

## Article

# An Explainable AI-Based Framework for Predicting Construction Firm Profitability and Capital Structure Considering Supply-Chain Volatility

Seong-Jun Ye <sup>1</sup> and Kyung-Tae Lee <sup>2,\*</sup> <sup>1</sup> Lotte Engineering & Construction Co., Ltd., Seoul 06515, Republic of Korea; yeastar1996@naver.com<sup>2</sup> Department of Architectural Engineering, Kangwon National University, Samcheok-si 25913, Republic of Korea

\* Correspondence: ktlee0422@kangwon.ac.kr; Tel.: +82-33-250-6221

## Abstract

Traditional financial models often fail construction firms by ignoring industry-specific volatility. This study proposes an Explainable AI (XAI) framework to predict Return on Assets (ROA) and Debt-to-Equity (D/E) ratios using KOSPI and KOSDAQ data (2015–2024). It compares a baseline financial dataset (Dataset 1) with an industry-augmented version (Dataset 2) that incorporates the Producer Price Index (PPI) and labor wages. Using ANN, LSTM, and RF, the study found that the RF model achieved the highest predictive power ( $R^2 = 0.7978$ ). Notably, Dataset 2 improved accuracy significantly, reducing the Mean Absolute Error (MAE) for ROA by approximately 56%. SHAP analysis revealed that rising rebar and steel prices (PPI\_RB) negatively impact both profitability and stability. In contrast, the concrete price index (PPI\_CO) showed a positive influence, reflecting firms' ability to manage costs or adjust prices. Crucially, material costs proved far more influential than labor costs. These findings indicate that supply-chain and raw-material-price indicators are salient predictors, suggesting that management may prioritize monitoring and mitigating supply-chain volatility, while causal validation remains beyond the scope of this study. Ultimately, by leveraging an XAI-based interpretation, this study provides a decision-making basis for formulating procurement and hedging strategies using raw material price indices, while also proposing an analytical framework that explains financial performance in the construction industry from a volatility-oriented perspective.

**Keywords:** construction finance; machine learning; explainable AI; supply-chain volatility; profitability prediction



Academic Editor: Phuong Hoang  
Dat Nguyen

Received: 22 January 2026

Revised: 15 February 2026

Accepted: 16 February 2026

Published: 19 February 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  
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

The fundamental objective of any enterprise is to maximize profit and enhance corporate value; thus, analyzing financial performance indicators is a pivotal research approach [1]. These indicators explain or predict a firm's financial health and diagnose its current operational status. However, while various metrics are employed to forecast bankruptcy or assess financial characteristics, most existing failure prediction models have been predominantly developed for the manufacturing sector. This creates a significant mismatch, as construction firms operate under unique financial structures—characterized by order-based production, progress-based revenue recognition, and high capital intensity—that manufacturing-centric models fail to capture [2,3]. Consequently, it is imperative to integrate industry-specific characteristics, such as competitive intensity and external market dynamics, into the analytical framework [4–6].

The construction industry is perpetually subject to financial pressures and is notoriously sensitive to economic fluctuations. Deterioration in the macro-environment often triggers a rapid downturn in the construction market, frequently leading to firm insolvency [7]. Furthermore, the sector is deeply intertwined with the financial system; particularly in South Korea, the structural reliance on Project Financing (PF) means that distress within the construction industry can precipitate broader instability across financial institutions [8]. Amidst these dynamics, the volatility of labor and material costs has intensified, directly impacted by unprecedented global events such as the COVID-19 pandemic and the Russia–Ukraine war. Consequently, there is an urgent need to incorporate these external drivers into financial impact assessments and to conduct a granular analysis of the specific degree of influence exerted by each factor [6].

Historically, traditional statistical methodologies—such as regression and factor analysis—were predominantly employed to classify and assess the influence of financial variables through empirical, inductive approaches [9]. However, recent research has shifted toward exploring the potential of machine learning (ML) models for performance analysis. In the construction industry, financial issues are often nonlinear and strongly affected by external shocks, leading to substantial interaction effects among variables; thus, conventional statistical approaches that focus on linear and static analyses have inherent limitations in capturing these complex relationships [2,6]. Unlike static statistical methods, ML offers the distinct advantage of integrating real-time data and continuously learning from new inputs, enabling a more dynamic and responsive assessment of shifting financial conditions [7]. Proposing an optimal prediction framework by evaluating various algorithms not only enhances predictive accuracy but also provides a powerful tool for informed decision-making and long-term financial sustainability. Furthermore, to address the drawback of conventional approaches—namely, that they may enable prediction but provide limited insight into why a given outcome occurs—incorporating an Explainable AI (XAI) framework allows stakeholders to interpret complex model outputs and derive actionable insights that were previously obscured. Through this, the proposed approach can support not only prediction but also variable-level decision-making in practical applications. While interpretability is vital for external stakeholders, XAI research has often overlooked their specific requirements. To address this, our framework provides explanations that allow firms to pinpoint the primary drivers of financial distress. Notably, counterfactual explanations offer specific strategies for recovery, while an evaluation of feature importance and synergistic interactions provides a broader perspective on the final model output [10].

Accordingly, this study is structured as follows. To predict the financial performance of construction firms, we first conducted an in-depth literature review to define input variables, incorporating additional statistical parameters that capture the unique characteristics of the construction sector. We then developed two distinct datasets: a baseline financial dataset and an industry-augmented dataset that reflects specific construction market dynamics. For predictive modeling, Return on Assets (ROA) and the Debt-to-Equity (D/E) Ratio were selected as dependent variables. Utilizing these datasets, we implemented three representative machine learning models: Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) RNNs, and Random Forest. Various validation metrics were employed to evaluate performance and identify the optimal model–dataset combination. Furthermore, to investigate the influence of independent variables, we applied SHapley Additive exPlanations (SHAP) as a representative XAI algorithm. This experimental approach not only highlights the importance of utilizing industry-specific variables but also provides empirical evidence of the distinct financial characteristics inherent in the construction industry compared to others.

## 2. Literature Review

### 2.1. Financial Risks Rooted in the Structural Characteristics of the Construction Industry

The construction industry possesses intrinsic characteristics that fundamentally distinguish it from manufacturing and other sectors. It is primarily an order-based industry where the form and content of each project vary according to individual client requirements, making standard cost estimation difficult and pre-production inventory impossible [8]. Furthermore, as a labor-intensive sector with relatively low fixed-asset investment, a significant portion of its assets consists of operation-related current assets—such as inventories (work-in-progress), loans, and advances—which are subject to high volatility [11]. These prolonged production cycles inevitably expose firms to significant cost fluctuations. The construction sector is highly sensitive to macroeconomic shifts and crises, including foreign exchange crises, the subprime mortgage crisis, and the recent global pandemic, coupled with surging raw material prices. In particular, the aggressive interest rate hikes by the U.S. Federal Reserve since 2022 have posed substantial threats to the construction market, underscoring the critical importance of financial risk management [12]. Following COVID-19, delays in material procurement and resource scarcities led to operational disruptions, directly impacting the financial structures of firms. Consequently, the emphasis on financial risk management has intensified, inextricably linking socio-cultural circumstances to corporate solvency [11,13].

These pressures result in immediate financial distress, particularly for small and medium-sized enterprises (SMEs). For sustainable operation, it is essential to manage factors that reflect the industry's unique financial profiles [14]. For instance, due to a high dependency on debt over equity, the Korean construction sector has faced severe financial health issues related to PF loans [13]. While high exchange rates and interest rates have played a major role, the impact of surging material and labor costs (inflation) has become increasingly prominent. Similar trends have been observed in other markets, such as Malaysia, where the price volatility of key materials like steel and cement significantly affected firms' financial status [5]. Therefore, given that economic downturns and rising interest rates can lead to a surge in financial costs and liquidity crises, it is imperative to incorporate both structural cost characteristics and financial sensitivities—beyond conventional variables—to accurately predict the financial performance of construction firms.

### 2.2. Key Indicators and Frameworks for Predicting Financial Performance in Construction

To evaluate the financial performance of construction firms, a diverse array of financial statement-based indicators is systematically utilized. Among these, profitability metrics such as Return on Assets (ROA) and Return on Equity (ROE) play a pivotal role due to their direct linkage with capital structure [15]. ROA represents a firm's capability to generate net income from its total assets, serving as a measure of how efficiently a construction enterprise utilizes its equipment, site assets, and overall resources to generate revenue and profit [16]. Simultaneously, ROE indicates the net income realized relative to shareholder equity. It evaluates the effectiveness of resource utilization based on profits that reflect the industry's unique cost structures, including selling, general, and administrative expenses and total sales [17]. Since these profitability indicators are directly influenced by the leverage structure—specifically, debt ratios and equity scale—a balanced combination of debt and equity is considered an essential element for ensuring long-term profitability and enhancing corporate value in project-based industries like construction [18].

Furthermore, liquidity acts as a crucial variable in predicting financial performance, as the cashability of assets directly impacts corporate value and investment capacity [19]. For construction firms, liquidity levels are often constrained by project delays, accounts receivable collection periods, and the composition of working capital, which can exacerbate

financial constraints. Interestingly, some studies suggest that excessive liquidity may lead to inefficient over-investment [16]. In addition to liquidity, Total Asset Turnover, which demonstrates the efficiency of revenue generation relative to total assets (including both debt and equity), serves as a key metric for evaluating the activity of construction firms. Due to the asset-intensive nature of industry—where equipment, site assets, and current assets are fixed in the production process—asset turnover directly reflects operational efficiency, construction execution capability, and revenue realization structures. Thus, along with profitability and leverage, asset turnover constitutes a major pillar of the activity-based variable system for predicting financial performance [20]. In this context, asset management efficiency, leverage levels, and liquidity indicators are established as the core framework for systematically forecasting construction financial performance. Currently, the industry's status is primarily monitored through financial performance indicators, compliance with ESG guidelines, maintenance of financial transparency, and assessments of operational sustainability [21,22].

### *2.3. Theoretical Background of Financial Ratio Analysis and Performance Prediction*

Predicting financial ratios serves as a critical analytical tool for identifying a firm's financial standing and management performance. These ratios have been extensively utilized to evaluate financial soundness and bankruptcy probability, providing indispensable information for corporate strategy formulation and investment decision-making. Consequently, a wide range of studies have been conducted in this field. For instance, Zeitun and Tian [1] analyzed the performance of Jordanian firms by employing ROA, ROE, and the market-to-book value ratio as dependent variables, while using the total debt-to-total assets ratio and total debt-to-total equity ratio as independent variables. Their research demonstrated that corporate debt ratios, particularly short-term debt, negatively impact ROA and EBITDA, yet exert a positive influence on market performance.

Additionally, Heikal et al. [23] examined automotive companies within the Indonesian stock market, utilizing ROA, ROE, Net Profit Margin, D/E Ratio, and the Current Ratio as key variables. Through multiple regression analysis and classical hypothesis testing, they identified that ROA and ROE have a significant and positive effect on profit growth. Furthermore, it was observed that while ROA contributes positively to a firm's profit growth, the Current Ratio may have a negative impact on it [24]. In the specific context of construction firms, research comparing healthy and distressed companies has been conducted using statistical methods. Lim and Choi [22] identified that the Equity Ratio, Total Asset Turnover Ratio, and Labor Equipment Ratio are pivotal indicators for maintaining stable management. Furthermore, Choi et al. [24] proposed a financial distress prediction model for construction companies by leveraging variables such as ROA, ROE, Total Asset Turnover, Debt Ratio, Current Ratio, and Return on Sales (ROS). Their study established an optimal predictive framework by comparing the performance of Support Vector Machines, Logistic Regression, and Decision Trees. Similarly, to predict bankruptcy among U.S. construction firms, Jang et al. [25] utilized an LSTM RNN model. Their analysis integrated accounting variables—including ROA, ROE, ROS, the ratio of liquid assets to net assets, and the ratio of working capital to total assets—with construction market indicators and macroeconomic variables. The results demonstrated that models incorporating market and macroeconomic variables exhibit superior predictive performance, highlighting the effectiveness of multi-dimensional data in forecasting. In recent years, the integration of XAI has increasingly facilitated decision-making by evaluating the influence of specific variables on model outputs and providing a clearer understanding of the underlying operational mechanisms of AI [10]. This approach is being actively adopted within the financial sector as well. For instance, studies have utilized the SHAP (SHapley

Additive exPlanations) methodology to identify the relative importance of various feature characteristics [26]. Furthermore, XAI has been widely employed in finance to bridge the performance gap where simple, interpretable models often lack the accuracy of their complex counterparts. Specifically, SHAP was applied to measure feature importance in financial distress prediction models [27]. Similarly, by applying SHAP to P2P lending evaluation, researchers have enhanced the interpretability of machine learning models, demonstrating their ability to reflect non-linearities and structural changes in the relationship between specific features and target variables [28]. Beyond traditional finance, these methodologies have expanded into real estate, where they are used to analyze how energy efficiency and carbon emission reduction characteristics affect transaction data, thereby supporting rational decision-making through variable impact analysis [29]. Consequently, XAI enhances the explicability of financial forecasting models and fosters trust by enabling stakeholders to understand the reasoning behind AI-driven judgments on both local and global scales [26].

