International Journal of Data and Network Science 10 (2026) 577–596

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Firm valuation using accounting-based capital structure and cash holdings: An explainable machine learning approach

Gihan M. Ali[a,b*]

[a]*Department of Accounting, College of Business Administration in Hawtat Bani Tamim, Prince Sattam bin Abdulaziz University, Saudi Arabia*
[b]*Department of Accounting, Faculty of Commerce, Damanhour University, Damanhour 22514, Egypt*

C H R O N I C L E

Received October 12, 2025
Received in revised format December 27, 2025
Accepted February 3 2026
Available online February 3 2026

Keywords:
Explainable Machine Learning
Super Learner
SHAP analysis
Cash holdings
Capital structure
COVID-19
Firm valuation
Emerging markets

A B S T R A C T

This study investigates the impact of cash holdings and capital structure on firm valuation in Egypt's emerging market, examining how COVID-19 altered investor perceptions. The research employs explainable machine learning (ML) to uncover non-linear financial thresholds that traditional valuation models overlook. Egyptian listed firms from 2015 to 2022 are analyzed using a Super Learner ensemble (Extremely Randomized Trees, Extreme Gradient Boosting, and a Linear Regression meta-learner) alongside SHapley Additive exPlanations (SHAP) and partial dependence analysis, with the Super Learner's performance compared against conventional methods in assessing financial policy effects on Tobin's Q. Three key findings emerge: (1) Leverage exhibits a non-linear relationship with valuation, where extreme levels (LEV > 1.2) unexpectedly enhance firm value, challenging trade-off theory; (2) Cash holdings demonstrate threshold effects, with optimal value at ~40% of assets and sharply increasing marginal benefits beyond this point; and (3) COVID-19 amplified these dynamics, elevating the liquidity premium while penalizing excessive debt. The Super Learner significantly outperformed traditional statistical and ML models ($R^2 = 0.572$ vs. 0.19-0.47). Practical implications suggest that investors and managers in emerging markets should adopt dynamic cash-debt optimization to avoid undervaluation, while policymakers can use ML-driven thresholds to design crisis-responsive regulations. This study contributes to the literature by (1) identifying non-linear thresholds that extend trade-off and pecking order theories, (2) introducing explainable ML to valuation research to balance accuracy and interpretability, and (3) providing novel evidence of COVID-19's structural impact on investor behavior in emerging economies.

1. Introduction

Firm valuation serves as a fundamental pillar of corporate finance, guiding investment decisions, mergers and acquisitions, and performance assessments (Basci, 2019; Brennan & Schwartz, 1984). Traditional valuation methods, such as discounted cash flow analysis, often rely on linear assumptions that fail to capture the intricate, non-linear relationships embedded in financial data. Recent advancements in machine learning (ML) have revolutionized this field by enabling data-driven predictions that surpass the accuracy of conventional techniques and even human analysts (Geertsema & Lu, 2019; Geertsema & Lu, 2023). Empirical evidence demonstrates ML's superiority: Koklev (2022) achieved an R^2 of 86.7% in predicting market capitalization using gradient-boosted decision trees, while Zhang et al. (2023) showed that ML models outperform traditional venture capital valuation methods by prioritizing investor-related features over patents. Despite these strides, significant gaps remain in understanding how firm-specific financial policies, particularly capital structure and cash holdings, interact to shape valuation, especially in emerging markets where institutional and macroeconomic volatility amplify financial constraints. The relationship between capital structure and firm value has long been debated, with theories ranging from Modigliani and Miller's (1958) irrelevance proposition to trade-off and pecking order models. Empirical studies suggest that moderate debt levels enhance firm value through tax shields, while excessive leverage erodes it due to financial distress risks (Mazumdar &

* Corresponding author
E-mail address: g.ali@psau.edu.sa (G. M. Ali)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2026.2.001

Mara, 2024). Similarly, cash holdings exhibit a non-linear impact: they provide critical liquidity during crises (Chang et al., 2017) but may also invite agency costs (Jensen & Meckling, 1976). Prior research identifies an optimal cash threshold, often around 20–30% of assets, that maximizes firm value, with deviations reducing worth (Habib et al., 2021; Martínez-Sola et al., 2013). However, these findings primarily derive from developed markets or select emerging economies like China, leaving regions such as the MENA underexplored. Moreover, the COVID-19 pandemic's disruptive effects on investor perceptions of financial policies remain poorly documented, despite evidence that crises reshape the relative value of liquidity and leverage.

This study addresses these gaps by analyzing Egyptian listed firms, a setting characterized by thin capital markets and macroeconomic instability, using a novel Super Learner ensemble that integrates Extremely Randomized Trees (Extra Trees), and Extreme Gradient Boosting (XGBoost) with a Linear Regression (LR) meta-learner. The analysis reveals three critical insights: first, leverage demonstrates a non-linear relationship with firm value (proxied by Tobin's Q), where extreme levels (LEV > 1.2) unexpectedly enhance valuation, challenging conventional trade-off theory; second, cash holdings exhibit a positive yet non-monotonic impact, with Tobin's Q showing gradual improvement until cash reserves surpass 40% of assets, at which point the valuation benefit accelerates markedly; third, the COVID-19 pandemic intensified these relationships, elevating the premium on cash reserves while simultaneously increasing the perceived risk associated with high leverage positions. Methodologically, the proposed Super Learner achieves superior predictive performance ($R^2 = 0.572$) compared to traditional models and other ensemble techniques ($R^2 = 0.08$–0.47). Through the application of SHapley Additive exPlanations (SHAP) and partial dependence plots, interpretability is enhanced, revealing that leverage, firm size, profitability, and cash holdings dominate feature importance. The research makes four key contributions. First, capital structure theories are extended through the identification of threshold effects in leverage and cash holdings. Second, novel insights are provided into an understudied emerging market, where financial constraints intensify valuation sensitivities. Third, empirical evidence is presented on how COVID-19 reshaped investor preferences, offering practical guidance for crisis-era financial management. Fourth, methodological rigor in valuation research is advanced through an interpretable ML framework that balances accuracy with transparency.

The paper is structured as follows: Section 2 presents theoretical framework; Section 3 reviews empirical literature; Section 4 details methodology; Section 5 presents results and discusses implications; and Section 6 concludes.

2. Theoretical Background

The relationship between capital structure and firm value has been widely examined in corporate finance through multiple theoretical lenses (Fig. 1). While Modigliani and Miller's propositions establish a foundational benchmark, real-world factors, including taxes, bankruptcy costs, and information asymmetries, drive firms to adopt strategies consistent with the Pecking Order, Trade-Off, and Agency theories. Behavioral perspectives like Signaling Theory and Market Timing Theory further influence financing decisions. Together, these frameworks explain variations in capital structure across firms and their impact on firm value. Similarly, theories of corporate cash holdings fall into two main categories: capital structure theories (Pecking Order and Trade-Off) and Agency Theory.

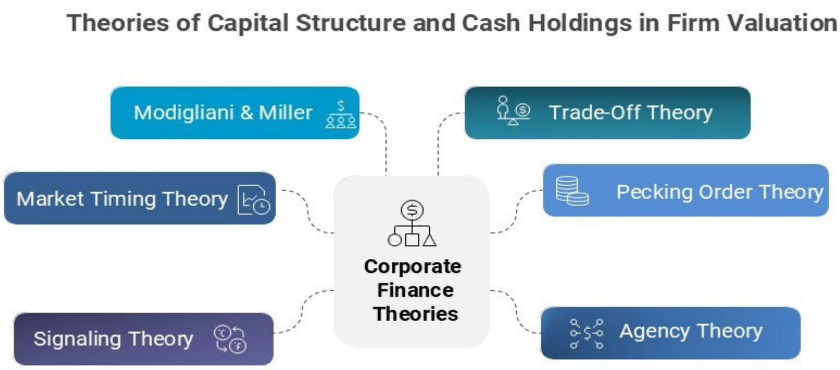

Fig. 1. Theories of Capital Structure and Cash Holdings in Firm Valuation.

2.1 Modigliani & Miller (MM)

The foundational work by Modigliani and Miller (1958) established that, under perfect capital market conditions, assuming no taxes, bankruptcy costs, or asymmetric information, a firm's value is independent of its capital structure (MM Proposition I). However, when corporate taxes are introduced, MM revised their theory in 1963, showing that debt financing becomes advantageous due to the tax deductibility of interest payments, thereby increasing firm value (MM Proposition II with taxes). Miller further extended this by incorporating personal taxes, arguing that the net tax benefit of debt depends on the relative tax rates of investors and corporations (Aljamaan, 2018; Gitman & Zutter, 2012).

2.2 Pecking Order Theory

The Pecking Order Theory, developed by Donaldson (1961) and formalized by Myers and Majluf (1984), presents a hierarchical approach to corporate financing that contrasts with Modigliani and Miller's capital structure irrelevance proposition. According to this framework, firms exhibit a distinct preference for internal financing sources - particularly retained earnings and cash reserves - before resorting to external funding. This behavior stems from information asymmetry concerns, where equity issuance may signal potential overvaluation to markets, potentially triggering adverse investor reactions (Brigham, 2016; Myers & Majluf, 1984). This financing hierarchy has profound implications for corporate cash management. Cash holdings serve as a financial buffer that allows firms to avoid costly external financing, particularly for undervalued firms facing information asymmetries. The theory highlights how growth opportunities and information gaps influence cash holding decisions, as companies balance the need for financial flexibility against financing costs. These dynamics create a direct link between capital structure choices and liquidity management in corporate financial strategy (Weidemann, 2018).

2.3 Trade-Off Theory

The Trade-Off Theory presents a framework where firms strategically balance competing financial considerations to optimize their capital structure and cash holdings. Regarding capital structure, the theory suggests companies weigh the benefits of debt financing, particularly interest tax shields, against potential costs including financial distress and bankruptcy risks (Paramasivan, 2009). This balancing act aims to minimize the weighted average cost of capital (WACC) while maximizing firm value. When applied to cash management, the Trade-Off Theory similarly involves a cost-benefit analysis. Firms determine optimal cash levels by evaluating the advantages of maintaining liquidity - such as avoiding transaction costs of external financing and preventing underinvestment - against the disadvantages, which include opportunity costs of idle funds and potential agency problems (Bates et al., 2009; Opler et al., 1999). This dual application of the theory demonstrates its versatility in explaining both capital structure decisions and corporate liquidity management.

2.4 Agency Theory

Agency Theory provides critical insights into corporate financial decisions by examining conflicts of interest between managers, shareholders, and creditors. At its core, the theory identifies how debt serves as a disciplinary mechanism by limiting free cash flow available for managerial misuse (Jensen & Meckling, 1976; Olokoyo, 2013). However, excessive leverage can impose restrictive covenants and increase financial risk, demonstrating the delicate balance in capital structure decisions. When applied to cash management, Agency Theory reveals several competing perspectives. The flexibility hypothesis argues that managers hold cash to maintain financial flexibility and avoid external discipline, with cash being valued positively by the market as it mitigates underinvestment risks, especially in firms with volatile cash flows or high growth opportunities (Acharya et al., 2007; Jensen, 1986). Conversely, the spending hypothesis highlights that weakly controlled managers may overinvest or spend cash on value-destroying projects, leading to higher cash holdings but lower firm value (Jensen & Meckling, 1976). The shareholder power hypothesis suggests that strong shareholder protection aligns interests, allowing higher cash holdings without fear of expropriation, making cash valuable when governance is strong (Harford et al., 2008; Kuan et al., 2011). The costly contracting theory notes that creditors may enforce higher cash holdings in risky firms through covenants, reducing the value of cash due to restricted investment freedom (Liu & Mauer, 2011). Lastly, the defense against hostile takeovers motive explains how managers hoard cash to deter takeovers, which can entrench management but also protect shareholder interests in certain contexts (Faleye, 2004). These agency perspectives collectively demonstrate how cash holdings reflect underlying governance quality and strategic priorities.

