DOI: https://doi.org/10.62985/j.huit_ojs.vol26.no3E.500

FORECASTING CORPORATE CASH HOLDINGS IN VIETNAM: COMPARING MACHINE LEARNING MODELS AND THE IMPACT OF THE WORLD UNCERTAINTY INDEX

Nguyen Binh Phuong Thuy

Ho Chi Minh City University of Industry and Trade

Email: *thuynbp@huit.edu.vn*

Received: 20 March 2026; Revised: 21 April 2026; Accepted: 8 May 2026

ABSTRACT

This study forecasts the cash holdings (CASH) of listed companies in Vietnam based on data from 2018-2025 using a machine learning approach. Using panel data with 4,752 observations, the study compares seven algorithms and identifies HistGradientBoosting as the optimal model (R^2 approx. 0.7145), far surpassing traditional linear regression (R^2 approx. 0.28). This result demonstrates strong non-linearity in liquidity management behavior in the Vietnamese market. Using SHAP and Permutation Importance, the study successfully elucidates the underlying predictive mechanisms. The main drivers are net working capital (NWC) and the current ratio (CR). Notably, short-term debt (STD), interest expense (IE), and the World Uncertainty Index for Vietnam (WUIVN) consistently showed positive effects, strongly reinforcing the Precautionary Motive. The study also found a moderating effect of size, indicating that small businesses are more sensitive and vulnerable to systemic shocks. The application of machine learning models not only improves forecasting accuracy but also provides a transparent view of corporate financial strategies in an uncertain economic environment.

Keywords: Cash holdings, machine learning, SHAP, World Uncertainty Index, HistGradientBoosting, Vietnam.

1. INTRODUCTION

The 2007-2008 Global Financial Crisis (GFC) marked the end of a relatively stable era for the global economy. This crisis showed how quickly systemic risk can spread and how it can freeze international financial markets. After the GFC, businesses increased their hedging activities and built larger cash reserves to protect against systemic shocks and keep liquidity [1]. The following economic downturn caused by the COVID-19 pandemic further increased instability. The pandemic led to supply chain disruptions, drops in cash flow, shrinking credit markets, and unprecedented policy uncertainty [2], [3]. In this environment, holding cash became a vital financial strategy, helping firms stay solvent and rely less on external financing [4]. This tendency to hoard cash reflects precautionary motives in response to higher risks. Although the immediate shocks from the COVID-19 pandemic have subsided, global macroeconomic uncertainty, as measured by the World Uncertainty Index (WUI), remains high relative to pre-crisis levels [5]. Additionally, [2] found that uncertainty regarding regulations, taxes, and public spending continues to negatively affect long-term investment and corporate financial planning. These factors suggest that market risk will remain high during the post-COVID-19 recovery, creating major challenges for liquidity management and financial decision-making. As a result, the high cash-holding behavior observed during the pandemic is likely to persist into the recovery period, as ongoing macroeconomic uncertainties continue to influence financial and investment activities. In this context, accurately forecasting corporate cash holdings is crucial. Finding the right level of cash reserves allows firms to manage market risks proactively while balancing liquidity and capital efficiency.

Traditional linear regression models often struggle to capture non-linear relationships and complex interactions among financial variables, economic instability, and firm-specific characteristics. In contrast, machine learning (ML) algorithms such as Random Forests, Gradient Boosting, XGBoost, and Neural Networks have shown better performance in forecasting financial volatility and modeling non-linear relationships between variables [6].

Although many international studies have examined corporate cash-holding behavior during crises, research in Vietnam remains scarce, especially that which combines market instability with advanced machine-learning-based forecasting techniques. To fill this gap, the present study, titled "Forecasting Corporate Cash Holdings in Vietnam: Comparing Machine Learning Models and the Impact of the World Uncertainty Index," was conducted. The findings are expected to provide valuable empirical evidence and significant practical insights. Specifically, the results will provide corporate managers with a scientific basis for developing flexible, risk-adjusted cash management policies, particularly in response to cash flow challenges arising from external shocks. Additionally, the research will aid policymakers in developing mechanisms to enhance businesses' financial resilience amid future volatility.

The rest of this paper is organized as follows. Section 2 summarizes the theoretical background and literature review, while Section 3 describes the data and research methodology. Section 4 presents the empirical results, and Section 5 concludes the paper.

2. LITERATURE REVIEW

Initially, empirical studies on cash holdings primarily focused on identifying the micro-factors and motivations that influence a company's decision to maintain cash. The study by Opler et al. is considered a foundational work, demonstrating that variables such as firm size, cash flow, and growth opportunities play a significant role in explaining the differences in cash holdings among companies [7]. The authors also showed that cash holding behavior primarily reflects a precautionary motive and is influenced by agency costs. Continuing this research direction, Almeida et al. expanded on this by examining the role of financial constraints. Their results showed that financially constrained companies tend to react more strongly to cash flow fluctuations, thus reinforcing the role of precautionary motives [8]. The research by Bates et al. reached a similar conclusion. Accordingly, the authors noted a significant increase in cash holdings by US businesses over the past two decades [9]. This trend is explained by changes in operational risk, investment in research and development, and an increase in free cash flow, thereby once again emphasizing the role of cash as a financial risk-hedging mechanism in a volatile economic environment.

Not only under normal conditions, but many empirical studies also indicate that cash holdings tend to increase significantly during periods of global crisis or before major systemic shocks. Specifically, the research results of Song and Lee showed that businesses in East Asian countries adjusted their financial structure by increasing the cash ratio by reducing investment and M&A activities during the 1997-1998 Asian financial crisis [10]. During the 2007-2008 global financial crisis, Lian et al. found, using data from over 8,600 observations between 1999 and 2009, that businesses in China increased their cash reserves to cope with uncertainty and credit restrictions [11]. In another context, Dimitropoulos and Koronios also observed an increase in cash levels of small and medium-sized enterprises (SMEs) in Denmark during the European sovereign debt crisis, with cash acting as a "liquidity buffer" to help businesses maintain their viability before the crisis [12]. However, not all studies show an increase in cash levels during a crisis. Tran [13], when analyzing Vietnamese corporate data during the 2008-2009 financial crisis, pointed out that although businesses had a precautionary incentive, the actual amount of cash held decreased because businesses were forced to use available funds to cover operating costs, maintain production, and cope with increased financial constraints during the crisis [13].

