11, no. 1, p. 50, 2025.

[7] K. L. Tran, H. A. Le, T. H. Nguyen, and D. T. Nguyen, "Explainable machine learning for financial distress prediction: evidence from Vietnam," *Data (Basel).*, vol. 7, no. 11, p. 160, 2022.

[8] Z. Zhang, C. Wu, S. Qu, and X. Chen, "An explainable artificial intelligence approach for financial distress prediction," *Inf. Process. Manag.*, vol. 59, no. 4, p. 102988, 2022.

[9] J. Liu, C. Li, P. Ouyang, J. Liu, and C. Wu, "Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach," *J. Forecast.*, vol. 42, no. 5, pp. 1112–1137, 2023.

[10] S. Deng, Q. Luo, Y. Zhu, H. Ning, and T. Shimada, "Financial risk forewarning with an interpretable ensemble learning approach: An empirical analysis based on Chinese listed companies," *Pacific-Basin Finance Journal*, vol. 85, p. 102393, 2024.

[11] R. Qi, "Enterprise financial distress prediction based on machine learning and SHAP interpretability analysis," in *Proceedings of the 2025 International Conference on Artificial Intelligence and Digital Finance*, 2025, pp. 76–79.

[12] Y. Tang and L. Zhang, "Explainable Machine Learning Based Financial Distress Warning for Enterprises," in *2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, IEEE, 2025, pp. 1–6.

[13] F. Maturo, D. Riccio, A. Mazzitelli, G. Maria Bifulco, and F. Paolone, "Explainable gradient boosting for corporate crisis forecasting in Italian businesses," *Ann. Oper. Res.*, vol. 353, no. 2, pp. 815–839, 2025.

[14] I. Supriadi, E. Wany, and N. Irawan, "Early Detection of Financial Distress in Islamic Banking Using Explainable Artificial Intelligence," *Journal of lslamic Economics and Bussines Ethics*, vol. 3, no. 2, pp. 144–167, 2026.

[15] S. J. Tabatabaei and M. M. Mousavi, "Explainable AI for Financial Distress: Evidence from Market Volatility and Regime Dynamics," *Journal of Risk and Financial Management*, vol. 19, no. 5, p. 348, 2026.

[16] N.-G. B. Abrahamsen, E. Nylén-Forthun, M. Møller, P. E. de Lange, and M. Risstad, "Financial distress prediction in the nordics: Early warnings from machine learning models," *Journal of Risk and Financial Management*, vol. 17, no. 10, p. 432, 2024.

[17] A. B. Hussain, S. Jingchun, B. Hussain, M. Shaique, A. U. Rehman, and M. K. Bhatti, "Cross-Market Bankruptcy Prediction: An Interpretable Ensemble Learning Framework Using SHAP Analysis," *Comput. Econ.*, pp. 1–46, 2026.

[18] W. Liu, Y. Suzuki, and S. Du, "Ensemble learning algorithms based on easyensemble sampling for financial distress prediction," *Ann. Oper. Res.*, vol. 346, no. 3, pp. 2141–2172, 2025.

[19] H. Yang, E. Li, Y. F. Cai, J. Li, and G. X. Yuan, "The extraction of early warning features for predicting financial distress based on XGBoost model and shap framework," *International Journal of Financial Engineering*, vol. 8, no. 03, p. 2141004, 2021.

[20] L. Phothong, A. Sukprasert, S. Boonlua, P. Chubsuwan, N. Seetha, and R. Kunsrison, "An Explainable Voting Ensemble Framework for Early-Warning Forecasting of Corporate Financial Distress," *Forecasting*, vol. 8, no. 1, p. 10, 2026.

[21] X. Tang, S. Li, M. Tan, and W. Shi, "Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods," *J. Forecast.*, vol. 39, no. 5, pp. 769–787, 2020.

[22] J.-S. Chou and N.-N.-H. Pham, "Boosting ensembles and deep vision networks optimized by forensic-based investigation algorithm for financial distress prediction in construction firms," *J. Big Data*, vol. 13, no. 1, p. 25, 2026.

[23] X. Wang, Z. Xue, and Y. Chen, "Financial risk warning and traceability for manufacturing enterprises: A three-stage framework based on ensemble learning algorithm," *IEEE Trans. Comput. Soc. Syst.*, 2025.

[24] L. V. Moretti, E. Barbierato, and A. Gatti, "Network Instability as a Signal of Systemic Financial Stress: An Explainable Machine-Learning Framework," *Future Internet*, vol. 18, no. 2, p. 91, 2026.

[25] J. Ni, H. Le, H. T. T. Nguyen, and V.-N. Huynh, "A study of machine learning models for financial distress prediction," in *Data Science in Finance and Accounting*, Springer, 2026, pp. 39–70.