Ultimately, while financial ratio research has been extensively conducted across various sectors, it is consistently argued that the relevance and influence of financial variables differ according to industry-specific characteristics and structures [6,30]. Particularly, the construction industry possesses a unique financial structure characterized by project-based operations and high sensitivity to fluctuations in raw material prices and labor costs. Therefore, a specialized approach that reflects these industry-specific factors must be pursued in financial analysis [31].

#### *2.4. Synthesis of Prior Studies and Research Gap*

Based on the analysis of the relevant literature, it was observed that while financial ratio analysis, capital structure research, and machine-learning-based analytics are currently being conducted in the construction industry, a critical methodological gap persists. Previous studies have predominantly relied on static financial indicators derived from accounting statements to explain firm performance and financial distress in construction firms [2,22,24]. Several studies have attempted to extend traditional financial models by incorporating macroeconomic or market-level variables, such as interest rates or market indices, to improve predictive performance [5,25]. However, these variables are typically treated as exogenous background factors and do not explicitly represent the operational cost mechanisms through which supply-chain disruptions affect firm-level profitability and capital structure.

Current AI-based forecasting frameworks primarily utilize standard accounting indicators, yet they frequently fail to structurally incorporate sector-specific risk mechanisms, relegating them to supplementary features [32]. Nevertheless, the adoption of advanced machine learning algorithms has yielded substantial gains in accuracy, particularly when addressing the complexities of medium- and long-term financial distress horizons within the construction sector [33]. However, these models frequently suffer from excessive abstraction, where critical elements of supply-chain volatility, such as material cost fluctuations and labor market instability, are subsumed into aggregate firm-level financial performance. A similar limitation is observed in the analysis of ESG-related financial outcomes. While recent cross-national comparative studies empirically demonstrate that liquidity, leverage, and profitability exert varying impacts on corporate resilience and ESG performance depending on the specific regulatory and market environments within the construction industry [6,8,34], these financial indicators are often conceptualized as either outcome-oriented or structurally static variables. Consequently, existing frameworks fail to provide a sufficient explanation of how exogenous cost shocks, such as volatility in raw

material prices or labor costs, propagate through operational margins and capital structures over time.

Taken together, the existing literature suggests that although machine learning techniques have substantially enhanced predictive accuracy in construction finance research, a clear methodological lacuna persists regarding the dynamic integration of supply-chain cost volatility into financial performance prediction frameworks. Thus, there is a critical need for a model capable of processing multifaceted variables to provide explainable insights into the relationship between external supply-chain shocks and internal financial indicators, thereby bridging the current methodological gap.

### 3. Research Design

The analytical framework of this study is designed based on the causal hypothesis that the financial performance of construction firms is structurally shaped by two core production inputs—labor and material costs—and that financial outcomes emerge through a cost pass-through and efficiency mechanism in which labor and material costs affect operating margins, which in turn influence capital structure and ultimately determine financial performance [18]. As shown in Figure 1, to empirically test this mechanism, two parallel datasets were constructed using financial statement data of listed construction firms: a baseline dataset (Dataset 1) composed of conventional financial indicators such as profitability, liquidity, leverage, and activity ratios, and an extended dataset (Dataset 2) that incorporates construction-specific labor and material cost indices. The difference in predictive performance between these two datasets enables identification of the marginal explanatory power of construction production inputs on corporate financial outcomes. To ensure a time-aware learning setup across all models, we reformulated the quarterly panel into a supervised learning dataset using a fixed lookback window of four quarters. Given the temporal nature of quarterly data, we used a time-series-aware cross-validation (10 splits; expanding window) within the training period to avoid training on future observations [35]. To examine the structural stability of the cost–performance relationship across different functional forms and temporal dependencies, multiple learning architectures—including RF, ANN, and LSTM models—were employed to predict financial performance. Despite the arguably limited volume of data, these models were implemented not only to compare analytical outcomes but also to characterize the data patterns inherent in the construction industry based on the model outputs. Model accuracy was evaluated using standard metrics such as RMSE, MAE, and  $R^2$ . Moreover, sensitivity analysis was introduced to validate the model. Finally, Shapley Additive Explanations (SHAP) were applied to decompose firm-level predictions into the contributions of labor costs, material costs, and other financial variables, thereby enabling quantitative identification of which production inputs drive financial vulnerability or performance at the firm level [9]. In this way, the proposed framework moves beyond pure predictive comparison and transforms construction firm financial performance into a production-input-based interpretable structure, offering an explainable financial risk analysis system that departs from traditional accounting-oriented construction finance research.

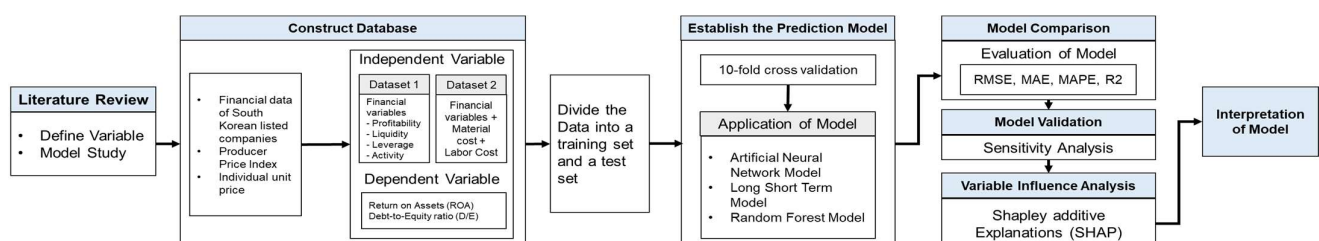


Figure 1. Research Progress.

### 3.1. Data Collection

To predict the financial performance of construction firms and to compare predictive accuracy across different modeling approaches, this study selected construction companies listed on the Korea Composite Stock Price Index (KOSPI) and the Korea Securities Dealers Automated Quotations (KOSDAQ) markets as the analytical sample. Using data from 2015 to 2024, twenty firms with complete financial records over the observation period were identified. Since the data are published on a quarterly basis, the relevant data could be collected accordingly, enabling the subsequent analysis. Two datasets were constructed: Dataset 1, which includes only financial variables, and Dataset 2, which augments the financial variables with non-financial cost variables related to material and labor costs. The difference in financial performance prediction accuracy between the two datasets was used to assess the explanatory contribution of construction cost structures.

Firm-level financial statement data were collected from the Data Analysis, Retrieval and Transfer System (DART), operated by the Financial Supervisory Service of Korea. The variables used in this study and their definitions are presented in Table 1. Following prior studies, financial indicators were categorized into four dimensions—profitability, liquidity, leverage, and activity—and corresponding variables were defined accordingly [6,16]. Profitability indicators included return on equity (ROE), selling and administrative expenses, and total sales, reflecting the firm’s capital-based earning capacity. While ROE and the dependent variable (ROA) share ‘net income’ as a common numerator, diagnostic tests were conducted to ensure statistical independence. The Pearson correlation between the two was 0.254, and the Variance Inflation Factor (VIF) was 1.114, significantly below the conventional threshold ( $VIF < 5\text{--}10$ ) [21]. Consequently, this suggests that the two indicators maintain statistical independence while providing complementary explanations of firm profitability from different perspectives. Liquidity indicators consisted of the current ratio and working capital to total assets, which capture the firm’s short-term solvency and asset liquidity. Leverage was represented by the debt ratio and shareholders’ equity, measuring the firm’s capital structure and financial stability. Activity was measured using total asset turnover to evaluate the efficiency with which firms utilize their assets.

**Table 1.** Description of the variables.

| Variable             | Category      | Variable                        | Description                                                                                                                                                  | Literature Review                                                                                                                     |
|----------------------|---------------|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| Independent Variable | Profitability | Return on Equity                | An indicator of how much profit a company generates with its equity; calculated by dividing net income by average equity.                                    | Cheng et al. (2014) [36], Tserng et al. (2015) [37], Jang et al. (2019) [25]                                                          |
|                      |               | Selling Expenses                | Total costs incurred in relation to the sale of goods or services (e.g., advertising, sales commissions).                                                    | Seo et al. (2024) [6]                                                                                                                 |
|                      |               | Administrative Expenses         | Total costs incurred from general management and maintenance activities (e.g., salaries, rent, communication expenses).                                      | Seo et al. (2024) [6], Ptok et al. (2018) [38]                                                                                        |
|                      | Liquidity     | Total Revenue                   | The total sales revenue achieved through core business activities represents the company’s external scale.                                                   | Graham et al. (1998) [39], Seo et al. (2024) [6]                                                                                      |
|                      |               | Current Ratio                   | An indicator to evaluate a company’s short-term debt-paying ability; calculated by dividing current assets by current liabilities.                           | Kangari et al. (1992) [40], Horta & Camanho (2013) [41], Cheng et al. (2014) [36], Tserng et al. (2015) [37], Jang et al. (2019) [25] |
|                      |               | Working Capital to Total Assets | The proportion of net working capital (current assets—current liabilities) within total assets, representing the company’s short-term financial flexibility. | Kim (2019) [42]                                                                                                                       |

Table 1. Cont.

| Variable             | Category                          | Variable                                                                                                                              | Description                                                                                                                                       | Literature Review                                                       |
|----------------------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Independent Variable | Leverage                          | Equity Ratio                                                                                                                          | An indicator representing the proportion of total equity within total assets, calculated by dividing total equity by total assets.                | Lim and Choi (2015) [22]                                                |
|                      |                                   | Total Equity                                                                                                                          | Net assets remain after subtracting total liabilities from total assets, consisting of capital contributed by shareholders and retained earnings. | Graham et al. (1998) [39]                                               |
|                      | Activity                          | Total Asset Turnover                                                                                                                  | An indicator of how efficiently a company uses its total assets to generate revenue.                                                              | Cheng et al. (2014) [36], Oh & Yeo (2019) [43], Jang et al. (2019) [25] |
|                      | Construction Material Input Costs | Producer Price Index (PPI) for Major Materials (Cement, Ready-mixed concrete, Concrete, Rebar & Steel bars, Metal products, Plastics) | Data on various construction materials based on the Producer Price Index (PPI) provided by the Bank of Korea.                                     | The Bank of Korea (2024) [44]                                           |
|                      | Labor Costs                       | Prevailing Wage by Trade (Rebar workers, Concrete workers, Formwork carpenters, Plasterers)                                           | Standard labor cost indices based on the prevailing wage rates published by the Construction Association of Korea (CAK).                          | Construction Association of Korea (2025) [45]                           |
|                      | Profitability                     | Return on Assets (ROA)                                                                                                                | The ratio of net income generated relative to the company's total assets (liabilities + equity).                                                  | Saputra (2022) [16], Choi et al. (2018) [24]                            |
| Dependent Variable   |                                   | Debt-to-Equity Ratio                                                                                                                  | A stability indicator showing the relationship between liabilities and equity; calculated by dividing total liabilities by total equity.          | Cheng et al. (2014) [36], Jang et al. (2019) [25]                       |

Raw material cost information was obtained from the Producer Price Index (PPI) published by the Bank of Korea through its Economic Statistics System [44]. To reflect the cost structure specific to construction firms, PPIs were selected for major construction materials, including cement (PPI\_CE), ready-mixed concrete (PPI\_RE), concrete products (PPI\_CO), reinforcing bars and steel rods (PPI\_RB), metal forms for construction (PPI\_CM), and plastic construction materials (PPI\_PL). These indices were used to analyze how fluctuations in material costs affect financial performance prediction. Moreover, rather than relying on a single aggregated price measure, the use of material-specific PPIs enabled the model to capture heterogeneous input-cost movements across different construction materials. Labor cost variables were constructed to reflect the labor-intensive nature of the construction industry and were aligned with the material cost variables. Four representative occupations with high daily wages and significant cost shares were selected: reinforcing steel workers (Labor\_RE), concrete workers (Labor\_CO), formwork carpenters (Labor\_FW), and plasterers (Labor\_PL). Labor cost data from 2015 to 2024 were collected from the Construction Wage Survey Reports published by the Construction Association of Korea.