Empirical research results prior to the unprecedented economic crisis caused by the COVID-19 pandemic also showed similar trends. Specifically, the study by Qin et al. (2020) found that under the impact of the pandemic, businesses with positive goodwill significantly increased their cash reserves to create a "liquidity shield" to help them withstand the pressure of declining cash flow and operational risks [4]. In Vietnam, the research by Van et al. (2022) also showed that financially constrained businesses increased their reserves and adjusted their capital structures more cautiously amid cash flow contraction due to COVID-19 [14]. A similar result was also found in the study by Dao (2021). Accordingly, evidence from listed companies in Vietnam during the 2017-2020 period shows a significant increase in cash holdings in 2020, reflecting a hedging motive against increased uncertainty and higher capital raising costs during the crisis [15]. The Covid-19 pandemic not only caused an immediate shock but also left a highly unstable macroeconomic environment, reflected in the significant increase in global instability indicators. According to Ahir et al. (2022), the WUI index peaked in Q2/2020, when Covid-19 was spreading worldwide. Although WUI has declined gradually, it remains significantly higher than in the pre-pandemic period, indicating persistent instability in the global economy [5]. From a financial market perspective, the VIX index also surged to a historical high in March 2020, rising by 497% from January 2020 levels, reflecting uncertainty and volatility in international stock markets [3]. Evidence suggests that when instability increases, businesses tend to increase cash reserves as a hedging strategy to mitigate the impact of cash flow fluctuations, operational risks, and difficulties in accessing credit [16], [17]. Research in Vietnam by Nguyen et al. (2025) also indicates that global instability, as measured by the WUI, significantly affects financial performance and business efficiency, suggesting that uncertainty shocks can strongly affect the liquidity strategies of domestic businesses [18]. This result underscores that the impact of macroeconomic shocks on cash holdings depends on market conditions, financial structures, and access to capital in each country.

Although previous studies have built a solid foundation on the factors influencing corporate cash holdings, several important gaps remain. First, most traditional studies rely primarily on linear regression models or classical econometric methods. While these models excel at interpreting causal relationships, they are limited in modeling the non-linear relationships and complex interactions between financial variables and macroeconomic factors. In a volatile economic environment, the relationship between macroeconomic risk and corporate cash-holding behavior is difficult to fully capture with conventional linear models.

Furthermore, studies on the impact of the COVID-19 economic crisis on cash holdings are quite extensive worldwide [3], [4]. However, in Vietnam, these studies have focused only on descriptive or regression analyses of the impact, and no research has applied ML algorithms to predict corporate cash holdings. While models such as Random Forests, Gradient Boosting, XGBoost, and Neural Networks have proven effective at forecasting complex financial fluctuations across many countries [6], having not been explored in cash-holding research in emerging markets such as Vietnam.

Therefore, this study aims to fill these gaps through three main objectives. First, the author applies modern ML algorithms such as Multiple Linear Regression, k-Nearest Neighbors, Support Vector Regression, Decision Tree, Random Forest, XGBoost, and LightGBM Artificial Neural Network to build a model predicting the cash holdings of listed companies, thereby verifying the ability to handle complex non-linear interactions that traditional models often miss. Next, by exploiting explainable AI techniques, the study examines the importance of features and clarifies which variables drive a company's liquidity decisions. In particular, in a volatile economic environment, the study analyzes the impact of macroeconomic instability, thereby providing a multidimensional view of how economic shocks affect the cash-holding behavior of listed companies in Vietnam.

The achievement of these objectives helps the research establish a distinct academic position and directly address the limitations present in previous ML applications. Specifically, while [19] mainly focused on comparing decision trees with traditional regression models, this study conducts a systematic analysis using 7 modern algorithms to identify the optimal forecasting structure for panel data in the Vietnamese market. Unlike the approach of [20], which focuses on micro-specific characteristics, this paper pioneers integrating the World Uncertainty Index for Vietnam (WUIVN) into a non-linear model to accurately capture businesses' risk-taking motives in a volatile macro environment. Finally, through the application of SHAP techniques, the research not only improves forecasting accuracy but also elucidates the underlying predictive mechanisms, connects technical results with classical financial theories to clarify the interaction mechanism between firm size and macro-shocks. This is an important aspect that previous ML studies have often overlooked or not explored in depth.

3. METHODOLOGY

3.1. Data

This study collected data from the annual financial reports of listed companies on the HOSE and HNX exchanges from 2018 to 2025. The author excluded financial companies and companies with missing data, negative leverage, or negative tangible assets to ensure sample consistency. After data processing, the author collected data on 594 companies during the study period, yielding 4,752 firm-year observations. Furthermore, to control for outliers in the Vietnamese market, the author Winsorized at the 1st and 99th percentiles for all financial variables. This normalization step not only reduces data noise but also optimizes ML algorithms for identifying complex, nonlinear relationships, minimizing the influence of extreme fluctuations.

3.2. Research framework

The general model for forecasting cash holdings (CASH) is established as follows:

$$CASH_{i,t} = f(Financial\_Factorsi, WUIVN_t) + \varepsilon_t$$

Where the function f represents the nonlinear ML algorithms. The variables and their definitions are shown in Table 1.

Table 1. Definition of variables and determinants of cash holdings

Explanatory variables	Definitions	Studies
CASH	The ratio of cash and cash equivalents to the total assets	Alam et al. [1], Özlem & Tan [20]
DIV	The ratio of total dividend payments to the total asset	Alam et al. [1], Song and Lee [10], Wu et al. [19], Özlem and Tan [20]
GROWTH	Percentage change in total assets	Alam et al. [1], Song and Lee [10], Özlem and Tan [20], Kim et al. [21]

Explanatory variables	Definitions	Studies
SIZE	The logarithm of its total assets.	Qin et al. [4], Lian et al. [11], Gao et al. [22]
CAPEX	The ratio of capital expenditure to the total assets	Alam et al. [1], Bates et al. [9], Lian et al. [11]
CF_ratio	The ratio of the sum of pre-tax income plus depreciation to the total assets	Alam et al. [1], Wu et al. [19], Özlem and Tan [20]
IE	The ratio of interest expense to the total assets	Özlem and Tan [20]
NWC	The ratio of Net working capital minus cash to total assets	Almeida et al. [8], Özlem and Tan [20]
STD	The ratio of short-term debt to the total assets	Özlem and Tan [20], Lozano and Yaman [23]
ROE	The ratio of net income to the total equity	Alam et al. [1], Özlem and Tan [20], Manoel et al. [24]
AR	The ratio of account receivable to the total assets	Alam et al. [1], Wu et al. [19], Özlem and Tan [20]
AP	The ratio of accounts payable to the total assets	Alam et al. [1], Wu et al. [19], Özlem and Tan [20]
CR	The ratio of current assets to current liabilities	Alam et al. [1], Manoel et al. [24]
LEV	The ratio of total liabilities to total assets	Alam et al. [1], Ozkan and Ozkan [25]
WUIVN	Annual average of quarterly data of World Uncertainty Index for Vietnam	Nguyen et al. [18]

