

Article

Systems-Oriented Explainable AI for Corporate Bankruptcy Early Warning: A Deep Learning Framework for Risk-Attribution Analysis

Chenxi Yang ^{1,2} and Guangfan Sun ^{3,4,*} ¹ JD.com, Inc., Beijing 101111, China; chenxiyang0521@outlook.com² Shandong Jianda Education Real Estate Co., Ltd., Jinan 250101, China³ Aviation Industry Development Research Center of China, Beijing 100029, China⁴ School of Business, Renmin University of China, Beijing 100872, China

* Correspondence: aurora_sun@ruc.edu.cn

Highlights

What are the main findings?

- This study advances systems methodology by modeling corporate bankruptcy as an outcome of interacting financial subsystems rather than isolated indicators, introducing a hierarchical architecture that explicitly captures cross-subsystem dynamics for risk-attribution analysis.

What are the implications of the main findings?

- The proposed framework achieves strong and stable predictive performance while providing auditable subsystem-level attributions, demonstrating that explainable AI can translate accurate bankruptcy warnings into actionable financial-risk insights without sacrificing interpretability.

Abstract

Corporate bankruptcy early warning has often been treated as a binary classification task, yet financial distress is better understood as the outcome of interacting financial conditions that must be interpreted within practical risk-management contexts. To address this issue, this study proposes a systems-oriented explainable artificial intelligence framework for corporate bankruptcy early warning. The framework implements a Hierarchical LFRM-MACI architecture in which the Local Feature Refinement Module (LFRM) refines representations within profitability, solvency, liquidity, efficiency, and growth/shareholder performance subsystems, and cross-subsystem attention models their interactions. The framework is evaluated on the public UCI Taiwanese Bankruptcy Prediction dataset under balanced and moderately imbalanced training settings. For reporting clarity, the benchmark methods are classified into traditional machine learning methods and deep learning methods; traditional tabular learners are included in the formal single-split empirical comparison, while the proposed method's contribution is positioned as a structured deep representation with an attribution workflow rather than as an overall superiority claim over traditional machine learning models. In the single 70%/30% validation split, the proposed model obtains ROC-AUC values of 0.9286 and 0.9269 under the 1:1.0 and 1:2.5 settings, respectively. In repeated 5-fold cross-validation with five repetitions, its ROC-AUC is 0.8940 [0.8777, 0.9103] under 1:1.0 and 0.9135 [0.9001, 0.9269] under 1:2.5. To examine the interpretability of the predictions, Permutation Feature Importance (PFI) and SHAP are applied to identify subsystem-level attribution patterns across major financial subsystems. The explanation results highlight influential predictors associated with leverage pressure, profitability and asset structure, liquidity, operating efficiency, and



Academic Editor: Cristian Păun

Received: 20 May 2026

Revised: 17 July 2026

Accepted: 20 July 2026

Published: 1 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

growth/shareholder performance. These findings indicate that explainable AI can support corporate bankruptcy early warning when predictive benchmarking is combined with transparent and auditable attribution analysis for financial decision-making.

Keywords: bankruptcy early warning; explainable artificial intelligence; corporate financial distress; risk-attribution analysis; deep learning; attention mechanism; decision support; financial risk management

1. Introduction

Corporate bankruptcy early warning is a core function of modern business and management systems, yet it is often reduced to a narrow predictive modeling task that ignores its inherent systemic nature. Financial distress does not arise from a single isolated indicator; it reflects interacting conditions across profitability, leverage, liquidity, operating efficiency, growth expectations, reporting practices, and governance responses. For this reason, bankruptcy prediction should be understood as part of a business and management system in which data, analytical models, organizational routines, auditors, regulators, creditors, and managers jointly shape risk recognition and intervention. This systems perspective is consistent with the need for holistic, systems-based research on socio-technical and business systems, where analytical tools should explain how interrelated elements form a decision-relevant whole rather than optimize a detached mathematical model.

Within such a system, corporate transparency and sustainable governance remain central because they shape investor confidence, disclosure quality, capital allocation, and long-term organizational resilience [1,2]. Bankruptcy is one of the most severe outcomes of financial distress, and reliable early-warning signals can support supervisory review, credit monitoring, and firm-level risk management before deterioration becomes irreversible. However, the boundary of the relevant system extends beyond accounting ratios alone. A bankruptcy early-warning model must translate heterogeneous financial signals into explanations that can be inspected, challenged, and incorporated into institutional decision processes. Therefore, predictive accuracy and interpretability must be treated as coupled system design requirements rather than competing technical objectives, as the value of an early warning system ultimately depends on its ability to be integrated into existing organizational decision-making processes.

Artificial intelligence (AI) provides a promising toolset for this problem because financial indicators exhibit strong nonlinearity, cross-feature dependence, and heterogeneous reporting patterns across firms [3–6]. Recent studies further suggest that AI-enabled analytics can improve organizational learning, resource allocation, and sustainability-oriented decision support [7–9]. Yet AI systems used in governance-sensitive contexts cannot be evaluated only by aggregate accuracy. If model outputs cannot be explained in financially meaningful terms, boards, auditors, regulators, and creditors have limited basis for trusting, contesting, or operationalizing them [10,11]. Explainability is therefore not merely a reporting add-on; it is a mechanism that connects the technical subsystem of machine learning with the social subsystem of accountability and decision making.

This tension is amplified in bankruptcy prediction, where class imbalance, heterogeneous firm characteristics, and interacting financial ratios can distort both model training and interpretation. Prior studies have advanced bankruptcy and financial-distress prediction through hybrid deep-learning architectures and post hoc explanation tools [12–16], yet there remains a need for models that jointly emphasize predictive robustness, controlled

imbalance evaluation, and explanation quality. In parallel, data-quality issues such as size-related bias in disclosure-linked datasets continue to raise concerns about the reliability of governance-oriented AI systems [17]. From a holistic systems analysis standpoint, these limitations indicate that model design, data imbalance, element interdependence, and decision transparency are not isolated issues but must be examined as interconnected components of a unified socio-technical early warning system.

To address these gaps, this study develops a systems-oriented explainable AI (XAI) framework for corporate bankruptcy prediction. The model implements a Hierarchical LFRM-MACI architecture that first refines local representations within financial subsystems and then models cross-subsystem interactions among subsystem-level tokens. The framework treats financial ratios as a category-structured indicator system rather than as unrelated covariates, allowing local refinement and cross-subsystem attention to model relations within and across profitability, solvency, liquidity, efficiency, and growth/shareholder performance subsystems. In this design, the systems-oriented contribution is not limited to arranging variables into categories. The architecture operationalizes systems thinking at three connected levels: local subsystem encoders represent within-subsystem coupling among financially related indicators; Cross-Subsystem MACI represents interaction among subsystem summaries; and the PFI-SHAP attribution workflow translates learned model behavior back into subsystem-level risk evidence. The framework therefore studies bankruptcy warning as an interaction-driven financial system state rather than as an isolated feature-ranking problem. We further use Permutation Feature Importance (PFI) and SHAP to interpret the learned behavior of the model in a mathematically disciplined way [11,18,19]. Using the publicly available Taiwanese Bankruptcy Prediction Dataset [20], we evaluate the framework under balanced and moderately imbalanced settings. The benchmark methods are classified into traditional machine learning methods (RandomForest, XGBoost, LightGBM, CatBoost, SVM, and LogisticRegression) and deep learning methods (Transformer, LSTM, GRU, CNN1D, CNN-LSTM, MLP, and the proposed Hierarchical LFRM-MACI). Traditional tabular learners are included as formal single-split empirical comparisons, and the proposed method is positioned as a structured deep representation with an attribution workflow rather than as an overall superiority claim over traditional machine learning methods. The main comparison emphasizes ROC-AUC, while the broader experimental record also reports F1, PR-AUC, balanced accuracy, specificity, MCC, Brier score, ECE, and attribution outputs. For the attribution analysis, we focus on stable within-category indicator ordering across financial subsystems, thereby keeping the interpretation aligned with what the normalized importance tables can validly support. In this study, the systems-oriented perspective is implemented through representation, evaluation, and interpretation. Financial ratios are organized into functionally related subsystems, modeled along a structured indicator axis, evaluated under controlled class-distribution settings, and interpreted through subsystem-level attribution patterns. This design links the predictive model with financially meaningful risk interpretation while retaining a transparent empirical workflow.

