



A dynamic asset allocation approach under technical analysis and machine learning classification

Gia Buu Luc & Chih-Chiang Wu

To cite this article: Gia Buu Luc & Chih-Chiang Wu (2026) A dynamic asset allocation approach under technical analysis and machine learning classification, *Applied Economics*, 58:33, 6753-6769, DOI: [10.1080/00036846.2025.2526177](https://doi.org/10.1080/00036846.2025.2526177)

To link to this article: <https://doi.org/10.1080/00036846.2025.2526177>

 View supplementary material 

 Published online: 06 Jul 2025.

 Submit your article to this journal 

 Article views: 159

 View related articles 

 View Crossmark data 

 Citing articles: 1 View citing articles 



A dynamic asset allocation approach under technical analysis and machine learning classification

Gia Buu Luc and Chih-Chiang Wu

College of Management, Yuan Ze University, Taoyuan, Taiwan

ABSTRACT

This research evaluates the effectiveness of machine learning models by comparing the performance of seven models in predicting the return movement of four asset types over different return periods. The assets include S&P 500 Futures, Gold Futures, US 10Y Treasury Bond Futures, and a risk-free investment. The study considers 1-week, 4-week, and 12-week forward return periods. The analysis assesses model accuracy and portfolio performance using data spanning two decades. ML-based portfolios modestly outperform an equal-weighted benchmark in terms of risk-adjusted metrics over longer horizons, even though the differences are generally insignificant. However, the study emphasizes the importance of considering diverse economic cycles to understand asset behaviour comprehensively. Nevertheless, the study did not find significant evidence that combining machine learning prediction with dynamic asset allocation can outperform passive portfolio investment. The study underscores the potential of non-linear models in short-term trading but also notes limitations in adaptiveness, threshold design, and model generalization under atypical conditions.

KEYWORDS

Classification; machine learning; model selection; technical analysis; asset allocation; trading algorithm

JEL CLASSIFICATION

C38; C52; C61; G11; G17

1. Introduction

Financial markets are complex systems characterized by significant volatility, interdependencies, and myriad external influences. Predictive modelling and portfolio construction within these markets have been subjects of intense academic and practical interest. Regarding the latter, asset allocation is a cornerstone of modern portfolio theory, embodying the strategic distribution of investment assets across diverse asset classes. Markowitz's (1952) seminal work laid the foundation for modern portfolio theory, which seeks to optimally balance risk and return by spreading investments across various asset categories.

While stocks and bonds have historically been the cornerstone of traditional portfolios, the enduring allure of gold as a store of value has prompted a re-evaluation of portfolio construction strategies. For example, Kinateder, Gurrib, and Choudhury (2024) argued that gold acts as a safe-haven asset during crises, including COVID-19, the Russia/Ukraine war, and the 2008 global financial crisis. Baur and

Lucey (2010) investigated the role of gold as a safe haven asset for stocks in the short term and found that gold exhibits characteristics of a safe haven during turbulent market periods, offering potential diversification benefits for investors. Sherman (1982), Faff and Hillier (2004), and Tully and Lucey (2007) also examined the nature of the gold market and its influence, while Akhtaruzzaman et al. (2021) updated the safe haven characteristics of gold during the COVID-19 pandemic. Erb and Harvey (2005) explored the strategic implications of incorporating a commodity asset into a conventional portfolio framework and found significant advantages to implementing an asset allocation overlay that tactically utilizes exposures to commodity futures. Similarly, regarding digital assets, Bitcoin and stable coins, Giudici, Leach, and Pagnottoni (2022) employed machine learning and volatility spillover analysis to show that basket-based stable coins can provide enhanced stability compared to single-currency-pegged coins, highlighting their potential role as a place to preserve value during market turbulence.

Meanwhile, financial markets widely use technical analysis to forecast future price movements based on historical price data. Brock, Lakonishok, and LeBaron (1992) examined the two standard technical analysis tools: moving average (MA) and trading range break, and found that technical trading rules have predictive power. Gurrib, Nourani, and Bhaskaran (2022) found that Fibonacci retracements capture energy stock prices better than energy cryptos, and that combining the technical analysis tool with price crossover strategies does not improve results. Gencay (1998) provided compelling evidence of non-linear predictability in stock market returns, utilized historical buy and sell signals generated by moving average rules, and highlighted the potential of moving average indicators to effectively forecast market movements, highlighting their relevance in capturing non-linear dynamics within financial markets. Lo, Mamaysky, and Wang (2000) suggested that technical analysis could add value to investment strategies, particularly in markets with predictable stock returns. Additionally, Y. Zhu et al. (2009) found that combining technical analysis with fixed allocation rules can enhance portfolio performance when stock returns exhibit predictability. Marshall, Nguyen, and Visaltanachoti (2017) compared time series momentum and moving average trading rules and found that MA rules frequently signal earlier, leading to meaningful return gains. Similarly, Avramov, Kaplanski, and Levy (2018) showed that technical analysis can significantly predict individual stock returns up to a one-year horizon. Ding et al. (2023) suggested that technical analysis can enhance trading strategies through sentiment-based predictions, and technical indicators indicate that market sentiment influences near-term and long-term return patterns. Additionally, Y. Fang and Xu (2003), Shynkevich (2016), and Lin (2018) support the predictive power of technical analysis.

Machine learning models, characterized by their ability to process and learn from large data volumes, are particularly adept at identifying patterns and trends that may not be evident through traditional analysis. They have introduced powerful tools capable of capturing intricate patterns and making sophisticated predictions, enhancing traditional financial models. Finance applications,

including Shi et al. (2022) credit risk assessment synthesis, Hizioglu's (2013) customer segmentation, Ali et al. (2022) fraud detection, and financial market predictions by Henrique, Sobreiro, and Kimura (2019). Gurrib (2022) also extensively reviewed the use of machine learning in portfolio management.

These capabilities can be leveraged in financial markets to predict asset prices, manage risk, optimize portfolios, and detect fraudulent activities. The flexibility and adaptability of machine learning algorithms make them suitable for dealing with the complex, non-linear nature of financial data. Several seminal works have laid the groundwork for machine learning applications in finance, demonstrating the potential of these advanced analytical tools. Gu et al. (2020) measured risk premiums, setting a foundation for machine learning applications in asset pricing. Bianchi et al. (2021) examined the ability of neural networks and extreme trees in forecasting bond returns, and Geertsema and Lu (2023) examined machine learning in relative valuation and peer firm selection.

Among the most compelling areas in which machine learning has shown promise is using technical analysis for asset allocation and trading strategies, using historical price data and trading volumes to predict future price movements. These models can continuously learn and adapt to new market conditions by integrating machine learning with technical indicators, enhancing their predictive power and robustness over time. Kamalov, Gurrib, and Rajab (2021) found that adding technical indicators to input features with machine learning equalized the effectiveness of price and returns in predicting the direction of price movement.

A study by Murray, Xia, and Xiao (2024) utilized machine learning to forecast stock returns based on historical performance, showing this method's predictive strength in predicting stock returns, especially large stocks with important nonlinearities and interactions in the forecasting function. Their findings demonstrate that machine learning models perform well in optimization and out-of-sample. S. Zhu et al. (2024) investigated factor investing in Chinese commodities markets and found that non-linear integration significantly enhances investment performance when combining various characteristics. Besides

predicting the returns of financial assets, some studies have also considered these models for solving classification problems of technical indicators in financial markets. Gerlein et al. (2016) suggested that combining technical indicators as inputs for machine learning classification can improve forex market profitability. Teixeira and de Oliveira (2010) suggested a combination method of technical analysis and the nearest neighbour classification for stock trading. Dash and Dash (2016) proposed a method that combines machine learning classification and prior-set rules to enhance the trading decisions. In addition to the growing body of machine learning research in finance, recent studies emphasized the importance of explainable AI (XAI), especially when dealing with complex models based on technical indicators. For example, Babaei et al. (2022) applied Shapley values to a dynamic Markowitz-based portfolio model for crypto assets, helping demystify black-box predictions and enhance investor trust in robo-advisors. Similarly, Ahelegbey and Giudici (2022) introduced NetVIX, a network-based volatility index that transparently decomposes global market risk into contagion and volatility effects, offering an interpretable framework by design. These contributions highlight the growing relevance of XAI in financial applications, both from post hoc and model-based perspectives. Beyond predictive accuracy, recent literature has emphasized the importance of robustness and explainability in machine learning models, particularly within technical analysis frameworks. In this regard, Giudici and Raffinetti (2023) introduced the SAFE AI paradigm, which evaluates models in four dimensions: Sustainability, Accuracy, Fairness, and Explainability. Their framework offers a structured approach to assessing the trustworthiness of financial AI systems, aligning with emerging regulatory expectations and helping bridge the gap between performance and transparency. This perspective is especially relevant when deploying machine learning in high-risk financial tasks such as crypto asset management or algorithmic trading.