2.5 Signaling Theory

The Signaling Theory suggests that financing decisions convey insider information to investors, issuing debt signals confidence in future cash flows, while issuing equity may indicate overvaluation (Ross, 2009).

2.6 Market Timing Theory

The Market Timing Theory (Baker & Wurgler, 2002) argues that firms adjust their capital structure opportunistically, issuing equity when stock prices are high and debt when they are low, rather than adhering to a fixed target ratio.

3. Literature Review

3.1 The Impact of Capital Structure on Firm Value

Capital structure, the composition of a firm's financing sources, is crucial for maximizing firm value and ensuring adequate resources for investments (Datta et al., 2013). The success of any firm largely depends on its capital structure choices. For small and medium enterprises (SMEs), capital structure decisions are particularly important, as they often rely more on internal funds and face liquidity constraints compared to larger firms (Dogra & Gupta, 2009). Various theories have emerged to explain capital structure decisions, including the trade-off between cheaper debt and increased risk, as well as the tax benefits of debt (Cerkovskis et al., 2022). Understanding these theories and the impact on firm value is crucial for firms, especially in emerging and underdeveloped economies, to identify the most appropriate blend of capital (Datta et al., 2013). The relationship between capital structure and firm value has been extensively studied across various industries and geographical contexts, with researchers employing diverse methodologies to examine this complex relationship. Recent empirical evidence demonstrates that moderate debt levels generally enhance firm value, while excessive leverage tends to have detrimental effects. Mazumdar and Mara (2024) conducted a comprehensive cross-sectional quantitative analysis examining the relationship between capital structure and firm value across multiple industries. Their study, which measured firm value through market capitalization and Tobin's Q, found that moderate debt levels provide tax benefits that increase firm value, but beyond a certain threshold, the costs of financial instability outweigh these benefits. Similarly, Uzliawati et al. (2018) focused specifically on Indonesian manufacturing firms, analyzing 101 companies from 2012 to 2015 through correlation analysis. Their results indicated positive relationships between debt-to-equity ratio (DER) and long-term debt-to-asset ratio (LDAR) with firm value, while showing a negative correlation for long-term debt-to-equity ratio (LDER). These findings collectively suggest that the composition and level of debt significantly influence firm valuation.

The impact of capital structure appears to vary considerably across different industries and economic contexts. Aggarwal et al. (2017) provided compelling evidence from the Indian hospitality sector, where their panel data analysis revealed that high leverage negatively impacts firm value, directly contradicting the Modigliani-Miller theorem. Their study employed pooled OLS, fixed effects, and random effects models to analyze both firm-level variables (leverage, size, profitability) and macroeconomic factors (GDP growth, inflation). In contrast, research in African markets presents different patterns. Antwi et al. (2012) conducted OLS regression analysis on Ghanaian firms and found long-term debt to be the primary determinant of firm value, while RONIC and Amadi (2021), using cross-sectional regression and Pearson correlation on Nigerian firms, observed positive effects of short-term and total debt but negative impacts from long-term debt. These regional variations highlight the importance of contextual factors in capital structure decisions.

Several studies have identified additional determinants that influence the capital structure-firm value relationship. Chowdhury and Chowdhury (2010) employed a Panel Corrected Standard Error model to analyze Bangladeshi firms, finding that while leverage negatively affected return on assets (ROA) due to bankruptcy risks, it positively influenced return on equity (ROE) by increasing shareholder returns. These findings suggest that managerial characteristics and risk profiles mediate the capital structure-firm value relationship. The literature also presents some contradictory findings that warrant consideration. Bui et al. (2023) conducted an extensive panel data analysis of Vietnamese firms using multiple estimation techniques (OLS, FEM, REM, GLS), revealing that while overall debt ratios positively impacted firm value, both short-term and long-term debt ratios showed negative effects. Akash et al. (2023) similarly found mixed results in their balanced panel analysis, with capital structure negatively affecting return on equity (ROE) but positively influencing operating profit margin (OPM). These contradictory outcomes suggest that the relationship between capital structure and firm value may be more complex than linear models can capture, potentially requiring threshold or non-linear approaches as demonstrated by Cheng et al. (2010) in their panel threshold regression analysis of Chinese firms.

In conclusion, the body of research consistently demonstrates that capital structure significantly impacts firm value, though the nature of this relationship varies by context, industry, and methodological approach. While existing research has significantly advanced our understanding of the relationship between capital structure and firm value, several critical gaps remain that limit the generalizability and practical applicability of findings. A key methodological limitation is the overreliance on linear models in many studies (e.g., RONIC & Amadi, 2021), which employ OLS or panel regressions that assume simple linear relationships, despite evidence from Cheng et al. (2010) demonstrating more complex, inverted-U-shaped relationships that suggest non-linear or threshold models would better capture reality. Furthermore, while some studies have examined emerging markets like Nigeria (RONIC & Amadi, 2021) and Bangladesh (Chowdhury & Chowdhury, 2010), there remains a notable lack of research on frontier markets with underdeveloped capital markets, which often face unique financial constraints and institutional challenges. To address these gaps, the current study applies ML techniques to model non-linear relationships in the emerging capital market of Egypt, thereby providing more nuanced insights into how capital structure affects firm value in understudied market contexts while overcoming the limitations of traditional linear approaches.

Based on the above, the following hypothesis is proposed:

H$_1$: *Incorporating capital structure (measured by the leverage ratio) into the proposed Super Learner model significantly increases its ability to predict firm value.*

3.2 The Impact of Cash Holdings on Firm Value

Cash holdings play a vital role in corporate financial management, serving multiple strategic purposes for firms. Maintaining adequate cash reserves ensures liquidity to meet short-term obligations and operational needs (Almeida et al., 2014). Companies hold cash as a precautionary buffer against financial distress and unexpected cash flow shocks, particularly important during economic downturns (Bates et al., 2009; Opler et al., 1999). Furthermore, cash reserves enable firms to seize profitable investment opportunities without relying on costly external financing (Ferreira & Vilela, 2004). However, excessive cash holdings may lead to agency problems, as managers might misuse funds for value-destroying projects or personal benefits, highlighting the importance of strong governance mechanisms (Dittmar & Mahrt-Smith, 2007; Jensen, 1986). The relationship between cash holdings and firm value has been extensively studied, with most research finding a non-linear, inverted U-shaped relationship. This suggests firms have an optimal cash level that maximizes value, where both excessive and insufficient cash reserves can be detrimental. Martínez-Sola et al. (2013) established this concave relationship using Generalized Method of Moments (GMM) on U.S. industrial firms, finding that deviations from the optimal cash level reduced firm value as measured by Tobin's Q. Similar findings emerged in emerging markets: Habib et al. (2021) applied fixed effects and GMM estimators to Chinese manufacturers, showing that managerial optimism influenced this relationship, while Nhan & Ha (2016) found an optimal cash threshold in Vietnamese firms, though state ownership had limited impact. The non-linear effect appears robust across contexts, with Anton & Nucu (2019) identifying precise optimal cash levels (27.06% of assets for Polish firms) using dynamic panel models. However, some studies emphasize how agency problems can make cash holdings value-destructive. Lee & Lee (2009) demonstrated through OLS and fixed effects regressions that excessive cash reduces firm value when corporate governance is weak, particularly in firms with large boards and low independence. This aligns with Luo & Hachiya (2005) findings from Japanese firms, where insider ownership and bank relationships moderated cash's negative effects. Cross-country analysis by Pinkowitz et al. (2006) further showed cash is valued less in countries with poor investor protection, highlighting governance's crucial role in determining whether cash creates or destroys value. The value impact of cash holdings also varies with firm circumstances. Chang et al. (2017) revealed through difference-in-differences models that cash became more valuable during the 2008 crisis, especially for constrained firms with strong governance. Similarly, Lau & Block (2012) found founder-managed firms maintained higher, more valuable cash reserves than family firms. These context-dependent effects suggest cash's valuation depends on both external factors like economic conditions and internal factors like ownership structure, with optimal cash policies varying accordingly.

Collectively, the literature demonstrates that cash holdings influence firm value through multiple channels, providing liquidity benefits up to a certain threshold but potentially enabling agency problems beyond that point. The precise nature of this relationship depends on governance quality, financial constraints, and the macroeconomic environment. Empirical research reveals several key findings. First, studies consistently identify a non-linear relationship between cash holdings and firm value, where deviations from optimal levels, whether excess or insufficient, reduce value (Habib et al., 2021; Nhan & Ha, 2016). Second, the value of cash is highly context-dependent: it is more valuable for financially constrained firms with limited access to external capital (Almeida et al., 2014), whereas weak corporate governance exacerbates the agency costs of holding excess cash (Pinkowitz et al., 2006). Third, during economic crises, such as the 2008 financial crisis, cash becomes particularly valuable as a hedge against uncertainty (Chang et al., 2017). These findings underscore the importance of adaptive cash management strategies that account for firm-specific conditions and macroeconomic dynamics.

Despite extensive research on cash holdings and firm value, several critical gaps persist in the literature. While studies have extensively examined developed markets (e.g., U.S., Japan) and select emerging economies (e.g., China, Vietnam), there remains insufficient focus on MENA markets, where unique financial constraints and institutional weaknesses may significantly alter cash-value dynamics. Furthermore, most studies aggregate firms across sectors, potentially masking important industry-specific nuances that could influence optimal cash policies. Another notable gap exists in understanding how extreme events affect cash valuation - while some research (Chang et al., 2017) has explored financial crises, the impact of unprecedented events like the COVID-19 pandemic on market perceptions of cash holdings remains underexamined, particularly regarding whether the crisis fundamentally changed how investors value corporate liquidity buffers. Finally, while methodological choices like GMM or fixed effects helping isolate these complex interactions, more research is needed on how ML models can better capture these nonlinear effects. These gaps present valuable opportunities for future research to provide more comprehensive insights into cash management strategies across different economic contexts and extraordinary circumstances.

Based on the above, the following hypothesis is proposed:

H$_2$: *Incorporating Cash Holdings into the proposed Super Learner model significantly increases its ability to predict firm value.*

3.3 Machine Learning Approaches to Firm Valuation

Recent advances in machine learning (ML) have transformed traditional firm valuation methods by enabling more accurate, data-driven predictions. Where conventional approaches like discounted cash flow analysis rely on linear assumptions, ML techniques capture complex, non-linear relationships in financial data. This review synthesizes key findings from contemporary research on ML applications in firm valuation, examining methodological innovations, empirical results, and remaining challenges in the field. The superiority of ML models over traditional valuation methods has been well-documented across multiple studies (Ali et al., 2022). P. G. Geertsema & Lu (2019) demonstrated that ML algorithms using only historical accounting data achieved a median absolute percentage error of 17.2% in firm valuation, outperforming both finance students and professional analysts. Their subsequent work (P. Geertsema & Lu, 2023) revealed that decision-tree-based ML models could reduce valuation errors by 5.6 to 31.4 percentage points compared to traditional multiples-based approaches while identifying key drivers consistent with discounted cash flow theory. Similarly, Koklev (2022) found Gradient Boosting Decision Trees (GBDT) achieved remarkable explanatory power (R^2 = 86.7%) in predicting market capitalization, substantially outperforming conventional econometric models. These findings collectively suggest that ML not only enhances valuation accuracy but also provides deeper insights into the determinants of firm value.