This study contributes to systems-oriented AI and financial risk analysis in three ways. First, it proposes a hierarchical subsystem-aware architecture that combines local feature refinement with cross-subsystem interaction modeling, thereby representing firm-level distress signals as an interacting financial system rather than as isolated indicators. Second, it combines predictive benchmarking with PFI and SHAP attribution to translate model behavior into financial attribution profiles. Third, it evaluates the framework under balanced and moderately imbalanced training settings, with traditional machine learning methods reported as reference models and deep learning methods used as the main comparison group for the proposed architecture.

2. Related Work

2.1. Corporate Transparency, Sustainability Reporting, and Governance

The literature on sustainability reporting underscores that governance quality and stakeholder engagement significantly enhance the decision usefulness of corporate disclosures [1,2]. AI technologies are increasingly embedded within these reporting processes, enabling firms to integrate and analyze ESG information at scale. For instance, Du et al. [8] examine how AI standardization participation fosters sustainable development among strategic emerging enterprises in China. Yao and Xu [21] demonstrate that AI adoption mediates the relationship between political connections, corporate social responsibility, and innovation performance. Similarly, Sun et al. [22] show that AI applications can support sustainable breakthrough innovation in Chinese enterprises, while Damnjanović et al. [7] discuss AI-enabled strategic transformation and sustainable outcomes in SMEs. Recent work on sustainable aviation manufacturing further links AI-enabled green innovation with advanced manufacturing contexts in China [9]. Collectively, these studies indicate that AI supports sustainable governance through organizational learning, innovation, and decision-support capacity.

2.2. AI and Machine Learning for Financial Risk Analysis

AI-based financial risk modeling increasingly combines deep representation learning with strong tabular-data learners for structured accounting-ratio datasets. Comprehensive surveys summarize the adoption of AI and machine learning across finance and economics [3,4,23–25]. Neural-network methods have also been applied in adjacent financial-prediction tasks such as intelligent stock prediction, further showing the relevance of deep representation learning for financial signal extraction [26]. For bankruptcy prediction based on firm-level financial indicators, ensemble-tree, margin-based, and neural tabular learners provide important comparison methods because they often perform strongly on structured, heterogeneous, and class-imbalanced data. Accordingly, the retrained experiments in this study classify comparator models into traditional machine learning methods and deep learning methods under the same feature representation and imbalance-control protocol. Traditional machine learning models are included in the formal single-split empirical comparison, while the proposed Hierarchical LFRM-MACI model is interpreted as a structured deep representation and attribution workflow without claiming overall superiority over traditional tabular learners. Recent studies integrate hybrid modules or multi-source data to improve bankruptcy prediction and risk evaluation [13,27,28]. Additional financial-distress prediction research has examined explainable AI, deep-learning models, country-specific bankruptcy prediction, corporate governance links, and systematic reviews of statistical and AI-based approaches [5,6,29–33]. These advances motivate a reporting design that separates traditional tabular-data references from the deep learning comparison group while preserving this paper's hierarchical local-refinement and cross-subsystem interaction pipeline for cross-feature dependency modeling and ordered local feature refinement.

2.3. Explainable AI and Corporate Governance

Explainable artificial intelligence (XAI) has emerged as a vital enabler of responsible AI and sustainable governance. As Rudin et al. [10] emphasize, interpretability should not be treated merely as a post hoc visualization problem but as a fundamental design principle of trustworthy AI. PFI and SHAP provide complementary global and local tools for quantifying model reasoning and enhancing transparency across domains [18,19]. In the financial domain, XAI techniques have been increasingly applied to credit scoring, fraud detection, and corporate bankruptcy prediction, improving both accountability and

stakeholder trust [11,12,14,16]. Recent credit-card risk identification research has also used improved generative adversarial modeling to support digital-economy financial risk management [34]. Moreover, ESG data integrity and disclosure transparency have been shown to interact with AI interpretability, underscoring the need for governance-aware data pipelines and explainability-integrated reporting systems [17]. These findings collectively motivate the design of AI systems whose interpretive outputs can be incorporated into risk management, audit, and sustainability reporting frameworks, thereby aligning predictive intelligence with corporate transparency and long-term governance objectives.

2.4. AI for Sustainable Innovation and Strategic Transformation

Beyond financial risk prediction, AI plays a broader role in supporting sustainable innovation, supply-chain management, and stakeholder engagement. Damnjanović et al. [7] demonstrate that AI-enabled strategic transformation is associated with sustainable outcomes in SMEs, while Sun et al. [9] place AI-enabled green innovation in the context of sustainable aviation manufacturing in China. These studies collectively highlight the governance potential of AI as both a technological and institutional enabler of sustainability. Building upon this perspective, the present work contributes a governance-aligned AI framework that emphasizes explainability as a prerequisite for sustainable transparency.

2.5. Positioning and Contributions

However, few studies jointly examine representation learning, class-imbalance evaluation, and model explanation within a unified bankruptcy early-warning workflow. Most existing work emphasizes either predictive accuracy or post hoc explanation, leaving limited evidence on how structured financial indicators, benchmark stability, and attribution analysis can be combined. Accordingly, this research makes three key contributions: (1) It proposes a hierarchical subsystem-aware architecture that combines local feature refinement with cross-subsystem interaction modeling, thereby representing firm-level distress signals as an interacting financial system rather than as isolated indicators; (2) It integrates global and local attribution mechanisms (PFI and SHAP) into a financial sub-system taxonomy, translating model outputs into risk dimensions for interpretation [10,18,19]; (3) It conducts repeated-validation experiments under different class imbalance conditions on the public Taiwanese Bankruptcy Prediction Dataset [20], reporting traditional machine learning methods as formal single-split comparison models and positioning the proposed architecture as a structured deep representation using ROC-AUC, calibration, and attribution-consistency evidence. Through this integration, the framework connects benchmarked prediction with transparent attribution analysis for corporate bankruptcy early warning.

3. Methodology

The methodological framework of this study addresses two complementary objectives: (i) to predict corporate bankruptcy from structured firm-level financial indicators, and (ii) to make the learned warning signal interpretable through subsystem-level and indicator-level attribution. The prediction model is a Hierarchical LFRM-MACI architecture. It represents each firm as a category-ordered financial indicator system, applies LFRM-based local refinement within the five financial subsystems, models cross-subsystem relations with attention, and then produces a bankruptcy probability through a gated prediction head. The interpretation component uses Permutation Feature Importance (PFI) and SHAP on the same trained model and feature grouping, so that predictive evaluation and attribution analysis are aligned.

3.1. Systemic Early Warning Framework Architecture

The proposed model receives a standardized vector of 95 financial indicators and returns a probability that the firm observation belongs to the bankrupt class. Figure 1 provides a high-level conceptual schematic of the proposed information flow. The figure is intended to orient the reader to the main modules and should not be read as a complete layer-by-layer or code-level implementation diagram. The pipeline consists of six stages: feature-specific tokenization, local subsystem encoding, LFRM-based local refinement, subsystem-specific attention pooling, Cross-Subsystem MACI, and gated prediction from deep- and raw-feature representations.

