Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.34592

Adaptive Portfolio Optimization under Market Regime Shifts: A Deep Reinforcement Learning Approach

Qianhui Ma

School of Business, Sun Yat-sen University, Shenzhen, China
Maqh23@mail2.sysu.edu.cn

Abstract. This study examines whether a reinforcement-learning allocation rule can remain usable when the return-generating process changes across market regimes. The portfolio problem is written as a Markov Decision Process in which the agent observes short- and medium-horizon price trends, volatility-spread information, current holdings, recent portfolio returns, and drawdown. A PPO policy is then trained to choose continuous portfolio weights. To keep the strategy implementable, the raw network output is mapped into a long-only, fully invested portfolio with an explicit upper-bound-style diversification constraint, and the reward deducts both a rolling-variance risk penalty and transaction costs. The empirical test uses ten representative China A-share stocks. The PPO agent is trained on 2015-2022 data, selected using 2023 validation performance, and evaluated only on the 2024-2025 out-of-sample period. Over 484 test trading days, PPO obtains a 10.24% annualized return and a Sharpe ratio of 0.6164. It does not dominate the equal-weight benchmark, which records 12.62% and 0.7399, but it substantially exceeds the rolling mean-variance maximum-Sharpe strategy, which records only 1.98% and 0.1976. These results suggest that PPO is a feasible adaptive allocation framework, while also showing that deep reinforcement learning should be judged against simple robust baselines rather than only against estimation-sensitive optimization benchmarks.

Keywords: portfolio optimization, reinforcement learning, PPO, market regime shifts

1. Introduction

Portfolio choice is not a one-shot calculation in actual trading. A manager observes new prices, absorbs the effect of past trades, pays trading costs, and then revises the next allocation. The decision therefore has a sequential structure: today's weight affects tomorrow's net value, risk exposure, and feasible rebalancing path. A static optimizer can be useful as a benchmark, but it is not designed to learn from this feedback loop or to react differently when market conditions move from a calm period to a volatile or policy-driven regime.

The starting point of the literature is the mean-variance framework of Markowitz [1], which makes the return-risk trade-off explicit and remains the standard reference model for portfolio construction. Its practical weakness is also well known: expected returns and covariances must be estimated from historical data, and these estimates are unstable when returns are non-normal, heavy-tailed, or non-stationary [2]. This helps explain why the simple 1/N rule can perform surprisingly

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.34592

well out of sample. DeMiguel, Garlappi, and Uppal [3] show that naive diversification is difficult to beat once estimation error is taken seriously. For a market such as China A-shares, where sector rotations and policy shocks are frequent, this concern is not only theoretical but directly affects backtest interpretation.

Deep reinforcement learning offers a different route. Rather than estimating a fixed mean vector and covariance matrix and then solving a separate optimization problem, a DRL agent learns a mapping from observed states to portfolio actions through repeated interaction with an environment. The success of deep reinforcement learning in high-dimensional sequential tasks was popularized by Mnih et al. [4]. In portfolio applications, Jang and Seong [5] use PPO-style policies in a multi-stock setting, while Sood et al. [6] compare DRL with mean-variance rules under changing regimes. Related machine-learning evidence in asset management also emphasizes that prediction and allocation can be linked, although not without overfitting risk [7, 8].

This paper follows that line of work but keeps the empirical question deliberately narrow. I do not claim that PPO is universally superior to simple diversification. Instead, the question is whether a PPO policy, trained only on pre-test historical data and evaluated without look-ahead information, can produce a reasonable long-only allocation rule when the test period contains clear market-regime changes. The model uses a compact state vector so that the economic content of the signals is interpretable: trend variables proxy for short-run momentum, volatility-spread variables proxy for changes in uncertainty, and portfolio-state variables record the consequences of previous decisions.

The action design also reflects practical portfolio constraints. The policy network is allowed to express continuous over- or under-weighting signals, but the final weights must be non-negative and sum to one. In addition, the mapping limits extreme single-stock concentration. This design is less flexible than an unconstrained neural policy, but it is closer to a realistic long-only portfolio and reduces the chance that an apparent backtest gain is merely the result of excessive concentration in one stock.

The reward function is specified to match the implementation objective. A high one-period return increases the reward, a high recent variance lowers it, and turnover is penalized through transaction costs. Thus the agent is not trained to chase raw returns alone. It is encouraged to trade only when the expected improvement is large enough to justify additional risk and rebalancing cost.

The empirical comparison is organized around three strategies. PPO is the proposed adaptive rule. Equal weight is the parameter-free benchmark and serves as a demanding baseline because it is robust to estimation error. Rolling mean-variance maximum-Sharpe allocation is included to represent the classical plug-in optimization approach. Comparing all three strategies helps distinguish two claims that are often conflated: whether reinforcement learning is operationally feasible, and whether it is strong enough to beat a simple robust benchmark in a difficult out-of-sample period.