Various ML techniques have proven effective for different valuation contexts. Tree-based methods like Random Forest (RF) and XGBoost have shown particular promise, with Chen et al. (2024) demonstrating that XGBoost models using 19 fundamental signals outperformed linear models by 27% in measuring firm quality. Neural networks have excelled in specific applications, as Zhang et al. (2020) showed Artificial Neural Networks (ANNs) outperformed other models by 18-19% in valuing energy firms. For early-stage ventures, R. Zhang et al. (2023) found neural networks optimized with differential evolution algorithms effectively predicted entrepreneurial firm valuations, surprisingly revealing that VC investor syndicate size was more influential than patent portfolios. The interpretability challenge of complex ML models has been addressed through techniques like SHAP values and permutation importance, with Koklev (2022) using these methods to identify comprehensive income as the most significant predictor of market capitalization. The application of ML in valuation extends across diverse industries and market contexts. Sector-specific analyses have yielded important insights, such as ANNs demonstrating superior performance in oil and power company valuations (Zhang et al., 2020) and Bayesian Additive Regression Trees (BART) showing significant accuracy (R^2 = 0.922-0.934) in predicting food processing firm performance (Saha et al., 2023). For unlisted companies, Vayas-Ortega et al. (2020) identified Bagging Trees, Support Vector Machine Regression, and Gaussian Process Regression as particularly effective when incorporating both DCF-based variables and industry-specific factors. Cross-country studies like Cakici et al. (2023) have demonstrated ML's global applicability, with models successfully predicting returns across 46 markets using 148 firm characteristics, though performance varied based on market size and idiosyncratic risk factors. Moreover, ML-based valuation models have demonstrated significant practical utility in investment decision-making. Trading strategies derived from ML valuations have generated substantial abnormal returns (Geertsema & Lu, 2023), while ML earnings forecasts have consistently outperformed analyst consensus estimates (Cao & You, 2024). The economic benefits extend to portfolio construction, with Chen et al. (2024) showing that ML-derived firm quality measures could enhance value investing strategies. These applications underscore the real-world impact of ML in financial analysis and asset pricing.

In conclusion, ML has fundamentally enhanced firm valuation through superior predictive accuracy and richer insights into value drivers. From tree-based methods to neural networks, diverse ML techniques have demonstrated effectiveness across industries and market conditions. Despite these advancements, several research gaps remain. The trade-off between model complexity and interpretability persists, suggesting need for further development of explainable ML techniques in valuation contexts. Emerging markets also represent understudied areas where ML applications could prove particularly valuable.4 Proposed System

4 Proposed Framework

4.1 General Context

This study presents a framework for predicting firm value using an explainable ML. The system operates in two phases: training and testing. As shown in Fig. 2, the methodology begins with data preprocessing of corporate annual report data, followed by feature extraction and validation. The study employs a Super Learner ensemble model combining Extra Trees and XGBoost algorithms, with Linear Regression as the meta-learner to enhance prediction accuracy. The model's outputs are interpreted through feature importance analysis and SHapley Additive exPlanations (SHAP) values.

4.2 Phase 1: Data Preparation

Step 1: Dataset Overview

This study analyzes non-financial companies listed on the Egyptian Stock Exchange from 2016 to 2022. Financial firms are excluded due to their unique financial structures, which differ significantly from non-financial entities. Additionally, companies presenting financial statements in foreign currencies are removed from the sample. The final dataset includes 104 firms, producing a total of 728 firm-year observations for the period examined.

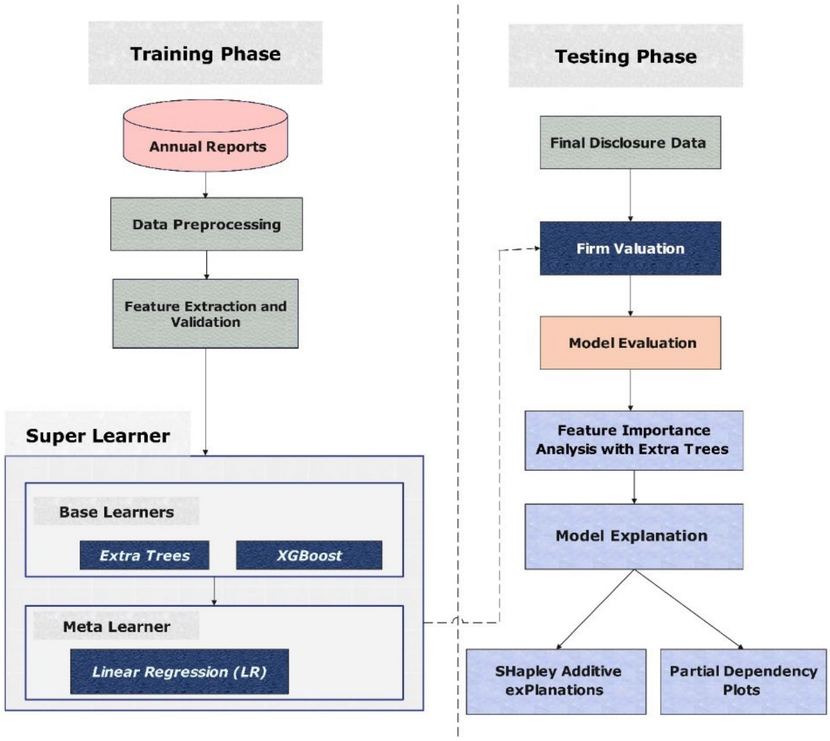

Fig. 2 The Proposed System.

Step 2: Data Preprocessing

Corporate data often contains missing entries, inaccuracies, and inconsistencies that can compromise model outcomes. As a result, data preprocessing is essential for improving the performance and accuracy of ML models. Missing data is handled using median imputation, and outliers are addressed through Winsorization by capping values at the 5th and 95th percentiles (Kwak & Kim, 2017). These steps yield a final dataset of 699 observations.

Step 3: Feature Extraction and Validation

This phase focuses on selecting the most impactful variables that influence firm valuation, measured by Tobin's Q. Drawing on prior research, 13 key features are identified to support accurate prediction. These variables are summarized in Table 1.

Table 1
Variables included in the Super Learner for Firm Valuation

Variables	Symbol	Measure
Firm Value	Tobin's Q	The ratio of a firm's total liabilities plus market capitalization to its total assets; The market capitalization is computed as the product of the year-end closing stock price and the quantity of shares outstanding.
Financial leverage	LEV	Debt ratio (Total liabilities / Total assets)
Cash Holdings	CASH	Cash ratio (Cash and equivalents / Total assets)
COVID-19	COVID	An indicator variable equal to 1 for the pandemic years (2019-2020) and 0 otherwise
Financial performance	ROA	Return on assets (Net income / Total assets)
Net Capital	NETCAP	Working capital ratio (Working capital / Total assets)
Capital Expenditures	CAPEX	Ratio of capital expenditures to total assets in a given fiscal year
Trade Ratio	TRADE	Accounts payable turnover (% change in accounts payable / Cost of goods sold)
Firm size	FSIZE	Natural logarithm of total assets
Dividend Payment ratio	DIV	The proportion of declared and distributed dividends relative to earnings before interest and taxes
Dividend Payment	DIVPAY	A binary indicator equal to 1 if the firm paid dividends and 0 otherwise
Firm Age	AGE	Years since company incorporation
Industry Effects	INDUSTRY	Industry fixed effects
Year Effects	YEAR	Year fixed effects (2016-2022)

To evaluate potential multicollinearity, I examined pairwise correlations and calculated variance inflation factors (VIF). Table 2 displays the correlation matrix for the independent variables. The correlation matrix reveals no evidence of severe multicollinearity, as all coefficients remain below the conventional threshold of 0.7. The strongest correlation appears between DIVPAY and DIV (0.55), which, although notable, does not reach levels that would typically bias regression estimates. Other moderately correlated pairs include LEV and NETCAP (-0.42) and CASH and ROA (0.41), all within acceptable ranges. While these relationships may indicate some conceptual overlap, particularly among financial variables, they are unlikely to substantially affect estimation results. For robustness, I computed VIF statistics (James et al., 2013), with all values well below the critical threshold of 10 (Fig. 3), confirming that multicollinearity does not pose significant concerns for the subsequent analysis.

Table 2
The Correlation Matrix

	FSIZE	LEV	CAPEX	CASH	ROA	NETCAP	DIVPAY	DIV	TRADE	AGE
FSIZE	1.0***	0.13***	-0.09*	-0.04	0.12**	-0.2***	0.29***	0.14***	0.09*	0.11**
LEV	0.13***	1.0***	0.13***	0.02	-0.21***	-0.42***	-0.11**	-0.16***	0.09*	0.21***
CAPEX	-0.09*	0.13***	1.0***	-0.02	-0.09*	-0.29***	-0.05	0.01	-0.02	0.05
CASH	-0.04	0.02	-0.02	1.0***	0.41***	0.31***	0.25***	0.18***	-0.08*	-0.04
ROA	0.12**	-0.21***	-0.09*	0.41***	1.0***	0.38***	0.39***	0.23***	-0.08*	-0.07
NETCAP	-0.2***	-0.42***	-0.29***	0.31***	0.38***	1.0***	0.17***	0.12**	-0.06	-0.08*
DIVPAY	0.29***	-0.11**	-0.05	0.25***	0.39***	0.17***	1.0***	0.55***	-0.04	-0.07
DIV	0.14***	-0.16***	0.01	0.18***	0.23***	0.12**	0.55***	1.0***	-0.08*	-0.08*
TRADE	0.09*	0.09*	-0.02	-0.08*	-0.08*	-0.06	-0.04	-0.08*	1.0***	-0.01
AGE	0.11**	0.21***	0.05	-0.04	-0.07	-0.08*	-0.07	-0.08*	-0.01	1.0***

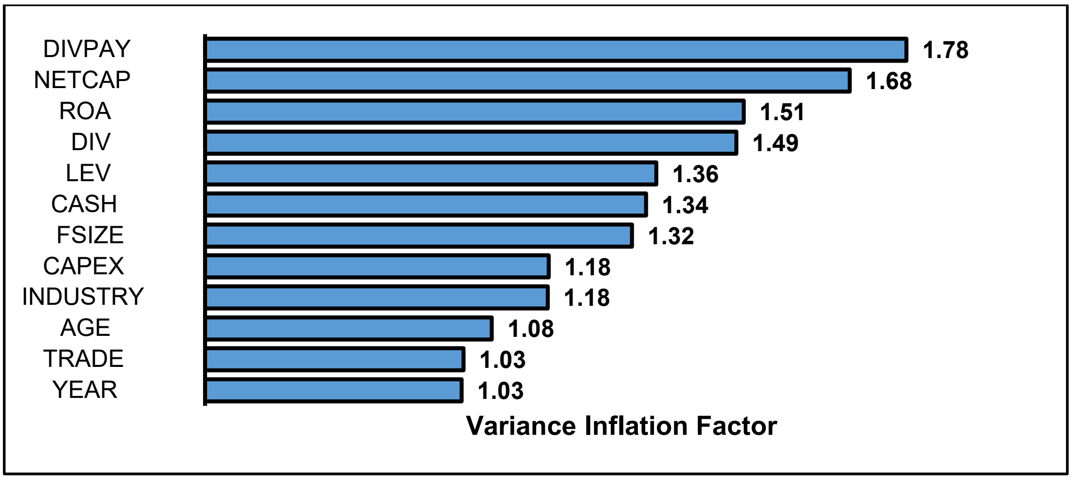

Fig. 3. VIF for the Study Variables.