Source: own research

3.3. Estimation Strategy

This study applies a multidimensional approach to forecast cash holdings, combining traditional financial theories and nonlinear ML algorithms. Based on the collected dataset, the author applies ML models sequentially, from traditional parametric models to modern nonparametric ML structures such as Multiple Linear Regression, k-Nearest Neighbors, Support Vector Regression, Decision Tree, Random Forest, HistGradientBoosting, and Multi-Layer Perceptron to predict cash holding values based on groups of firm characteristic variables [1], [19]. After model construction, the study evaluates and compares the predictive performance of models at each stage using R-squared and Mean Absolute Error (MAE), thereby identifying the most suitable model under different economic conditions. In addition, to ensure the model's explanatory power, the author used the Permutation Importance technique to rank factors based on the degree of model performance degradation when variables are randomly shuffled [1], [19]. Next, the Shapley Additive exPlanations (SHAP) method was applied to estimate the marginal contribution of each variable to the forecast. This technique not only quantifies the impact of corporate financial characteristics but also clearly identifies their direction of impact, thereby clarifying the predictive logic of ML algorithms [1].

Specifically, the study evaluates the role of the macroeconomic instability index (WUIVN) using the SHAP Summary Plot and the SHAP Dependence Plot. The SHAP Summary Plot displays the range and distribution of the impact of input variables on the model's forecast results [26]. This chart provides the author with an overall view of the contribution of input variables to the forecasting process, particularly the macroeconomic instability variable. Meanwhile, the SHAP Dependence Plot is used to further analyze the variability and spread of SHAP values when an input variable changes [26]. This approach allows the author to verify how WUIVN interacts nonlinearly with characteristic variables, such as firm size (SIZE). Incorporating SHAP into the research methodology clarifies the transmission mechanism of macroeconomic shocks into cash hedging behavior, thereby providing in-depth empirical evidence from a financial-theory perspective that traditional regression methods often miss due to the assumption of linearity.

4. RESULTS

4.1. Descriptive Statistics

Before training the ML algorithms, the study performed descriptive statistics to assess the distribution characteristics of the dataset comprising 594 listed companies in Vietnam during the period 2018–2025. Table 2 presents basic statistics, including the mean, standard deviation, maximum (max), and minimum (min) for the core variables used in the final model.

Table 2. Descriptive Statistics

	Obs	Mean	Std. Dev.	Min	Max
CASH	4752	0.084	0.089	0.001	0.468
DIV	4752	0.031	0.046	0	0.281
GROWTH	4752	0.079	0.218	-0.348	1.173
SIZE	4752	27.781	1.63	24.489	32.087
CAPEX	4752	0.036	0.056	0	0.318
CF_ratio	4752	0.097	0.087	-0.097	0.417
IE	4752	-0.013	0.013	-0.053	0
NWC	4752	0.164	0.223	-0.325	0.736
STD	4752	0.145	0.143	0	0.563
ROE	4752	0.103	0.109	-0.286	0.46
AR	4752	0.247	0.184	0.012	0.815
AP	4752	0.058	0.075	0	0.425
CR	4752	2.682	3.401	0.377	23.746
LEV	4752	0.458	0.218	0.03	0.887
WUIVN	4752	0.118	0.185	0	0.598

Source: own research

The Descriptive Statistics results in Table 2 show that the average cash and cash equivalents (CASH) ratio of the sample companies is 0.084. However, the relatively high standard deviation (0.089) and the range of 0.001 to 0.468 reflect strong differentiation in liquidity management strategies among companies. The WUIVN index also recorded significant volatility during the study period (standard deviation of 0.185 and a peak of 0.598), thereby clearly capturing important economic shocks, such as the COVID-19 pandemic and intertwined policy changes. In addition, characteristic financial variables such as Net Working Capital (NWC), Current Ratio (CR), and Financial Leverage (LEV) exhibit very wide dispersion (for example, CR has an average of 2.682 and a standard deviation of up to 3.401). The diversity and high variability of all 4,752 observations ensure the representativeness of the data sample, creating ideal conditions for ML algorithms to identify complex, nonlinear patterns.

4.2. Correlation Analysis

Figure 1 shows a heatmap of the correlation matrix for 15 features included in the ML model. The graph shows that most independent variables have low to medium correlation coefficients, within a safe range. The highest correlation coefficient was recorded between return on equity (ROE) and cash flow ratio (CF_ratio) at 0.81. Although the correlation is relatively strong for financial performance, this level remains below the threshold for severe multicollinearity. Furthermore, unlike traditional linear regression models, which are highly susceptible to multicollinearity, the nonlinear ML algorithms used in this study (especially tree models) can automatically process and branch based on correlated variables without distorting the overall predictive structure. On the dependent variable side, cash holdings (CASH) showed a notable positive correlation with interest expense (IE) (0.30) and cash flow ratio (CF_ratio) (0.24), while having a negative correlation with short-term debt (STD) (-0.21). The macro variable WUIVN showed a very low linear correlation with most microfinance variables, suggesting that the impact of macro uncertainty on cash may not follow traditional linear patterns, thereby reinforcing the need to apply sophisticated nonlinear ML models in the following section.