Regarding machine learning classification combined with technical analysis in asset allocation, this research aims to evaluate the effectiveness of various machine learning models in predicting the

performance of four financial assets – S&P500 Futures, Gold Futures, and US 10Y Treasury Bond Futures in combination with a risk-free asset – across differing forward return intervals. Additionally, it investigates the effects of these predictions on portfolio construction and overall performance. The study hypothesizes that applying machine learning within a dynamic asset allocation strategy incorporating technical indicators can surpass a passive investment's performance, i.e. an equal-weighted portfolio.

This study is divided into several key analyses. First, it investigates the descriptive statistics of these financial assets from March 2002 to April 2024 and during a subsample period from January 2019 to April 2024. This enables a comparative analysis of long-term trends and recent market behaviours, providing insights into these assets' stability and volatility.

The second focus is the performance evaluation of seven machine learning models: neural networks (NN), random forest, stochastic gradient boosting (SGB), support vector machine (SVM), logistic regression, bagged decision trees (Bagging), and K-nearest neighbour (KNN). These models are assessed based on their predictive accuracy, precision, recall, and F1-scores over three forward return periods: 1-week, 4-week, and 12-week.

Lastly, this study investigates the asset weights within these portfolios, providing a detailed allocation breakdown among S&P 500 Futures, US 10Y Treasury Bond Futures, Gold Futures, and US Treasury Bill (Risk-Free) components. By integrating the performance metrics and asset weights, it seeks to identify patterns and potential correlations, offering a comprehensive understanding of portfolio dynamics influenced by machine learning model predictions.

As part of the evaluation process, this study assesses the performance of portfolios constructed using machine learning model-generated signals, comparing them against a traditional equal-weighted strategy. In addition to determining risk-adjusted returns through the Sharpe ratio, the study incorporates a broader set of performance metrics,¹ including Rank Graduation Accuracy (RGA) and

¹Computational performance metrics for all models are reported in Appendix Tables A2–A4. We thank the Associate Editor for this suggestion. Our results indicate that computational efficiency was found to be comparable and reasonable across models when executed on standard hardware configurations.

Area Under the Curve (AUC), to evaluate predictive quality. To ensure robustness, the analysis is extended across different market regimes, including stress-testing during the COVID-19 market crash in 2020 and the Russia – Ukraine war in 2022, capturing the models' stability under volatile conditions.

Moreover, this research endeavours to transcend mere portfolio construction and performance evaluation by considering dynamic hedging techniques based on asset allocation technical indicators. By employing technical analysis tools with machine learning methods, the study proposes an alternative method for dynamic asset allocation and optimizing risk management strategies.

The rest of the paper is structured as follows. [Section II](#) describes the data employed in the analysis. [Section III](#) outlines the research methodology. [Section IV](#) presents the empirical results. [Section V](#) examines the robustness of the results. [Section VI](#) concludes the study.

II. Data

This research utilized data retrieved from Yahoo Finance, encompassing OHLCV (Open, High, Low, Close, Volume) data for four financial assets: S&P 500 Futures, Gold Futures, US 10-Year Treasury Bond Futures, and the 13-week Treasury Bill, which serves as the risk-free asset. The data collection period spans January 2001 to April 2024. Period weekly forward return signals and various technical indicators were subsequently computed as dependent and independent variables, respectively, as suggested by Dash and Dash (2016).

Rather than predicting the precise returns, this study focuses on forecasting return signs, thereby simplifying the implementation of machine learning techniques. To achieve this, the zero threshold was chosen for the simplistic generation of labels for forward returns. Forward returns less than or equal to zero were labelled -1 as a selling signal, and those greater than zero were labelled 1 as a buying signal.

Y. Zhu et al. (2009) suggested that the moving average with a fixed allocation rule is more effective than a complex strategy with prior assumptions. To streamline the calculations in this research, simple

moving averages (SMAs) with periods of 5, 10, 20, 30, and 50 days were employed. Additionally, a widely adopted method for generating buying signals involves observing the occurrence of a crossover where the prices surpass the MA. This occurrence suggests potential buying signals, indicative of recent price movements exhibiting greater strength compared to historical trends, thereby signalling a favourable buying opportunity. Conversely, a selling signal is identified when the MA intersects below the current price, signalling a probable sell signal. Thus, the differences between asset prices and their corresponding prices minus SMAs (denoted as `minus_MA5`, `minus_MA10`, `minus_MA20`, `minus_MA30`, `minus_MA50`) were utilized as independent variables. These differences directly affect return signals and are standard practice in defining buy, sell, and hold signals.

Moreover, this study incorporates the Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), and Stochastic Oscillator as oscillator indicator groups based on Wu and Diao (2015). This suggests that applying this group of indicators can improve the short-term prediction of stock market price changes.

The technical indicator calculation utilizes the Python TA-Lib² package. The data are divided into two distinct subsets: one for training and testing the machine learning models and the other for constructing the portfolio strategy. The models will be trained on historical data from December 2001 to December 2018. Following this training phase, the data from January 2019 to January 2024 will be used to construct the portfolio. This portfolio construction phase incorporates the predicted signal labels generated by the machine learning models, utilizing the most recent predictions to inform investment strategies. This approach leverages historical patterns learned by the models to make informed decisions based on recent market conditions.

[Tables 1 and 2](#) present the descriptive statistics for all assets during the training period (2001–2018) and the portfolio construction period (2019–2024), respectively. For S&P 500 Futures, forward returns in the recent sample are significantly higher: 0.003, 0.010, and 0.029 for 1-week, 4-week, and 12-week

²<https://pypi.org/project/TA-Lib/>.

Table 1. Descriptive statistics for all assets during the full sample period(Dec-2001 - dec-2018).

	Mean	S.D.	Min	Max	Kurtosis	Skewness
S&P500 Futures						
Ret_1wrf	0.001	0.023	-0.175	0.103	6.715	-0.902
Ret_4wrf	0.004	0.043	-0.279	0.174	6.236	-1.401
Ret_12wrf	0.012	0.076	-0.456	0.257	6.078	-1.682
Minus_MA5	3.171	32.105	-178.05	89.9	4.405	-1.418
Minus_MA10	7.57	48.556	-272.275	142.15	4.058	-1.423
Minus_MA20	17.191	69.972	-333.487	219.338	3.575	-1.414
Minus_MA30	26.665	87.942	-392.9	273.575	3.219	-1.419
Minus_MA50	44.289	119.881	-462.44	347.63	2.285	-1.32
MACD_Hist	0.087	9.035	-41.094	26.906	1.827	-0.63
RSI	55.981	12.109	14.616	90.635	0.402	-0.567
Perc_K	67.557	27.466	3.398	99.077	-0.88	-0.677
Perc_D	67.642	26.505	6.559	98.924	-0.872	-0.678
Gold Futures						
Ret_1wrf	0.002	0.025	-0.13	0.111	3.301	-0.602
Ret_4wrf	0.007	0.049	-0.205	0.175	1.482	-0.287
Ret_12wrf	0.021	0.073	-0.233	0.265	0.307	0.013
Minus_MA5	2.232	28.336	-161.3	132.98	4.144	-0.645
Minus_MA10	4.942	40.615	-186.49	191.62	2.333	-0.102
Minus_MA20	10.243	57.136	-249.365	239.86	2.528	-0.25
Minus_MA30	15.602	70.699	-316.53	293.303	2.717	-0.344
Minus_MA50	27.004	92.917	-374.758	363.882	1.948	-0.331
MACD_Hist	-0.07	7.504	-27.935	22.872	1.195	-0.126
RSI	54.709	11.248	22.531	84.049	-0.167	0.02
Perc_K	57.657	29.98	0.147	102.999	-1.247	-0.297
perc_D	57.562	28.973	1.509	99.883	-1.252	-0.293
US 10Y Treasury Bond Futures						
Ret_1wrf	0	0.008	-0.04	0.032	1.668	-0.136
Ret_4wrf	0.001	0.017	-0.069	0.087	2.443	0.216
Ret_12wrf	0.002	0.027	-0.097	0.118	1.216	0.153
Minus_MA5	0.031	1.084	-3.544	6.019	2.299	0.257
Minus_MA10	0.061	1.649	-5.303	10.07	3.14	0.376
Minus_MA20	0.117	2.292	-6.815	11.689	1.267	0.342
Minus_MA30	0.189	2.758	-7.904	12.875	0.857	0.444
minus_MA50	0.35	3.235	-6.889	12.87	-0.052	0.469
MACD_Hist	-0.006	0.33	-1.037	1.451	1.078	-0.089
RSI	50.112	10.822	19.832	74.75	-0.253	-0.113
Perc_K	49.972	29.492	2.049	99.75	-1.368	-0.004
Perc_D	49.892	28.458	2.289	97.051	-1.348	0

horizons, compared to 0.001, 0.004, and 0.012 previously. Standard deviations remain essentially unchanged, except for the 4-week horizon, which increased from 0.043 to 0.055, indicating increasing volatility. Kurtosis for 1-week and 12-week returns declined, suggesting fewer extreme values, while the 4-week kurtosis increased markedly. Skewness became more negative for 1-week and 4-week returns, indicating greater left-tail risk.