4.3 Phase 2: Model Development

Step 1: Super Learner Model Specifics

This study employs ensemble learning to construct a predictive model, leveraging the principle that combining multiple weak learners can yield greater accuracy and robustness than individual models by compensating for their respective limitations while capitalizing on their collective strengths (Mohammed & Kora, 2023; Thabet et al., 2024). Ensemble methods are broadly classified into two categories: homogeneous and heterogeneous ensembles. Homogeneous ensembles employ a single algorithm across different data subsets, implemented through either (1) bagging techniques such as RF and Extra Trees, which train models in parallel, or (2) boosting algorithms like Adaptive Boosting (AdaBoost), Gradient Boosting Regression Trees (GBR), Light Gradient-Boosting Machine (LightGBM), Categorical Gradient Boosting (CatBoost), and XGBoost, which sequentially train models to iteratively correct errors. Conversely, heterogeneous ensembles integrate diverse ML models to improve predictive accuracy and generalization, employing either voting methods (majority, average, or performance-weighted aggregation) or meta-learning strategies where a secondary meta-learner optimally combines base model outputs into a stacked ensemble, thereby enhancing performance in complex regression and classification tasks.

The Super Learner, introduced by Van der Laan et al. (2007), is a stacking-based ensemble algorithm that overcomes traditional ML limitations through optimal combination of base models via meta-learning, achieving asymptotically optimal performance (Dey & Mathur, 2023; Naimi & Balzer, 2018). Its meta-learner dynamically weights predictions from diverse algorithms while excluding poorly performing models, maximizing accuracy and robustness while reducing overfitting (Balagopal et al., 2025). This approach has proven especially effective for complex prediction tasks where no single model dominates, demonstrating superior performance across diverse domains, including spatial prediction, clay compression index, and bank efficiency (Davies & Van Der Laan, 2016; Díaz & Spagnoli, 2024; Thabet et al., 2024). By leveraging each algorithm's strengths while mitigating weaknesses, the Super Learner consistently outperforms both individual models and traditional ensembles.

The Super Learner algorithm, as illustrated in Fig. 4, implements a sophisticated two-tiered architecture to construct an optimal predictive model. To prevent overfitting, the corporate data was divided into a training set (80%) and a validation set (20%). In the initial phase, base learners (Extra Trees and XGBoost in this study) are trained on the complete training dataset. The algorithm employs a k-fold cross-validation scheme, where the training data is systematically partitioned - with each fold serving iteratively as a testing fold while the remaining data trains the base models. This process generates out-of-sample predictions (denoted as Z_1, Z_2, Z_3 in the figure) that are aggregated into a stacked prediction matrix, while validation data predictions (V_1, V_2, V_3) are computed in parallel. These cross-validated outputs serve as training data for the meta-learner, which in this implementation is a linear regression model. The linear regression meta-learner determines the optimal combination weights by minimizing the prediction error on the validation outputs, leveraging its inherent properties of interpretability and computational efficiency (Thabet et al., 2024; Van der Laan et al., 2007).

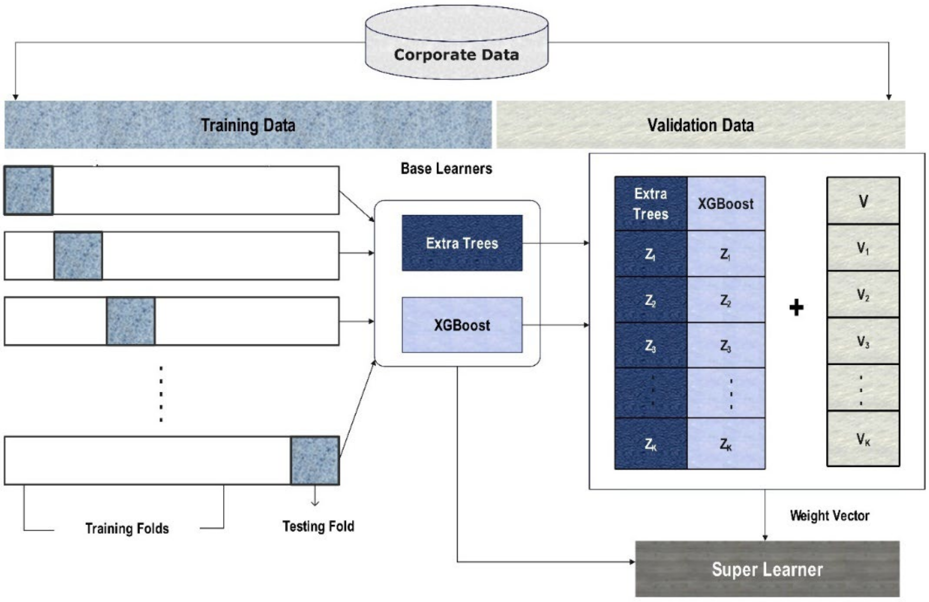

Fig. 4. Structure of Super Learner Ensemble Model with Two Base Learners: Extra Trees and XGBoost.

The final Super Learner ensemble is formed by applying these optimized linear combination weights to the base learners when trained on the full dataset. This architecture, particularly through the use of linear regression as the meta-learner, ensures robust performance by: (1) providing a principled framework for weighting base learner contributions, (2) maintaining model interpretability through linear coefficients, and (3) achieving computational efficiency in the combination phase (Polley & Laan, 2010). The figure's tabular representation effectively captures this data flow, highlighting the transformation from base learner predictions (Z_i) through to the linearly-combined validation outputs (V_i) that inform the final ensemble construction.

Step 2: Super Learner Development

The base learners for the Super Learner ensemble were selected through a comprehensive evaluation process employing 10-fold cross-validation to assess the performance of multiple candidate algorithms. Table 3 ranks their performance based on R^2, with Extra Trees, CatBoost, LightGBM, and XGBoost Regressors emerging as the top-performing models. The Super Learner framework was systematically evaluated using LR as the meta-learner to determine the optimal combination of base algorithms. As shown in Table 4, the Extra Trees & XGBoost configuration achieved superior performance (R^2=0.572) compared to alternative ensemble structures while maintaining model parsimony. The comparative analysis examined multiple configurations, including three-model ensembles (Extra Trees, CatBoost, & XGBoost: R^2=0.562; Extra Trees, CatBoost, & LightGBM: R^2=0.538), four-model combinations (Extra Trees, CatBoost, LightGBM, & XGBoost : R^2=0.517), and simpler two-model variants (Extra Trees & CatBoost: R^2=0.529; Extra Trees & LightGBM: R^2=0.529).

Table 3
Performance of base regression models obtained by k-fold cross validation.

Ensemble Models	R^2	Linear & ML Models	R^2
Extra Trees	0.47	Ridge	0.19
CatBoost	0.45	LR	0.19
LightGBM	0.43	Bayesian Ridge	0.19
XGBoost	0.41	LASSO	0.19
RF	0.38	KNN	0.18
GBR	0.37	ANN	0.17
Bagging Regressor	0.32	SVR-Linear	0.08

The experimental results reveal three significant advantages of the Extra Trees & XGBoost configuration. First, it demonstrates an 8.2% performance improvement over alternative two-model combinations while maintaining greater computational efficiency than more complex ensembles. Second, the selected configuration achieves a 22% enhancement over the best-performing individual base learner. Third, the analysis uncovered a unique synergistic effect between XGBoost's gradient boosting approach and Extra Trees' randomization strategy, which proves particularly effective for financial prediction tasks. This complementary combination of algorithmic approaches - integrating XGBoost's sequential error correction with Extra Trees' ensemble diversity - accounts for the configuration's superior predictive performance in firm valuation.

Table 4
Performance of various Super Learner combinations

Base Learners Combinations	R^2
Extra Trees, CatBoost, LightGBM, XGBoost	0.517
Extra Trees, CatBoost, LightGBM	0.538
Extra Trees, CatBoost, XGBoost	0.562
Extra Trees, CatBoost	0.529
Extra Trees, LightGBM	0.529
Extra Trees, XGBoost	0.572

Step 3: Feature Importance Analysis

Identifying the primary drivers of firm valuation is essential for informed decision-making. This study utilizes Extra Trees's feature importance evaluation to quantify each predictor's contribution to Tobin's Q prediction, enabling stakeholders to recognize the most influential variables.

Step 4: Model Explanation

SHAP (SHapley Additive exPlanations) is a widely used framework for interpreting ML models by quantifying the influence of each feature on model predictions. Based on cooperative game theory, it uses Shapley values to determine how much each feature contributes to a specific prediction. SHAP represents the relationship between features and predictions as a weighted combination of binary inputs, with values calculated through an exhaustive analysis of all possible feature subsets and their impact on the output. This method is particularly effective for structured data with a moderate number of variables, supporting both global insights (via average absolute SHAP values) and local, instance-level interpretations.

In this study, TreeSHAP, a version of SHAP tailored for tree-based algorithms is used. Unlike traditional feature importance methods, such as permutation importance, which assess relevance by measuring the drop in performance when features are removed, TreeSHAP calculates exact Shapley values by exploiting the structure of decision trees. Furthermore, Partial Dependence Plots (PDPs) are used to visualize predictor effects. This dual approach provides both mathematical rigor (through Shapley values) and intuitive visualization (via PDPs), making the model's decisions transparent to all stakeholders while maintaining predictive accuracy.

5. Experimental Results and Discussion

This section begins with a descriptive overview of the variables involved in the study. It proceeds to evaluate the predictive capability of the proposed model by comparing it with several regression models using various performance metrics. Furthermore, the importance of the chosen features is analyzed to enhance decision-making. Finally, SHAP values and partial dependence plots are utilized to offer a more comprehensive interpretation of the model's predictions.

5.1 Descriptive Analysis

As shown in Table 5, the descriptive statistics provide insights into the distribution and variability of the dataset's variables (n = 699). Leverage shows moderate variation, with firms on average financing half their assets with debt (mean = 0.49). The wide range and high upper quartile (0.66) reflect diverse capital structures, from conservative to highly leveraged. Cash holdings are positively skewed; while the average is 10%, the median is lower, indicating many firms hold little cash, though some retain up to 49% of assets as cash reserves. This highlights the differing liquidity strategies across firms.

Firm Size (FSIZE) has a relatively tight range (mean = 9.03, std = 0.72), suggesting consistent firm scale. Return on Assets (ROA) ranges from negative to positive, suggesting performance disparity. Net Capital (NETCAP) also varies widely, with negative minimum values. Capital Expenditure (CAPEX) is also right-skewed, with low medians.

Trade (TRADE) displays high dispersion and extreme values, implying heterogeneity in trade exposure. Firm Age (AGE) ranges broadly from 15 to 82 years, with a mean of about 38 years. Dividend-related variables (DIVPAY and DIV) show binary and proportion-based behavior, respectively. Although 53% of firms distribute dividends (as indicated by the mean of DIVPAY), the median dividend yield (DIV) is zero, suggesting that a substantial number of firms do not pay dividends, while a minority offer relatively high yields, reaching up to 1.58. Finally, Tobin's Q (TOBINQ) has a wide range and high standard deviation (std = 1.50), highlighting substantial variability in market valuation across firms.