The results should therefore be read cautiously. PPO achieves positive out-of-sample performance and clearly outperforms the rolling mean-variance strategy, but it remains below the equal-weight portfolio in both annualized return and Sharpe ratio. This pattern is informative. It suggests that the PPO framework can learn useful allocation behavior, but it also confirms that estimation-free diversification remains highly competitive when the market regime shifts.

The core research question is accordingly refined as follows: under long-only and transaction-cost constraints, can PPO learn an allocation policy that is more robust than a rolling mean-variance rule during regime shifts, and how far is it from the equal-weight benchmark that requires almost no estimation?

DOI: 10.54254/2755-2721/2026.34592

2. Set up

This section describes the reinforcement-learning environment used in the empirical exercise. PPO is adopted because it can update a stochastic continuous-action policy while limiting excessively large policy changes through its clipped objective. In the portfolio setting, this property is useful because abrupt policy updates may translate into unstable weights and unnecessary turnover.

The portfolio problem is formulated as an MDP. At trading step t, the agent observes state s_t, selects an allocation action, receives the next-period market return, pays transaction costs when weights are adjusted, and then moves to state s_{t+1}. The five components of the environment are reported below: the state vector, the action-to-weight mapping, the transition rule, the reward, and the discount factor.

2.1. State space $\mathbb{S}$

The empirical universe contains M industry-representative stocks. This size is intentionally modest: it is large enough to require cross-sectional allocation, yet small enough to make the reinforcement-learning experiment transparent and reproducible. The state vector combines market information with portfolio information so that the agent observes both external price conditions and the internal position it already holds.

2.1.1. Market State

The market state consists of a trend factor $\mathbf{X}_t^{(1)}$ and a volatility spread factor $\mathbf{X}_t^{(2)}$.

1. **Trend Factor** $\mathbf{X}_t^{(1)} = \left(X_{1,t}^{(1)}, \ldots, X_{M,t}^{(1)}\right)$: Captures the price momentum of assets, defined as the relative deviation between the 5-day and 10-day moving average prices:

$$X_{i,t}^{(1)} = \frac{\frac{1}{5}\sum_{k=1}^{5} P_{i,t-k} - \frac{1}{10}\sum_{k=1}^{10} P_{i,t-k}}{\left|\frac{1}{10}\sum_{k=1}^{10} P_{i,t-k}\right| + \varepsilon}$$

• where $P_{i,k}$ is the closing price of asset i at time k, and $\varepsilon > 0$ is a constant for numerical stability. $X_{i,t}^{(1)} > 0$ indicates that the short-term trend is stronger than the medium-term, reflecting upward momentum.

2. **Volatility Spread Factor** $\mathbf{X}_t^{(2)}$: Captures the volatility term structure of asset returns, defined as the difference between short-term and long-term volatility:

$$X_{i,t}^{(2)} = \sqrt{\frac{1}{4}\sum_{j=0}^{4}\left(r_{i,t-j} - \bar{r}_{i,t}^{(5)}\right)^2} - \sqrt{\frac{1}{19}\sum_{j=0}^{19}\left(r_{i,t-j} - \bar{r}_{i,t}^{(20)}\right)^2}$$

• where $r_{i,j}$ is the daily return of asset i at time j, and $\bar{r}_{i,t}^{(k)}$ is the arithmetic mean return over the past k trading days. $X_{i,t}^{(2)} > 0$ indicates that recent volatility exceeds the historical level,

DOI: 10.54254/2755-2721/2026.34592

implying rising market uncertainty.

2.1.2. Portfolio state

The portfolio state describes the current allocation structure and recent performance characteristics:

- **Current Weights** $\mathbf{X}_t^{(3)} = (w_{1,t}, \ldots, w_{M,t})$, satisfying $\sum_{i=1}^{M} w_{i,t} = 1$ and $w_{i,t} \geq 0$.
- Portfolio Return $X_t^{(4)} = \sum_{i=1}^{M} w_{i,t} r_{i,t}$, reflecting the current period's return.
- **Maximum Drawdown** $X_t^{(5)} = \frac{\max_{\tau \in [0,t]} V_\tau - V_t}{\max_{\tau \in [0,t]} V_\tau}$, where V_t is the portfolio's net asset value at time t , capturing historical extreme downside risk.

Combining these features, the complete state vector is represented as:

$$S_t = \left(\mathbf{X}_t^{(1)},\ \mathbf{X}_t^{(2)},\ \mathbf{X}_t^{(3)},\ X_t^{(4)},\ X_t^{(5)} \right) \in \mathbb{R}^{3M+2}$$

For $M = 10$, the state dimension is 32.

2.2. Action space $\mathbb{A}$

Given the PPO algorithm's capacity to handle high-dimensional continuous action spaces directly, this paper avoids discretizing the asset allocation proportions. We define the raw output of the policy network at time t as the vector $\mathbf{u}_t \in \mathbb{R}^M$.