For Gold Futures, mean returns remained stable, while volatility declined slightly. Kurtosis fell for the 4-week and 12-week returns, with skewness remaining close to zero, reflecting symmetric distributions. U.S. 10-Year Treasury Bond Futures showed near-zero mean returns and stable volatility. Kurtosis decreased notably for the 4-week and 12-week horizons, while skewness remained close to zero throughout, indicating no significant shift in return asymmetry.

In comparing the training (2001–2018) and portfolio construction (2019–2024) periods, S&P 500 Futures exhibited increased returns and volatility. Meanwhile, Gold and Bond Futures displayed more stable behaviour, with generally reduced tail risks and consistent distributional symmetry.

III. Methodology

This study adopts diverse machine learning algorithms to address the classification task in financial prediction and portfolio construction, with each chosen based on their proven financial application strengths and complementary characteristics. This selection balances model interpretability, computational efficiency, and the ability to capture non-linear relationships within financial data.

Table 2. Descriptive statistics for all assets during the constructing portfolio period(Jan-2019 - jan-2024).

	Mean	S.D.	Min	Max	Kurtosis	Skewness
S&P500 Futures						
Ret_1wf	0.003	0.026	-0.128	0.111	5.549	-1.251
Ret_4wf	0.01	0.055	-0.339	0.183	9.219	-1.913
Ret_12wf	0.029	0.08	-0.287	0.277	2.201	-1.033
Minus_MA5	17.499	103.489	-539.3	207.2	4.067	-1.446
Minus_MA10	37.413	150.784	-722.8	311.15	4.02	-1.556
Minus_MA20	71.062	196.974	-714.213	430.6	1.768	-1.224
Minus_MA30	102.183	234.937	-650.933	527.75	0.937	-1.097
Minus_MA50	164.186	309.899	-666.345	632.61	-0.111	-0.699
MACD_Hist	1.815	27.184	-109.08	48.665	1.73	-1.055
RSI	57.218	10.792	23.726	76.504	-0.245	-0.452
Perc_K	69.618	27.635	4.394	98.971	-0.841	-0.733
Perc_D	69.315	26.749	9.652	98.406	-0.839	-0.717
Gold Futures						
Ret_1wf	0.002	0.021	-0.105	0.1	3.579	-0.087
Ret_4wf	0.007	0.038	-0.084	0.156	0.462	0.357
Ret_12wf	0.023	0.064	-0.133	0.174	-0.531	0.18
Minus_MA5	5.848	37.208	-124.16	135.98	0.682	0.142
Minus_MA10	13.455	53.729	-116.49	215.61	-0.024	0.247
Minus_MA20	28.322	74.953	-161.02	278.645	-0.215	0.072
Minus_MA30	42.107	89.714	-159.817	335.807	-0.041	-0.036
Minus_MA50	67.843	105.901	-183.604	415.018	-0.116	0.158
MACD_Hist	0.576	10.793	-22.173	26.628	-0.717	0.009
RSI	55.755	10.316	32.784	81.289	-0.342	0.096
Perc_K	55.83	30.099	3.098	99.238	-1.352	-0.141
Perc_D	55.967	29.257	4.69	97.698	-1.377	-0.135
US 10Y Treasury Bond Futures						
Ret_1wf	0	0.008	-0.025	0.036	1.86	0.481
Ret_4wf	-0.002	0.017	-0.048	0.054	0.446	0.191
Ret_12wf	-0.005	0.031	-0.091	0.076	-0.131	0.077
Minus_MA5	-0.068	1.097	-2.987	3.872	0.539	0.25
Minus_MA10	-0.135	1.792	-4.491	5.37	0.274	0.111
Minus_MA20	-0.306	2.793	-6.886	7.343	-0.188	0.038
Minus_MA30	-0.503	3.619	-9.005	7.796	-0.586	-0.029
Minus_MA50	-0.825	5.187	-12.22	9.687	-0.811	-0.035
MACD_Hist	0.005	0.344	-0.647	0.816	-0.931	0.354
RSI	48.651	15.714	17.898	82.608	-0.883	0.045
Perc_K	48.616	31.933	2.796	97.804	-1.602	0.075
Perc_D	48.662	31.294	5.99	96.752	-1.622	0.074

Logistic regression

Logistic regression (LR) is a common linear classification model that estimates the probability of a binary outcome based on a linear combination of input features. Due to its simplicity and interpretability, it is frequently employed as a benchmark in financial classification tasks (Lessmann et al. 2015). The core idea involves transforming the linear combination of the input variables into a probability value using the logistic function. This function, also known as the sigmoid function, is defined as $\sigma(z) = \frac{1}{1+e^{-z}}$, where z is the linear combination of the input variables and their corresponding coefficients. The logistic function maps any real-valued number into the interval (0, 1), which can be interpreted as a probability. The LR model can be expressed as: $P(y = 1|X) = \sigma(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n)$, where $P(y = 1|X)$ is the probability that the dependent

variable y equals one given the input vector X , which represents each technical indicator in this study, and $\beta_0, \beta_1, \dots, \beta_n$ are the model parameters.

K-Nearest neighbor

KNN has been used in financial applications for trend prediction and trading signal classification due to its adaptability (Teixeira and de Oliveira 2010). Classified under the instance-based or lazy learning algorithms, KNN refrains from explicitly constructing a model during the training phase. Instead, it retains all available instances within its memory and utilizes a similarity measurement to classify new cases. The core concept underlying KNN is straightforward: objects exhibiting similar characteristics are likely to belong to the same class. The algorithm computes the degree of similarity (or distance) between the input data point and all

other instances in the training set. Subsequently, it identifies the K-nearest neighbours (i.e. the data points with the smallest distances) and assigns the majority class amongst these neighbours to the input data point.

Ensemble method

Bagged decision trees

Bagging aggregates predictions from multiple individual decision trees to enhance predictive performance. It achieves this by training each decision tree on a random training data subset, drawn with replacement. This training data diversity mitigates overfitting and improves generalization. During prediction, the final output is determined through a voting mechanism, wherein each tree contributes equally to the final decision. Mathematically, the prediction process in bagging can be expressed as: $\hat{y} = \text{mode}\{y_{\text{tree}_1}, y_{\text{tree}_2}, \dots, y_{\text{tree}_n}\}$. Here, $\hat{y}$ denotes the predicted class label, and y_{tree_i} represents the output of the i-th decision tree. While simpler than ensemble methods like random forests or gradient boosting, bagging offers a solid baseline with fewer hyperparameters and lower computational cost. It is included here to enable direct comparisons among ensemble techniques, assessing the trade-off between model simplicity and predictive robustness in financial classification tasks.

Random Forest

Given training dataset: $D = \{(x_i, y_i)\}_{i=1}^n$, where x_i represents the feature vector and y_i represents the class label, the random forest algorithm can be described as follows. In the training phase, it generates B bootstrap samples D_b from the original dataset D . For each bootstrap sample D_b , it trains a decision tree T_b using D_b . At each node, it randomly selects m features from the total p features and splits the node using the feature that provides the best split according to a chosen criterion. In the prediction phase, for a new input instance x, each of the B trained decision trees T_b provides a class prediction $\hat{y}_b(x)$. The final predicted class $\hat{y}(x)$ is determined by majority voting among the individual predictions. This method is particularly suitable for financial classification tasks due to its high

predictive accuracy and robustness to overfitting (Caparrini, Arroyo, and Mansilla 2024; Chu 2023). It also provides feature-important measures, aiding model explainability.

Stochastic gradient boosting

SGB combines boosting's strengths and randomness' benefits to produce a more robust and accurate model. Initially, the model begins with a base prediction, typically the mean of the target variable. In each iteration, a new decision tree is trained to fit the residual errors of the previous model to minimize a specified loss function. SGB introduces stochastic sampling by selecting a random subset of the training data with replacement at each iteration. Additionally, it employs feature subsampling by randomly selecting a feature subset for each decision tree split. The final model aggregates all the individual trees, weighted by a learning rate that scales their contributions. It typically achieves superior predictive performance in structured datasets and has gained popularity in financial applications (Fang et al. 2020).

Support vector machine

The core idea behind support vector machine (SVM) is to identify the hyperplane that best separates the data into distinct classes. In a binary classification scenario, this hyperplane is defined as the one that maximizes the margin between the two classes. The margin is the distance between the hyperplane and the nearest data points from each class, known as support vectors. By maximizing this margin, SVM ensures the model has the best generalization capability on unseen data. To handle cases where data is not linearly separable, SVM employs the kernel trick, transforming the original feature space into a higher-dimensional space where a hyperplane can effectively separate the classes. Commonly used kernels include linear, polynomial, radial basis function (RBF), and sigmoid. SVMs can model non-linear relationships using kernel functions and have been shown to perform well in financial time series classification and regime-switching models (Cao and Tay 2003).