Table 5
Descriptive Statistics

	LEV	CASH	FSIZE	ROA	NETCAP	CAPEX	AGE	TRADE	DIV	DIVPAY	TOBINQ
count	699	699	699	699	699	699	699	699	699	699	699
mean	0.49	0.10	9.03	0.06	0.20	0.22	38.43	0.05	0.25	0.54	1.60
std	0.33	0.13	0.72	0.10	0.29	0.23	18.80	0.98	0.42	0.50	1.50
min	0.04	0.00	7.84	-0.11	-0.33	0.00	15.00	-2.87	0.00	0.00	0.36
0.25	0.25	0.02	8.47	0.00	0.01	0.03	23.00	-0.07	0.00	0.00	0.79
0.50	0.48	0.05	9.02	0.04	0.18	0.15	33.00	0.05	0.00	1.00	1.08
0.75	0.66	0.13	9.51	0.11	0.37	0.32	52.00	0.28	0.35	1.00	1.70
max	1.37	0.49	10.35	0.29	0.82	0.76	82.00	2.26	1.58	1.00	6.61

5.2 Analysis of Experimental Results

To evaluate the predictive capability of the proposed model, its performance is compared with several other algorithms. These benchmarks are grouped into four categories: (1) other heterogenous ensemble approaches, (2) homogenous ensemble approaches, (3) LR and individual ML models, and (4) a variant of the proposed model without the capital structure and cash holdings variables.

5.2.1 Performance Comparison of the Super Learner with Other Heterogenous Ensemble Approaches

This section compares the predictive performance of the Super Learner with other ensemble methods in forecasting firm value, using R^2 as the evaluation metric. As illustrated in Fig. 5, the Super Learner clearly outperforms both the Voting Ensemble (R^2: 0.451) and the Stacked Ensemble (R^2: 0.391), despite leveraging identical base models. Specifically, the Super Learner achieves improvements of 20.8% over stacking and 11.5% over voting. These results emphasize the advantages of the Super Learner's adaptive weighting strategy, which adjusts model contributions dynamically, unlike the fixed-weight approach of voting or the sequential design of stacking.

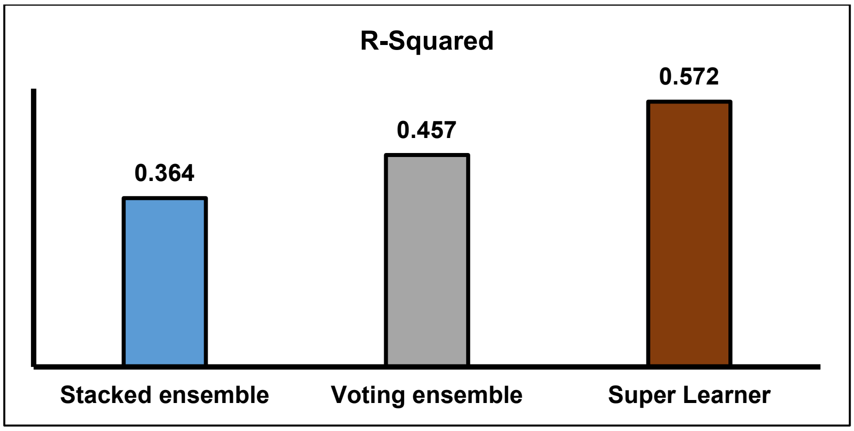

Fig. 5. Bar Plots Showing the R^2 Results for Various Heterogenous Ensemble Approaches Against the Super Learner Model for Firm Valuation.

5.2.2 Performance Comparison of the Super Learner with Homogenous Ensemble Approaches

The Super Learner's performance was benchmarked against seven established ensemble methods for firm valuation using R^2 metrics as shown in Fig. 6.

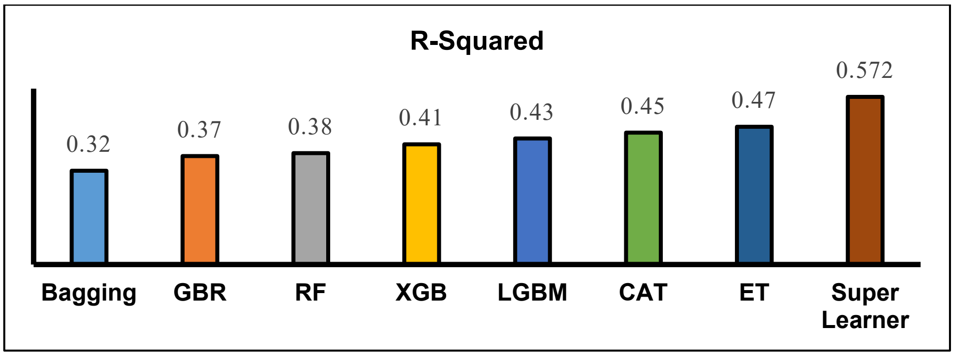

Fig. 6 Bar Plots Showing the R^2 Results for Various Homogenous Ensemble Models Against the Super Learner Model for Firm Valuation.

Among traditional ensemble approaches, Bagging (R^2=0.32) and RF (R^2=0.38) demonstrated baseline predictive capabilities, while gradient-boosted implementations including XGBoost (R^2=0.41) and LightGBM (R^2=0.43) showed incremental improvements. The strongest individual ensemble performers were CatBoost (R^2=0.45) and Extra Trees (R^2=0.47), representing the state-of-the-art in standalone ensemble techniques.

The Super Learner significantly outperformed all comparative models with an R^2 of 0.572, representing a 22% improvement over the best individual ensemble (Extra Trees) and more substantial gains of 39-79% against other variants. This performance advantage stems from the Super Learner's unique meta-learning architecture, which dynamically combines and weights predictions from diverse constituent models. Unlike homogeneous ensembles that rely on a single algorithmic approach, the Super Learner's ability to integrate multiple learning paradigms allows it to better capture complex patterns in financial data while mitigating individual model biases.

5.2.3 Comparison of the Super Learner with Linear Models and Individual ML Models

This section benchmarks Super Learner against linear and ML models for firm valuation using R^2. The Super Learner's predictive accuracy significantly outperforms both linear and nonlinear alternatives (Table 3, Fig. 7). Linear models (LR, Ridge, BR, LASSO) show limited effectiveness with R^2 values of 0.19, indicating linear assumptions fail to capture the data's complexity. Individual ML models perform worse, with SVR (R^2=0.08), ANN (R^2=0.17), and KNN (R^2=0.18) demonstrating sensitivity to hyperparameters or inherent structural limitations. The Super Learner demonstrates superior predictive performance (R^2=0.572), achieving a 201% improvement over the best individual model. This gap highlights its ability to overcome standalone models' weaknesses through meta-learning, combining diverse models to enhance predictive robustness.

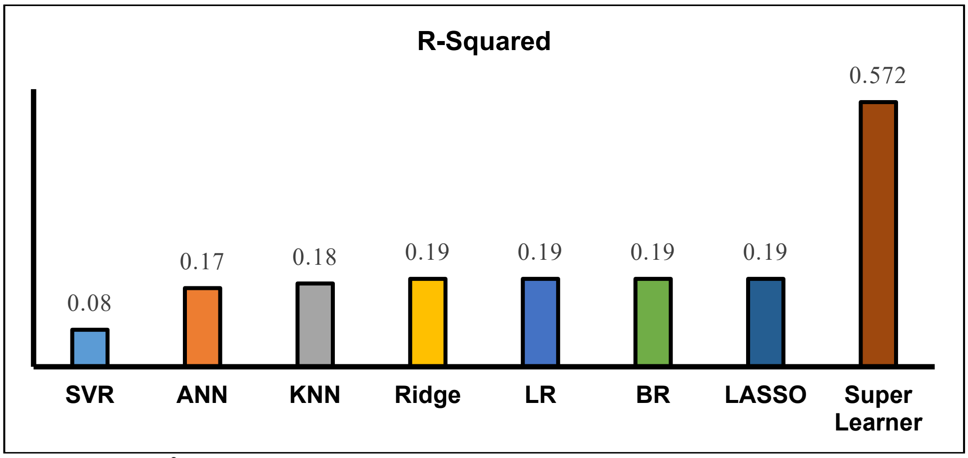

Fig. 7. Bar Plots Showing the R^2 Results for Individual ML Models and LR Models Against the Super Learner Model for Firm valuation.

5.2.4 Impact of Capital Structure and Cash holdings on Super Learner Performance

The study systematically evaluates how excluding key independent variables affects the Super Learner's predictive accuracy. As demonstrated in Table 6 and Fig. 8, removing the leverage ratio (LEV) causes a severe performance degradation, with R^2 decreasing from 0.572 to 0.282 - a 50.7% reduction. This substantial drop is supported by an extremely significant t-statistic of 36.07 ($p \approx 0$), indicating LEV's critical role in explaining firm value variance. This analysis provides robust empirical support for H1, confirming that leverage ratio significantly enhances predictive performance. However, excluding cash holdings (CASH) results in a more modest 3.0% R^2 reduction (0.572 to 0.555), though still statistically significant ($t = 2.29$, $p = 0.023$). These results provide support for H2, confirming that cash holdings significantly enhance predictive performance. From a practical perspective, these findings suggest that: (1) leverage ratios should be prioritized in firm valuation systems, and (2) while cash positions contribute meaningful information, their exclusion is less detrimental to model accuracy. This performance improvement underscores the value of including both variables in comprehensive financial valuation.

Table 6
Results of the Super Learner with and without Leverage and Cash Holdings

Regressor	R-Squared	t-test
Super Learner	0.572	36.07**
Super Learner (Excluding LEV)	0.282	P-value= 0.000
Super Learner (Excluding CASH)	0.555	2.29* P-value= 0.023

Statistical significance at the 0.05 and 0.01 levels is denoted by * and **, respectively.

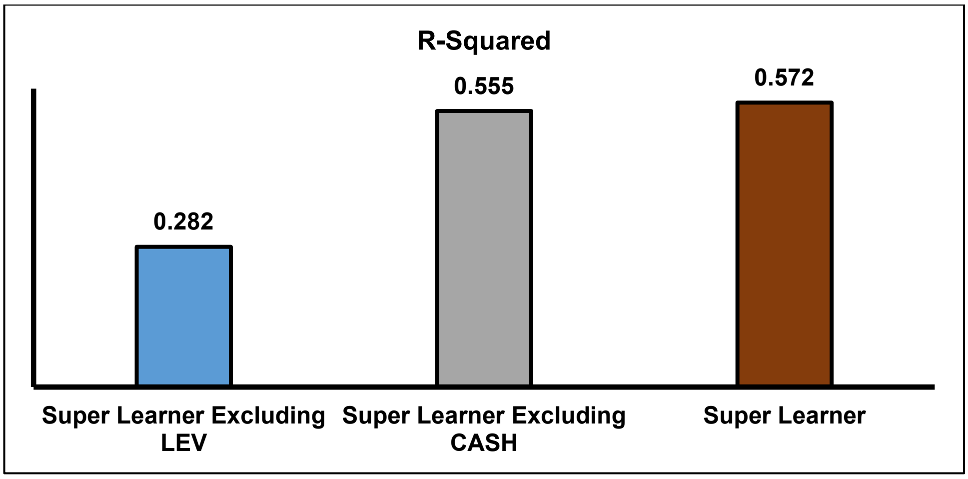

Fig. 8. Bar Plots Showing the R^2 Results for Super Learner Model Excluding Leverage and Cash Holdings Against the Full Model for Firm valuation.