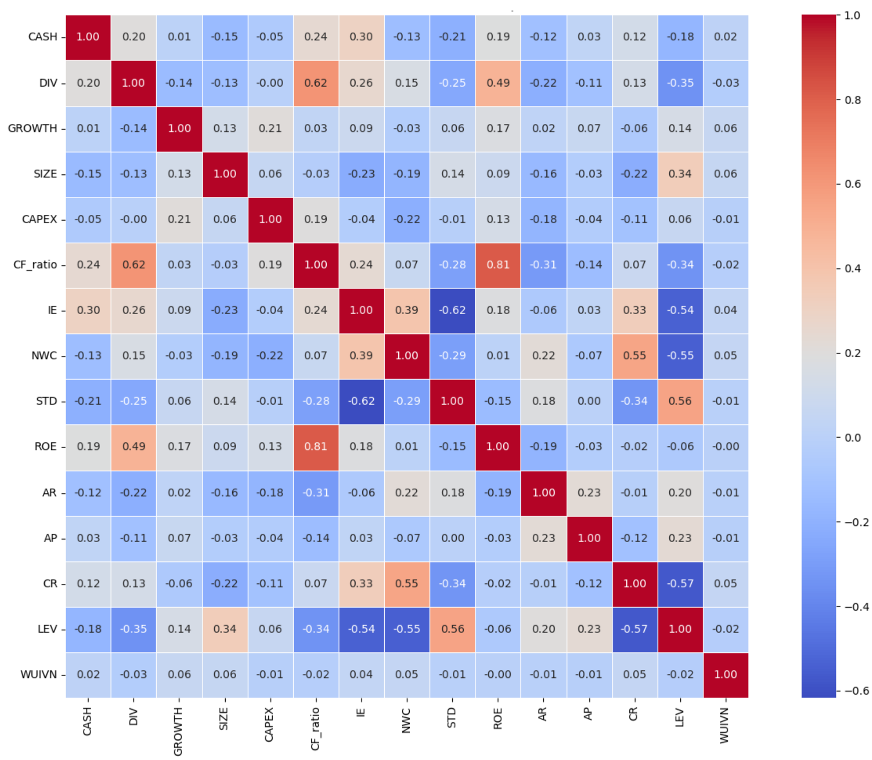

Figure 1. Pearson correlation matrix heatmap of the input variables.

Source: own research

4.3. Model Performance Evaluation

To build an ML model capable of predicting businesses' cash holdings using business characteristics, especially the WUIVN, the author applied ML algorithms to identify and rank variables that significantly influence liquidity management behavior, while accounting for non-linear effects often overlooked by traditional models. However, the reliability of these advanced algorithms depends fundamentally on a rigorous training and evaluation framework. Consequently, to ensure the model's objectivity and generalizability in panel data, this study employs a Group Shuffle Split strategy rather than conventional random splitting [27]. Specifically, 80% of the businesses (475 companies) were included in the training set, and the remaining 20% (119 companies) were kept separate as the test set. Furthermore, stability assessment was performed using 5-fold group cross-validation, ensuring that data from the same business never appear simultaneously in both the training and cross-validation sets within the same fold [28].

Table 3. Predictive performance of machine learning models

Model	Train R^2	Val R^2 (Mean ± SD)	Test R^2	Train MAE	Val MAE (Mean ± SD)	Test MAE
Boosting (HGB)	0.9356	0.6324 ± 0.0553	0.6907	0.0161	0.0347 ± 0.0029	0.0345
Random Forest	0.9552	0.5333 ± 0.0400	0.5873	0.0125	0.0406 ± 0.0034	0.0418
SVR	0.0492	0.0167 ± 0.0564	0.0608	0.0695	0.0700 ± 0.0034	0.0774
k-NN	0.5604	0.1114 ± 0.0779	0.1614	0.0394	0.0575 ± 0.0033	0.0631
Dec. Tree	1	0.0608 ± 0.1203	0.2338	0	0.0546 ± 0.0040	0.0565
MLR	0.2492	0.2151 ± 0.0486	0.2753	0.055	0.0558 ± 0.0035	0.0599
ANN (MLP)	0.0677	0.1902 ± 0.1014	0.0789	0.0599	0.0554 ± 0.0049	0.0644

Source: own research

Table 3 presents the detailed performance results of the seven ML algorithms. Experimental results indicate that the HistGradientBoosting (HGB) model achieved the best predictive performance. Specifically, the HGB model's R^2 score was 0.6907, significantly outperforming the traditional multivariate linear regression (MLR) model, which achieved an R^2 of 0.2753. The test MAE of the HGB model was also the lowest (0.0345), indicating an average deviation of only about 3.45% between predicted and actual values. Notably, the results also clearly revealed the weaknesses of the individual base models. For instance, the Decision Tree model achieved a perfect score on the training set (Train R^2 = 1) but suffered a significant drop in Validation (Val R^2 = 0.0608 ± 0.1203). This result indicates overfitting, which the HGB algorithm successfully corrected using the ensemble mechanism.

To provide a more visual and insightful understanding of the robustness of the models when faced with different data sets, Figures 2 and 3 illustrate boxplots of the distributions of R^2 and MAE scores from 5-fold group cross-validation. Each dot on the plot represents the result of one fold.

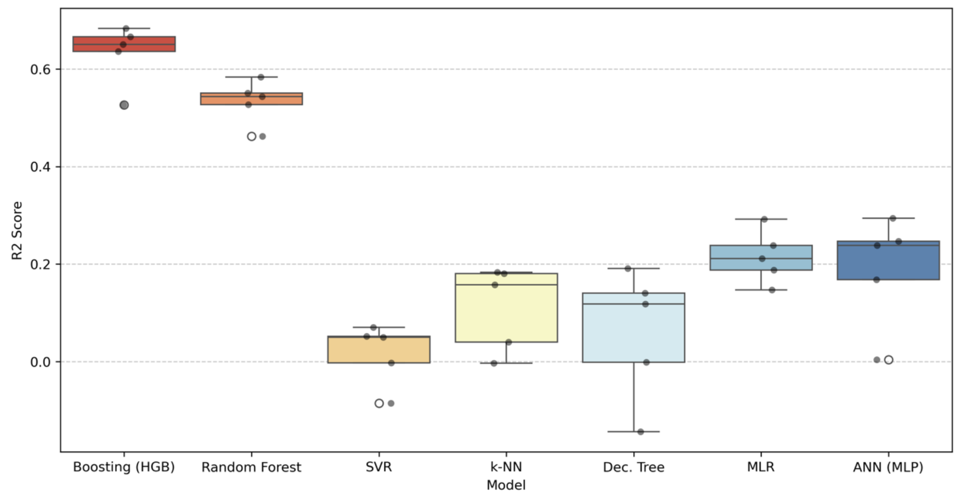

Figure 2. Distribution of the predictive performance R^2 of machine learning models after 5-Fold Group Cross-Validation.