Neural Network

Neural networks (NN) are powerful models that capture complex nonlinear patterns through a layered architecture. They are widely used in deep learning for finance, especially in high-frequency trading and algorithmic strategies (Bao et al. 2017). While less interpretable, their predictive power makes them a strong candidate in pattern recognition within noisy financial data. A typical NN is structured into three layers: an input layer, one or more hidden layers, and an output layer. Each layer consists of multiple neurons, and the neurons in adjacent layers are connected. The input layer receives the raw data, which is then processed through the hidden layers where non-linear transformations are applied, and finally, the output layer provides the classification results. The learning process of a NN involves adjusting the connection weights to minimize prediction errors. This is achieved through the backpropagation process, which utilizes gradient descent optimization to iteratively update the weights based on the error gradient.

Hyperparameter tuning

Hyperparameter tuning is a crucial step in the machine learning workflow that significantly affects models' performance and accuracy. Unlike model parameters, which are learned from the data during training, hyperparameters are set before the learning process and control the behaviour of the learning algorithm. Optimized hyperparameters can drastically improve the model's accuracy, generalization, and convergence speed. Conversely, poor hyperparameter choices can cause underfitting or overfitting, where the model either fails to capture the underlying pattern or captures noise as if it were a pattern. Thus, hyperparameter tuning is indispensable for ensuring that a model performs well on the training data and unseen data, thereby enhancing its predictive power and robustness.

To address concerns of overfitting and ensure model generalizability, time-aware cross-validation using TimeSeriesSplit with five splits was

implemented to preserve temporal order and avoid look-ahead bias (Baillie, Kapetanios, and Papailias 2014). This method, implemented using Pedregosa et al. (2011) scikit-learn package,³ is utilized alongside the described machine learning models. By randomly exploring a wide range of hyperparameter values, random search can efficiently identify promising configurations without the exhaustive computational expense associated with grid search. All models were optimized using a randomized hyperparameter search with cross-validation, followed by retraining on the full training set using the best-found parameters. This approach was applied consistently across all classifiers.

Model evaluation

The RGA implemented in this study is based on pairwise comparison. It calculates the proportion of all possible pairs of observations where the relative order of the predicted values matches that of the actual values. The formula is:

$$RGA = \frac{\text{Number of correctly ordered pairs}}{\text{Total number of pairs}}. \quad (1)$$

Specifically, for each sample pair, if the differences in true and predicted values have the same sign, the pair is considered correctly ordered; if either difference is zero, the pair contributes 0.5 to the correct count. The total RGA score is the average correctness across all unique pairs.

AUC provides a robust measure of a model's ability to discriminate between classes, independent of any specific classification threshold. It represents the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. Here, AUC is computed based on either class probability estimates or decision function scores. This implementation ensures that each model's scoring mechanism is used optimally for evaluation. The AUC is defined as:

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx. \quad (2)$$

³https://scikit-learn.org/stable/supervised_learning.html.

Portfolio construction

After generating the predicted signals from machine learning models, the signals are integrated into a trading strategy utilizing technical indicators. As delineated earlier, these signal labels are defined as follows: -1 for selling signals and 1 for buying signals. At a given trade period t , an asset labelled -1 will be sold by reducing its value by an up-down limit percentage (udl) and accounting for a percentage of transaction fee (fee). Conversely, an asset labelled 1 will be bought by increasing its value by the same up-down limit percentage with a transaction fee. For example, if Gold Futures is assigned a label of -1 at trade period t , its value will be adjusted as follows: $\text{Gold Value}_t \times (1 - udl - fee)$.

The trading strategy is executed weekly for a strategy with a 1-week forward return, every 4 weeks for a 4-week forward return, and every 12 weeks for a 12-week forward return. Additionally, the portfolio rebalancing period differs based on the forward return interval: it is performed every 24 weeks for the 1-week forward return strategy and every 52 weeks for the 4-week and 12-week forward return strategies. This trading strategy is applied exclusively to three assets: S&P 500 Futures, Gold Futures, and US 10Y Treasury Bond Futures. The risk-free rate, represented by the 13-week Treasury Bill, is utilized to manage the investment funds allocated to these strategic assets. Specifically, if the total value of the strategic assets increases, the corresponding amount is withdrawn from the risk-free fund. Conversely, if the total value of the strategic assets decreases, funds are deposited back into the risk-free asset. If the fund's balance falls below zero, no transactions are executed. Thus, the strategy does not allow short-selling.

This strategy's performance is evaluated by comparison against an equal-weighted portfolio comprising of the same four assets: S&P 500 Futures, Gold Futures, US 10Y Treasury Bond Futures, and US 13-week Treasury Bond. The equal-weighted portfolio rebalancing period aligns with the forward return intervals: weekly for the 1-week forward return, every 4 weeks for the 4-week forward return, and every 12 weeks for the 12-week forward return. This alignment ensures that the number of trades matches the strategy based on machine learning predicted signals.

Both the strategic portfolio and the equal-weighted portfolio and the equal-weighted portfolio incur a transaction cost of 0.1% of the trading amount per trade. This cost is factored into the portfolio's performance to provide a realistic assessment of its efficacy. Moreover, the value adjustment for each asset in the portfolio, given the predicted signal at the trade period t , is calculated as follows: $\text{Asset Value}_t = \text{Asset Value}_{t-1} \times (1 + \text{label}_t \times udl - |\text{label}_t \times fee|)$. In this formula, label represents the predicted label for the asset at trade period t with the values -1 or 1 , corresponding to sell or buy signals, respectively, while udl is the up-down limit percentage, specified as 0.05 . Lastly, fee represents the transaction fee percentage.

IV. Empirical results

Performance metrics of machine learning models

Table 3 shows the comparative performance across models and forecast horizons, revealing several meaningful insights. For the 1-week forward return, most models performed close to random, with RGA values approximating 0.5 across assets, indicating the difficulty of capturing short-term market noise. However, KNN stood out on Gold Futures with the highest RGA (0.533) and AUC (0.572) for 1-week, suggesting potential localized patterns in short-term gold price movements that simpler, instance-based models can exploit. In the 4-week horizon, NN and random forest consistently performed well across all three assets, with NN achieving the highest AUC for T-Bond Futures (0.606), and random forest showing strong RGA across S&P500 (0.538) and Gold (0.508), likely due to their capacity to model nonlinearities and capture intermediate trends. Notably, SVM outperformed all other models on T-Bond Futures in terms of AUC (0.662), underscoring its strength in handling medium-term volatility in fixed-income markets. However, in the 12-week horizon, all models generally struggled, especially with S&P500 and Gold, where AUC values were substantially lower, possibly due to structural shifts or decaying predictor relevance. An exception was SVM on Gold Futures, with an impressive AUC of 0.582 , hinting at persistent long-term

commodity pricing patterns that it could capture. These results suggest that model performance greatly depends on asset characteristics and forecast horizons, with medium-term windows offering the most reliable signals, especially for fixed-income and commodity markets.

Portfolio performance

Portfolio evaluation

The analysis of the performance metrics in Table 4 reveals that, across most strategies, the weekly return differences relative to the equal-weighted portfolio are statistically insignificant, as indicated by low t -statistics.⁴ In Panel A, all strategies report weekly returns between 0.10% and 0.15%, with corresponding t -statistics ranging from -0.15 to 0.33 . None of these values exceed conventional thresholds for statistical significance (e.g. $t > 1.96$), suggesting that no strategy exhibits a reliably superior short-term return compared to the equal-weighted baseline. This outcome indicates the inherent difficulty of extracting meaningful predictive signals over short-term horizons in financial markets, where noise often dominates signals.

Despite this, there are notable patterns in other risk-adjusted performance measures that offer insights into the potential value of these strategies. In particular, LR stands out in Panel A, delivering the highest Sharpe ratio (8.750%) and the lowest maximum drawdown (-5.2%), with a modest 38.83% total return. While the weekly return is not statistically significant, the enhanced risk-adjusted profile suggests that this strategy may better manage downside volatility and capital preservation, even if its return-generating capacity is statistically insignificant. Similarly, SGB consistently performs well in Panels A and B, posting relatively higher Sharpe ratios and lower drawdowns compared to most other models.

Moreover, when the investment horizon is extended to 12 weeks, some strategies begin to show more consistent improvements. For example, random forest in Panel C achieves a higher total return (33.81%) and a 9.61% Sharpe ratio, outperforming the equal-weighted portfolio by

1.03 percentage points in risk-adjusted terms. While still statistically insignificant by return alone ($t = 0.17$), these Sharpe and Sortino ratio improvements over longer horizons hints at ML models' potential value in medium-term tactical asset allocation, where short-term noise is smoothed and broader structural patterns are more accessible.