The dependent variables were return on assets (ROA) and the D/E ratio. Among the commonly used financial performance indicators, ROA and ROE are the most widely applied measures of profitability [42,44]. ROA was selected because it is particularly appropriate for the construction industry, which typically operates with relatively low fixed capital and thus relies on asset-based performance evaluation. The D/E ratio was selected as a key indicator of financial soundness, as capital structure and leverage exert a substantial influence on corporate financial performance. Together, these two dependent variables allow for a comprehensive assessment of both profitability and financial stability.

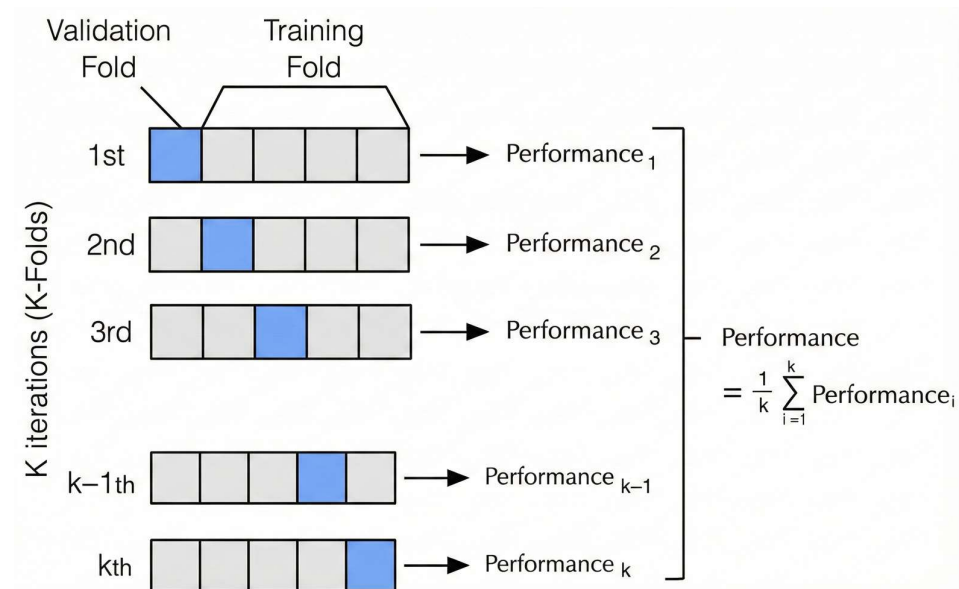
### 3.2. Data Pre-Processing

During the data preprocessing phase, a standardization (scaling) technique was employed to eliminate scale discrepancies among the variables. This study utilized the Stan-

StandardScaler from the Scikit-learn library, which transforms the data to have a mean of 0 and a variance of 1 for each variable, as expressed in Equation (1). This standardization process ensures that no single variable exerts a disproportionate influence during model training, thereby establishing comparability across all features [46].

$$Z = (X - u)/S \quad (Z: \text{Standardized Coefficient}, X: \text{data}, u: \text{average}, S: \text{standard deviation}) \quad (1)$$

To evaluate the model's generalization performance and mitigate the risk of overfitting, a time-series aware K-fold cross-validation approach ( $K = 10$ ) was implemented [47]. Unlike standard K-fold methods that rely on random shuffling, this approach strictly adheres to the chronological order of the observations, thereby ensuring that no future information is used to predict past outcomes. As illustrated in Figure 2, the dataset is partitioned into 10 sequential subsets; at each iteration, the training set consists solely of data points that occurred prior to the validation set, following an expanding-window scheme. In this design, each fold's validation set effectively serves as a temporally separated test set, meaning that the train-test temporal separation is inherently enforced at every iteration without requiring a separate hold-out set. The reported performance metrics represent the average across all 10 folds, providing a reasonable estimate of out-of-sample predictive accuracy.



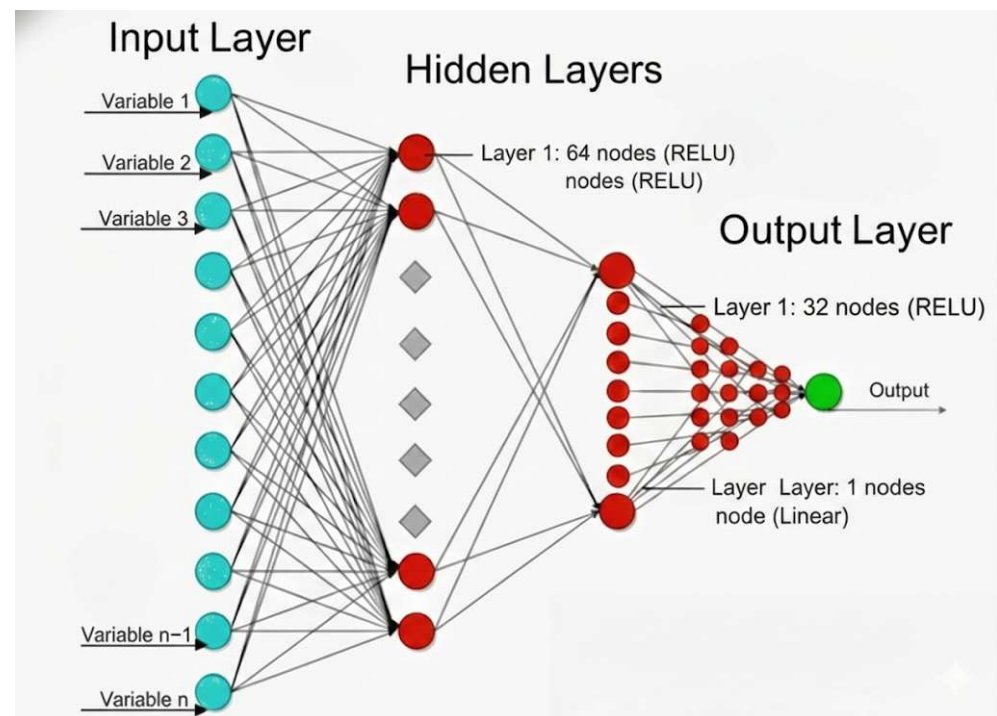
**Figure 2.** Schematic algorithm process of K-fold cross-validation.

Additionally, all preprocessing parameters, including standardization coefficients, were fit only on the training folds and then applied to the corresponding validation fold, thereby avoiding potential information leakage at the data processing level [48]. Through this procedure, each sub-dataset serves as the validation set exactly once within a temporally consistent sequence, minimizing variance in model accuracy and offering a reliable assessment of predictive performance.

All models in this study were implemented using Python 3.11 within the Google Colab environment. Specifically, we developed Artificial Neural Network (ANN), Recurrent Neural Network (RNN), and Random Forest (RF) models leveraging the Scikit-learn and TensorFlow libraries. These representative artificial intelligence algorithms facilitate the identification of independent variables that significantly influence the dependent variable through a comparative analysis of their outputs. The computational experiments were conducted on a workstation equipped with an Intel Xeon Silver 4216 CPU (@ 2.1 GHz; Intel Corp., Santa Clara, CA, USA), 192 GB of RAM, and an NVIDIA TITAN RTX GPU (NVIDIA Corp., Santa Clara, CA, USA), with a total storage capacity of 10 TB.

### 3.3. Data Analysis

The ANN model utilized in this research is inspired by the biological neural networks of the human brain. It is structured into three primary components: an input layer, hidden layers, and an output layer, with the specific architecture illustrated in Figure 3. The process involves receiving variable-related data at the input layer, conducting training through the hidden layers, and subsequently generating results at the output layer [21]. A key advantage of ANN is its robustness in modeling complex and non-linear relationships between input and output variables that are otherwise difficult to define. Furthermore, its ability to process both continuous and discrete variables contributes to its superior predictive performance across diverse datasets.



**Figure 3.** ANN Model Architecture.

LSTM RNNs were specifically developed to capture sequential and temporal patterns within time-series data. The implementation of this model enables the identification of temporal dependencies within the financial performance outcomes of long-term construction projects. The structure of this model is presented in Figure 4. LSTM was originally proposed to address the challenge of maintaining short-term memory over longer periods [49]. In this architecture, standard hidden neurons are replaced by LSTM blocks. These blocks consist of memory cells and specialized gates that play a crucial role in learning long-range dependencies by precisely controlling the storage and flow of information [25]. Each gate utilizes a mechanism to determine whether to accept information at each cell; these cells are composed of sigmoid neural net layers and pointwise multiplication operations to facilitate effective information transmission.

The Random Forest model, as shown in Figure 5, is an ensemble learning method that enhances predictive performance by aggregating the results of multiple Decision Trees (base learners). Its objective is to build a robust model by combining several base algorithms through bagging (bootstrap aggregating) and randomness. This involves generating multiple training subsets via resampling with replacement and training independent decision trees on each. This process effectively reduces model variance and prevents overfitting [50]. During node splitting, the optimal criteria are selected from a random subset of features,

increasing tree diversity and reducing correlation. Beyond high predictive accuracy, RF provides a feature importance evaluation, which is instrumental in identifying the key factors influencing the predictions.

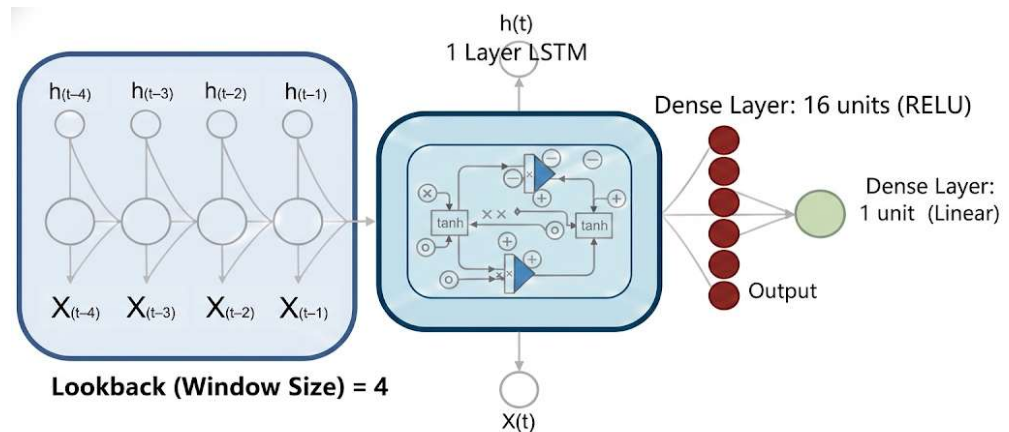


Figure 4. LSTM RNN Architecture.