Source: own research

Figure 2 shows that the HGB model not only has the highest median but also a compact box. This demonstrates that the HGB's forecasting performance remains consistently high, with little fluctuation, regardless of which company group is selected as the validation set. The Random Forest model ranks second, but its distribution box is significantly lower. Conversely, algorithms like SVR or Decision Tree show extremely large dispersion, a long box, and even negative R^2 values, meaning the model's forecasting performance is worse than using the mean. The fact that modern ML models achieve high R^2 values confirms their ability to "learn" and track the fluctuating patterns of corporate cash flow more effectively than simple linear models. In particular, the box plot shows a very narrow range of variation for the HGB model with a low standard deviation (0.0553), indicating high reliability and stability of the algorithm across different validation folds.

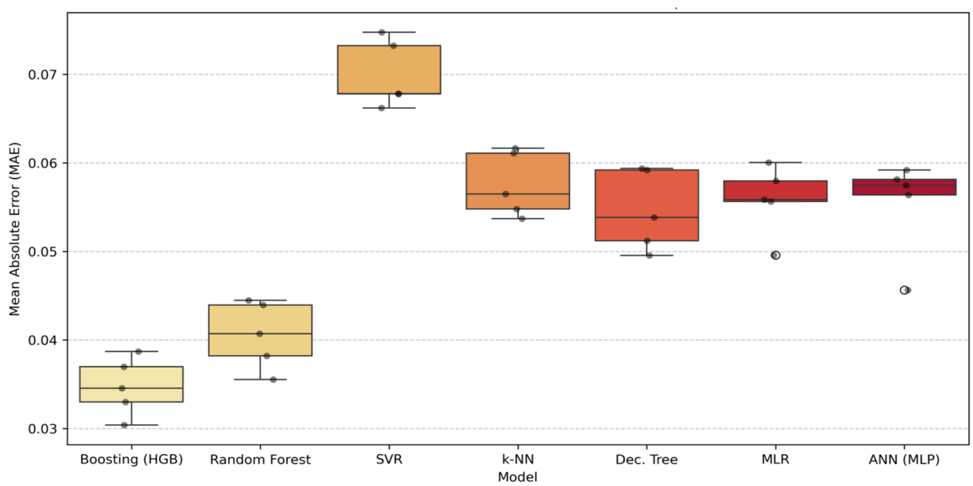

Figure 3. Distribution of the Mean Absolute Error after 5-Fold Group Cross-Validation

Source: own research

In addition, the models' performance was evaluated using the MAE metric, as shown in Figure 3. In this study, MAE serves as a quantitative measure of the margin of error between predicted and actual cash holdings. Experimental results show that the HGB model achieved the lowest MAE of 0.0345, which is significantly lower than that of the MLR model (0.0599). In the context of corporate liquidity management, this is equivalent to an average forecast error of only about 3.45%, an extremely accurate margin given the highly volatile nature of financial data. More importantly, the very small variation in error across five-fold cross-validation further reinforces the model's consistent reliability. Mathematically, this demonstrates that the HGB's forecasting accuracy is truly robust despite structural heterogeneity across company clusters. In stark contrast, traditional parametric models and nonlinear neural networks exhibit significantly higher baseline error ranges (0.055-0.060) and prominent outliers. From a practical perspective, the presence of these extreme outliers is a serious problem in financial forecasting; it implies that, while these baseline algorithms may perform acceptably on average, they suffer from severe local forecasting failures, leading to the complete misjudgment of the cash-holding behavior of certain groups of businesses.

In summary, both the quantitative measures in Table 3 and the visual illustration in the boxplots provide strong statistical evidence that HistGradientBoosting is the most suitable, stable, and reliable algorithm for simulating complex nonlinear relationships in corporate liquidity management data. Therefore, HGB was chosen as the core model for the following steps.

4.4. Hyperparameter Optimization

To optimize forecasting performance and ensure time-efficient processing, the study used a hierarchical model parameter tuning strategy. Instead of allocating resources to all algorithms, the Randomized Search method [29] combined with Group 5-Fold cross-validation was applied only to the HGB model, which demonstrated the best baseline performance. This tuning process took approximately 15.5 seconds, effectively meeting the computational cost requirements for large-scale panel data analysis [30]. The optimization results not only improve accuracy but also reveal the complexity of the research dataset. The optimized model creates an expanded training space through an ensemble of 500 unconstrained-depth trees. This architecture shows that the relationships between factors affecting cash reserves are non-linear and highly interactive. The impact of the WUIVN variable further complicates these relationships. Additionally, to minimize the risk of overfitting, the algorithm automatically calibrated the learning rate to a conservative level of 0.05. Simultaneously, a very strict L2 penalty of 10.0 was applied. This combination creates a strict penalty mechanism. As a result, the model maintains sensitivity to systematic changes while remaining robust enough to filter out market noise. The refined HGB model significantly improved its out-of-sample forecasting ability. The R^2 coefficient on the test set increased to 0.7145, and the MAE narrowed to 0.0333. This optimized version of the model provides a reliable quantitative basis for analyzing and interpreting economic effects in the next step.

4.5. Model Interpretation

To determine the contribution of each input factor to the HGB model's accuracy, the study used the Permutation Feature Importance technique. Using this technique, the values of each variable in the test dataset were randomly shuffled, and the resulting increase in MAE was measured. This process was repeated 10 times for each variable to eliminate random error.

Figure 4 illustrates the distribution of the importance of the input features, combined with error bars representing the standard deviation over 10 iterations. The analysis results from Figure 4 show a very sharp differentiation in the roles of the variable groups. Specifically, the fundamental indicators that directly reflect short-term financial health completely dominate the model's predictive ability. The Current Ratio (CR) and Net Working Capital (NWC) variables have the greatest impact, resulting in the largest increase in MAE (above 0.07) when the data is shuffled. Following closely behind is the Financial Leverage (LEV) index. This is entirely consistent with traditional financial theory and previous empirical studies, confirming that intrinsic characteristics are the foundational determinants of a company's liquidity [19], [20]. Other financial factors, such as Interest Expense (IE), Cash Flow Variance (STD), and Accounts Receivable (AR), show moderate contributions. The dominance of these micro-variables in the ML model stems from their high variability across the entire sample of 4,752 observations, which creates many effective "splits" that reduce forecasting errors. Meanwhile, WUIVN is a macroeconomic variable with only 8 values over 8 years and is measured for all companies. Therefore, the WUIVN feature contributes little to the increase in the model's predictive accuracy.