The lack of statistical significance in weekly returns suggests that ML-based strategies may not provide robust alpha at short horizons, especially in the presence of transaction costs and market frictions. However, improvements in risk-adjusted performance, lower drawdowns, and better volatility targeting imply that these strategies could remain valuable parts of a broader portfolio framework focused on capital preservation and smoother equity curves. Future research may benefit by integrating macroeconomic or fundamental signals, using ensemble learning across multiple time horizons, or dynamically adjusting strategies in response to regime changes, thereby enhancing both the statistical and economic significance of results.

V. Robustness analysis

Model evaluation

To evaluate the robustness and generalizability of our machine learning classifiers, out-of-sample testing was conducted on two exogenous crisis periods in Tables 5 and 6: the COVID-19 pandemic (February – June 2020) and the early Ukraine-Russia war (February – June 2022). All models were exclusively trained on December 2001 to December 2018 data, ensuring that these high-volatility periods were fully evaluated. This temporal separation allows for a stringent assessment of each model's ability to generalize under abrupt regime shifts.

To guard against overfitting during the model development phase, a time-aware cross-validation strategy was employed using the TimeSeriesSplit technique with five folds. This preserved the financial data temporal sequence, mitigating look-ahead bias and ensuring realistic evaluation conditions. Hyperparameter optimization was conducted via randomized search,

⁴Appendix Table A1 shows how machine learning portfolio strategies allocate capital across four assets over various forecast horizons.

Table 3. Performance metrics of machine learning models for S&P500 Futures, Gold Futures, and US 10Y Treasury Bond Futures.

	S&P500 Futures		Gold Future		T-Bond Futures	
	RGA	AUC	RGA	AUC	RGA	AUC
Panel A: 1-week forward return						
Neural Network	0.542	0.588	0.497	0.545	0.497	0.520
Random Forest	0.497	0.494	0.490	0.477	0.500	0.495
Logistic Regression	0.498	0.485	0.502	0.521	0.519	0.543
Bagging	0.463	0.427	0.491	0.503	0.498	0.528
SGB	0.516	0.537	0.476	0.483	0.499	0.534
SVM	0.500	0.502	0.500	0.473	0.511	0.476
KNN	0.469	0.447	0.533	0.572	0.484	0.479
Panel B: 4-week forward return						
Neural Network	0.512	0.512	0.460	0.397	0.521	0.606
Random Forest	0.538	0.534	0.508	0.491	0.510	0.541
Logistic Regression	0.500	0.420	0.498	0.438	0.509	0.672
Bagging	0.494	0.548	0.522	0.540	0.519	0.544
SGB	0.514	0.466	0.488	0.471	0.504	0.549
SVM	0.500	0.430	0.501	0.455	0.516	0.662
KNN	0.520	0.507	0.454	0.396	0.510	0.555
Panel C: 12-week forward return						
Neural Network	0.499	0.479	0.430	0.280	0.478	0.463
Random Forest	0.496	0.208	0.489	0.422	0.491	0.462
Logistic Regression	0.500	0.298	0.493	0.437	0.474	0.486
Bagging	0.471	0.317	0.496	0.466	0.472	0.437
SGB	0.497	0.301	0.483	0.398	0.504	0.485
SVM	0.500	0.596	0.466	0.582	0.499	0.485
KNN	0.471	0.309	0.495	0.452	0.499	0.510

Table 4. Portfolios performance evaluation.

	Weekly Return	Sharpe Ratio	Sortino Ratio	Max Drawdown	Total Return
Panel A: 1-week forward return					
Neural Network	0.11% (−0.07)	7.95% (−0.78)	−2.68% (−0.50)	−4.13%	29.06%
Random Forest	0.12% (0.01)	7.71% (−0.92)	−4.99% (−1.25)	−3.95%	30.79%
SGB	0.12% (0.04)	7.70% (−0.95)	−4.96% (−1.23)	−5.07%	31.63%
SVM	0.12% (0.02)	5.92% (−2.49)	−4.09% (−1.01)	−4.75%	31.05%
Logistic Reg	0.15% (0.31)	8.75% (−0.14)	−0.57% (0.21)	−5.20%	38.83%
Bagging	0.11% (−0.11)	6.54% (−1.82)	−6.86% (−1.81)	−4.44%	27.86%
KNN	0.11% (−0.13)	7.72% (−0.90)	−4.51% (−1.12)	−2.88%	27.62%
Equal Weighted	0.12%	8.91%	−1.22%	−3.99%	31.92%
Panel B: 4-week forward return					
Neural Network	0.14% (0.28)	9.57% (0.54)	−1.97% (−0.18)	−5.47%	37.22%
Random Forest	0.13% (0.12)	8.63% (−0.20)	−3.56% (−0.71)	−3.70%	33.01%
SGB	0.13% (0.13)	8.93% (0.03)	−3.29% (−0.61)	−3.77%	33.36%
SVM	0.14% (0.25)	9.30% (0.31)	−4.03% (−0.90)	−5.26%	36.38%
Logistic Reg	0.13% (0.19)	9.54% (0.51)	−3.24% (−0.59)	−5.56%	34.80%
Bagging	0.10% (−0.17)	7.24% (−1.29)	−4.14% (−0.90)	−3.07%	26.59%
KNN	0.12% (0.10)	8.84% (−0.05)	−3.44% (−0.66)	−3.66%	32.67%
Equal Weighted	0.12%	8.89%	−1.42%	−4.00%	31.68%
Panel C: 12-week forward return					
Neural Network	0.12% (0.07)	8.71% (0.10)	−1.87% (−0.07)	−4.02%	31.48%
Random Forest	0.13% (0.17)	9.61% (0.81)	−1.42% (0.08)	−4.86%	33.81%
SGB	0.13% (0.14)	9.35% (0.62)	−0.95% (0.23)	−4.66%	33.10%
SVM	0.13% (0.15)	9.26% (0.54)	−1.07% (0.19)	−4.92%	33.47%
Logistic Reg	0.12% (0.12)	9.28% (0.56)	−1.71% (−0.02)	−4.85%	32.59%
Bagging	0.13% (0.14)	9.46% (0.70)	−1.09% (0.18)	−4.65%	33.08%
KNN	0.13% (0.13)	9.38% (0.63)	−1.49% (0.05)	−4.87%	32.99%
Equal Weighted	0.11%	8.58%	−1.65%	−3.92%	31.26%

and final models were retrained on the full training set before testing on the stress periods. While these design elements substantially reduced overfitting risk, concerns remain in the context of short test windows (~20 weeks) and elevated noise during crises, which can cause performance metrics to become unstable.

Gold futures displayed the most consistent predictive performance across both stress regimes. During the COVID-19 period, the KNN classifier achieved an AUC of 0.75 and RGA of 0.54–0.57 for 1-week and 4-week periods, while LR delivered an AUC of 0.87 for the latter period. Similarly, during the Ukraine-Russia war, SVM

Table 5. Performance metrics of machine learning models for S&P500 Futures, Gold Futures, and US 10Y Treasury Bond Futures during COVID-19 period.

	S&P500 Futures		Gold Future		T-Bond Futures	
	RGA	AUC	RGA	AUC	RGA	AUC
Panel A: 1-week forward return						
Neural Network	0.500	0.404	0.500	0.533	0.500	0.788
Random Forest	0.512	0.385	0.550	0.689	0.545	0.635
Logistic Regression	0.500	0.221	0.557	0.622	0.510	0.615
Bagging	0.421	0.313	0.550	0.644	0.564	0.663
SGB	0.493	0.510	0.521	0.556	0.502	0.538
SVM	0.500	0.471	0.500	0.422	0.557	0.668
KNN	0.598	0.721	0.571	0.744	0.517	0.409
Panel B: 4-week forward return						
Neural Network	0.500	0.588	0.507	0.778	0.510	0.735
Random Forest	0.500	0.544	0.536	0.700	0.538	0.603
Logistic Regression	0.500	0.544	0.636	0.867	0.500	0.471
Bagging	0.519	0.787	0.471	0.467	0.467	0.250
SGB	0.500	0.647	0.479	0.533	0.510	0.515
SVM	0.500	0.426	0.500	0.233	0.500	0.647
KNN	0.500	0.397	0.543	0.750	0.526	0.463
Panel C: 12-week forward return						
Neural Network	0.500	0.700	0.500	NA	0.512	0.300
Random Forest	0.500	0.750	0.500	NA	0.505	0.700
Logistic Regression	0.500	0.050	0.500	NA	0.471	0.350
Bagging	0.500	0.500	0.500	NA	0.505	0.600
SGB	0.500	0.950	0.500	NA	0.464	0.150
SVM	0.500	0.100	0.500	NA	0.469	0.200
KNN	0.500	0.175	0.500	NA	0.521	0.900

AUC values are reported as NA when only one class label is present in the test set.

Table 6. Performance metrics of machine learning models for S&P500 Futures, Gold Futures, and US 10Y Treasury Bond Futures during the beginning period of Ukraine-Russia war.