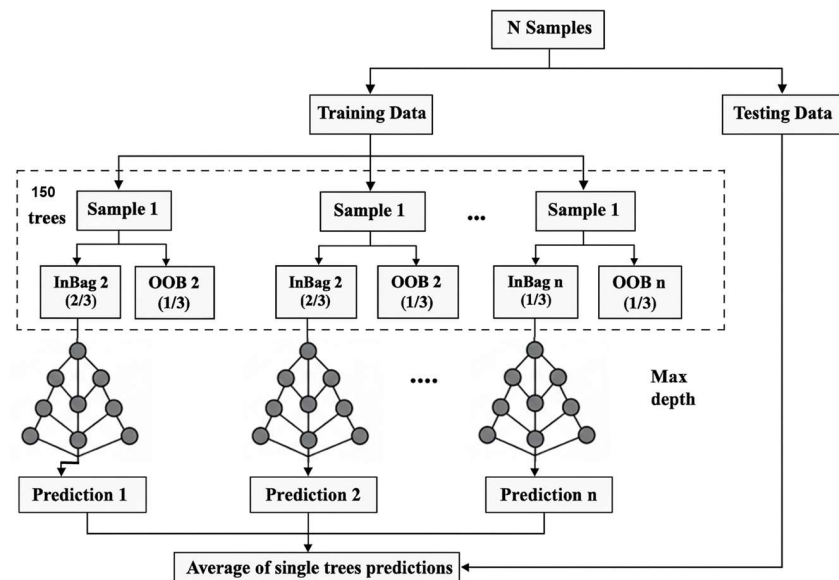


Figure 5. RF model Architecture.

To ensure model transparency and reproducibility, the hyperparameters for each model were determined through a systematic Grid Search process. The search space included varying learning rates (0.01 to 0.0001), batch sizes (16 to 64), and structural configurations. For example, for the ANN, the number of hidden layers was tested from 1 to 3, with neurons ranging from 16 to 128 per layer. Similarly, the RF model was optimized by fine-tuning the number and depth of its individual decision trees. In the case of the LSTM, the initial lookback window was set to 4 quarters to align with the annual reporting cycle of the construction industry, as quarterly financial data inherently reflects annual cyclicity. This domain-informed selection was subsequently validated through sensitivity analysis across alternative windows of 2 and 6 quarters. The final hyperparameter combination for each model was selected at the point where errors were minimized, or performance metrics were optimized within the validation dataset; this approach was specifically designed to maximize generalization performance and mitigate the risk of overfitting while maintaining robust predictive power.

### 3.4. Model Evaluation and Validation

To evaluate the performance and compare the accuracy of the prediction models, we utilized metrics including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and the Coefficient of Determination ( $R^2$ ) [51]. These performance metrics play a critical role in verifying the accuracy and predictive power of a model by measuring the discrepancies between the predicted values and the actual observed values.

Mean Absolute Error (MAE) directly reflects the average of the absolute errors between the predicted values ( $\hat{y}_k$ ) and the observed values ( $y_k$ ); therefore, a smaller MAE indicates superior predictive performance [52]. While MAE is a robust metric that is less sensitive to outliers and allows for a straightforward evaluation of a model's predictive capability, it does not account for the direction of the errors, which can sometimes lead to an underestimation of the actual error. Consequently, it is essential to utilize other metrics in conjunction with MAE.

$$MAE = (1/n) \sum |\hat{y}_k - y_k| \quad (2)$$

Mean Absolute Percentage Error (MAPE) represents the average percentage error between predicted and actual values, making it particularly useful for comparing data across different scales. MAPE enables an intuitive assessment of prediction accuracy by calculating errors as percentages, allowing for consistent evaluation regardless of the data's magnitude. However, it has a limitation in that it can produce disproportionately large errors when the actual values are close to zero [53].

$$MAPE = \frac{100}{n} \sum_{k=1}^n \left| \frac{y_k - \hat{y}_k}{y_k} \right| \quad (3)$$

Root Mean Squared Error (RMSE) is the square root of the Mean Squared Error and provides the standard deviation of prediction errors. Since it takes the square root after averaging the squared errors, it restores the error to the same units as the original data, allowing for an intuitive interpretation [54]. A lower RMSE value indicates higher predictive accuracy. Because RMSE is sensitive to outliers, using it in conjunction with MAE allows for a more precise evaluation of model performance [55].

$$RMSE = \sqrt{\frac{1}{n} \sum (y_k - \hat{y}_k)^2} \quad (4)$$

Lastly, the Coefficient of Determination ( $R^2$ ) is a metric representing the correlation between predicted and actual values, evaluating how well independent variables explain the dependent variable in a regression model. A value closer to 1 indicates higher explanatory power. However,  $R^2$  has a characteristic where its value increases as the number of independent variables grows, even if those variables do not significantly affect the dependent variable. This can lead to an overestimation of the model's actual utility. Therefore, judging the validity of a regression model solely based on  $R^2$  has limitations; to address this, the Adjusted  $R^2$  was utilized for evaluation.

To ensure the robustness of the predictive models and address potential sensitivity to temporal parameters, a sensitivity analysis was conducted during the validation phase by systematically varying the lookback window size. While the primary analysis utilizes a 4-quarter window to capture annual cyclicity, alternative windows of 2 and 6 quarters were additionally applied to verify the models' stability across diverse temporal horizons. This analysis was specifically performed on the LSTM and RF models, as their architectures are structurally more sensitive to window configuration: LSTM processes sequential dependencies through its gating mechanism, while RF's tree-splitting structure is directly

affected by the dimensionality of lag-based features. In contrast, the ANN treats the flattened lookback inputs as a static feature vector without encoding temporal order; thus, changes in window length merely alter the input dimensionality rather than fundamentally modifying the model's learning dynamics, making a separate window-based sensitivity analysis less informative for this architecture.

Following this evaluation, the influence of independent variables on the model's predictions was assessed. Given that an analysis of influence is required based on the flexibility of each model's explanation and representation, a model-agnostic method was applied. This approach treats the original model as a black box, deriving explanations generated post hoc without relying on the internal structure of the model [56]. Among various methods, this study employed the SHAP (SHapley Additive exPlanations) technique.

SHAP is a game theory-based method that excels in quantitatively clarifying and visualizing the contribution of individual input features to the machine learning model's predicted values [9]. SHAP values dynamically demonstrate how interactions between features affect the probability of risk and the role each feature plays at an individual level [57]. Therefore, the main contribution of this study is the proposal of an explainable model that predicts corporate profitability based on each variable, while further identifying which material- and labor-related cost factors exert direct influence on these predictions by applying an ensemble learning model and SHAP to analyze the key risk factors underlying the outcome.

## 4. Results

### 4.1. Comparative Analysis of Predictive Capabilities Introduced by Machine Learning

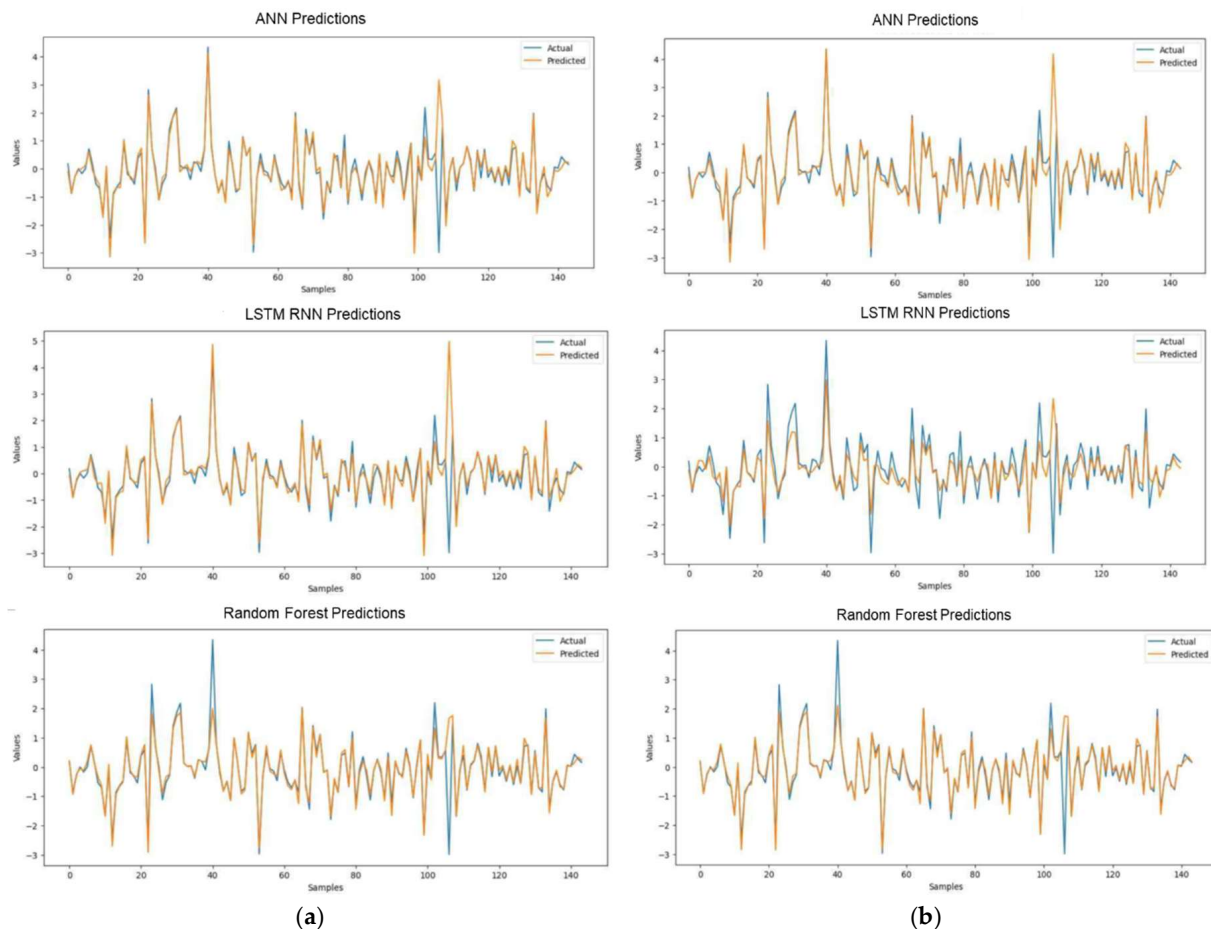
This study compared the predictive performance of ANN, LSTM, and Random Forest algorithms for predicting the Return on Assets (ROA) and D/E Ratio of construction firms using Dataset 1 and Dataset 2.

The hyperparameters used to obtain the ROA prediction results in this study are summarized in Table 2. Moreover, the performance results of each model for ROA prediction are presented in Figure 6a,b, derived from Dataset 1 and Dataset 2, respectively. The evaluation of each model's performance is detailed in Table 3.

**Table 2.** Hyperparameter applied in model for ROA prediction.

| Applied Model | Dataset   | Hyperparameter                                                                                                                                                               |
|---------------|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ANN           | Dataset 1 | Lookback Window: 4, Layer 1: 64 nodes (ReLU), Layer 2: 32 nodes (ReLU), Output Layer: 1 node (Linear), Optimizer: Adam (Learning Rate: 0.001), Epochs: 150, Batch Size: 32   |
|               | Dataset 2 |                                                                                                                                                                              |
| LSTM RNN      | Dataset 1 | Lookback Window: 4, Layer 1: 64, Layer 2: 32, Activation: tanh, Dense Layer: 16, Dense Activation: ReLU, Optimizer: Adam (Learning Rate: 0.001), Epochs: 150, Batch Size: 32 |
|               | Dataset 2 |                                                                                                                                                                              |
| RF            | Dataset 1 | Lookback Window: 4, n_estimators: 200, max_depth: 15, min_samples_split: 2, min_samples_leaf: 1, random_state: 42, bootstrap: True                                           |
|               | Dataset 2 |                                                                                                                                                                              |

The experimental results demonstrate that the Random Forest (RF) model exhibited the most superior performance in predicting ROA across all scenarios. The RF model recorded the highest Coefficient of Determination ( $R^2$ ) and the lowest error rates (RMSE and MAE) for both datasets. Notably, in Dataset 2, which incorporated material and labor costs, the RF model achieved an  $R^2$  of 0.7978, proving its ability to explain approximately 80% of the variance in ROA. In contrast, the ANN model maintained a respectable performance with an  $R^2$  in the 0.61 range. However, the LSTM model, characterized by its gated structure, showed relatively lower predictive power ( $R^2$  0.52–0.53) within the data environment of this study, indicating limitations in predicting ROA based on static data.