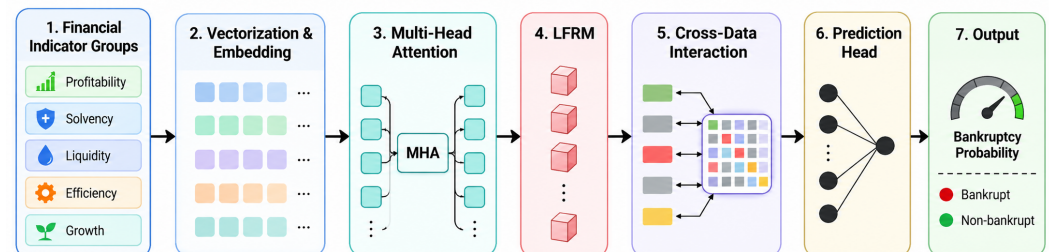


Figure 1. High-level conceptual schematic of the proposed Hierarchical LFRM-MACI workflow for bankruptcy prediction and attribution analysis. The figure summarizes the main information flow and should not be read as a complete layer-by-layer implementation diagram.

3.1.1. Data Representation and Financial Subsystem Ordering

To evaluate the proposed systemic early warning framework, we use the public Taiwanese Bankruptcy Prediction Dataset from the UCI Machine Learning Repository [20] as a standardized testbed for corporate financial system analysis. The released data contain 6819 firm observations, 95 quantitative financial variables, 1 binary target variable (Bankrupt, where 1 indicates bankruptcy and 0 indicates non-bankruptcy), and no missing values. Because the repository provides structured cross-sectional accounting ratios rather than longitudinal time series, each firm is represented as an ordered set of financial-system elements. The 95 predictors are grouped into five non-overlapping financial subsystems used consistently throughout this manuscript: profitability, solvency, liquidity, efficiency, and growth/shareholder performance. The default implementation order is 21 profitability indicators, 23 solvency indicators, 25 liquidity indicators, 13 efficiency indicators, and 13 growth/shareholder performance indicators. This ordering is fixed before training and is used to provide financial structure for local encoding and attribution; it is not a calendar-time order. Accordingly, the feature axis used by LFRM is the fixed within-subsystem indicator axis defined in this study. Indicators are placed in a consistent functional grouping and order to enable local modeling; this ordering is an implementation choice rather than a time axis. The transform-domain pathway provides an auxiliary representation of local feature variation and assigns no predefined economic or causal meaning to individual transform coefficients.

Formally, the dataset is represented as

$$\mathcal{X} = \{x_{i,m} \mid i = 1, \dots, N; m = 1, \dots, M\}, \quad (1)$$

where $x_{i,m}$ denotes the value of the m -th financial variable of firm i , N is the number of observations, and $M = 95$ is the number of predictors. All input features are standardized using training-set statistics only:

$$\tilde{x}_{i,m} = \frac{x_{i,m} - \mu_m}{\sigma_m}, \quad (2)$$

where μ_m and σ_m are computed from the corresponding training partition or training fold. After normalization and category-based reordering, each firm is represented as

$$\mathbf{x}_i = [\tilde{x}_{i,1}, \tilde{x}_{i,2}, \dots, \tilde{x}_{i,M}]^\top \in \mathbb{R}^M. \quad (3)$$

Let $\mathcal{I}_r$ denote the index set for subsystem $r \in \{1, \dots, 5\}$. The index sets are disjoint and satisfy $\cup_{r=1}^5 \mathcal{I}_r = \{1, \dots, M\}$.

3.1.2. Feature-Specific Tokenization

The model does not use a single shared linear projection for all variables. Instead, each financial indicator has a feature-specific value vector, a feature identity embedding, and a group identity embedding. For feature m in subsystem $g(m)$, the input token is

$$\mathbf{t}_{i,m} = \tilde{x}_{i,m} \mathbf{w}_m + \mathbf{b}_m + \mathbf{e}_m + \mathbf{u}_{g(m)}, \quad (4)$$

where $\mathbf{w}_m \in \mathbb{R}^d$ is the feature-specific value weight, $\mathbf{b}_m \in \mathbb{R}^d$ is a feature-specific bias, $\mathbf{e}_m \in \mathbb{R}^d$ is a learnable feature identity embedding, and $\mathbf{u}_{g(m)} \in \mathbb{R}^d$ is the learnable subsystem embedding. Layer normalization and dropout are then applied to obtain the token matrix

$$\mathbf{T}_i = \text{Dropout}(\text{LayerNorm}([\mathbf{t}_{i,1}, \dots, \mathbf{t}_{i,M}])), \quad \mathbf{T}_i \in \mathbb{R}^{M \times d}. \quad (5)$$

This tokenizer preserves the identity of individual accounting ratios while making group membership explicit to the network.

3.1.3. Local Subsystem Encoding and LFRM

Each financial subsystem is processed by its own branch. For subsystem r , the branch first selects the token subset $\mathbf{T}_{i,r} = \mathbf{T}_i[\mathcal{I}_r]$ and passes it through a pre-normalized Transformer block with residual connections. For a local hidden matrix $\mathbf{Z}$, one block is written as

$$\mathbf{Z}' = \mathbf{Z} + \gamma_a \text{MHA}(\text{LayerNorm}(\mathbf{Z})), \quad (6)$$

$$\mathbf{Z}'' = \mathbf{Z}' + \gamma_f \text{FFN}(\text{LayerNorm}(\mathbf{Z}')), \quad (7)$$

where $\text{MHA}(\cdot)$ is multi-head self-attention, $\text{FFN}(\cdot)$ is a GELU feed-forward network, and γ_a and γ_f are learnable LayerScale parameters. This local encoder allows each subsystem to learn dependencies among financially related indicators before cross-subsystem interaction is introduced.

The Local Feature Refinement Module (LFRM) is then applied within each subsystem branch. LFRM supplements the local attention representation with a lightweight transform-domain pathway operating along the fixed within-subsystem indicator axis. The RFFT/IRFFT operations provide an auxiliary representation of local feature variation, while a bounded learnable gain and residual gate control the refinement process. This design does not assume periodic structure over time or assign predefined economic meanings to individual transform coefficients. For a subsystem representation $\mathbf{Z}_{i,r} \in \mathbb{R}^{M_r \times d}$, LFRM first computes

$$\mathbf{U}_{i,r} = \text{GELU}(\mathbf{W}_p \text{LayerNorm}(\mathbf{Z}_{i,r}) + \mathbf{b}_p). \quad (8)$$

An RFFT is applied along the fixed within-subsystem indicator axis:

$$\mathbf{S}_{i,r} = \text{RFFT}(\mathbf{U}_{i,r}). \quad (9)$$

The transform coefficients are reweighted by a bounded learnable gain,

$$\mathbf{G}_r = 1 + \delta \tanh(\mathbf{A}_r), \quad \tilde{\mathbf{S}}_{i,r} = \mathbf{S}_{i,r} \odot \mathbf{G}_r, \quad (10)$$

where δ limits the maximum gain change, and $\mathbf{A}_r$ is learned. The locally refined representation is mapped back to the feature domain by an IRFFT:

$$\hat{\mathbf{U}}_{i,r} = \text{IRFFT}(\tilde{\mathbf{S}}_{i,r}). \quad (11)$$

Finally, a learned residual gate controls how much auxiliary transform-domain information is added to the local attention representation:

$$\mathbf{L}_{i,r} = \text{LayerNorm}(\mathbf{Z}_{i,r} + \text{Dropout}(\sigma(\boldsymbol{\eta}_r) \odot \hat{\mathbf{U}}_{i,r})). \quad (12)$$

Thus, LFRM performs bounded transform-domain reweighting within each financial subsystem, while the residual path preserves the local attention representation. Figure 2 summarizes this transform-domain local refinement at a high conceptual level; the equations above provide the implementation details.