	S&P500 Futures		Gold Future		T-Bond Futures	
	RGA	AUC	RGA	AUC	RGA	AUC
Panel A: 1-week forward return						
Neural Network	0.500	0.575	0.539	0.450	0.522	0.750
Random Forest	0.543	0.688	0.556	0.575	0.565	0.558
Logistic Regression	0.500	0.275	0.509	0.458	0.500	0.575
Bagging	0.500	0.392	0.517	0.563	0.543	0.658
SGB	0.500	0.438	0.500	0.525	0.496	0.492
SVM	0.500	0.333	0.500	0.542	0.500	0.583
KNN	0.500	0.650	0.587	0.729	0.587	0.725
Panel B: 4-week forward return						
Neural Network	0.500	0.588	0.448	0.562	0.500	0.706
Random Forest	0.500	0.476	0.494	0.381	0.496	0.447
Logistic Regression	0.500	0.412	0.383	0.152	0.500	0.929
Bagging	0.550	0.635	0.444	0.338	0.481	0.535
SGB	0.485	0.400	0.461	0.362	0.506	0.447
SVM	0.500	0.435	0.418	0.210	0.500	0.929
KNN	0.500	0.500	0.483	0.329	0.517	0.565
Panel C: 12-week forward return						
Neural Network	0.500	0.667	0.478	0.500	0.500	0.681
Random Forest	0.500	0.278	0.539	0.675	0.500	0.347
Logistic Regression	0.500	0.069	0.504	0.500	0.500	0.694
Bagging	0.500	0.444	0.509	0.663	0.500	0.292
SGB	0.500	0.493	0.509	0.625	0.500	0.375
SVM	0.500	0.069	0.500	1.000	0.500	0.444
KNN	0.500	0.597	0.483	0.438	0.509	0.778

reached an AUC of 1 for 12-weeks and KNN maintained strong RGA (0.729) and AUC (0.587) for 1-week. These results suggest that gold futures retained a relatively stable behavioural structure amid global shocks, particularly in short horizons, where flight-to-safety dynamics

often dominate. The consistency in both rank-based and probabilistic metrics supports these models' suitability for forecasting directional trends in gold under stress.

In contrast, S&P 500 futures exhibited limited learnability during both crisis periods, with most

classifiers clustering around chance-level AUC values. For example, during the Ukraine-Russia war, NN and LR both yielded AUCs of 0.575 and 0.275, respectively, for 1-week, with corresponding RGA values stagnant at 0.500. These results suggest that directional predictability substantially diminished in equity markets, possibly due to market participants' abrupt and heterogeneous reactions during crises. Similar degradation was observed during COVID-19, reinforcing the idea that models trained on relatively stable precrisis dynamics may not capture chaotic equity movements under duress.

Treasury bond futures showed mixed results. While performance varied across classifiers and horizons, certain models, such as KNN during COVID-19, reported high AUC scores (e.g. 0.9 at the 12-week horizon). Additionally, LR and SVM achieved an RGA of 0.929 at the 4-week horizon during the Ukraine-Russia war. However, such high AUCs warrant cautious interpretation, especially in short-term testing windows. Given the sparse observations and heightened variance during these periods, occasional overperformance may reflect noise fitting rather than generalizable signal extraction.

Taken together, the use of held-out crisis periods for testing and time-aware validation during

training provides a robust defence against overfitting. Yet, the potential for spurious patterns remains, particularly when extreme performance metrics are observed in small datasets. Across all assets, gold futures' stability stands out, with KNN demonstrating credible robustness in both RGA and AUC. In contrast, equity markets remain difficult to forecast during crises, and bond futures offer selective predictability with model- and horizon-specific caveats. These findings underscore the importance of model diversification and asset-specific strategies when applying machine learning in volatile financial regimes.

Portfolio performance

To evaluate the stability and reliability of the machine learning-based technical analysis strategies, their performance under two distinct crisis periods are examined in [Tables 7 and 8](#): the COVID-19 market crash (February to June 2020) and the initial Ukraine-Russia war period (February to June 2022). By comparing these results with the overall performance observed during the full sample period from January 2019 to April 2024, we aim to assess the degree to which these strategies remain robust under extreme market stress.

Table 7. Portfolios performance evaluation for COVID-19 period Feb 2020 - Jun 2020.

	Weekly Return	Sharpe Ratio	Sortino Ratio	Max Drawdown	Total Return
Panel A: 1-week forward return					
Neural Network	0.09% (−0.13)	23.90% (1.13)	6.80% (−2.99)	−4.13%	1.71%
Random Forest	0.19% (0.04)	25.51% (1.63)	10.97% (4.34)	−3.95%	3.86%
SGB	0.16% (−0.00)	20.48% (−0.03)	7.53% (−1.35)	−5.07%	3.07%
SVM	0.17% (0.00)	17.68% (−0.85)	6.30% (−2.78)	−4.75%	3.23%
Logistic Reg	0.16% (0.00)	18.82% (−0.50)	6.53% (−2.56)	−5.20%	3.13%
Bagging	0.04% (−0.19)	17.73% (−0.76)	2.02% (−10.07)	−4.44%	0.56%
KNN	0.17% (0.00)	26.85% (1.88)	9.99% (2.74)	−2.88%	3.38%
Equal Weighted	0.16%	20.60%	nan%	−3.99%	3.23%
Panel B: 4-week forward return					
Neural Network	0.17% (0.01)	22.36% (0.43)	8.25% (0.07)	−5.47%	3.31%
Random Forest	0.16% (−0.00)	24.36% (1.02)	8.63% (0.70)	−3.70%	3.16%
SGB	0.17% (0.02)	24.39% (0.99)	8.73% (0.84)	−3.77%	3.38%
SVM	0.17% (0.01)	23.31% (0.67)	8.25% (0.07)	−5.26%	3.31%
Logistic Reg	0.15% (−0.02)	23.04% (0.59)	7.73% (−0.78)	−5.56%	2.96%
Bagging	0.13% (−0.05)	22.36% (0.41)	7.23% (−1.66)	−3.07%	2.52%
KNN	0.16% (−0.01)	24.34% (1.03)	8.40% (0.32)	−3.66%	3.07%
Equal Weighted	0.16%	20.97%	nan%	−4.00%	3.18%
Panel C: 12-week forward return					
Neural Network	0.14% (−0.01)	22.14% (0.56)	7.75% (−0.08)	−4.02%	2.82%
Random Forest	0.15% (−0.00)	22.37% (0.61)	7.79% (−0.01)	−4.86%	2.98%
SGB	0.15% (0.00)	21.93% (0.48)	7.78% (−0.03)	−4.66%	3.02%
SVM	0.15% (0.00)	22.20% (0.56)	7.78% (−0.03)	−4.92%	3.02%
Logistic Reg	0.15% (−0.01)	22.61% (0.69)	7.72% (−0.13)	−4.85%	2.87%
Bagging	0.15% (−0.00)	21.88% (0.47)	7.79% (−0.01)	−4.65%	2.98%
KNN	0.15% (−0.00)	22.19% (0.55)	7.80% (0.01)	−4.87%	2.99%
Equal Weighted	0.15%	20.25%	nan%	−3.92%	3.01%

Table 8. Portfolios performance evaluation for Ukraine-Russia war period Feb 2022 - Jun 2022.

	Weekly Return	Sharpe Ratio	Sortino Ratio	Max Drawdown	Total Return
Panel A: 1-week forward return					
Neural Network	−0.33% (−0.06)	−0.04% (−0.18)	−2.13% (−0.20)	−4.13%	−7.01%
Random Forest	−0.35% (−0.12)	−4.15% (−1.48)	−4.57% (−0.79)	−3.95%	−8.02%
SGB	−0.30% (−0.00)	−2.26% (−0.90)	−2.56% (−0.33)	−5.07%	−7.49%
SVM	−0.35% (−0.10)	−0.25% (−0.24)	−1.65% (−0.09)	−4.75%	−7.82%
Logistic Reg	−0.30% (0.00)	2.15% (0.50)	0.14% (0.42)	−5.20%	−7.24%
Bagging	−0.35% (−0.11)	−4.43% (−1.49)	−3.84% (−0.67)	−4.44%	−7.86%
KNN	−0.40% (−0.21)	−2.49% (−0.86)	−4.55% (−0.76)	−2.88%	−8.36%
Equal Weighted	−0.30%	0.54%	−1.33%	−3.99%	−6.69%
Panel B: 4-week forward return					
Neural Network	−0.35% (−0.10)	2.60% (0.63)	−0.90% (0.13)	−5.47%	−7.56%
Random Forest	−0.33% (−0.05)	−2.07% (−0.76)	−3.50% (−0.54)	−3.70%	−7.18%
SGB	−0.30% (0.01)	−0.62% (−0.32)	−3.66% (−0.58)	−3.77%	−6.67%
SVM	−0.36% (−0.13)	1.59% (0.33)	−1.74% (−0.09)	−5.26%	−7.82%
Logistic Reg	−0.35% (−0.09)	2.76% (0.67)	−2.71% (−0.33)	−5.56%	−7.51%
Bagging	−0.32% (−0.03)	−2.47% (−0.85)	−4.03% (−0.66)	−3.07%	−6.99%
KNN	−0.33% (−0.07)	−2.16% (−0.78)	−3.69% (−0.57)	−3.66%	−7.30%
Equal Weighted	−0.30%	0.47%	−1.39%	−4.00%	−6.73%
Panel C: 12-week forward return					
Neural Network	−0.31% (−0.01)	0.23% (0.08)	−1.60% (0.05)	−4.02%	−6.84%
Random Forest	−0.31% (−0.01)	2.19% (0.66)	−1.29% (0.13)	−4.86%	−6.84%
SGB	−0.29% (0.04)	2.15% (0.65)	0.15% (0.52)	−4.66%	−6.46%
SVM	−0.31% (−0.01)	1.98% (0.61)	0.07% (0.48)	−4.92%	−6.84%
Logistic Reg	−0.29% (0.04)	2.22% (0.68)	0.30% (0.55)	−4.85%	−6.46%
Bagging	−0.28% (0.06)	2.24% (0.68)	−1.12% (0.18)	−4.65%	−6.30%
KNN	−0.31% (−0.01)	2.04% (0.62)	−1.34% (0.12)	−4.87%	−6.84%
Equal Weighted	−0.31%	−0.05%	−1.79%	−3.92%	−6.77%