5.2.5 Feature Importance Analysis Using Extra Trees

Understanding the key drivers of firm valuation is essential for making data-driven financial decisions. The Extra Trees model was employed to assess the predictive power of various financial and operational features as shown in Fig. 9. The analysis reveals that LEV (leverage ratio) is the most significant predictor, contributing 20.6% of explanatory power, underscoring the critical role of capital structure decisions in firm value. Following closely are FSIZE (firm size, 9.9%) and CASH (cash holdings, 8.8%), highlighting the influence of financial stability and liquidity management. Other notable predictors include CAPEX (capital expenditures, 8.7%), reflecting the impact of long-term investments, and ROA (return on assets, 8.3%), reinforcing the importance of operational efficiency. Additionally, TRADE (trading activity, 7.7%) and NETCAP (net working capital, 6.7%) indicate the significance of liquidity and working capital management.

Interestingly, DIV (dividend payout, 3.5%) and DIVPAY (dividend payment indicator, 2.3%) have relatively lower predictive power, suggesting that while dividend policies matter, they are less dominant compared to leverage, firm size, and cash holdings. AGE (firm age, 8.5%) also plays a role, implying that more established firms may exhibit different financial behaviors than younger ones. In summary, this analysis emphasizes that leverage control, firm size, liquidity management, and operational efficiency are the most impactful levers for firm valuation. Managers should focus on these areas to strengthen financial stability, while investors may use these insights to assess a company's long-term viability.

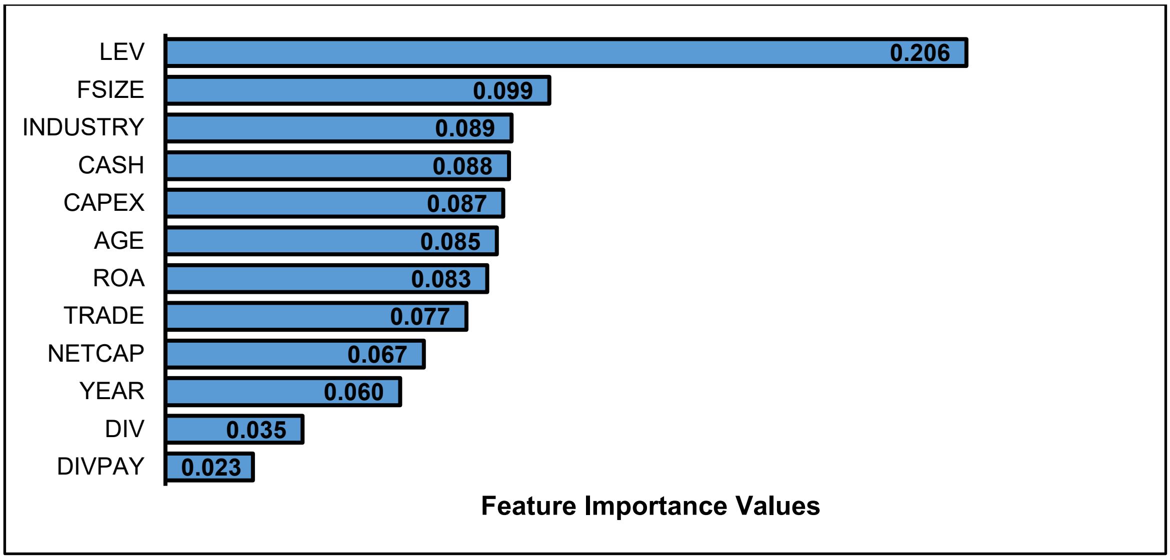

Fig. 9. Feature Importance Analysis using Extra Trees Model

5.2.6 Model Explanation with SHapley Additive exPlanations

The SHAP analysis offers a detailed perspective on the key factors influencing firm valuation by measuring both the strength and direction of each variable's impact. The SHAP summary plot (Fig. 10) visually ranks variables by their impact on firm value predictions, with the most influential at top. Each dot represents an observation: right-side dots increase predictions (positive effect), while left-side dots decrease them (negative effect). Color indicates variable values (blue=low, red=high), revealing how different ranges affect outcomes. Regarding our key variables of interest, the leverage ratio (LEV) demonstrates particularly strong influence, with scattered dots indicating a nonlinear relationship. On the other hand, cash holdings (CASH) show a more consistently positive effect, underscoring liquidity's role as a fundamental driver of financial valuation.

Fig. 10. Global Interpretability Plot of the Super Learner Model.

5.2.7 Partial Dependence Plots: Analyzing the Effects of Capital Structure and Cash Holdings on Firm Value

Fig. 11 demonstrates the partial dependency plots that examine the effects of capital structure and cash holdings on firm value. The first partial dependence plot illustrates the marginal effect of leverage ratio (LEV) on the predicted firm value, proxied by Tobin's Q. The plot reveals a nonlinear relationship: at lower and moderate leverage levels (LEV < 1.0), the marginal effect on Tobin's Q remains relatively stable and close to the average prediction. However, once leverage ratio surpasses approximately 1.2, there is a steep, exponential increase in Tobin's Q, indicating that firms with very high leverage are associated with substantially higher valuations.

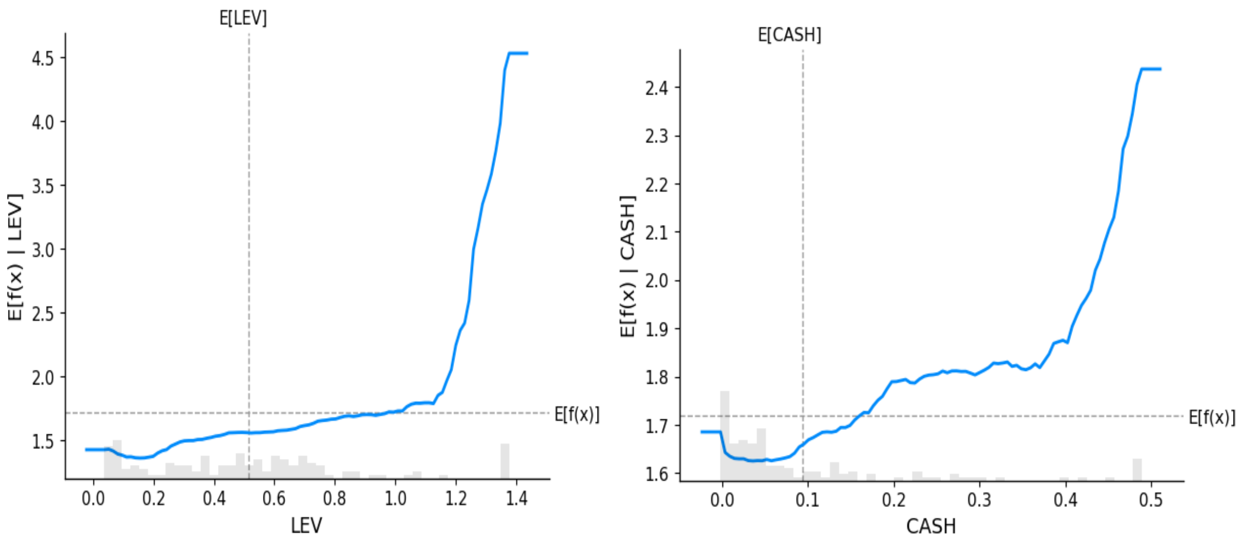

Fig. 11. Partial Dependence Plots for Leverage and Cash Holdings.

This pattern suggests that while moderate leverage may not significantly influence firm value, extremely leveraged firms tend to be rewarded with higher market valuations, potentially due to perceived growth opportunities, tax shields, or aggressive investment strategies. Nevertheless, this effect appears concentrated among a small subset of firms, as indicated by the sparse data density in the high-LEV range. Overall, the dependency plot underscores that leverage exerts a threshold-based influence on firm value, minimal at conventional levels, but markedly positive beyond a critical point.

The second partial dependence plot reveals a positive and nonlinear relationship between cash holdings and Tobin's Q. At very low cash levels, the predicted firm value starts below the average, indicating that firms with minimal liquidity may be undervalued or face operational constraints. As cash holdings increase, Tobin's Q rises gradually at first and then accelerates sharply beyond the 40% mark, suggesting that firms with substantial cash reserves are valued significantly higher by the market.

This trend supports the view that cash enhances firm value by improving financial flexibility, reducing distress risk, and enabling timely investment opportunities. The inflection point and steep upward slope at higher cash ratios imply that excess cash may be especially valuable in uncertain or capital-constrained environments. However, the histogram shows that most firms cluster around lower cash levels, meaning the strong positive effect of high cash holdings on value applies to a relatively small subset of firms.

5.2.8 The Moderating Role of COVID-19

Existing research reveals a gap in understanding how extreme events influence the valuation of corporate financial policies such as cash holdings and leverage. While some studies (e.g., Chang et al., 2017) examine financial crises, little attention has been given to how unprecedented shocks like the COVID-19 pandemic shape market perceptions of liquidity and leverage. It remains unclear whether such crises have fundamentally altered investor attitudes toward these financial strategies.

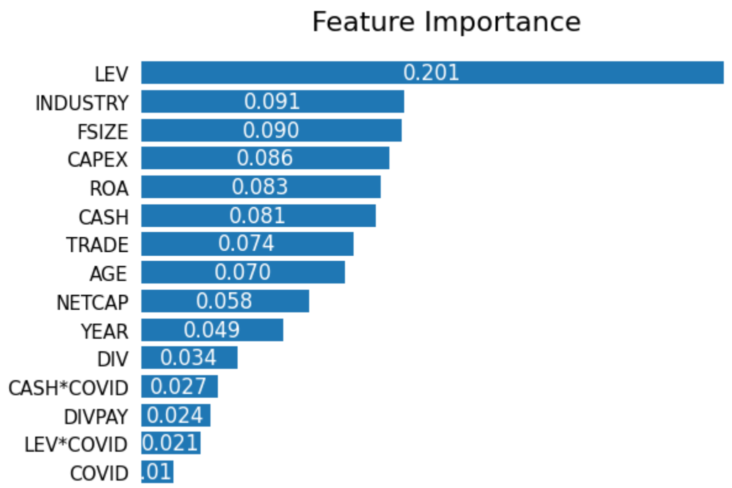

Fig. 12. Feature Importance Analysis using Extra Trees Model Including COVID as a Moderator for Leverage and Cash

Feature importance analysis (Fig. 12) shows that COVID alone has minimal predictive power for firm value (importance = 0.01). However, the interaction terms, CASHCOVID (0.027) and LEVCOVID (0.021), are more impactful. This suggests that COVID acted as a moderator, amplifying the effects of financial structure on firm value. Firms with larger cash reserves were rewarded during the crisis, while those with higher leverage faced increased risk. SHAP summary plots (Fig. 13) reinforce this interpretation. The CASHCOVID interaction contributes positively to model output, indicating that cash became more valuable as a buffer during the pandemic. Conversely, LEVCOVID displays a slight negative skew, implying that leverage was perceived as more detrimental under COVID conditions.