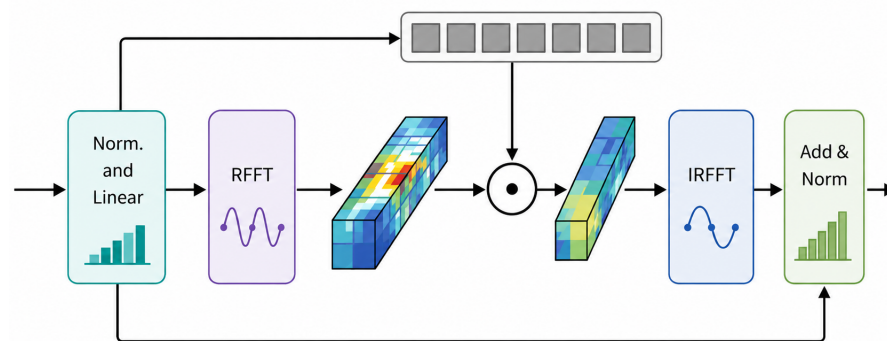


Figure 2. High-level conceptual schematic of the Local Feature Refinement Module (LFRM). The figure illustrates transform-domain local refinement over a fixed within-subsystem indicator axis and is not intended as a complete layer-by-layer code diagram.

Each subsystem branch then uses attention pooling to produce one subsystem token. The pooling weight for local feature position m in subsystem r is

$$\alpha_{i,r,m} = \text{Softmax}_m(\mathbf{v}^\top \tanh(\mathbf{W}_a \text{LayerNorm}(\mathbf{L}_{i,r,m}))), \quad (13)$$

and the subsystem token is

$$\mathbf{c}_{i,r} = \sum_{m=1}^{M_r} \alpha_{i,r,m} \mathbf{L}_{i,r,m}. \quad (14)$$

The model therefore compresses each subsystem into a learned summary token while retaining indicator-level attention weights for later inspection.

3.1.4. Cross-Subsystem MACI and Gated Prediction

The Cross-Subsystem MACI module models interactions among the five subsystem tokens rather than applying attention over all 95 raw feature positions at once. The subsystem tokens are stacked as

$$\mathbf{C}_i = [\mathbf{c}_{i,1}, \dots, \mathbf{c}_{i,5}] \in \mathbb{R}^{5 \times d}. \quad (15)$$

A learnable class token and subsystem position embeddings are added before cross-subsystem attention:

$$\mathbf{R}_i^{(0)} = [\mathbf{c}_{\text{cls}}; \mathbf{C}_i + \mathbf{P}], \quad (16)$$

where $\mathbf{P} \in \mathbb{R}^{5 \times d}$ stores the subsystem position embeddings. The resulting six-token sequence is passed through pre-normalized Transformer blocks with residual attention and feed-forward layers. The output class token is used as the deep cross-subsystem representation:

$$\mathbf{h}_i^{\text{deep}} = \mathbf{R}_{i,0}^{(L)}. \quad (17)$$

This design makes MACI a cross-subsystem interaction module: it learns how profitability, solvency, liquidity, efficiency, and growth/shareholder performance summaries jointly shape the bankruptcy warning signal.

A raw-feature branch is included to preserve information that may be weakened by subsystem compression. It applies layer normalization and a small multilayer perceptron directly to the standardized 95-dimensional vector:

$$\mathbf{h}_i^{\text{raw}} = \text{MLP}_{\text{raw}}(\text{LayerNorm}(\mathbf{x}_i)). \quad (18)$$

The deep and raw representations are combined through a gated fusion layer:

$$\mathbf{g}_i = \sigma(\mathbf{W}_g[\mathbf{W}_d \mathbf{h}_i^{\text{deep}}; \mathbf{W}_r \mathbf{h}_i^{\text{raw}}] + \mathbf{b}_g), \quad (19)$$

$$\mathbf{h}_i = \text{LayerNorm}(\text{Dropout}(\mathbf{g}_i \odot \mathbf{W}_d \mathbf{h}_i^{\text{deep}} + (1 - \mathbf{g}_i) \odot \mathbf{W}_r \mathbf{h}_i^{\text{raw}})). \quad (20)$$

A feed-forward prediction head maps $\mathbf{h}_i$ to a logit z_i and bankruptcy probability p_i :

$$p_i = \sigma(z_i) = \sigma(\text{Head}(\mathbf{h}_i)). \quad (21)$$

The binary decision is obtained with a probability threshold θ :

$$\hat{y}_i = \begin{cases} 1, & p_i \geq \theta, \\ 0, & p_i < \theta, \end{cases}$$

where $\hat{y}_i = 1$ indicates bankruptcy, and $\hat{y}_i = 0$ indicates non-bankruptcy.

3.1.5. Loss Function, Variants, and Threshold Selection

The model is trained with weighted binary cross-entropy to account for the rare bankrupt class:

$$\mathcal{L}_{\text{main}} = -\frac{1}{N} \sum_{i=1}^N [w_1 y_i \log(p_i) + w_0 (1 - y_i) \log(1 - p_i)], \quad (22)$$

where $y_i \in \{0, 1\}$ is the released bankruptcy label, and w_1 and w_0 are class weights. When auxiliary subsystem heads are used, each contextual subsystem token also predicts the same bankruptcy label. The total training loss is

$$\mathcal{L} = \mathcal{L}_{\text{main}} + \lambda_{\text{aux}} \mathcal{L}_{\text{aux}}, \quad (23)$$

where λ_{aux} controls the contribution of the auxiliary subsystem losses.

The optimal decision threshold θ_{opt} is selected by Youden's J statistic within the corresponding validation split or validation fold:

$$\theta_{\text{opt}} = \arg \max_{\theta} [\text{Sensitivity}(\theta) + \text{Specificity}(\theta) - 1]. \quad (24)$$

Consequently, threshold-dependent metrics are interpreted as validation or repeated-validation metrics under the stated split protocol.

3.2. System-Level Attribution and Interpretation Pipeline

Beyond accurate system state identification, a corporate financial early warning system should indicate why a firm is predicted to fall into bankruptcy. This study uses one unified attribution pipeline for all PFI and SHAP analyses reported in Section 5. The same trained proposed model, feature grouping, and imbalance settings are used throughout the interpretability analysis.

3.2.1. Global System-Level Attribution via Permutation Feature Importance (PFI)

Permutation Feature Importance measures how much model loss increases when one input feature is randomly permuted. For feature x_d , the PFI score is computed as the increase in binary cross-entropy (BCE) loss after permutation:

$$I_d = L_{\text{BCE}}(f(X^{\pi_d}), y) - L_{\text{BCE}}(f(X), y), \quad (25)$$

where X^{π_d} denotes the validation matrix after permuting feature d , $f(\cdot)$ is the trained model, and y is the bankruptcy label. A larger positive value indicates that the feature has stronger global influence on the prediction task.

3.2.2. Local Attribution via SHAP

SHAP is used to summarize local feature contributions for individual predictions. In the final attribution pipeline, SHAP values are estimated with KernelExplainer using 100 explained samples and 50 background samples. The SHAP output is then aggregated to identify recurring local contribution patterns across the validation observations. This local explanation complements PFI because it describes contribution structure rather than only global loss sensitivity.

3.2.3. Preprocessing and Financial Taxonomy

All PFI and SHAP results are organized according to the five financial indicator groups in Table 1: profitability, solvency, liquidity, efficiency, and growth/shareholder performance indicators. Within each category, method, and imbalance setting, the largest score is set to 1.000, and the remaining scores are normalized relative to that maximum. This category-wise normalization is the only normalization used for the reported PFI/SHAP summaries. Accordingly, the PFI/SHAP tables in Section 5 report relative profile strength from this single final attribution pipeline.

Table 1. Study-specific reclassification of the 95 Taiwanese bankruptcy predictors into five analytical groups, based on the detailed UCI variable appendix and used here for analytical convenience.

Category	Count	Integrated UCI Variables
Profitability and Earnings Quality	21	X10 Operating Income/Capital; X11 Pretax Income/Capital; X38 Operating Expenses/Net Sales; X39 Research and Development Expenses/Net Sales; X40 Effective Tax Rate; X51–X60 return, gross-margin, operating-margin, pretax-income, net-income, and non-operating-income ratios; X64–X69 total income to total expenses, total expenses to assets, net income to total assets, gross profit to sales, net income to stockholder’s equity, and the negative-income indicator.