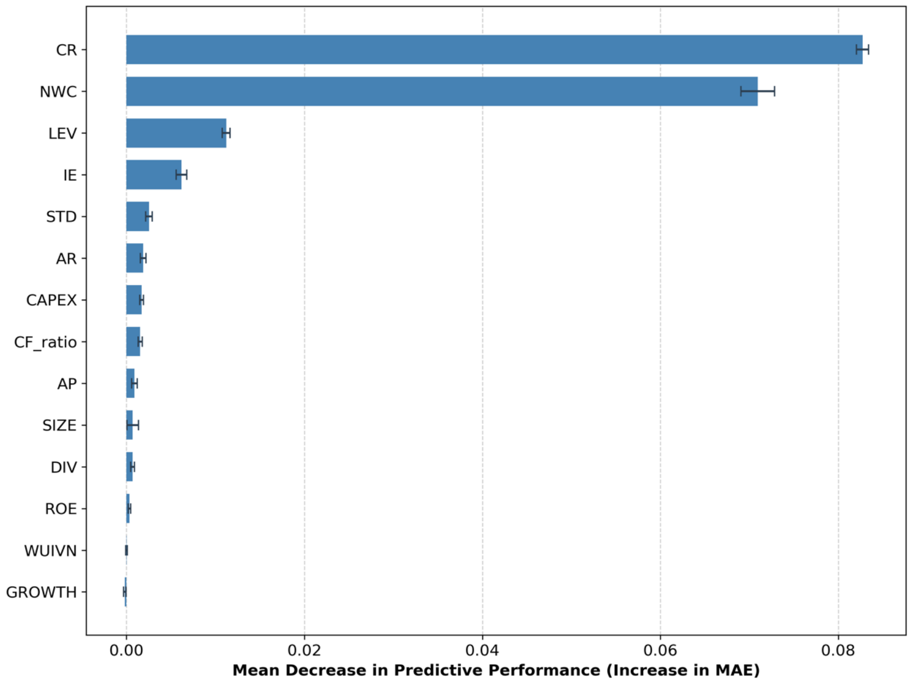

Figure 4. Ranking of the importance of factors influencing cash holdings using the Permutation Importance method.

Source: own research

Consequently, to thoroughly analyze the roles of economic instability and company characteristic variables in the model, the author used the SHAP Summary Plot in Figure 5, which provides a more multidimensional and detailed view of how these factors affect businesses' cash holdings. Accordingly, variables with high importance for forecasting cash holdings, such as NWC, are represented by red dots (high NWC values) concentrated on the left side of the zero axis in Figure 5. This indicates a strong inverse relationship. As other components of working capital (such as inventory and accounts receivable) increase, businesses tend to reduce their cash holdings due to substitution effects or misappropriated capital. Meanwhile, the current ratio (CR) variable shows red dots (high CR values) concentrated to the right of the zero axis. This result indicates that businesses with higher current ratios hold more cash. The interest expense (IE) and short-term debt (STD) variables also show similar results, with red dots clearly located to the right of the zero axis. This shows that as short-term debt increases, businesses face significant payment pressure and refinancing risk. To ensure they do not become insolvent when debts come due, companies tend to "hoard" cash as a safety buffer. Furthermore, when interest costs are high, holding a large cash balance helps businesses ensure timely interest payments, avoid violating loan terms, and maintain creditworthiness with banks. Conversely, the size of the business (SIZE) variable, typically represented by red dots to the left of the zero axis, indicates good access to capital, allowing large corporations to reduce their need to hold cash.

Meanwhile, the WUIVN variable contributes relatively little to the model's total $R^{2,}$ according to the Permutation Importance analysis in Figure 4. The SHAP plot shows a consistent pattern: data points with high WUIVN values (red) are clearly concentrated to the right of the zero axis. This indicates that as macroeconomic instability increases, the marginal contribution (SHAP value) of this variable is positive, or in other words, when economic instability increases, businesses tend to increase their cash holdings. This finding provides solid empirical evidence supporting the precautionary motive in a volatile market like Vietnam.

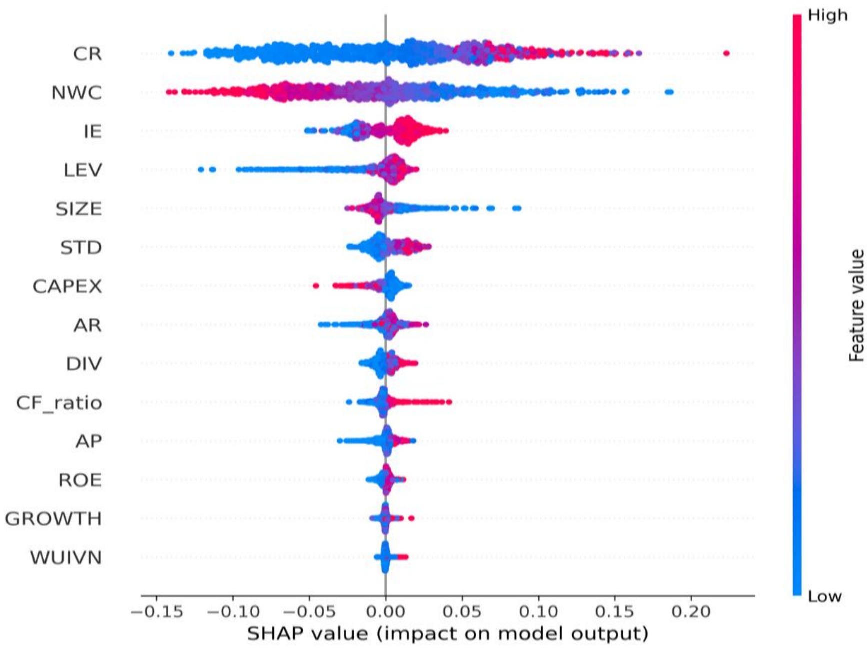

Figure 5. The trend and extent of each factor's contribution to cash holdings are shown in the SHAP Summary Plot.