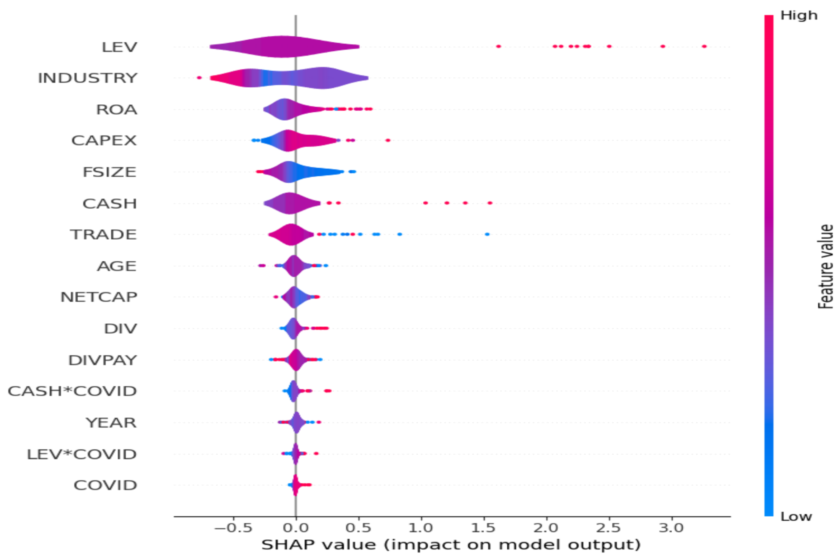

Fig. 13. Global Interpretability Plot of the Super Learner Model Including COVID as a Moderator for Leverage and Cash

As shown in Fig. 14, the partial dependence plot for CASH*COVID shows a clear upward trend: as the interaction increases, so does expect firm value. At low levels, firm value remains below average, but rises sharply for firms with higher cash reserves during the pandemic. This underscores the enhanced importance of liquidity in times of crisis. A similar pattern emerges in the LEV*COVID plot as demonstrated in Fig. 14. While leverage has a minimal effect on firm value under normal conditions, its impact becomes increasingly positive as COVID intensity rises. The steep upward trend beyond a certain threshold suggests that during the pandemic, leverage may have been used effectively or supported by external measures such as government stimulus. Together, these findings highlight how COVID-19 reshaped the role of key financial policies. The crisis emphasized the value of financial flexibility, both in maintaining liquidity and in managing leverage, as firms navigated an uncertain economic landscape.

.

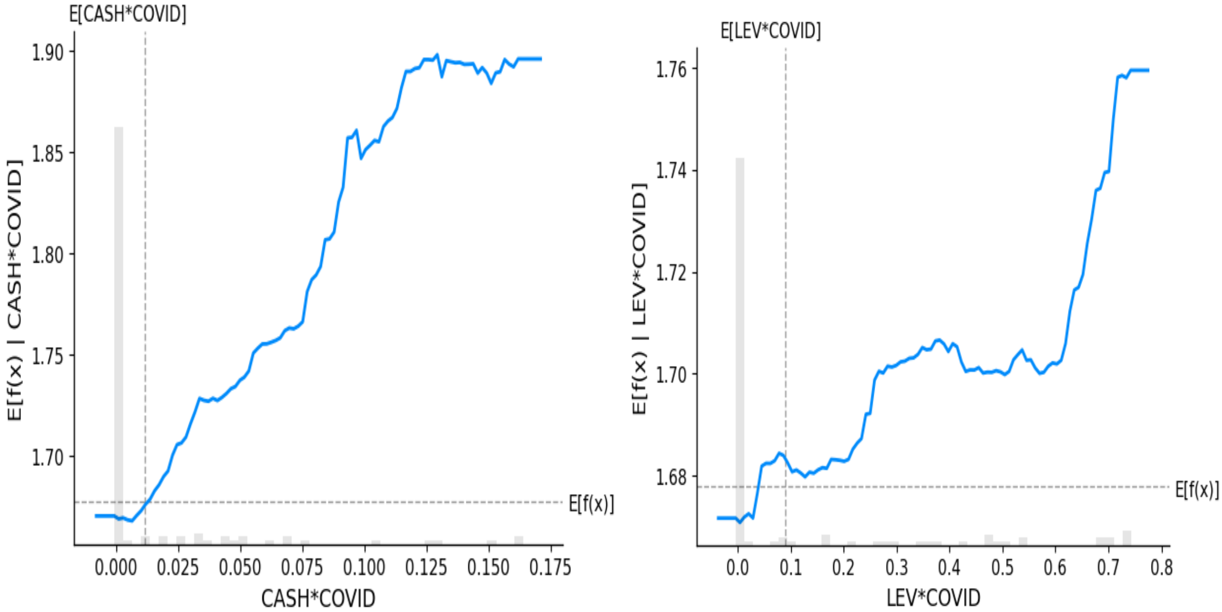

Fig. 14. Partial Dependence Plots for COVID as a Moderator for Leverage and Cash

.

5.3 Discussion of the Results

The results of the study extend existing theoretical and empirical literature on capital structure, cash holdings, and firm valuation. The finding that leverage (LEV) is the most influential predictor of firm value (contributing 20.6% to explanatory power) resonates with empirical studies like Antwi et al. (2012), who found long-term debt to be the primary determinant of firm value. The nonlinear relationship observed, where moderate leverage has minimal impact but extreme leverage (LEV > 1.2) sharply increases Tobin's Q, supports the threshold or non-linear relationship suggested by Cheng et al. (2010). This mirrors Modigliani and Miller's (1963) revised proposition that debt enhances firm value through tax shields.

The significant role of cash holdings (CASH) in firm valuation (8.8% importance) aligns with the Pecking Order Theory's emphasis on liquidity as a buffer against asymmetric information (Myers & Majluf, 1984) and the Trade-Off Theory's cost-benefit analysis of cash reserves (Opler et al., 1999). The positive, nonlinear effect of cash, where higher reserves (>40% of assets) disproportionately boost Tobin's Q, echoes Martínez-Sola et al. (2013), Anton and Nucu (2019), and Habib et al. (2021), who identified optimal cash thresholds. This supports the precautionary motive (Bates et al., 2009) while also reflecting agency concerns (Jensen, 1986), as the histogram shows most firms hold lower cash levels, limiting the prevalence of excessive reserves. The performance drop when excluding CASH (3% reduction in R^2) suggests its role is secondary to leverage, consistent with (Almeida et al., 2014), who noted cash's context-dependent value.

The COVID-19 pandemic's moderating effects on cash and leverage interactions further validate crisis-era studies (Chang et al., 2017). The heightened value of cash during the pandemic (CASHCOVID's positive SHAP values) underscores its role as a financial buffer, aligning with (Chang et al., 2017) argument that liquidity becomes more valuable during crises. Conversely, the negative skew of LEVCOVID reflects increased distress risks during crisis, supporting Pinkowitz et al. (2006)'s findings on leverage's vulnerability in volatile markets. These results address a critical gap in the literature by quantifying how extreme events reshape financial policy valuations, extending Chang et al. (2017)'s crisis framework to a novel context.

The Super Learner's superior performance ($R^2 = 0.572$) over traditional models (e.g., LR, Random Forest) highlights the value of ensemble methods in capturing nonlinear relationships, as advocated by (Geertsema & Lu, 2023) and Koklev (2022). The model's feature importance rankings, prioritizing leverage, firm size, and cash, mirror empirical findings from diverse markets (e.g., Aggarwal et al., 2017; Uzliawati et al., 2018), while its adaptability to Egyptian data addresses a regional research gap identified by (RONIC & Amadi, 2021). The methodological innovation of combining Extra Trees and XGBoost aligns with Chen et al. (2024)'s demonstration of tree-based models' efficacy in financial prediction, while SHAP explanations enhance interpretability.

In summary, the study's results robustly integrate theoretical frameworks (Trade-Off, Pecking Order, and Agency theories) with empirical evidence, while advancing the literature through ML applications and crisis-era insights. The findings underscore the contextual nature of capital structure and cash holdings, emphasizing the need for dynamic, data-driven valuation approaches in emerging markets.

6. Conclusion

This study investigated the relationship between capital structure, cash holdings, and firm value using an explainable ML framework applied to Egyptian non-financial firms from 2016 to 2022. The Super Learner ensemble model, combining Extra Trees and XGBoost algorithms, addressed key limitations in prior research by capturing nonlinear relationships while maintaining interpretability through SHAP analysis. The findings contribute to financial theory while offering practical insights for emerging market contexts.

The results provide robust support for the Trade-Off Theory, demonstrating that capital structure significantly influences firm value in a nonlinear fashion. The study identified threshold effects where extreme leverage levels (LEV > 1.2) generated disproportionately higher valuations, potentially through enhanced tax shields or growth opportunities. This finding extends Modigliani and Miller's (1963) tax-adjusted propositions by quantifying specific inflection points. For cash holdings, the results align with Pecking Order Theory, showing that reserves exceeding 40% of assets yielded substantial valuation premiums, though with relatively lower explanatory power (8.8%) compared to leverage (20.6%). The COVID-19 pandemic emerged as an important moderator, amplifying both the value of cash reserves (supporting the precautionary buffer arguement) and the risks associated with high leverage during crisis periods.

The proposed Super Learner approach demonstrated superior predictive power ($R^2 = 0.572$) compared to traditional methods, representing a 22% improvement over the best individual model and 201% over linear models. The integration of SHAP values and partial dependence plots successfully addressed interpretability challenges commonly associated with ML applications in accounting and finance. This methodological framework provides a foundation for future financial research investigating complex, nonlinear relationships.

The findings offer actionable insights for multiple stakeholders. Corporate managers should prioritize capital structure decisions while recognizing threshold effects in financial policies. Investors can enhance firm valuation by monitoring extreme leverage positions and cash reserve levels, particularly during market crises. Policymakers may consider developing capital market infrastructure and crisis-response mechanisms informed by these results.

The study is limited to Egyptian-specific context. Future research should expand to cross-country comparisons. Sector-specific analyses and examination of diverse crisis types would further enrich understanding of these relationships.
This study advances emerging market finance literature by successfully integrating advanced ML techniques with established financial theory. The approach balances methodological innovation with practical relevance, providing a foundation for future research while offering immediate value to financial practitioners and policymakers navigating complex market environments.

Acknowledgment

The author extends her appreciation to Prince Sattam bin Abdulaziz University, Saudi Arabia for funding this research work through the project number (PSAU/2025/R/1447).

Data Availability

The dataset is available from the author upon reasonable request.

Competing Interest

The author has no relevant financial or non-financial interests to disclose.

References

Acharya, V. V, Almeida, H., & Campello, M. (2007). Is cash negative debt? A hedging perspective on corporate financial policies. *Journal of Financial Intermediation*, *16*(4), 515–554.

Aggarwal, D., Padhan, P. C., Aggarwal, D., & Padhan, P. C. (2017). Impact of Capital Structure on Firm Value: Evidence from Indian Hospitality Industry. *Theoretical Economics Letters*, *7*(4), 982–1000. https://doi.org/10.4236/TEL.2017.74067

Akash, R. S. I., Khan, M. I., & Shear, F. (2023). The Corporate financial policy and the firm value. *International Journal of Business and Economic Affairs*, *8*(3), 65–74.

Ali, G. M., Gomaa, I. I., & Darwish, S. M. (2022). An Intelligent Model for Success Prediction of Initial Coin Offerings. *IEEE Access*, *10*, 58589–58602. https://doi.org/10.1109/ACCESS.2022.3178369

Aljamaan, B. E. (2018). Capital structure: Definitions, determinants, theories and link with performance literature review. *European Journal of Accounting, Auditing and Finance Research*, *6*(2), 49–72.

Almeida, H., Campello, M., Cunha, I., & Weisbach, M. S. (2014). Corporate liquidity management: A conceptual framework and survey. *Annu. Rev. Financ. Econ.*, *6*(1), 135–162.