Table 1. *Cont.*

Category	Count	Integrated UCI Variables
Solvency and Capital Structure	23	X1 Cost of Interest-bearing Debt; X5–X9 interest-burden, liability, and contingent-liability ratios; X18 Current Liability to Assets; X22 Current Liabilities/Liability; X24–X37 current-, long-term-, and total-liability structure ratios together with leverage and interest-coverage measures; X63 Retained Earnings to Total Assets.
Liquidity and Cash Flow Stability	25	X2 Cash Reinvestment Ratio; X3 Current Ratio; X4 Acid Test; X12–X17 working-capital, quick-asset, cash, and current-liability coverage ratios; X19–X21 operating-funds and inventory-liquidity ratios; X23 Working Capital/Equity; X44 Cash Flow Per Share; X77–X87 cash, working-capital, no-credit-interval, and operating cash-flow ratios.
Efficiency and Asset Utilization	13	X45–X50 sales, operating income, fixed assets per employee, and total assets to GNP price; X70–X76 inventory/accounts receivable to equity and turnover ratios for total assets, receivables, inventory, fixed assets, and equity.
Growth and Shareholder Performance	13	X41–X43 Book Value Per Share variants; X61 EPS-Net Income; X62 Pretax Income Per Share; X88–X95 realized gross profit, operating income, net income, continuing operating income, total asset, total equity, and return-on-total-asset growth indicators.

4. Data

4.1. Dataset Description

This study utilizes the public Taiwanese Bankruptcy Prediction dataset hosted by the UCI Machine Learning Repository and documented by Pham et al. [20] as a standardized testbed for corporate financial early warning system validation. The released benchmark contains 6819 firm observations, 95 predictors, 1 binary target variable, and no missing values. The original dataset contains 220 bankrupt observations (3.23%) and 6599 non-bankrupt observations (96.77%). Under the stratified 70%/30% split used for the single-split experiment, the training partition contains 154 bankrupt and 4619 non-bankrupt observations, while the validation partition contains 66 bankrupt and 1980 non-bankrupt observations. The financial system elements are derived from Taiwanese firms and are organized into the five analytical subsystems used consistently in this study: profitability, solvency, liquidity, efficiency, and growth/shareholder performance. Leverage-related, cash-flow, and efficiency-related indicators are incorporated within these five subsystem groups rather than treated as additional subsystems. Because the repository exposes both a short variable table and a detailed “Additional Variable Information” appendix, all systemic regrouping and indicator naming in this manuscript follow the appendix notation (X1–X95) while preserving the original binary target. Table 1 summarizes how the full set of 95 predictors is systematically organized in this study, for analytical convenience, into the five analytical groups used throughout this article.

4.2. Data Preprocessing

Prior to system training and validation, a standardized system input preprocessing procedure was applied to ensure the consistency and reliability of the dataset:

- Variable regrouping: The 95 raw predictors were reordered according to the five-category framework in Table 1 so that related accounting ratios occupy contiguous positions in the model input.
- Missing values: No imputation step was required because the released UCI dataset reports no missing values.

- Feature normalization: All numerical predictors were standardized using z-score normalization based on the training-set statistics only, preventing information leakage across dataset splits.
- Data splitting: The dataset was divided using a stratified random split of 70% training and 30% validation observations, maintaining the original bankruptcy proportion in each subset.

Winsorization thresholds and z-score statistics are estimated only from the training partition, or from the training folds in repeated validation, and then applied unchanged to the corresponding validation data. We did not apply log transformations or robust scaling in the reported benchmark because the objective was to preserve comparability with the released financial-ratio benchmark after leakage-controlled winsorization and standardization. These preprocessing steps reduce information leakage and improve comparability across firm observations while remaining faithful to the released benchmark.

4.3. Target Variable Definition

The dependent variable is the original repository label, where $y = 1$ denotes a bankrupt firm observation and $y = 0$ denotes a non-bankrupt observation. We therefore do not reconstruct bankruptcy events from auxiliary accounting rules or future-period heuristics; the released binary label is used directly throughout training, validation, and interpretation.

4.4. Data Relevance to Sustainability and Governance

The chosen dataset aligns with the sustainability-oriented perspective of this study. Corporate bankruptcy risk is often intertwined with deficiencies in corporate governance systems, transparency, and systemic risk management practices. By using structured accounting ratios and cash-flow indicators to predict and explain distress patterns, this framework provides potential decision-support value for early warning, accountability-oriented review, and informed financial decision-making.

5. Results and Discussion

5.1. System Validation Experimental Setup

All predictive experiments were conducted using the preprocessed Taiwanese Bankruptcy Prediction dataset described in the Section 4. Model performance was evaluated using both a stratified 70%/30% training-validation split and repeated stratified cross-validation with five folds and five repetitions. For each split, resampling, preprocessing, threshold selection, and model evaluation were performed within the corresponding training and validation data to avoid information leakage. The updated Hierarchical LFRM-MACI run used fixed 200-epoch training without early stopping, and the validation-best checkpoint was retained for reporting. For reproducibility, all random splitting and resampling operations used random state 42. The full Hierarchical LFRM-MACI configuration used embedding dimension 64, four attention heads, feed-forward width 128, one local subsystem encoder layer, two Cross-Subsystem MACI layers, dropout 0.20, auxiliary subsystem heads, and the raw-feature branch. The neural-network training configuration used the AdamW optimizer with learning rate 1×10^{-3} , batch size 128, weighted binary cross-entropy, 200 epochs, no early stopping, and validation-best checkpoint selection. The two training imbalance settings were retained: bankrupt-to-non-bankrupt ratios of 1:1.0 (balanced) and 1:2.5 (moderate imbalance). The resampling algorithm was applied exclusively to the training partition or training folds through random majority-class under-sampling to the target bankrupt-to-non-bankrupt ratio; validation observations preserved their original class distribution. The documented software environment includes Python 3.12.7, NumPy 1.26.4, pandas 2.2.2, scikit-learn 1.5.1, imbalanced-learn 0.12.3, and SciPy

1.13.1, with PyTorch 2.0.1 used for neural models and XGBoost, LightGBM, and CatBoost used for the corresponding boosting baselines.

For clarity, the evaluated methods are organized into two categories. Traditional machine learning methods, including RandomForest, XGBoost, LightGBM, CatBoost, SVM, and LogisticRegression, are retained as tabular-data reference models [35–40]. Deep learning methods, including Transformer, LSTM, GRU, CNN1D, CNN-LSTM, MLP, and the proposed Hierarchical LFRM-MACI model, form the main comparison group for assessing the proposed representation-learning architecture [41–46]. The single-split experiment reports F1, ROC-AUC, PR-AUC, balanced accuracy, specificity, MCC, Brier score, and ECE. A subsequent repeated-validation summary focuses only on the proposed method, because its purpose is to document the stability of Hierarchical LFRM-MACI across the two imbalance settings rather than to compare it directly with traditional machine learning methods. Together, these metrics describe discrimination, precision–recall behavior, confusion-matrix balance, and calibration under rare-event bankruptcy screening.

5.2. System State Identification Performance and Robustness Analysis

The first objective of this study is to evaluate whether the proposed hierarchical local-refinement and cross-subsystem interaction framework provides a competitive deep learning representation for bankruptcy early warning. This subsection reports the single-split results for traditional machine learning references and deep learning methods under the two retained imbalance settings, followed by the repeated-validation ROC-AUC summary for the proposed method. Traditional machine learning models are reported to contextualize the tabular benchmark, whereas the performance claim for Hierarchical LFRM-MACI is made within the deep learning and representation-learning group.