**Figure 6.** Result of ROA Prediction by applying various models. (a) ROA Prediction from Dataset 1; (b) ROA Prediction from Dataset 2.

**Table 3.** Model Performance for ROA Prediction.

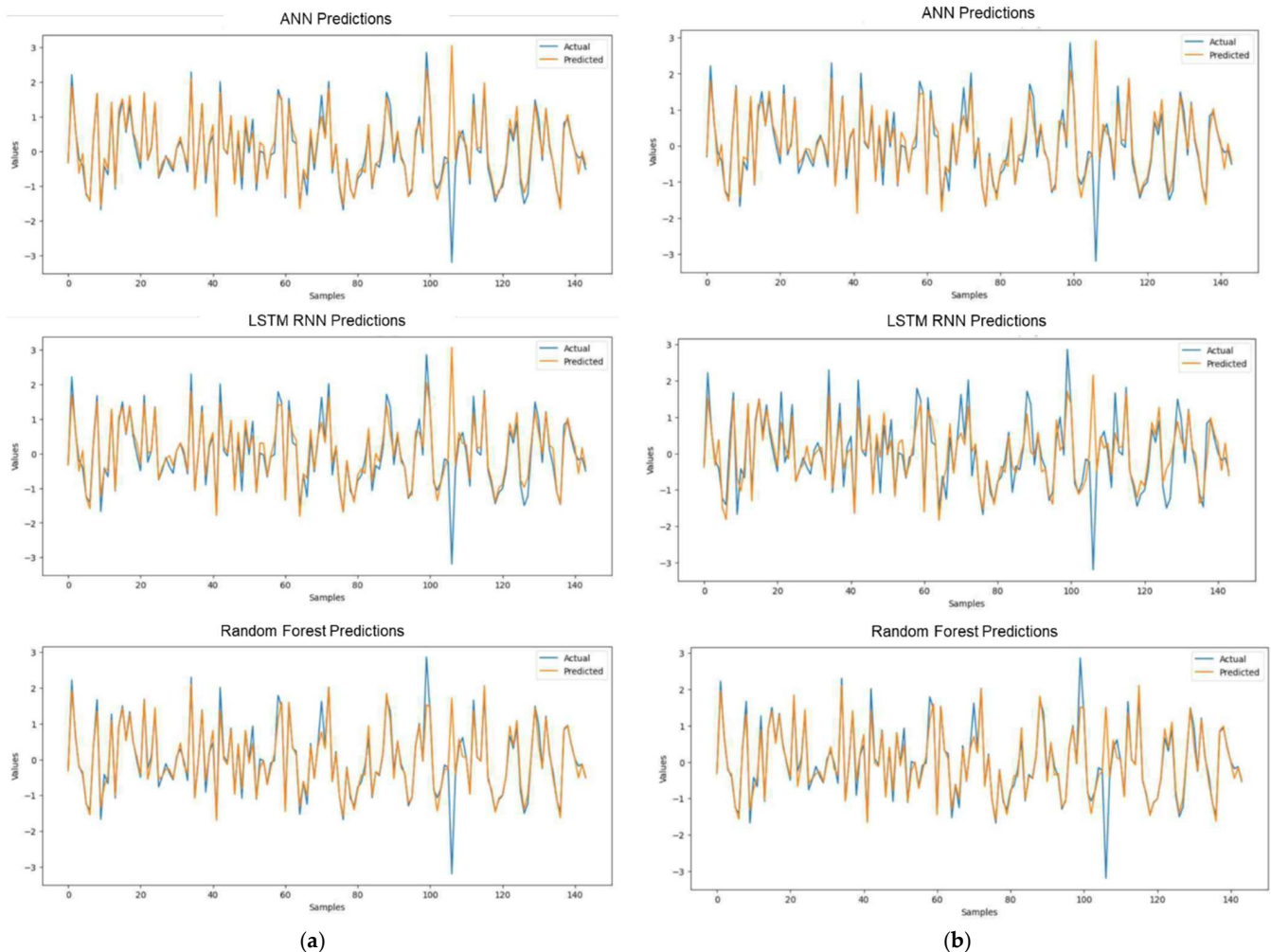
| Applied Model | Dataset   | RMSE   | MAE    | R <sup>2</sup> | Adj-R <sup>2</sup> | MAPE   |
|---------------|-----------|--------|--------|----------------|--------------------|--------|
| ANN           | Dataset 1 | 0.6171 | 0.2099 | 0.6186         | 0.5930             | 1.6829 |
|               | Dataset 2 | 0.6226 | 0.2043 | 0.6118         | 0.5497             | 1.2527 |
| LSTM RNN      | Dataset 1 | 0.6860 | 0.2236 | 0.5279         | 0.4962             | 1.5039 |
|               | Dataset 2 | 0.6829 | 0.2413 | 0.5331         | 0.4572             | 1.5515 |
| RF            | Dataset 1 | 0.4667 | 0.1574 | 0.7819         | 0.7673             | 1.1678 |
|               | Dataset 2 | 0.4522 | 0.1730 | 0.7978         | 0.7649             | 0.5125 |

Analysis of performance variations based on dataset composition confirmed that predictive accuracy improved significantly when actual operational cost variables—such as material and labor costs—were included alongside financial indicators (Dataset 2). For the RF model specifically, the error metric (MAE) decreased from 1.1678 in Dataset 1 to 0.5125 in Dataset 2, representing a reduction of approximately 56%. This suggests that in industries where the cost of goods sold (COGS) is a dominant factor in determining ROA, information on material and labor costs serves as essential variables for accurate profitability prediction.

Similarly, the algorithms were applied and evaluated for D/E ratio prediction, and the corresponding hyperparameters for each dataset are reported in Table 4. As a result, the predictive performance results for the D/E Ratio of each model are presented in Figure 7 and Table 4 below. Figure 7a illustrates the analysis results based on Dataset 1, while Figure 7b represents the results based on Dataset 2.

**Table 4.** Hyperparameter applied in model for D/E prediction.

| Applied Model | Dataset   | Hyperparameter                                                                                                                                                               |
|---------------|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ANN           | Dataset 1 | Lookback Window: 4, Layer 1: 64 nodes (ReLU), Layer 2: 32 nodes (ReLU), Output Layer: 1 node (Linear), Optimizer: Adam (Learning Rate: 0.001), Epochs: 1000, Batch Size: 32  |
|               | Dataset 2 |                                                                                                                                                                              |
| LSTM RNN      | Dataset 1 | Lookback Window: 4, Layer 1: 64, Layer 2: 32, Activation: tanh, Dense Layer: 16, Dense Activation: ReLU, Optimizer: Adam (Learning Rate: 0.001), Epochs: 100, Batch Size: 32 |
|               | Dataset 2 |                                                                                                                                                                              |
| RF            | Dataset 1 | Lookback Window: 4, n_estimators: 150, max_depth: 4, max_features: sqrt, min_samples_split: 10, min_samples_leaf: 5, random_state: 42, bootstrap: True                       |
|               | Dataset 2 |                                                                                                                                                                              |

**Figure 7.** Result of D/E Prediction by applying various models. (a) D/E Prediction from Dataset 1; (b) D/E Prediction from Dataset 2.

In the prediction of the Debt-to-Equity (D/E) Ratio, the Random Forest (RF) model once again demonstrated the highest predictive accuracy. As shown in Table 5, the RF model outperformed other models across all evaluation metrics. Notably, in the Dataset 2 environment, it recorded an  $R^2$  of 0.7978, precisely predicting the capital structure variations of firms at an approximately 80% confidence level.

The ANN model also exhibited respectable performance in predicting the debt ratio. In Dataset 2, the ANN recorded an  $R^2$  of 0.7513 and an RMSE of 0.5015, showing the second-highest reliability after the RF model. In contrast, the LSTM model remained within the  $R^2$  0.66 range, appearing relatively inefficient for predicting static financial structures. Similarly, the analysis of Dataset 2 revealed a reduction in the model's error

margins. This indicates that the debt ratio is not explained solely by financial indicators but is directly influenced by profitability through operational cost management efficiency, such as material and labor costs. This supports the premise that the additional material and labor cost variables are significant independent variables that contribute substantially to predicting the debt ratio, rather than serving as unnecessary noise. Specifically, the RF model's lowest RMSE (0.4521) signifies the minimal discrepancy between actual corporate debt levels and the model's predicted values, thereby enhancing its potential as a tool for corporate financial risk management in the future.

**Table 5.** Model Performance of D/E Prediction.

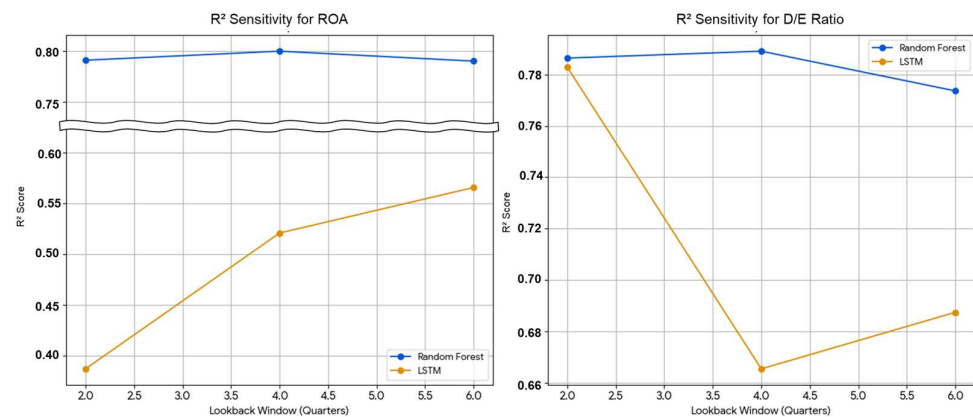
| Applied Model | Dataset   | RMSE   | MAE    | R <sup>2</sup> | Adj-R <sup>2</sup> | MAPE   |
|---------------|-----------|--------|--------|----------------|--------------------|--------|
| ANN           | Dataset 1 | 0.5084 | 0.2055 | 0.7444         | 0.7272             | 0.8008 |
|               | Dataset 2 | 0.5015 | 0.1807 | 0.7513         | 0.7109             | 0.5398 |
| LSTM RNN      | Dataset 1 | 0.5780 | 0.2354 | 0.6695         | 0.6474             | 0.8748 |
|               | Dataset 2 | 0.5819 | 0.3017 | 0.6652         | 0.6108             | 1.1255 |
| RF            | Dataset 1 | 0.4636 | 0.1655 | 0.7875         | 0.7732             | 0.4629 |
|               | Dataset 2 | 0.4521 | 0.1731 | 0.7978         | 0.7649             | 0.5126 |

Based on the analysis of these results, the Random Forest (RF) model demonstrated superior performance in predicting both ROA (a profitability indicator) and the Debt-to-Equity Ratio (a financial soundness indicator). In particular, a significant improvement in predictive power across all models was observed when operational cost elements—specifically material and labor costs—were integrated (Dataset 2) into the traditional financial indicators (Dataset 1).