Anton, S. G., & Nucu, A. E. A. (2019). Firm value and corporate cash holdings. Empirical evidence from the Polish listed firms. *E a M: Ekonomie a Management*, *22*(3), 121–134. https://doi.org/10.15240/TUL/001/2019-3-008

Antwi, S., Mills, E. F. E. A., & Zhao, X. (2012). Capital structure and firm value: Empirical evidence from Ghana. *International Journal of Business and Social Science*, *3*(22).

Baker, M., & Wurgler, J. (2002). Market timing and capital structure. *The Journal of Finance*, *57*(1), 1–32.

Balagopal, G., Wijeratne, L., Waczak, J., Hathurusinghe, P., Iqbal, M., Kiv, D., Aker, A., Lee, S., Agnihotri, V., Simmons, C., & Lary, D. J. (2025). Calibration of Low-Cost LoRaWAN-Based IoT Air Quality Monitors Using the Super Learner Ensemble: A Case Study for Accurate Particulate Matter Measurement. *Sensors*, *25*(5). https://doi.org/10.3390/s25051614

Basci, e. S. (2019). Firm valuation concept and discounted cash flow method: a comparison of stock markets. *Annals of Spiru Haret University. Economic Series*, *19*(2), 51–60. https://doi.org/10.26458/1922

Bates, T. W., Kahle, K. M., & Stulz, R. M. (2009). Why do US firms hold so much more cash than they used to? *The Journal of Finance*, *64*(5), 1985–2021.

Brennan, M. J., & Schwartz, E. S. (1984). Optimal Financial Policy and Firm Valuation. *The Journal of Finance*, *39*(3), 593–607. https://doi.org/10.1111/J.1540-6261.1984.TB03647.X

Brigham, E. F. (2016). *Financial management: Theory and practice*. Cengage Learning Canada Inc.

Bui, T., Ngoc, X. H., Nguyen, K. T., Pham, 2023, Bui, T. N., Nguyen, X. H., & Pham, K. T. (2023). The Effect of Capital Structure on Firm Value: A Study of Companies Listed on the Vietnamese Stock Market. *International Journal of Financial Studies 2023, Vol. 11, Page 100*, *11*(3), 100. https://doi.org/10.3390/IJFS11030100

Cakici, N., Fieberg, C., Metko, D., & Zaremba, A. (2023). Machine learning goes global: Cross-sectional return predictability in international stock markets. *Journal of Economic Dynamics and Control*, *155*, 104725. https://doi.org/10.1016/J.JEDC.2023.104725

Cao, K., & You, H. (2024). Fundamental analysis via machine learning. *Financial Analysts Journal*, *80*(2), 74–98.

Cerkovskis, E., Gajdosikova, D., & Ciurlau, C. F. (2022). CAPITAL STRUCTURE THEORIES: REVIEW OF LITERATURE. *Economic & Managerial Spectrum/Ekonomicko-Manažérske Spektrum*, *16*(1).

Chang, Y., Benson, K., & Faff, R. (2017). Are excess cash holdings more valuable to firms in times of crisis? Financial constraints and governance matters. *Pacific-Basin Finance Journal*, *45*, 157–173.

Chen, C., Ke, B., & Zhao, Q. (2024). Enhancing Firm Quality Measurement Using Machine Learning: The Importance of Theory-motivated Data Engineering. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4434498

Cheng, Y.-S., Liu, Y.-P., & Chien, C.-Y. (2010). Capital structure and firm value in China: A panel threshold regression analysis. *African Journal of Business Management*, *4*(12), 2500.

Chowdhury, A., & Chowdhury, S. P. (2010). Impact of capital structure on firm's value: Evidence from Bangladesh. *Business and Economic Horizons*, *3*(3), 111–122.

Datta, R., Chowdhury, T., & Mohajan, H. (2013). *Reassess of capital structure theories*.

Davies, M. M., & Van Der Laan, M. J. (2016). Optimal Spatial Prediction Using Ensemble Machine Learning. *International Journal of Biostatistics*, *12*(1), 179–201. https://doi.org/10.1515/IJB-2014-0060/ASSET/GRAPHIC/IJB-2014-0060_FIGURE2.GIF

Dey, R., & Mathur, R. (2023). Ensemble Learning Method Using Stacking with Base Learner, A Comparison. *Lecture Notes in Networks and Systems*, *727 LNNS*, 159–169. https://doi.org/10.1007/978-981-99-3878-0_14

Díaz, E., & Spagnoli, G. (2024). A super-learner machine learning model for a global prediction of compression index in clays. *Applied Clay Science*, *249*, 107239. https://doi.org/10.1016/J.CLAY.2023.107239

Dittmar, A., & Mahrt-Smith, J. (2007). Corporate governance and the value of cash holdings. *Journal of Financial Economics*, *83*(3), 599–634.

Dogra, B., & Gupta, S. (2009). An empirical study on capital structure of SMEs in Punjab. *IUP Journal of Applied Finance*, *15*(3), 60.

Donaldson, G. (1961). *Corporate debt capacity: A study of corporate debt policy and the determination of corporate debt capacity*. Beard Books.

Faleye, O. (2004). Cash and corporate control. *The Journal of Finance*, *59*(5), 2041–2060.

Ferreira, M. A., & Vilela, A. S. (2004). Why do firms hold cash? Evidence from EMU countries. *European Financial Management*, *10*(2), 295–319.

Geertsema, P. G., & Lu, H. (2019). Machine Valuation. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3447683

Geertsema, P., & Lu, H. (2023). Relative valuation with machine learning. *Journal of Accounting Research*, *61*(1), 329–376.

Gitman, L. J., & Zutter, C. J. (2012). Principles of managerial finance 13th edition. *Massachusetts: Addison-Wesley Publishing Company*.

Habib, A., Bhatti, M. I., Khan, M. A., & Azam, Z. (2021). Cash Holding and Firm Value in the Presence of Managerial Optimism. *Journal of Risk and Financial Management 2021, Vol. 14, Page 356*, *14*(8), 356. https://doi.org/10.3390/JRFM14080356

Harford, J., Mansi, S. A., & Maxwell, W. F. (2008). Corporate governance and firm cash holdings in the US. *Journal of Financial Economics*, *87*(3), 535–555.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, Issue 1). Springer.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review*, *76*(2), 323–329.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Koklev, P. S. (2022). Business Valuation with Machine learning. *Finance: Theory and Practice*, *26*(5), 132–148.

Kuan, T.-H., Li, C.-S., & Chu, S.-H. (2011). Cash holdings and corporate governance in family-controlled firms. *Journal of Business Research*, *64*(7), 757–764.

Kwak, S. K., & Kim, J. H. (2017). Statistical data preparation: management of missing values and outliers. *Korean Journal of Anesthesiology*, *70*(4), 407. https://doi.org/10.4097/KJAE.2017.70.4.407

Lau, J., & Block, J. H. (2012). Corporate cash holdings and their implications on firm value in family and founder firms. *Corporate Ownership and Control*, *9*(4), 309–326.

Lee, K.-W., & Lee, C.-F. (2009). Cash holdings, corporate governance structure and firm valuation. *Review of Pacific Basin Financial Markets and Policies*, *12*(03), 475–508.

Liu, Y., & Mauer, D. C. (2011). Corporate cash holdings and CEO compensation incentives. *Journal of Financial Economics*, *102*(1), 183–198.

Luo, Q., & Hachiya, T. (2005). Corporate governance, cash holdings, and firm value: evidence from Japan. *Review of Pacific Basin Financial Markets and Policies*, *8*(04), 613–636.

Martínez-Sola, C., García-Teruel, P. J., & Martínez-Solano, P. (2013). Corporate cash holding and firm value. *Applied Economics*, *45*(2), 161–170.

Mazumdar, M. A. A., & Mara, D. . O. (2024). Analysing the Impact of Capital Structure on Firm Value: A Study in Corporate Finance. *International Journal for Research in Applied Science and Engineering Technology*, *12*(10), 1191–1199. https://doi.org/10.22214/IJRASET.2024.64754

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, *48*(3), 261–297.

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/J.JKSUCI.2023.01.014

Myers, S. C., & Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, *13*(2), 187–221.

Naimi, A. I., & Balzer, L. B. (2018). Stacked generalization: an introduction to super learning. *European Journal of Epidemiology*, *33*, 459–464.

Nhan, D. T. T., & Ha, P. (2016). Cash Holding, State Ownership and Firm Value: The Case of Vietnam. *International Journal of Economics and Financial Issues*, *6*(6), 110–114. https://dergipark.org.tr/en/pub/ijefi/issue/31991/352812

Olokoyo, F. O. (2013). Capital structure and corporate performance of Nigerian quoted firms: A panel data approach. *African Development Review*, *25*(3), 358–369.

Opler, T., Pinkowitz, L., Stulz, R., & Williamson, R. (1999). The determinants and implications of corporate cash holdings. *Journal of Financial Economics*, *52*(1), 3–46.

Paramasivan, C. (2009). *Financial management*. New Age International.

Pinkowitz, L., Stulz, R., & Williamson, R. (2006). Does the Contribution of Corporate Cash Holdings and Dividends to Firm Value Depend on Governance? A Cross-country Analysis. *The Journal of Finance*, *61*(6), 2725–2751. https://doi.org/10.1111/J.1540-6261.2006.01003.X

Polley, E., & Laan, M. van der. (2010). Super Learner In Prediction. *U.C. Berkeley Division of Biostatistics Working Paper Series*. https://biostats.bepress.com/ucbbiostat/paper266

RONIC, H. I., & Amadi, C. (2021). THE NEXUS BETWEEN CAPITAL STRUCTURE AND FIRMS VALUE: EVIDENCE FROM NIGERIA. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3918974

Ross, S. A. (2009). Corporate finance: Core principles & applications. *(No Title)*.

Saha, D., Young, T. M., & Thacker, J. (2023). Predicting firm performance and size using machine learning with a Bayesian perspective. *Machine Learning with Applications*, *11*, 100453. https://doi.org/10.1016/J.MLWA.2023.100453

Thabet, H. H., Darwish, S. M., & Ali, G. M. (2024). Measuring the efficiency of banks using high-performance ensemble technique. *Neural Computing and Applications*, 1–19. https://doi.org/10.1007/S00521-024-09929-Y/TABLES/5

Uzliawati, L., Yuliana, A., Januarsi, Y., & Santoso, M. I. (2018). Optimisation of capital structure and firm value. *European Research Studies Journal*, *21*(2), 705–713. https://doi.org/10.35808/ERSJ/1034

Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology*, *6*(1).

Vayas-Ortega, G., Soguero-Ruiz, C., Rodríguez-Ibáñez, M., Rojo-Álvarez, J. L., & Gimeno-Blanes, F. J. (2020). On the Differential Analysis of Enterprise Valuation Methods as a Guideline for Unlisted Companies Assessment (II): Applying Machine-Learning Techniques for Unbiased Enterprise Value Assessment. *Applied Sciences 2020, Vol. 10, Page 5334*, *10*(15), 5334. https://doi.org/10.3390/APP10155334

Weidemann, J. F. (2018). A state-of-the-art review of corporate cash holding research. *Journal of Business Economics*, *88*, 765–797.

Zhang, C., Zhang, H., & Liu, D. (2020). A Contrastive Study of Machine Learning on Energy Firm Value Prediction. *IEEE Access*, *8*, 11635–11643. https://doi.org/10.1109/ACCESS.2019.2953807

Zhang, R., Tian, Z., McCarthy, K. J., Wang, X., & Zhang, K. (2023). Application of machine learning techniques to predict entrepreneurial firm valuation. *Journal of Forecasting*, *42*(2), 402–417. https://doi.org/10.1002/FOR.2912