5.2.1. Single-Split Validation Results

Table 2 reports the single 70%/30% validation results under the balanced 1:1.0 training imbalance setting. Table 3 reports the corresponding results under the moderate 1:2.5 setting. Both tables separate traditional machine learning methods from deep learning methods. Traditional tabular learners are reported as formal empirical comparison models, while the proposed Hierarchical LFRM-MACI model is interpreted as a structured deep representation within the deep learning group. The same validation set contains 2046 firm observations, including 66 bankrupt and 1980 non-bankrupt observations.

Table 2. Single-split validation results under the balanced 1:1.0 training imbalance setting, grouped by traditional machine learning references and deep learning methods.

Model	F1	ROC-AUC	PR-AUC	Bal. Acc.	Spec.	MCC	Brier	ECE
<i>Traditional machine learning references</i>								
RandomForest	0.2864	0.9422	0.3660	0.8806	0.8520	0.3559	0.1038	0.1971
XGBoost	0.2519	0.9480	0.3949	0.8952	0.8056	0.3372	0.1153	0.1595
LightGBM	0.2864	0.9486	0.4314	0.8866	0.8490	0.3588	0.1264	0.1483
CatBoost	0.2788	0.9450	0.3870	0.8957	0.8369	0.3574	0.1140	0.1617
SVM	0.2383	0.9171	0.2826	0.8649	0.8056	0.3123	0.1193	0.2156
LogisticRegression	0.2095	0.8994	0.2274	0.8525	0.7657	0.2850	0.1256	0.1791
<i>Deep learning methods</i>								
Transformer	0.2289	0.8721	0.1446	0.8146	0.8263	0.2812	0.1698	0.2937
LSTM	0.1536	0.8435	0.1570	0.7619	0.7056	0.1998	0.2044	0.3362
GRU	0.2418	0.8979	0.2558	0.8323	0.8313	0.2993	0.1412	0.2282
CNN1D	0.2470	0.8833	0.2271	0.8174	0.8470	0.2962	0.1395	0.2717
CNN-LSTM	0.2068	0.8783	0.3005	0.8129	0.7924	0.2636	0.1756	0.3560
MLP	0.2456	0.8603	0.3173	0.8399	0.8313	0.3058	0.1254	0.2230
Proposed (Hierarchical LFRM-MACI)	0.2274	0.9286	0.2627	0.8755	0.7813	0.3094	0.1075	0.1491

Note: Text in italics denotes category headings that group models into traditional machine learning references and deep learning methods.

The single-split results are interpreted in two categories. Traditional machine learning methods are retained as formal single-split comparison models for the tabular benchmark. Because gradient-boosting and ensemble-tree models are highly competitive and outperform the proposed model on several metrics, this manuscript does not claim overall superiority over traditional machine learning methods. Within the deep learning group under the 1:1.0 setting, Hierarchical LFRM-MACI obtains the highest ROC-AUC (0.9286), balanced accuracy (0.8755), MCC (0.3094), lowest Brier score (0.1075), and lowest ECE (0.1491). Under the 1:2.5 setting, the proposed model again obtains the highest ROC-AUC (0.9269), balanced accuracy (0.8672), lowest Brier score (0.1114), and lowest ECE (0.1648) within the deep learning group. These results support the proposed architecture as a discrimination- and calibration-oriented deep learning model for structured bankruptcy early warning, while the traditional machine learning results document the competitiveness of established tabular learners.

Table 3. Single-split validation results under the moderate 1:2.5 training imbalance setting, grouped by traditional machine learning references and deep learning methods.

Model	F1	ROC-AUC	PR-AUC	Bal. Acc.	Spec.	MCC	Brier	ECE
<i>Traditional machine learning references</i>								
RandomForest	0.2592	0.9377	0.3448	0.8694	0.8298	0.3301	0.0565	0.1087
XGBoost	0.2905	0.9495	0.4038	0.8881	0.8520	0.3626	0.0599	0.0809
LightGBM	0.2926	0.9429	0.4215	0.8889	0.8535	0.3645	0.0813	0.0906
CatBoost	0.2743	0.9480	0.4540	0.8879	0.8364	0.3503	0.0749	0.1096
SVM	0.2302	0.9157	0.2793	0.8604	0.7965	0.3041	0.0681	0.1286
LogisticRegression	0.1905	0.8729	0.2045	0.8273	0.7455	0.2588	0.1120	0.1639
<i>Deep learning methods</i>								
Transformer	0.2222	0.8871	0.2043	0.8167	0.8152	0.2770	0.1365	0.2263
LSTM	0.2308	0.8536	0.1939	0.7705	0.8591	0.2618	0.1610	0.2826
GRU	0.2337	0.9011	0.2996	0.8510	0.8081	0.3021	0.1611	0.2655
CNN1D	0.2308	0.8894	0.2383	0.8212	0.8242	0.2857	0.1473	0.2670
CNN-LSTM	0.3542	0.9086	0.3808	0.8432	0.9136	0.3899	0.1284	0.2171
MLP	0.2799	0.8951	0.3409	0.8480	0.8626	0.3356	0.1124	0.2189
Proposed (Hierarchical LFRM-MACI)	0.2542	0.9269	0.2937	0.8672	0.8253	0.3253	0.1114	0.1648

Note: Text in italics denotes category headings that group models into traditional machine learning references and deep learning methods.

5.2.2. Repeated Validation Results

Table 4 reports the repeated-validation ROC-AUC of the proposed Hierarchical LFRM-MACI model under the two retained imbalance settings. Traditional machine learning models are not included in this table, because Table 4 is intended to document within-model repeated-validation stability for the proposed deep learning architecture. Each setting was evaluated using 5-fold cross-validation with five repetitions.

Table 4. Repeated-validation ROC-AUC of the proposed Hierarchical LFRM-MACI model under two training imbalance settings.

Training Ratio	ROC-AUC	95% CI	Rounds
1:1	0.8940 ± 0.0394	[0.8777, 0.9103]	25
1:2.5	0.9135 ± 0.0324	[0.9001, 0.9269]	25

The repeated-validation results document the proposed model's discrimination across the two retained imbalance settings. The ROC-AUC changes from 0.8940 under the balanced 1:1.0 setting to 0.9135 under the moderate 1:2.5 setting, indicating that the 1:2.5 resampling design was associated with higher repeated-validation ROC-AUC in this experiment. Together with the single-split results, this repeated-validation evidence supports a more cautious interpretation of Hierarchical LFRM-MACI as a subsystem-aware deep learning model with a structured attribution workflow.

These findings are used to support the stability and interpretability of the proposed deep learning representation, while the single-split tables retain traditional machine learning methods as formal empirical comparison models.

5.2.3. Statistical Testing Summary

Because repeated fold-level predictions under an identical protocol were not available for all comparison methods, a valid paired significance test could not be conducted. Accordingly, this manuscript does not use paired-test results to support pairwise superiority claims. The repeated-validation means and confidence intervals in Table 4, together with the single-split deep learning comparisons, are used as the primary quantitative basis for interpreting the proposed model's stability.

5.3. Systemic Risk Driver Identification and Subsystem Importance Analysis

5.3.1. Experimental Setup and Evaluation Design

This subsection reports the final PFI and SHAP outputs generated by the unified attribution pipeline described in the Section 3. The interpretation emphasizes stable subsystem-level attribution patterns across the two retained imbalance regimes: 1:1.0 (balanced) and 1:2.5 (moderate). PFI and SHAP summarize model behavior at global and local levels and identify predictors associated with the learned bankruptcy classifier. Signed SHAP plots, dependence plots, and firm-level explanation cases can further extend this analysis. Because the attribution scores are normalized within each category and method, the results support within-category and within-summary comparisons. They should not be read as absolute cross-category importance magnitudes.

5.3.2. Category-Wise Importance Summary

Table 5 reports the category-wise means of within-category normalized PFI and SHAP scores. These values characterize the distribution and concentration of attribution scores within each financial category. Because each category is normalized independently, the reported means should not be interpreted as direct cross-category importance rankings.

Table 5. Category-wise means of normalized PFI and SHAP importance under two imbalance regimes.