Finally, to ensure the reliability of the proposed framework, a sensitivity analysis was conducted by varying the lookback window settings for the LSTM and RF models. As noted in Section 3.4, the ANN was excluded from this analysis because its fully connected architecture processes the lookback window as an unordered feature vector, rendering its performance comparatively invariant to window length adjustments. As illustrated in Figure 8, the RF model demonstrated remarkably stable predictive performance, maintaining an R<sup>2</sup> value within a consistent range of 0.78 to 0.80 for Debt-to-Equity (D/E) ratio predictions, even as the window size was adjusted from 2 to 6 quarters. In contrast, the LSTM model exhibited significant fluctuations in performance depending on the window size. This suggests that, given the current data scale, ensemble-based RF models are more effective than deep learning architectures at capturing the specific types of temporal dependencies inherent in this dataset. These findings confirm that adopting a 4-quarter setting as the primary analytical index is statistically robust; it avoids overfitting and remains unbiased toward specific parameter configurations, thereby accurately reflecting actual industry volatility.

Quantitatively, the Random Forest model substantiated its optimal explanatory power by achieving an R<sup>2</sup> of 0.7978 for both indicators. The superiority of the RF model stems from its structural robustness and its fit for the dataset's specific properties. By utilizing an ensemble approach that averages the outputs of multiple trees, RF provides a framework that effectively dampens errors and prevents overfitting. Unlike models that rely solely on numerical optimization, RF's ability to split data based on relative ranks allows it to learn nonlinear interactions more effectively. This makes the model particularly resilient to potential distortions arising from the heterogeneous relationships between financial ratios and supply-chain cost variables, thereby leading to more reliable performance. Furthermore, the ANN model also evidenced that real-world cost data is a key driver in predicting financial structures, as its Mean Absolute Error (MAE) for the debt ratio decreased from 0.2055 to 0.1807 upon the inclusion of cost variables. These findings

highlight the critical role of operational cost information in enhancing the accuracy of corporate financial performance forecasting. Even with a relatively limited dataset, this approach enabled a comparison between the ANN's nonlinear baseline and the LSTM's temporal-dependence baseline, thereby examining the extent to which construction-specific variables improve predictive performance. However, the relatively lower predictive power of the LSTM RNN indicates that the dataset in this study may be more strongly governed by the heterogeneity of firm-level financial structures rather than by temporal dependencies. This comparative setup allows us to isolate the incremental contribution of construction-specific cost variables across model classes.



**Figure 8.** Sensitivity Analysis results.

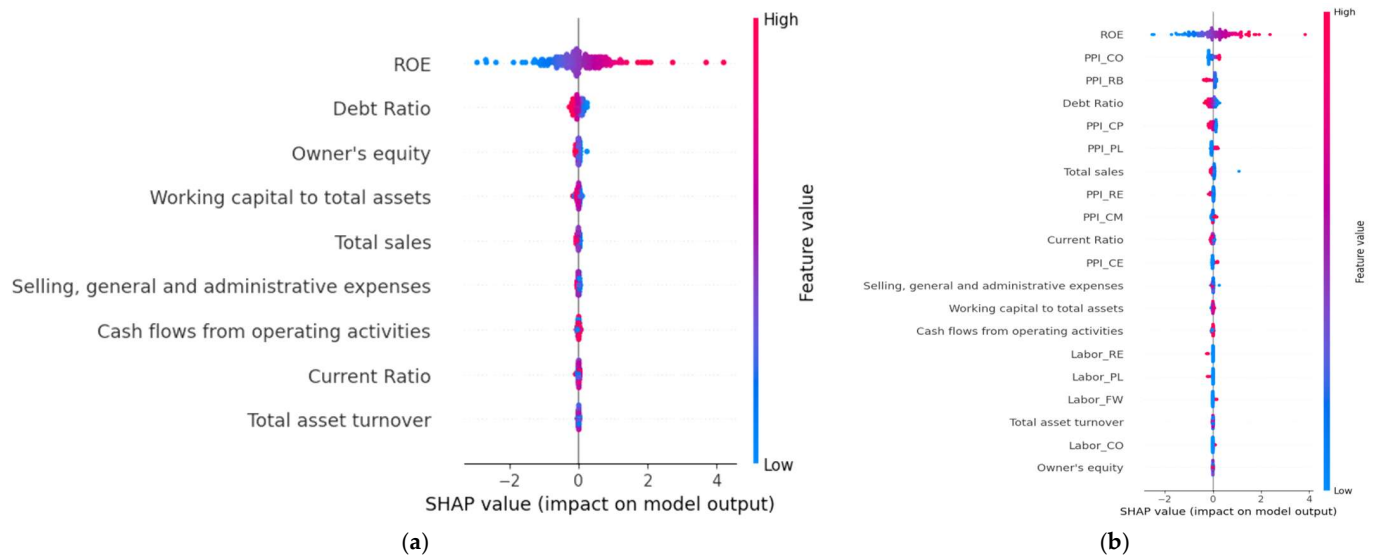
#### 4.2. Variable Influence Analysis Using SHAP Analysis

Random Forest performs predictions through multiple decision trees and allows for the evaluation of variable importance. However, as variable importance alone has limitations in identifying whether a variable's impact on the dependent variable is positive or negative, SHAP analysis was conducted to analyze the contribution of each independent variable to the model's predictions. In the SHAP summary plots, each dot represents an individual data point, while the color indicates the feature value (high or low). The X-axis, representing the SHAP value, serves as a key metric showing the extent of each variable's influence on the model's predicted outcome.

The ROA analysis results for Dataset 1 (Financial Variables) are shown in Figure 9a. The analysis identified Return on Equity (ROE) as the most dominant variable. Higher ROE values (indicated in pink) exerted a strong positive (+) influence on ROA predictions, implying that a firm's superior capability in utilizing equity is proportional to its overall asset-based profit generation capacity. Although the two metrics are structurally linked through their shared reliance on net income, the minimal multicollinearity confirms that ROE provides distinct informational value rather than simple redundancy. Specifically, ROE functions as a key antecedent determining ROA from the perspective of equity efficiency, highlighting the robust and integrated nature of these financial performance measures. The debt-to-equity ratio emerged as the second most critical variable; an increase in the debt ratio up to a certain level acted as a leverage effect, contributing slightly positively (+) to ROA. Furthermore, a larger equity size was confirmed to have a positive impact on financial performance, proving that a firm's foundational capital strength serves as the basis for asset management efficiency.

In Figure 9b, which represents Dataset 2 (Added Cost Variables), while the significance of ROE remained high, the contributions of the newly introduced material cost variables were prominent. Notably, an increase in the Producer Price Index for Concrete (PPI\_CO) had a positive influence on the predicted values. This indicates the financial sensitivity and cost-pass-through mechanisms of construction firms during periods of rising costs.

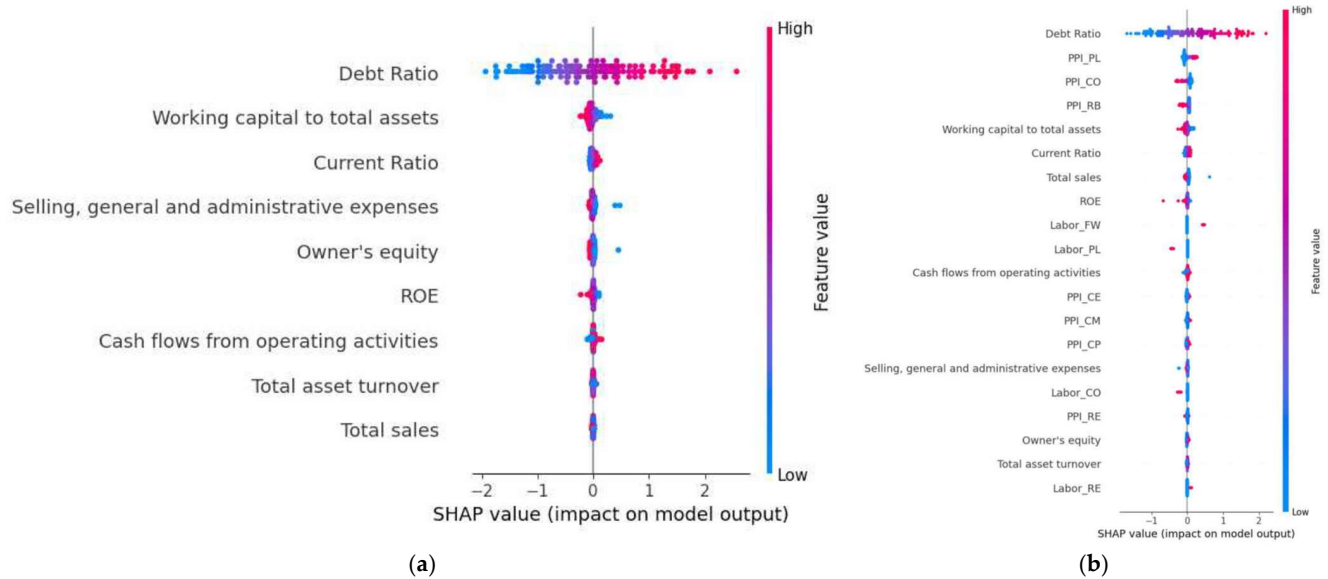
Rather than directly proving management capability, the positive SHAP value of PPI\_CO reflects the degree of financial exposure and the firm's structural resilience in adjusting to specific material price volatility. Conversely, the index for rebar and steel bars (PPI\_RB) could be explained as a major cost pressure factor, exerting a negative (−) impact on ROA. Interestingly, in Dataset 2, a higher debt ratio was found to have a negative impact on ROA, suggesting that when cost variables are controlled, increased financial risk becomes a direct threat to profitability. Meanwhile, although labor-related variables contributed to the model, their influence was relatively marginal compared to material costs or capital indicators.



**Figure 9.** SHAP Summary Plots for ROA Prediction Models. (a) Dataset 1 (Financial Variables Model); (b) Dataset 2 (Integrated Financial and Construction Cost Variables Model).

Analysis of Dataset 1 revealed that the debt-to-equity ratio itself acted as the most dominant variable in the prediction model, while the influence of other financial indicators remained relatively low. As shown in Figure 10a, a higher debt-to-equity ratio (indicated in red) exhibited a clear linear relationship, strongly driving up the predicted leverage indicator. Conversely, financial soundness indicators such as Working Capital to Total Assets (WC/TA) and the current ratio recorded negative (−) SHAP values as their values increased. This reflects the model's interpretation that higher corporate liquidity leads to a lower assessment of the firm's leverage risk.

In Dataset 2 (Figure 10b), while the debt-to-equity ratio maintained its position as the most influential factor, the material cost variable group recorded the second-highest contribution, demonstrating a correlation between financial structure and cost indicators. Increases in the Producer Price Index for construction plastics (PPI\_PL) and concrete (PPI\_CO) exerted a positive (+) influence on the predicted values, reflecting the high financial sensitivity and distinct cost-pass-through mechanisms of firms during periods of rising input prices. Conversely, the rise in prices for rebar and steel bars (PPI\_RB) was found to make a negative (−) contribution to the debt-to-equity ratio, identifying it as a specific threat to financial stability. This divergence in the direction of influence among various material indices suggests that the model is capturing item-specific transmission effects of supply-chain shocks rather than merely functioning as a proxy for a uniform macroeconomic inflation trend. Regarding labor costs, indicators such as those for formwork carpenters (Labor\_FW) showed a slight positive (+) contribution; however, the overall contribution remained lower than that of material costs, confirming that the substantial impact of labor costs on financial structure is relatively limited.



**Figure 10.** SHAP Summary Plots for D/E Prediction Models. (a) Dataset 1 (Financial Variables Model); (b) Dataset 2 (Integrated Financial and Construction Cost Variables Model).