Source: own research

Furthermore, the SHAP Dependence Plot (Figure 6) provides important empirical evidence on how businesses respond to macroeconomic instability depending on their asset size. Specifically, during periods of economic steadiness, indicated by a low WUIVN value (< 0.1), the marginal impact of this variable on cash holdings is negligible, even resulting in a negative SHAP value. This suggests that, under stable conditions, the precautionary motive to hold cash decreases, and companies will hold less cash to increase capital efficiency. Conversely, when WUIVN surges (especially at 0.60), the SHAP value turns positive and increases sharply. This indicates that businesses activate their emergency cash reserve mechanism only when they recognize sufficiently significant signals of macroeconomic instability that threaten their financial security. Moreover, the color differentiation of the SIZE variable reveals a sharp contrast regarding resilience. At the peak of macroeconomic instability

(WUIVN = 0.60), the data points in the highest deciles of firm size (indicated by red hues) exhibit the highest SHAP value above 0.0075. By contrast, the blue dots (small businesses) are located at the bottom, around 0.0000-0.0050. This indicates that when macroeconomic instability increases, companies with smaller asset bases often face tighter credit constraints, making it difficult for them to increase their cash holdings. Large companies, on the other hand, have better resources and access to capital, allowing them to proactively build large cash reserves as a precaution. This evidence not only enriches the theory of the Contingency Mechanism but also offers important governance implications: liquidity support policies during periods of instability should prioritize small businesses, which are the most vulnerable to market fluctuations.

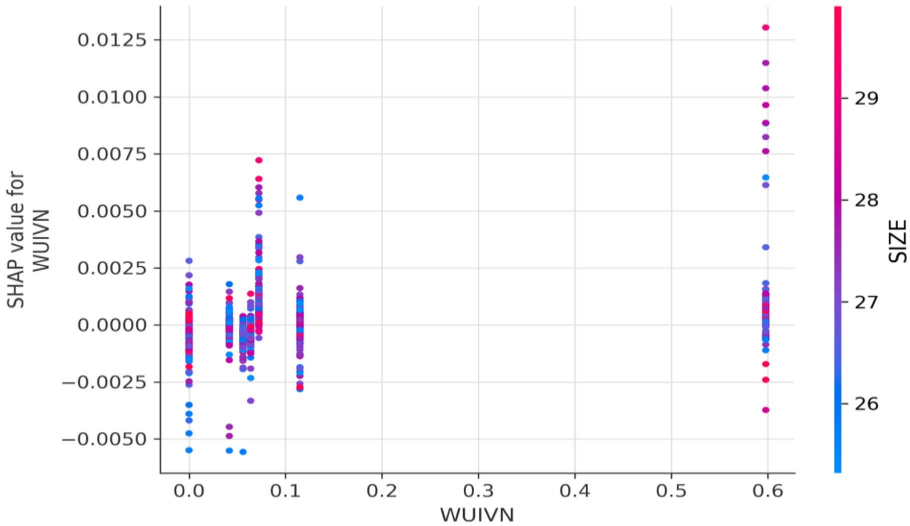

Figure 6. SHAP dependence plot of World Uncertainty Index for Vietnam (WUIVN) on Cash Holdings (CASH), colored by firm size (SIZE).

Source: own research

5. CONCLUSION

This study employs a multidimensional, modern approach to forecast the cash holdings (CASH) of listed companies in Vietnam during 2018–2025. By combining traditional financial theory with ML algorithms, the study highlights several key findings.

Empirical results confirm the robust predictive capabilities of modern nonlinear ML algorithms, particularly ensemble structures such as HistGradientBoosting (R^2 approx. 0.7145), over the traditional linear regression (MLR) approach (R^2 approx. 0.2753). This difference demonstrates that cash-holding behavior in the Vietnamese market exhibits complex interactions and strong nonlinearities.

The research results show that intrinsic characteristics, such as net working capital (NWC), current ratio (CR), firm size (SIZE), and cash flow ratio (CF_ratio), are important factors in cash forecasting. These results are entirely consistent with classical financial theories, confirming that internal factors are the primary determinants of a company's liquidity.

WUIVN's importance is relatively low due to its macroeconomic data characteristics. However, WUIVN does not have a linear impact, but acts as a "trigger" when the market exceeds a safety threshold. In particular, the study discovered a moderating effect of SIZE, indicating that small businesses are significantly more sensitive and vulnerable to systemic shocks. Based on this evidence, liquidity support policies during crises should prioritize smaller entities rather than broad cash-flow easing.

Despite achieving promising results in applying ML models to decode liquidity management behavior, the research still has certain limitations, opening up potential avenues for future development. Specifically, the current study focuses only on companies listed on the Vietnamese stock market. Therefore, the generalizability of the results may be limited when applied to SMEs or unlisted companies, which have significantly different financial structures and access to capital. Furthermore, the study period (2018–2025) includes several extreme shocks such as the COVID-19 pandemic and complex geopolitical shifts. While ensemble models like HistGradientBoosting handle structural breaks well, the lack of validated scenarios under long-term "new normal" economic conditions may affect the sustainability of future forecasts.

REFERENCES

[1] A. Alam, N. Ansar, S. F. Abbas, and Z. U. Rehman, "From data to decisions: Role of machine learning in predicting cash holdings of manufacturing firms in Pakistan," *Journal of Finance and Accounting Research*, vol. 7, no. 1, pp. 78–110, May 2025, doi: https://doi.org/10.32350/jfar.71.04

[2] D. Altig, S. Baker, J. M. Barrero, N. Bloom, P. Bunn, S. Chen, *et al.*, "Economic uncertainty before and during the COVID-19 pandemic," *Journal of Public Economics*, vol. 191, Art. no. 104274, Nov. 2020, doi: https://doi.org/10.1016/j.jpubeco.2020.104274

[3] S. R. Baker, N. Bloom, S. J. Davis, K. Kost, M. Sammon, and T. Viratyosin, "The unprecedented stock market reaction to COVID-19," *The Review of Asset Pricing Studies*, vol. 10, no. 4, pp. 742–758, July 2020, doi: https://doi.org/10.1093/rapstu/raaa008

[4] X. Qin, G. Huang, H. Shen, and M. Fu, "COVID-19 pandemic and firm-level cash holding—moderating effect of goodwill and goodwill impairment," *Emerging Markets Finance and Trade*, vol. 56, no. 10, pp. 2243–2258, July 2020, doi: https://doi.org/10.1080/1540496X.2020.1785864

[5] H. Ahir, N. Bloom, and D. Furceri, "The world uncertainty index," National Bureau of Economic Research, Cambridge, MA, Working Paper No. w29763, Feb. 2022, doi: https://www.nber.org/papers/w29763

[6] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geoscience Frontiers*, vol. 12, no. 1, pp. 469–477, Jan. 2021, doi: https://doi.org/10.1016/j.gsf.2020.03.007