(a) Balanced Setting (1:1.0)		
Category	Mean PFI	Mean SHAP
Profitability	0.297	0.238
Solvency	0.379	0.454
Liquidity	0.277	0.499
Efficiency	0.243	0.735
Growth/Shareholder Performance	−0.001	0.690
(b) Moderate Imbalance (1:2.5)		
Category	Mean PFI	Mean SHAP
Profitability	0.351	0.269
Solvency	0.384	0.439
Liquidity	0.365	0.473
Efficiency	0.260	0.397
Growth/Shareholder Performance	0.120	0.603

Note: Values are normalized category-level summaries from the PFI and SHAP analysis. PFI is based on the increase in BCE loss after permutation, whereas SHAP is based on mean absolute local contribution magnitude. The two columns therefore describe complementary attribution profiles and should not be converted into raw cross-category effect sizes.

5.3.3. Representative High-Ranking Indicators

Table 6 presents representative high-ranking indicators from the balanced 1:1.0 attribution summaries. Representative PFI indicators include Research and development expense rate, Inventory Turnover Rate, and Cash flow rate, while representative SHAP indicators include Borrowing dependency, Interest Expense Ratio, and Cash Flow to Total Assets. Because the scores are normalized within financial categories, the reported order is descriptive and should not be interpreted as a strict global ranking across all 95 predictors.

Table 6. Representative high-ranking indicators from the balanced 1:1.0 PFI and SHAP attribution summaries.

Reported Order	PFI Indicator	SHAP Indicator
1	Research and development expense rate	Borrowing dependency
2	Inventory Turnover Rate (times)	Interest Expense Ratio
3	Cash flow rate	Cash Flow to Total Assets
4	Current Assets/Total Assets	Non-industry income and expenditure/revenue
5	Realized Sales Gross Profit Growth Rate	Debt ratio %

Note: The table presents representative high-ranking indicators from the balanced-setting attribution summaries. Because the scores were normalized within financial categories, the reported order should not be interpreted as a strict global ranking across all 95 predictors.

Governance Implications

The final interpretability results support an indicator-level governance interpretation. Across imbalance settings, the model highlights variables linked to debt pressure, borrowing dependence, cash-flow stability, operating turnover, and growth/shareholder performance. These signals can be monitored by decision makers and reviewed alongside accounting expertise, audit routines, and risk-management procedures. Because the reported PFI and SHAP outputs are normalized importance profiles, they should be interpreted as warning-signal summaries rather than causal estimates or signed risk effects. Future work should report raw attribution magnitudes, signed local explanation plots, and firm-level explanation cases to distinguish importance strength from directional risk contribution.

6. Conclusions

This paper presents a systemic explainable AI framework for corporate bankruptcy early warning within business and management systems. The proposed Hierarchical LFRM-MACI architecture tokenizes firm-level financial indicators, applies LFRM-based local refinement within five financial subsystems, models cross-subsystem relations through MACI, and combines deep subsystem information with a raw-feature branch through gated fusion. This design makes the model consistent with this paper's systems-oriented premise: financial distress is represented as a structured interaction among profitability, solvency, liquidity, efficiency, and growth/shareholder performance rather than as a set of unrelated accounting ratios.

The updated benchmark is organized into traditional machine learning references and deep learning methods under balanced and moderately imbalanced training regimes. In the single 70%/30% validation split, Hierarchical LFRM-MACI obtains ROC-AUC values of 0.9286 under the 1:1.0 setting and 0.9269 under the 1:2.5 setting. Within the deep learning group, these results give the proposed model the highest ROC-AUC and the lowest Brier score and ECE under both imbalance settings. In repeated 5-fold cross-validation with five repetitions, its ROC-AUC is 0.8940 [0.8777, 0.9103] under 1:1 and 0.9135 [0.9001, 0.9269] under 1:2.5. These results support the proposed architecture as a structured deep learning model with an accompanying explanation pipeline, while the

single-split tables transparently report traditional machine learning methods as formal empirical comparison models without claiming overall superiority over them.

The interpretability analysis adds a complementary view of the learned risk structure. Using PFI and SHAP, the updated analysis identifies subsystem-level attribution patterns involving solvency and capital pressure, profitability and asset structure, liquidity and cash-flow stability, operating efficiency, and growth/shareholder performance [18,19]. Representative high-ranking PFI indicators under the balanced setting include Research and development expense rate, Inventory Turnover Rate, and Cash flow rate, whereas the SHAP ranking is led by Borrowing dependency, Interest Expense Ratio, and Cash Flow to Total Assets. Because the attribution summaries are normalized within category and method, these findings are interpreted as relative importance profiles rather than absolute cross-category rankings. The framework therefore provides potential decision-support value by converting model behavior into transparent signals that can support earnings monitoring, liquidity review, capital-structure assessment, and efficiency control [2,10].

Several limitations should be noted. First, the empirical evaluation is based on a single public Taiwanese cross-sectional benchmark, so external validation in other countries, private-firm samples, and different accounting regimes remains necessary. A comparable external firm-level bankruptcy dataset with harmonized indicator definitions and bankruptcy labels was not incorporated into the present design; the reported evidence is therefore limited to the standardized UCI benchmark. Second, the current model uses structured financial ratios and does not incorporate additional governance, ESG, macroeconomic, market, or textual disclosure variables. Third, LFRM depends on a fixed within-subsystem indicator ordering as an implementation choice; future work should test sensitivity to alternative economically plausible feature orderings and compare against permutation-invariant or learned-ordering variants. Fourth, the attribution analysis reports normalized PFI and SHAP rankings; future work should report raw attribution magnitudes, signed local explanations, and firm-level cases. Finally, future versions of the framework should further strengthen the deep learning architecture while preserving its hierarchical interpretability and subsystem-level attribution structure.

Future work can extend the proposed system by integrating additional subsystems, including governance and ESG variables, macroeconomic conditions, market signals, and textual disclosures. It should also report raw attribution magnitudes and signed local explanations to support more comprehensive systemic risk assessment and governance decision-making.

Author Contributions: Conceptualization, C.Y. and G.S.; methodology, C.Y. and G.S.; software, C.Y.; validation, C.Y. and G.S.; formal analysis, C.Y.; investigation, C.Y.; resources, G.S.; data curation, C.Y.; writing—original draft preparation, C.Y.; writing—review and editing, G.S.; supervision, G.S.; funding acquisition, C.Y. and G.S. All authors have read and agreed to the published version of this manuscript.

Funding: This paper is supported by Youth Program of Hubei Provincial Natural Science Foundation under grant 2026AFB092; Fundamental Research Funds for the Central Universities under grant 104972026RSCbs0227; Zhejiang Provincial Philosophy and Social Sciences Project under grant 23NDJC237YB; and the Outstanding Innovative Talents Cultivation Funded Programs 2022 of Renmin University of China.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Taiwanese Bankruptcy Prediction dataset used in this study is publicly available from the UCI Machine Learning Repository. Preprocessing, splitting, resampling, and validation procedures are described in the Sections 4 and 5.1. Code and processed split files are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the public dataset provider for making the benchmark openly accessible. During the preparation of this manuscript, the authors used GPT-5.5 to assist with language polishing and figure presentation refinement. The authors reviewed and revised all AI-assisted content and take full responsibility for the accuracy, integrity, and final content of the manuscript.