## 5. Discussion

### 5.1. Academic and Methodological Implications

The primary objective of this study was to construct a financial performance prediction model that reflects the unique characteristics of the construction industry. To this end, specific financial variables were selected and supplemented by industry-differentiated factors such as material and labor costs. Methodologically, this research is original in that it not only enhanced predictive accuracy through various algorithms but also employed SHAP (SHapley Additive exPlanations) analysis to provide interpretability. SHAP analysis transcends simple feature importance rankings by quantitatively elucidating the contribution and directionality (positive [+] or negative [−]) of individual variables—such as fluctuations in steel prices—on the predicted values of dependent variables (ROA and Debt-to-Equity Ratio). By facilitating feature-level interpretability, financial performance is conceptualized not merely as a result of static ratios based on conventional standards but as a dynamic system that responds to supply-chain volatility. This approach allows for the holistic integration of cost fluctuations into the financial evaluation framework. However, it should be noted that the feature importance derived from SHAP values represents the predictive relevance and marginal contribution of each variable within the model's correlational framework, rather than establishing a direct causal link or interventional effect.

The high precision and stability demonstrated by the Random Forest (RF) model signify its effectiveness in capturing complex, non-linear interactions between financial and industry-specific data [58]. This study confirms that the ensemble mechanism of RF, which combines numerous decision trees, is highly effective at reducing variance and identifying complex nonlinear patterns [46]. The superior stability of RF suggests that its recursive partitioning structure is better suited for datasets with high heterogeneity between financial and cost indicators. In contrast, the performance of LSTM—a model designed for long-term sequence learning—was suboptimal. It is inferred that the specific structure of our data, marked by restricted time-series horizons and dominant firm-level differences, prevented LSTM from effectively manifesting its strengths in temporal dependency modeling. The capacity of RF to accurately capture these inter-linkages makes it a particularly valuable tool for stakeholders in both the real estate and financial sectors for strategic decision-making [59].

This study, focusing on listed construction firms in the Republic of Korea, presents the predictive importance of supply-chain cost volatility for financial performance using a SHAP-based approach. However, the magnitude and direction of variable effects are likely to vary across countries depending on regulatory regimes, procurement practices, and macroeconomic conditions [46]. Similarly, cross-country comparative studies have reported that the effects of liquidity, leverage, and profitability differ according to institutional factors such as the level of ESG institutionalization and broader market contexts [34]. Therefore, to enhance the generalizability of the model, mapping the input features to country-specific material price indices and wage indices when applying it to other countries would allow the analytical framework of this study to be extended and utilized as a context-tailored prediction and diagnostic model.

### 5.2. Differentiation and Practical Value

This study distinguishes itself from previous research by selecting variables tailored to the construction sector [60]. A key differentiator is the significant reduction in prediction errors achieved by incorporating material and labor costs.

By quantitatively proving that non-financial, operational variables such as material price indices (Cement, Rebar, Concrete PPI) and labor wages substantially improve model performance (improving RMSE and  $R^2$ ), this research provides a clear empirical basis for financial experts and managers. It provides a quantitative basis for identifying firm-level vulnerabilities to supply-chain shocks. These findings should be interpreted as risk indicators reflecting the transmission of raw material price volatility into financial performance, rather than definitive evidence of internal procurement efficiency.

Notably, the heterogeneous directional impact observed across different material indices—where rebar (PPI\_RB) exerted a negative influence while concrete (PPI\_CO) showed a positive association—suggests that the model captures item-specific cost transmission mechanisms rather than merely functioning as a proxy for a uniform upward inflationary trend [6,8]. Despite the traditional labor-intensive nature of construction, labor-related variables (Labor) showed relatively marginal influence in ROA prediction. In contrast, material-related factors made substantial contributions; specifically, an increase in the Rebar and Steel Bar index (PPI\_RB) negatively affected ROA. It shows that construction firm financials can be understood as a dynamic system that is sensitive to supply-chain volatility, in which cost shocks driven by specific input items exert direct pressure on profitability, thereby extending the study's knowledge boundary. In debt ratio prediction, the rise in Concrete (PPI\_CO) and Construction Plastic (PPI\_PL) indices positively influenced predicted values, suggesting that material price volatility is a critical factor in corporate leverage and financial risk assessment. Overall, while prior studies have predominantly interpreted construction performance through financial ratios, this study adds domain validity by explicitly incorporating construction-specific cost drivers—material and labor costs—and by operationalizing item-level PPIs as grouped volatility indicators [6]. However, as the analysis period includes incidents of heightened supply-chain disruption such as the COVID-19 pandemic and geopolitical shocks, the observed effects may reflect event-amplified patterns rather than stable structural relationships [8,61]. Accordingly, the inferred feature effects and gains in predictive performance would not be assumed time-invariant or universally generalizable. Consistent with approaches in related research, it is appropriate to apply the proposed framework conditionally, by considering cross-country institutional and market contexts as well as period-specific shocks and by conducting localization prior to use.

The strategic implications for key stakeholders to support decision-making, derived from these results, are as follows. For project owners and financial institutions, the findings

suggest that the evaluation of externally audited construction firms can be enhanced by incorporating structural resilience against cost volatility over traditional financial ratios. This facilitates the adoption of a comprehensive risk management framework for project financing (PF) or cost shocks, establishing a novel model for holistic assessment [8]. From the perspective of contractors, the application of SHAP-based feature importance enables the identification of factors highly sensitive to cost fluctuations, thereby serving as a specialized financial tool for differentiated strategic management. Consequently, the model explanations suggest that supply-chain and raw material price variables are consistently influential, highlighting related risk management as significant indicators for preliminary risk screening. The results imply that construction companies need to prioritize material supply stability and derivative risk hedging strategies over simple labor cost reductions to secure long-term profitability.

## 6. Conclusions

This study aimed to propose a machine-learning-based framework capable of precisely predicting the financial performance of construction firms by reflecting industry-specific characteristics. To achieve this, we conducted an analysis that integrated traditional financial indicators with actual operational cost data—material costs (PPI) and labor costs—which are the primary determinants of real-world construction expenses. By directly applying various algorithms, this research identified the most suitable models for the construction industry and clarified the influence of specific variables.

The results demonstrate that among the ANN, LSTM, and RF models, the RF model recorded the highest predictive accuracy for both ROA and the debt-to-equity ratio, proving it to be the optimal model. Notably, the inclusion of construction-specific material and labor costs significantly reduced the MAE, suggesting that operational cost data are essential independent variables for predicting the financial soundness of construction companies. This approach overcomes the limitations of previous insolvency research, which has been predominantly focused on the manufacturing sector. Furthermore, by incorporating XAI, the study secured model interpretability by visualizing the grounds for predictions and identifying the contribution and directionality of each variable. A key discovery was that, despite the traditional labor-intensive nature of construction, the fluctuations in material costs—specifically rebar and concrete price indices—exerted a far more dominant influence on actual financial performance than labor costs. This implies that management and financial stakeholders keep concentrating on supply-chain cost indicators for risk management and hedging strategies. Moreover, these findings, grounded in the proposed predictive model and analytical framework, can be used for ex-ante assessments of future changes in financial performance. In particular, they provide a basis for scenario-based examination of how a firm's financial performance may be affected under varying levels of material and labor cost volatility. With further research to extend the framework, its applicability could be expanded beyond retrospective financial performance evaluations by leveraging real-time cost indicators or long-horizon time-series data—potentially enabling future-oriented financial performance forecasting and the proactive identification (screening) of externally audited firms.

Nevertheless, this study has limitations, such as not reflecting various external factors—including project progress status and Gross Domestic Product (GDP)—that could influence a firm's financial state. Additionally, while the impact on cost structure and performance may vary by firm size or type (e.g., general vs. specialty contractors), such classifications were not fully incorporated. Future research should address these limitations to more clearly elucidate the relationship between cost structures and financial performance. In particular, future research should build on SHAP-based local interpretations to identify

the causal effects of changes in supply-chain management on financial performance, thereby evaluating whether such changes exert a causal impact on financial outcomes. In addition, analyzing material-specific PPIs enabled an interpretation of their explanatory power; however, their predictive importance may partially function as a proxy for broader inflation or macro-level time trends inherent in price indices. Consequently, the interpreted influence of PPI should be understood as a measure of financial exposure to cost shocks. To decouple systemic inflation from genuine firm-level transmission effects, future research must utilize deflated financial metrics and explicit inflation-control parameters. Such advancements are expected to enhance the financial stability of construction firms and provide highly reliable tools that support strategic decision-making by proactively managing financial risks associated with cost volatility. Lastly, while this study identifies salient predictors through XAI, it does not explicitly establish causal pathways between supply-chain shocks and financial outcomes. The SHAP-based insights reflect statistical associations captured by the machine learning model. Therefore, future research should employ quasi-experimental designs, such as instrumental variables or difference-in-differences, to rigorously validate the causal mechanisms underlying the observed predictive patterns.

**Author Contributions:** Conceptualization, S.-J.Y. and K.-T.L.; methodology, K.-T.L.; software, S.-J.Y.; validation, S.-J.Y. and K.-T.L.; formal analysis, S.-J.Y.; investigation, K.-T.L.; resources, S.-J.Y.; data curation, S.-J.Y.; writing—original draft preparation, S.-J.Y.; writing—review and editing, K.-T.L.; visualization, K.-T.L.; supervision, K.-T.L.; project administration, K.-T.L.; funding acquisition, K.-T.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. RS-2024-00351295).

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** Author Seong-Jun Ye was employed by the company Lotte Engineering & Construction Co., Ltd. The remaining authors 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.

## Abbreviations

The following abbreviations are used in this manuscript:

|       |                                     |
|-------|-------------------------------------|
| XAI   | Explainable Artificial Intelligence |
| ANNs  | Artificial Neural Networks          |
| LSTM  | Long Short-Term Memory              |
| RF    | Random Forest                       |
| SHAP  | Shapley Additive exPlanations       |
| PPI   | Producer Price Index                |
| ROA   | Return On Assets                    |
| D/E   | Debt-to-Equity Ratio                |
| ROE   | Return on Equity                    |
| MAE   | Mean Absolute Error                 |
| MAPE  | Mean Absolute Percentage Error      |
| RMSE  | Root Mean Squared Error             |
| $R^2$ | Coefficient of Determination        |

## References

1. Zeitun, R.; Tian, G.G. Capital Structure and Corporate Performance: Evidence from Jordan. Available online: <https://ssrn.com/abstract=2496174> (accessed on 14 September 2014).
2. Lee, S.; Namgoong, D.; Han, S.H. Financial Model Assessment for Predicting Construction Company Bankruptcy. *Proc. Korean Soc. Civ. Eng. Conf.* **2024**, 432–433.