During the COVID-19 shock characterized by heightened volatility and sharp directional movements, most machine learning strategies delivered remarkable performance gains. Models such as random forest, SGB, and KNN achieved high Sharpe and Sortino ratios, alongside positive total returns, substantially outperforming equal-weighted benchmarks. For example, the KNN strategy applied to 1-week forward returns exhibited a Sharpe Ratio of 26.85%, compared to 20.6% for the equal-weighted portfolio. These performance differentials were both economically significant and statistically robust. Such outcomes suggest that machine learning models are highly effective at identifying short-term mispricing exploitable through technical signals during environments dominated by volatility clustering and momentum breakdowns.

Sharply contrasting, the same models' performance deteriorated markedly during the early phase of the Ukraine-Russia war. All model-driven strategies posted negative average weekly returns across 1-week, 4-week, and 12-week forecast horizons. The magnitude of underperformance was particularly notable for KNN and random forest, both of which showed total returns lower than −8% over 1-week forward return periods, worse than that of the equal-

weighted portfolio. Except for the LR Sharpe and Sortino ratios during this period, which were generally close to zero or negative, no strategy demonstrated statistically significant superiority over the benchmark. The equal-weighted portfolio, although also negative in return, experienced smaller drawdowns and exhibited relatively greater stability.

The divergent results between the two crisis periods highlight the regime-dependence of machine-learning-based strategies. While these models adapt well to structurally volatile regimes like the COVID-19 crash, where technical indicators may capture shifts in momentum or volatility, they appear fragile in environments driven by geopolitical shocks. War introduces discontinuities and non-stationary behaviour in price dynamics that are not easily captured by models trained on historical data dominated by endogenous market behaviour. Unlike volatility-driven crises, geopolitical events often trigger regime shifts that disrupt the statistical regularities upon which technical models rely.

Furthermore, the decline in performance during the war period suggests potential overfitting to pre-crisis dynamics or insufficient sensitivity to external macro-political drivers. This observation, combined with the full-period findings, highlights the need for hybrid modelling frameworks that

incorporate both technical signals and sentiment analysis, macroeconomic indicators, or regime-switching components to improve resilience across diverse market conditions.

In sum, the robustness analysis indicates that while machine learning-based technical strategies exhibit strong performance across normal and volatility-induced crises, they are less effective under geopolitical uncertainty. This asymmetry underscores the importance of contextual model evaluation and advocates for more adaptive and multifactorial portfolio construction approaches when deploying such models in live trading environments.

VI. Conclusion

This study evaluates the efficacy of machine learning models at predicting directional stock movements for the purpose of constructing portfolio strategies based on technical analysis signals. Drawing on a comprehensive dataset spanning January 2019 to April 2024, the performance of several prominent classifiers: NNs, random forest, SGB, SVM, LR, bagging, and KNN, is assessed through both predictive accuracy metrics and real-world portfolio outcomes. The evaluation incorporates not only the full-period performance but also robustness tests during two major geopolitical and financial crises: the COVID-19 market crash and the early Ukraine-Russia war.

Over the full period, most machine learning-based portfolios delivered slightly higher risk-adjusted returns than the equal-weighted benchmark. However, t-statistics indicated that these outperformance levels were not statistically significant, limiting confidence in the combination of ML methods and the asset allocation strategy. ML methods like LR and SVM showed consistently strong performance, particularly at medium-return horizons for fixed-income assets.

Model robustness varied across regimes. During the COVID-19 crash, ML-based strategies driven by endogenous volatility exhibited better results than the baseline strategy, KNN and random forest for a shorter period. In contrast, during the Ukraine-Russia war, an exogenous shock,

these models underperformed, failing to provide alpha. This reflects a key limitation: technical-feature-based models may struggle with structural breaks and unforeseen external disruptions.

The divergence in model performance across regimes highlights important insights. First, machine learning models trained solely on price and volume-based technical features may lack sufficient flexibility to adapt to regime shifts introduced by external shocks. Second, model selection and portfolio performance must be evaluated on both average or in-sample accuracy, and on stability across tail risk environments. Third, incorporating ensemble diversity and robust evaluation frameworks, including out-of-sample testing under various market conditions, is critical to ensuring model resilience.

An additional methodological limitation lies in the use of fixed thresholds for labelling up and down return signals. While this approach offers simplicity, it lacks rigorous empirical justification and may negatively affect classification accuracy and generalizability. Fixed thresholds may fail to capture the dynamic nature of financial markets, particularly during regime changes or extreme volatility. Future research could investigate the effects of varying these thresholds and consider adaptive or data-driven methods to enhance the signal labelling process.

Moreover, advancing the practical utility of machine learning in finance will require a comprehensive and interdisciplinary framework that integrates insights from finance, economics, and computer science. By addressing these methodological and conceptual limitations, future research can significantly improve the robustness, interpretability, and adaptability of predictive models.

In conclusion, this study provides strong empirical support for the utility of machine learning in technical portfolio construction, particularly under standard and volatility-dominated conditions. However, its limitations under geopolitical uncertainty and reliance on fixed assumptions caution against overreliance solely on static pattern recognition. A robust investment framework will likely require a synthesis of predictive modelling, contextual awareness, adaptive learning, and real-world constraints to navigate modern financial market complexities.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research is supported in part by the National Science and Technology Council of Taiwan.