Conflicts of Interest: Author Chenxi Yang were employed by the company JD.com, Inc and Shandong Jianda Education Real Estate Co., Ltd. The remaining author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Alshareef, M.N. Artificial Intelligence-Enhanced Environmental, Social, and Governance Disclosure Quality and Financial Performance Nexus in Saudi Listed Companies Under Vision 2030. *Sustainability* **2025**, *17*, 7421. [CrossRef]
2. Delegkos, A.E.; Skordoulis, M.; Kalantonis, P. Business Strategies and Corporate Reporting for Sustainability: A Comparative Study of Materiality, Stakeholder Engagement, and ESG Performance in Europe. *Sustainability* **2025**, *17*, 8814. [CrossRef]
3. Ahmed, S.; Alshater, M.M.; Ammari, A.E.; Hammami, H. Artificial intelligence and machine learning in finance: A bibliometric review. *Res. Int. Bus. Financ.* **2022**, *61*, 101646. [CrossRef]
4. Nazareth, N.; Ramana Reddy, Y.V. Financial applications of machine learning: A literature review. *Expert Syst. Appl.* **2023**, *219*, 119640. [CrossRef]
5. Zhang, Z.; Wu, C.; Qu, S.; Chen, X. An Explainable Artificial Intelligence Approach for Financial Distress Prediction. *Inf. Process. Manag.* **2022**, *59*, 102988. [CrossRef]
6. Geng, R.; Bose, I.; Chen, X. Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining. *Eur. J. Oper. Res.* **2015**, *241*, 236–247. [CrossRef]
7. Damjanović, A.M.; Rašković, M.; Janković, S.D.; Jevtić, B.; Skoropad, V.N.; Marković, Z.D.; Lukić-Vujadinović, V.; Injac, Z.; Marinković, S. AI-Enabled Strategic Transformation and Sustainable Outcomes in Serbian SMEs. *Sustainability* **2025**, *17*, 8672. [CrossRef]
8. Du, Y.; Hao, G.; Zhu, H. How Does Participation in AI Standardisation Affect the Sustainable Development of Strategic Emerging Enterprises Under the Background of Uncertainty? Evidence from China. *Sustainability* **2025**, *17*, 7817. [CrossRef]
9. Sun, G.; Song, Y.; Xiao, J.; Xu, D. Artificial Intelligence and Sustainable Aviation Manufacturing: A Perspective from Green Innovation in China. *Sustainability* **2026**, *18*, 4298. [CrossRef]
10. Rudin, C.; Chen, C.; Chen, Z.; Huang, H.; Semenova, L.; Zhong, C. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Stat. Surv.* **2022**, *16*, 1–85. [CrossRef]
11. Zhang, T.; Zhu, W.; Wu, Y.; Wu, Z.; Zhang, C.; Hu, X. An explainable financial risk early warning model based on the DS-XGBoost model. *Financ. Res. Lett.* **2023**, *56*, 104045. [CrossRef]
12. Arora, N.; Kaur, P.D. A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. *Appl. Soft Comput.* **2020**, *86*, 105936. [CrossRef]
13. Liang, Y.; Zhang, J.J.; Li, H.; Liu, X.C.; Hu, Y.; Wu, Y.; Zhang, J.; Liu, Y.; Wu, Y. DeRisk: An effective deep learning framework for credit risk prediction over real-world financial data. *arXiv* **2023**, arXiv:2308.03704.
14. Liao, S.; Liu, Z. Enterprise financial influencing factors and early warning based on decision tree model. *Sci. Program.* **2022**, *2022*, 6260809. [CrossRef]
15. Sehgal, S.; Mishra, R.K.; Deisting, F.; Vashisht, R. On the determinants and prediction of corporate financial distress in India. *Manag. Financ.* **2021**, *47*, 1428–1447. [CrossRef]
16. Tong, L.; Tong, G. A novel financial risk early warning strategy based on decision tree algorithm. *Sci. Program.* **2022**, *2022*, 4648427. [CrossRef]
17. Dobrick, J.; Klein, C.; Zwergel, B. Size bias in Refinitiv ESG data. *Financ. Res. Lett.* **2023**, *55*, 104014. [CrossRef]
18. Fisher, A.; Rudin, C.; Dominici, F. All Models Are Wrong, but Many Are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81. [CrossRef] [PubMed]

19. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
20. Pham, H.; Chu, T.; Le, T.M.; Tran, H.M.; Tran, H.T.K.; Yen, K.N. Comprehensive Evaluation of Bankruptcy Prediction in Taiwanese Firms Using Multiple Machine Learning Models. *Int. J. Technol.* **2025**, *16*, 289–309. [\[CrossRef\]](#)
21. Yao, M.; Xu, M. Exploring the Roles of Corporate Social Responsibility and Artificial Intelligence Adoption in the Impact of Political Connections on Innovation Performance. *Sustainability* **2025**, *17*, 7883. [\[CrossRef\]](#)
22. Sun, Z.; Wu, X.; Dong, Y.; Lou, X. How Does Artificial Intelligence Application Enable Sustainable Breakthrough Innovation? Evidence from Chinese Enterprises. *Sustainability* **2025**, *17*, 7787. [\[CrossRef\]](#)
23. Torres, J.F.; Hadjout, D.; Sebaa, A.; Martínez-Álvarez, F.; Troncoso, A. Deep learning for time series forecasting: A survey. *Big Data* **2021**, *9*, 3–21. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Hu, Z.; Zhao, Y.; Khushi, M. A survey of forex and stock price prediction using deep learning. *Appl. Syst. Innov.* **2021**, *4*, 9. [\[CrossRef\]](#)
25. Reis, P.; Serra, A.P.; Gama, J. The role of deep learning in financial asset management: A systematic review. *arXiv* **2025**, arXiv:2503.01591.
26. Shahrour, M.H.; Dekmak, M. Intelligent Stock Prediction: A Neural Network Approach. *Int. J. Financ. Eng.* **2023**, *10*, 2250016. [\[CrossRef\]](#)
27. Ding, Y.; Yan, C. Corporate financial distress prediction: Based on multi-source data and feature selection. *arXiv* **2024**, arXiv:2404.12610.
28. Liu, S. LTR-Net: A deep learning-based approach for financial data prediction and risk evaluation in enterprises. *PLoS ONE* **2025**, *20*, e0328013. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Elhoseny, M.; Metawa, N.; Sztano, G.; El-Hasnony, I.M. Deep Learning-Based Model for Financial Distress Prediction. *Ann. Oper. Res.* **2025**, *345*, 885–907. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Nugroho, D.S.; Dewayanto, T. Application of Statistics and Artificial Intelligence for Corporate Financial Distress Prediction Models: A Systematic Literature Review. *J. Model. Manag.* **2025**, *20*, 1999–2023. [\[CrossRef\]](#)
31. Hamdi, M.; Mestiri, S.; Arbi, A. Artificial Intelligence Techniques for Bankruptcy Prediction of Tunisian Companies: An Application of Machine Learning and Deep Learning-Based Models. *J. Risk Financ. Manag.* **2024**, *17*, 132. [\[CrossRef\]](#)
32. Charalambakis, E.C.; Garrett, I. On Corporate Financial Distress Prediction: What Can We Learn from Private Firms in a Developing Economy? Evidence from Greece. *Rev. Quant. Financ. Account.* **2019**, *52*, 467–491. [\[CrossRef\]](#)
33. Yousaf, U.B.; Ullah, I. Corporate Governance and Financial Distress: A Review of the Theoretical and Empirical Literature. *Int. J. Financ. Econ.* **2024**, *29*, 1627–1679. [\[CrossRef\]](#)
34. Zhang, Y.; Yuqi, Z.; Liang, Y.; Mu, S.; Zhang, Z.; Chen, X. AD-VGF: An Improved Generative Adversarial Network for Credit Card Risk Identification and Management in Digital Economy. *Cybern. Syst.* **2025**. [\[CrossRef\]](#)
35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
36. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: New York, NY, USA, 2016; pp. 785–794. [\[CrossRef\]](#)
37. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; pp. 3146–3154.
38. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. In Proceedings of the Advances in Neural Information Processing Systems 31, Montreal, QC, Canada, 3–8 December 2018; pp. 6638–6648.
39. Cortes, C.; Vapnik, V. Support-vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
40. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013. [\[CrossRef\]](#)
41. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017.
42. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [\[CrossRef\]](#)
44. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)

45. Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Saenko, K.; Darrell, T. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2625–2634. [[CrossRef](#)]
46. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, 323, 533–536. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.