[7] T. Opler, L. Pinkowitz, R. M. Stulz, and R. Williamson, "The determinants and implications of corporate cash holdings" *Journal of Financial Economics*, vol. 52, pp. 3–46, April 1999, doi: https://doi.org/10.1016/S0304-405X(99)00003-3

[8] H. Almeida, M. Campello, and M. Weisbach, "The cash flow sensitivity of cash," *Journal of Finance*, vol. 59, no. 4, pp. 1777–1804, Nov. 2004, doi: https://doi.org/10.1111/j.1540-6261.2004.00679.x

[9] T. Bates, K. Kahle, and R. M. Stulz, "Why do US firms hold so much more cash than they used to?" *Journal of Finance*, vol. 64, pp. 1985–2021, Sep. 2009, doi: https://doi.org/10.1111/j.1540-6261.2009.01492.x

[10] K. R. Song and Y. Lee, "Long-term effects of a financial crisis: Evidence from cash holdings of East Asian firms," *Journal of Financial and Quantitative Analysis*, vol. 47, pp. 617–641, Feb. 2012, doi: https://doi.org/10.1017/S0022109012000142

[11] Y. Lian, M. Sepehri, and M. Foley, "Corporate cash holdings and financial crisis: An empirical study of Chinese companies," *Eurasian Business Review*, vol. 1, no. 2, pp. 112–124, Aug. 2011, doi: https://doi.org/10.14208/BF03353801

[12] P. Dimitropoulos, K. Koronios, A. Thrassou, and D. Vrontis, "Cash holdings, corporate performance and viability of Greek SME: Implications for stakeholder relationship management," *EuroMed Journal of Business*, vol. 15, no. 3, pp. 333–348, Nov. 2019m doi: https://doi.org/10.1108/EMJB-08-2019-0104

[13] Q. T. Tran, "Corporate cash holdings and financial crisis: New evidence from an emerging market," *Eurasian Business Review*, vol. 10, pp. 271–285, Sep. 2019, doi: https://doi.org/10.1007/s40821-019-00134-9

[14] V. T. T. Van, D. N. Hung, N. N. Tram, and A. L. Hoang, "Cash flow and external financing in the Covid pandemic context and financial constraints," *Journal of Organizational Behavior Research*, vol. 7, no. 2, pp. 109–119, Jan. 2022, doi: https://doi.org/10.51847/p5cjIAXsr4

[15] N. T. T. Dao, "Cash holdings and over-investments during Covid-19 pandemic: The evidence from Vietnam," *Universal Journal of Accounting and Finance*, vol. 9, no. 6, pp. 1273–1279, 2021, doi: https://doi.org/10.13189/ujaf.2021.090607

[16] E. Demir and O. Ersan, "Economic policy uncertainty and cash holdings: Evidence from BRIC countries," *Emerging Markets Review*, vol. 33, pp. 189–200, Dec. 2017, doi: https://doi.org/10.1016/j.ememar.2017.08.001

[17] H. V. Phan, N. H. Nguyen, H. T. Nguyen, and S. P. Hegde, "Policy uncertainty and corporate cash holdings," *Journal of Business Research*, vol. 95, pp. 71–82, Feb. 2019, doi: https://doi.org/10.1016/j.jbusres.2018.10.001

[18] T. H. P. Nguyen, G. L. Diep, H. H. G. Bao, and H. P. Le, "Does the world uncertainty index impact firm performance? New evidence from Vietnam," *Asia-Pacific Financial Markets*, pp. 1–25, Aug. 2025, doi: https://doi.org/10.1007/s10690-025-09562-2

[19] H.-C. Wu, J.-H. Chen, and P.-W. Wang, "Cash holdings prediction using decision tree algorithms and comparison with logistic regression model," *Cybernetics and Systems*, vol. 52, no. 8, pp. 689–704, Sep. 2021, doi: https://doi.org/10.1080/01969722.2021.1976988

[20] Ş. Özlem and O. F. Tan, "Predicting cash holdings using supervised machine learning algorithms," *Financial Innovation*, vol. 8, no. 1, Art. no. 44, May 2022, doi: https://doi.org/10.1186/s40854-022-00351-8

[21] H. J. Kim, S. H. Han, and S. Mun, "Analyzing the effects of terrorist attacks on the value of cash holdings," *Finance Research Letters*, vol. 45, Art. no. 102171, Mar. 2022, doi: https://doi.org/10.1016/j.frl.2021.102171

[22] H. Gao, J. Harford, and K. Li, "Determinants of corporate cash policy: Insights from private firms," *Journal of Financial Economics*, vol. 109, no. 3, pp. 623–639, Sep. 2013, doi: https://doi.org/10.1016/j.jfineco.2013.04.008

[23] M. B. Lozano and S. Yaman, "The European financial crisis and firms' cash holding policy: An analysis of the precautionary motive," *Global Policy*, vol. 11, pp. 84–94, Jan. 2020, doi: https://doi.org/10.1111/1758-5899.12768

[24] A. A. S. Manoel, M. B. da Costa Moraes, D. F. L. Santos, and M. F. Neves, "Determinants of corporate cash holdings in times of crisis: Insights from Brazilian sugarcane industry private firms," *International Food and Agribusiness Management Review*, vol. 21, no. 2, pp. 201–218, Nov. 2018, doi: https://doi.org/10.22434/IFAMR2017.0062

[25] A. Ozkan and N. Ozkan, "Corporate cash holdings: An empirical investigation of UK companies," *Journal of Banking & Finance*, vol. 28, no. 9, pp. 2103–2134, Sep. 2004, doi: https://doi.org/10.1016/j.jbankfin.2003.08.003

[26] S. Mangalathu, S. H. Hwang, and J. S. Jeon, "Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach," *Engineering Structures*, vol. 219, Art. no. 110927, Sep. 2020, doi: https://doi.org/10.1016/j.engstruct.2020.110927

[27] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA: O'Reilly Media, 2022, doi: https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/

[28] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, *et al.*, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, Dec. 2017, doi: https://doi.org/10.1111/ecog.02881

[29] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. 2, Mar. 2012, doi: https://dl.acm.org/doi/10.5555/2188385.2188395

[30] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *The Review of Financial Studies*, vol. 33, no. 5, pp. 2223–2273, Feb. 2020, doi: https://doi.org/10.1093/rfs/hhaa009