References

- Ahelegbey, D. F., and P. Giudici. 2022. "netVIX—A Network Volatility Index of Financial Markets." *Physica A: Statistical Mechanics and Its Applications* 594:127017. <https://doi.org/10.1016/j.physa.2022.127017>.
- Akhtaruzzaman, M., S. Boubaker, B. M. Lucey, and A. Sensoy. 2021. "Is Gold a Hedge or a safe-Haven Asset in the COVID-19 Crisis?" *Economic Modelling* 102:105588. <https://doi.org/10.1016/j.econmod.2021.105588>.
- Ali, A., S. A. Razak, S. H. Othman, T. A. E. Eisa, A. Al-Dhaqm, M. Nasser, T. Elhassan, H. Elshafie, and A. Saif. 2022. "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review." *Applied Sciences* 12 (19): 9637. <https://doi.org/10.3390/app12199637>.
- Avramov, D., G. Kaplanski, and H. Levy. 2018. "Talking Numbers: Technical versus Fundamental Investment Recommendations." *Journal of Banking & Finance* 92:100–114. <https://doi.org/10.1016/j.jbankfin.2018.05.005>.
- Babaei G., P. Giudici, E. Raffinetti. 2022. "Explainable Artificial Intelligence for Crypto asset allocation." *Finance Research Letters* 47 (B): 102941. <http://dx.doi.org/10.1016/j.frl.2022.102941>.
- Baillie, R. T., G. Kapetanios, and F. Papailias. 2014. "Bandwidth Selection by cross-Validation for Forecasting Long Memory Financial Time Series." *Journal of Empirical Finance* 29:129–143. <https://doi.org/10.1016/j.jempfin.2014.04.002>.
- Bao, W., J. Yue, Y. Rao, and B. Podobnik. 2017. "A Deep Learning Framework for Financial Time Series Using Stacked Autoencoders and long-Short Term Memory." *PLoS One* 12 (7): e0180944. <https://doi.org/10.1371/journal.pone.0180944>.
- Baur, D. G., and B. M. Lucey. 2010. "Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold." *Financial Review* 45 (2): 217–229. <https://doi.org/10.1111/j.1540-6288.2010.00244.x>.
- Bianchi, D., M. Büchner, A. Tamoni, and S. Van Nieuwerburgh. 2021. "Bond Risk Premiums with Machine Learning Edited by Stijn Van Nieuwerburgh." *Review of Financial Studies* 34 (2): 1046–1089. <https://doi.org/10.1093/rfs/hhaa062>.
- Brock, W., J. Lakonishok, and B. LeBaron. 1992. "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns." *Journal of Finance* 47 (5): 1731–1764. <https://doi.org/10.1111/j.1540-6261.1992.tb04681.x>.
- Cao, L. J., and F. E. H. Tay. 2003. "Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting." *IEEE Transactions on Neural Networks* 14 (6): 1506–1518. <https://doi.org/10.1109/TNN.2003.820556>.
- Caparrini, A., J. Arroyo, and J. E. Mansilla. 2024. "S&P 500 Stock Selection Using Machine Learning Classifiers: A Look into the Changing Role of Factors." *Research in International Business and Finance* 70:102336. <https://doi.org/10.1016/j.ribaf.2024.102336>.
- Chu, B. 2023. "Technical Analysis with Machine Learning Classification Algorithms: Can it Still 'Beat' the buy-And-Hold Strategy?" SSRN Working Paper No. 4765615. <https://doi.org/10.2139/ssrn.4765615>.
- Dash, R., and P. K. Dash. 2016. "A Hybrid Stock Trading Framework Integrating Technical Analysis with Machine Learning Techniques." *The Journal of Finance and Data Science* 2 (1): 42–57. <https://doi.org/10.1016/j.jfds.2016.03.002>.
- Ding, W., K. Mazouz, O. A. Gwilym, and Q. Wang. 2023. "Technical Analysis as a Sentiment Barometer and the cross-Section of Stock Returns." *Quantitative Finance* 23 (11): 1617–1636. <https://doi.org/10.1080/14697688.2023.2244991>.
- Erb, C. B., and C. R. Harvey. 2005. *The Tactical and Strategic Value of Commodity Futures*. Cambridge, Mass. USA: National Bureau of Economic Research.
- Faff, R., and D. Hillier. 2004. "An International Investigation of the Factors That Determine Conditional Gold Betas." *Financial Review* 39 (3): 473–488. <https://doi.org/10.1111/j.0732-8516.2004.00085.x>.
- Fang, M., L. Kuo, F. Shih, and S. Taylor. 2020. "Sector Categorization Using Gradient Boosted Trees Trained on Fundamental Firm Data." *Algorithmic Finance* 8 (3–4): 91–99. <https://doi.org/10.3233/AF-200308>.
- Fang, Y., and D. Xu. 2003. "The Predictability of Asset Returns: An Approach Combining Technical Analysis and Time Series Forecasts." *International Journal of Forecasting* 19 (3): 369–385. [https://doi.org/10.1016/S0169-2070\(02\)00013-4](https://doi.org/10.1016/S0169-2070(02)00013-4).
- Geertsema, P., and H. Lu. 2023. "Relative Valuation with Machine Learning." *Journal of Accounting Research* 61 (1): 329–376. <https://doi.org/10.1111/1475-679X.12464>.
- Gencay, R. 1998. "The Predictability of Security Returns with Simple Technical Trading Rules." *Journal of Empirical Finance* 5 (4): 347–359. [https://doi.org/10.1016/S0927-5398\(97\)00022-4](https://doi.org/10.1016/S0927-5398(97)00022-4).
- Gerlein, E. A., M. McGinnity, A. Belatreche, and S. Coleman. 2016. "Evaluating Machine Learning Classification for Financial Trading: An Empirical Approach." *Expert Systems with Applications* 54:193–207. <https://doi.org/10.1016/j.eswa.2016.01.018>.
- Giudici, P., T. Leach, and P. Pagnottoni. 2022. "Libra or Librae? Basket Based Stablecoins to Mitigate Foreign Exchange Volatility Spillovers." *Finance Research Letters* 44:102054. <https://doi.org/10.1016/j.frl.2021.102054>.

- Giudici, P., and E. Raffinetti. 2023. "SAFE Artificial Intelligence in Finance." *Finance Research Letters* 56:104088. <https://doi.org/10.1016/j.frl.2023.104088>.
- Gu, S., B. Kelly, D. Xiu, and A. Karolyi. 2020. "Empirical Asset Pricing via Machine Learning." Edited by Andrew Karolyi." *Review of Financial Studies* 33 (5): 2223–2273. <https://doi.org/10.1093/rfs/hhaa009>.
- Gurrib, I. 2022. "Machine Learning and Portfolio Management: A Review." *Annals of Mathematics and Computer Science* 5:31–43.
- Gurrib, I., M. Nourani, and R. K. Bhaskaran. 2022. "Energy Crypto Currencies and Leading U.S. Energy Stock Prices: Are Fibonacci Retracements Profitable?" *Financial Innovation* 8 (1): 8. <https://doi.org/10.1186/s40854-021-00311-8>.
- Henrique, B. M., V. A. Sobreiro, and H. Kimura. 2019. "Literature Review: Machine Learning Techniques Applied to Financial Market Prediction." *Expert Systems with Applications* 124:226–251. <https://doi.org/10.1016/j.eswa.2019.01.012>.
- Hiziroglu, A. 2013. "Soft Computing Applications in Customer Segmentation: State-Of-Art Review and Critique." *Expert Systems with Applications* 40 (16): 6491–6507. <https://doi.org/10.1016/j.eswa.2013.05.052>.
- Kamalov, F., I. Gurrib, and K. Rajab. 2021. "Financial Forecasting with Machine Learning: Price Vs Return." *Journal of Computer Science* 17 (3): 251–264. <https://doi.org/10.3844/jcssp.2021.251.264>.
- Kinateder, H., I. Gurrib, and T. Choudhury. 2024. "Navigating Crises: Gold's Role as a Safe Haven for U.S. Sectors." *Finance Research Letters* 69:106210. <https://doi.org/10.1016/j.frl.2024.106210>.
- Lessmann, S., B. Baesens, H. V. Seow, and L. C. Thomas. 2015. "Benchmarking state-Of-The-Art Classification Algorithms for Credit Scoring: An Update of Research." *European Journal of Operational Research* 247 (1): 124–136. <https://doi.org/10.1016/j.ejor.2015.05.030>.
- Lin, Q. 2018. "Technical Analysis and Stock Return Predictability: An Aligned Approach." *Journal of Financial Markets* 38:103–123. <https://doi.org/10.1016/j.finmar.2017.09.003>.
- Lo, A. W., H. Mamaysky, and J. Wang. 2000. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation." *Journal of Finance* 55 (4): 1705–1765. <https://doi.org/10.1111/0022-1082.00265>.
- Markowitz, H. M. 1952. "Portfolio Selection." *The Journal of Finance* 7 (1): 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>.
- Marshall, B. R., N. H. Nguyen, and N. Visaltanachoti. 2017. "Time Series Momentum and Moving Average Trading Rules." *Quantitative Finance* 17 (3): 405–421. <https://doi.org/10.1080/14697688.2016.1205209>.
- Murray, S., Y. Xia, and H. Xiao. 2024. "Charting by Machines." *Journal of Financial Economics* 153:103791. <https://doi.org/10.1016/j.jfineco.2024.103791>.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.
- Sherman, E. J. 1982. "Gold: A Conservative, Prudent Diversifier." *The Journal of Portfolio Management* 8 (3): 21–27. <https://doi.org/10.3905/jpm.1982.408850>.
- Shi, S., R. Tse, W. Luo, S. D'Addona, and G. Pau. 2022. "Machine Learning- Driven Credit Risk: A Systemic Review." *Neural Computing & Applications* 34 (17): 14327–14339. <https://doi.org/10.1007/s00521-022-07472-2>.
- Shynkevich, A. 2016. "Predictability in Bond Returns Using Technical Trading Rules." *Journal of Banking & Finance* 70:55–69. <https://doi.org/10.1016/j.jbankfin.2016.06.010>.
- Teixeira, L. A., and A. L. I. de Oliveira. 2010. "A Method for Automatic Stock Trading Combining Technical Analysis and Nearest Neighbor Classification." *Expert Systems with Applications* 37 (10): 6885–6890. <https://doi.org/10.1016/j.eswa.2010.03.033>.
- Tully, E., and B. M. Lucey. 2007. "A Power GARCH Examination of the Gold Market." *Research in International Business and Finance* 21 (2): 316–325. <https://doi.org/10.1016/j.ribaf.2006.07.001>.
- Wu, M., and X. Diao. 2015. "Technical Analysis of Three Stock Oscillators Testing MACD, RSI and KDJ Rules in SH & SZ Stock Markets." 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), 1, 320–323. IEEE. <https://doi.org/10.1109/ICCSNT.2015.7490760>.
- Zhu, S., C. Zhou, H. Liu, and Y. Ren. 2024. "Commodity Factor Investing via Machine Learning." *Pacific-Basin Finance Journal* 83:102231. <https://doi.org/10.1016/j.pacfin.2023.102231>.
- Zhu Y., G. Zhou. 2009. "Technical Analysis: An asset allocation Perspective on the Use of Moving Averages." *Journal of Financial Economics* 92 (3): 519–544.