3. Sinha, S.; Samanta, P.K. Determinants of capital structure of Indian manufacturing firms: A Hierarchical linear modelling approach. In Proceedings of the India Finance Conference (IFC), Kolkata, India, 20–22 December 2018; pp. 20–22.
4. Harris, M.; Raviv, A. The theory of capital structure. *J. Financ.* **1991**, *46*, 297–355. [\[CrossRef\]](#)
5. Lai, H.Y.; Abdul Aziz, A.-R.; Chan, T.K. Effect of the global financial crisis on the financial performance of public listed construction companies in Malaysia. *J. Financ. Manag. Prop. Constr.* **2014**, *19*, 246–263. [\[CrossRef\]](#)
6. Seo, W.; Kim, B.; Bang, S.; Kang, Y. Identifying key financial variables predicting the financial performance of construction companies. *J. Constr. Eng. Manag.* **2024**, *150*, 04024007. [\[CrossRef\]](#)
7. Balogun, E.D.; Ogunsola, K.O.; Ogunmokun, A.S. Developing an advanced predictive model for financial planning and analysis using machine learning. *IRE J.* **2022**, *5*, 320–328.
8. Jang, J.H.; Hong, R.L.; Lee, K.T.; Kim, J.H. A Comprehensive Approach to Capturing the Impact and Identifying Countermeasures of the COVID-19 Pandemic at Construction Sites in the Republic of Korea. *Buildings* **2023**, *14*, 30. [\[CrossRef\]](#)
9. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
10. Öcal, M.E.; Oral, E.L.; Erdis, E.; Vural, G. Industry financial ratios—Application of factor analysis in Turkish construction industry. *Build. Environ.* **2007**, *42*, 385–392. [\[CrossRef\]](#)
11. Alsharef, A.; Banerjee, S.; Uddin, S.J.; Albert, A.; Jaselskis, E. Early impacts of the COVID-19 pandemic on the United States construction industry. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1559. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Choi, J.-S.; Roh, S.-H. Estimation of financial soundness models and stress tests for specialty construction companies by size and industry. *J. Real Estate Anal.* **2023**, *29*, 91–105. [\[CrossRef\]](#)
13. Choi, M.K. A study on the closing status and rehabilitation cases of construction companies—Focusing on cases from 2020 to 2024. *Rehabil. Law* **2025**, *30*, 39–81.
14. Karmaker, C.L.; Al Aziz, R.; Palit, T.; Bari, A.M. Analyzing supply chain risk factors in the small and medium enterprises under fuzzy environment: Implications towards sustainability for emerging economies. *Sustain. Technol. Entrepreneursh* **2023**, *2*, 100032. [\[CrossRef\]](#)
15. Dey, R.K.; Hossain, S.Z.; Rahman, R.A. Effect of corporate financial leverage on financial performance: A study on publicly traded manufacturing companies in Bangladesh. *Asian Soc. Sci.* **2018**, *14*, 124–133. [\[CrossRef\]](#)
16. Saputra, F. Analysis effect return on assets (ROA), return on equity (ROE) and price earning ratio (PER) on stock prices of coal companies in the Indonesia Stock Exchange (IDX) period 2018–2021. *Dinasti Int. J. Econ. Financ. Account.* **2022**, *3*, 82–94. [\[CrossRef\]](#)
17. Languju, O.; Mangantar, M.; Tasik, H.H.D. The Influence of Return on Equity, Company Size, Price Earning Ratio and Capital Structure on Property and Real Estate Company Values Listed on the Indonesia Stock Exchange. *J. Berk. Ilm. Efisiensi* **2016**, *16*, 387–398.
18. Nukala, R.V.; Prasada, S.S.R.R. The impact of capital structure on firm performance: Evidence from Indian construction companies. *Theor. Appl. Econ.* **2021**, *28*, 261–274.
19. Claessens, S. Capital and liquidity requirements: A review of the issues and literature. *Yale J. Reg.* **2014**, *31*, 735.
20. Anuar, R.B.; Jais, M.B.; Tinggi, M. The Impact of Non-Current Assets on the Performance of Firms in Malaysian Construction Sector. *Int. J. Acad. Res. Account. Financ. Manag. Sci.* **2021**, *11*, 57–79. [\[CrossRef\]](#)
21. Kyung-Tae, L.; Min-Seok, K.; Hee-Jeong, K.; Ju-Hyung, K. A model to predict occupational safety and health management expenses in construction applying multi-variate regression analysis and deep neural network. *J. Archit. Inst. Korea* **2021**, *37*, 217–226. [\[CrossRef\]](#)
22. Lim, H.K.; Choi, J.H. Derivation of core financial ratios for evaluating the management soundness of domestic construction companies. *J. Archit. Inst. Korea Struct. Eng.* **2015**, *31*, 117–126. [\[CrossRef\]](#)
23. Heikal, M.; Khaddafi, M.; Ummah, A. Influence Analysis of Return on Assets (ROA), Return on Equity (ROE), Net Profit Margin (NPM), Debt to Equity Ratio (DER), and Current Ratio (CR) on Profit Growth in Automotive Companies in Indonesia Stock Exchange. *Int. J. Acad. Res. Bus. Soc. Sci.* **2014**, *4*, 101–114. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Choi, S.Y.; Kim, S.S.; Kim, J.J.; Kim, J.H. Financial distress prediction model for construction companies using machine learning. *J. Archit. Inst. Korea* **2018**, *34*, 53–60.
25. Jang, Y.; Jeong, I.B.; Cho, Y.K.; Ahn, Y. Predicting business failure of construction contractors using long short-term memory recurrent neural network. *J. Constr. Eng. Manag.* **2019**, *145*, 08219003. [\[CrossRef\]](#)
26. Černevičienė, J.; Kabašinskas, A. Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artif. Intell. Rev.* **2024**, *57*, 216. [\[CrossRef\]](#)
27. Kaka, A.; Lewis, J. Development of a company level cash flow forecasting model. *Constr. Manag. Econ.* **2003**, *21*, 693–705. [\[CrossRef\]](#)
28. Ariza-Garzón, M.J.; Arroyo, J.; Caparrini, A.; Segovia-Vargas, M.J. Explainability of a machine learning granting scoring model in peer-to-peer lending. *IEEE Access* **2020**, *8*, 64873–64890. [\[CrossRef\]](#)

29. Konhäuser, K.; Werner, T. Uncovering the financial impact of energy-efficient building characteristics with eXplainable artificial intelligence. *Appl. Energy* **2024**, *374*, 123960. [\[CrossRef\]](#)
30. Michaelas, N.; Chittenden, F.; Poutziouris, P. Financial policy and capital structure choice in U.K. SMEs: Empirical evidence from company panel data. *Small Bus. Econ.* **1999**, *12*, 113–130. [\[CrossRef\]](#)
31. Yıldırım, M.; Okay, F.Y.; Özdemir, S. Big data analytics for default prediction using graph theory. *Expert Syst. Appl.* **2021**, *176*, 114840. [\[CrossRef\]](#)
32. Janappriya, J.; Pabasara, W.; Zora, V.; Malindu, S. Artificial intelligence for predicting insolvency in the construction industry—A systematic review and empirical feature derivation. *Buildings* **2025**, *15*, 2988. [\[CrossRef\]](#)
33. Jeong, J.; Kim, C. Comparison of machine learning approaches for medium-to-long-term financial distress predictions in the construction industry. *Buildings* **2022**, *12*, 1759. [\[CrossRef\]](#)
34. Martin, H.; Zhou, Y.; Raman, R. Financial Metrics and Environment, Social, Governance (ESG) Performance: A Cross Border Comparison of China and the UK Construction Industries. *Buildings* **2025**, *15*, 1236. [\[CrossRef\]](#)
35. Shahrabadi, S.; Adão, T.; Peres, E.; Morais, R.; Magalhães, L.G.; Alves, V. Automatic Optimization of Deep Learning Training through Feature-Aware-Based Dataset Splitting. *Algorithms* **2024**, *17*, 106. [\[CrossRef\]](#)
36. Cheng, M.Y.; Hoang, N.D.; Limanto, L.; Wu, Y.W. A novel hybrid intelligent approach for contractor default status prediction. *Knowl.-Based Syst.* **2014**, *71*, 314–321. [\[CrossRef\]](#)
37. Tserng, H.P.; Chen, P.C.; Huang, W.H.; Lei, M.C.; Tran, Q.H. Prediction of default probability for construction firms using the logit model. *J. Civ. Eng. Manag.* **2014**, *20*, 247–255. [\[CrossRef\]](#)
38. Ptok, A.; Jindal, R.P.; Reinartz, W.J. Selling, general, and administrative expense (SGA)-based metrics in marketing: Conceptual and measurement challenges. *J. Acad. Mark. Sci.* **2018**, *46*, 987–1011. [\[CrossRef\]](#)
39. Graham, J.R.; Lemmon, M.L.; Schallheim, J.S. Debt, leases, taxes, and the endogeneity of corporate tax status. *J. Financ.* **1998**, *53*, 131–162. [\[CrossRef\]](#)
40. Kangari, R.; Farid, F.; Elgharib, H.M. Financial performance analysis for construction industry. *J. Constr. Eng. Manag.* **1992**, *118*, 349–361. [\[CrossRef\]](#)
41. Horta, I.M.; Camanho, A.S. Company failure prediction in the construction industry. *Expert Syst. Appl.* **2013**, *40*, 6253–6257. [\[CrossRef\]](#)
42. Kim, B.S. Prediction of corporate bankruptcy in the construction industry using financial variables. *J. Korea Acad.-Ind. Coop. Soc.* **2019**, *20*, 115–122.
43. Oh, J.G.; Yeo, G.T. An analysis of Financial Factors' Characteristic for Global Shipping Companies using Panel Regression Analysis. *J. Digit. Converg.* **2019**, *17*, 65–73.
44. The Bank of Korea. *Economic Outlook Report (November 2024)*; Bank of Korea: Seoul, Republic of Korea, 2024.
45. Construction Association of Korea. Report on the Construction Industry Wage Survey for the Second Half of 2024. 2025. Available online: [https://www.cak.or.kr/lay1/bbs/S1T41C42/A/14/view.do?article\\_seq=5320](https://www.cak.or.kr/lay1/bbs/S1T41C42/A/14/view.do?article_seq=5320) (accessed on 14 September 2024).
46. Nindartin, A.; Park, S.J.; Lee, K.T.; Kim, J.H.; Rostiyanti, S.F. Prediction of cost contingency in construction projects by introducing machine learning algorithms. *J. Civ. Eng. Manag.* **2025**, *31*, 860–880. [\[CrossRef\]](#)
47. Min, J.H.; Lee, Y.C. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Syst. Appl.* **2005**, *28*, 603–614. [\[CrossRef\]](#)
48. Rodriguez, J.D.; Perez, A.; Lozano, J.A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
51. Mahmoodzadeh, A.; Mohammadi, M.; Farid Hama Ali, H.; Hashim Ibrahim, H.; Nariman Abdulhamid, S.; Nejati, H.R. Prediction of safety factors for slope stability: Comparison of machine learning techniques. *Nat. Hazards* **2022**, *111*, 1771–1799. [\[CrossRef\]](#)
52. Taiwo, B.O.; Fissha, Y.; Hosseini, S.; Khishe, M.; Kahraman, E.; Adebayo, B.; Sazid, M.; Adesida, P.A.; Famobuwa, O.V.; Faluyi, J.O.; et al. Machine learning based prediction of flyrock distance in rock blasting: A safe and sustainable mining approach. *Green Smart Min. Eng.* **2024**, *1*, 346–361. [\[CrossRef\]](#)
53. Shehadeh, A.; Alshboul, O.; Al Mamlook, R.E.; Hamedat, O. Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. *Autom. Constr.* **2021**, *129*, 103827. [\[CrossRef\]](#)
54. Sirsikar, R.A.; Chowdhury, R. Machine Learning Approaches in Prediction of Railway Bridge Dynamic Response. In *National Conference on Futuristic Structural Engineering*; Springer Nature: Singapore, 2024; pp. 111–129.
55. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [\[CrossRef\]](#)
56. Belle, V.; Papantonis, I. Principles and practice of explainable machine learning. *Front. Big Data* **2021**, *4*, 688969. [\[CrossRef\]](#)

57. Timilsina, M.S.; Sen, S.; Uprety, B.; Patel, V.B.; Sharma, P.; Sheth, P.N. Prediction of HHV of fuel by Machine learning Algorithm: Interpretability analysis using Shapley Additive Explanations (SHAP). *Fuel* **2024**, *357*, 129573. [[CrossRef](#)]
58. Shoar, S.; Chileshe, N.; Edwards, J.D. Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression. *J. Build. Eng.* **2022**, *50*, 104102. [[CrossRef](#)]
59. Gala, D.M.; Pawar, B.; Band, G.; Dalal, A.; Chakraborty, A.; Das, U.; Patil, B.V. Evaluating the Effectiveness of Random Forest versus Decision Tree Algorithms in Predictive Analytics for Enhancing Sustainable Development Strategies in Real Estate, Infrastructure, and Construction: A Machine Learning Perspective. In Proceedings of the 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA), Pune, India, 25–26 October 2024; pp. 1–6. [[CrossRef](#)]
60. De Wet, J.H.; Du Toit, E. Return on equity: A constant companion in the world of financial performance measurement. *South. Afr. Bus. Rev.* **2007**, *11*, 56–69. [[CrossRef](#)]
61. Park, S.J.; Lee, K.T.; Im, J.B.; Kim, J.H. The need for smart architecture caused by the impact of COVID-19 upon architecture and city: A systematic literature review. *Sustainability* **2022**, *14*, 7900. [[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.