To prevent the policy from concentrating excessively on a single asset and to maintain portfolio diversification, we apply a constrained action mapping to transform the raw signals into valid portfolio weights. Specifically, let Δ be the maximum deviation parameter. The mapping is implemented in two steps:

$$\begin{aligned} \delta_{t,j} &= \Delta \cdot \tanh\left(u_{t,j}\right), \quad \forall j = 1, \ldots, M \\ a_{t,j} &= \frac{\mathrm{clip}\left(\frac{1}{M} + \delta_{t,j},\, 0,\, 1\right)}{\sum_{k=1}^{M} \mathrm{clip}\left(\frac{1}{M} + \delta_{t,k},\, 0,\, 1\right)} \end{aligned}$$

where $\mathrm{clip}\left(x, 0, 1\right) = \min\left(\max\left(x, 0\right), 1\right)$ is the clipping function. Centered at the equal weight $1/M$, this mapping allows the agent to actively overweight or underweight each asset within the range $\left[\frac{1}{M} - \Delta, \frac{1}{M} + \Delta\right]$ while strictly satisfying non-negativity (no short-selling) and normalization constraints. We set $\Delta = 0.25$, meaning the upper bound for a single stock's weight is approximately 35%, which effectively prevents extreme concentration.

2.3. State transition function $\mathbb{T}$

The state transition is driven jointly by the exogenous evolution of the market environment and the endogenous updating of the portfolio state:

$$S_{t+1} = f\left(S_t, \mathbf{a}_t, \varepsilon_{t+1}\right)$$

DOI: 10.54254/2755-2721/2026.34592

In implementation, market state variables are obtained by rolling forward true historical data sequences as the exogenous environment. The portfolio state is updated deterministically based on the current weights $\mathbf{a}_t$ and actual market return feedback. The weights undergo a passive drift between two rebalancing periods:

$$w_{i,t+1}^{\text{passive}} = \frac{a_{i,t}\left(1 + r_{i,t+1}\right)}{\sum_{k=1}^{M} a_{k,t}\left(1 + r_{k,t+1}\right)}$$

2.4. Reward function $\mathbb{R}$

The reward function directly determines the agent's optimization objective. In dynamic portfolio management, the agent must strike a balance among maximizing absolute returns, controlling net value volatility, and reducing portfolio turnover. We construct the following linear risk-adjusted reward function:

$$R_t = r_{p,t} - \lambda \sigma_{p,t}^2 - \eta c_t$$

where each component is defined as follows:

• **Portfolio Return** $r_{p,t}$: The pre-tax return of the portfolio at time t , $r_{p,t} = \sum_{i=1}^{M} a_{i,t} r_{i,t}$.

• **Risk Penalty** $\sigma_{p,t}^2$: The rolling variance of portfolio returns over the past W trading days ($W = 20$). $\lambda \geq 0$ is the risk aversion coefficient, encouraging the agent to actively reduce risk exposure during highly volatile periods.

• **Transaction Cost** c_t : The sum of absolute differences between the target weights and the beginning-of-period passive weights, $c_t = \sum_{i=1}^{M} \left| a_{i,t} - w_{i,t}^{\text{passive}} \right|$, where η is the two-way transaction fee rate.

The net value update of the portfolio identically deducts the actual transaction costs: $V_{t+1} = V_t \cdot \left(1 + r_{p,t} - \eta c_t\right)$.

2.5. Discount factor γ

The discount factor gamma determines how much the agent values future rewards relative to the current reward. I set gamma = 0.99. This value keeps a multi-period investment horizon in the objective, but it still discounts distant outcomes whose signal-to-noise ratio is low in non-stationary financial data. In practice, this choice also helps stabilize value-function estimation during PPO training.

3. Empirical study

The empirical exercise is designed as a chronological backtest rather than a random-split prediction task. The model is trained, selected, and tested in the same order in which an investor would have observed the data. This design is important because the central concern of the paper is performance after regime changes, not in-sample fit.

DOI: 10.54254/2755-2721/2026.34592

3.1. Data description and preprocessing

Data Source and Asset Pool Construction The asset pool includes M = 10 China A-share stocks, each chosen to represent a different sector: Food & Beverage, Insurance, Banking, Pharmaceuticals, Machinery, Basic Chemicals, New Energy, Public Utilities, Communications, and Consumer Retail. The purpose is not to build the broadest possible market index. Instead, the sample creates a controlled allocation problem with meaningful sector heterogeneity while keeping the state and action dimensions manageable. Table 1 lists the selected stocks.

Table 1. Composition of the empirical asset pool

No.	Stock Code	Stock Name	Sector
1	600519	Kweichow Moutai	Food & Beverage
2	601318	Ping An Insurance	Insurance
3	600036	China Merchants Bank	Banking
4	600276	Hengrui Medicine	Pharmaceuticals
5	600031	SANY Heavy Industry	Machinery
6	600309	Wanhua Chemical	Basic Chemicals
7	601012	LONGi Green Energy	New Energy
8	600900	China Yangtze Power	Public Utilities
9	600050	China Unicom	Communications
10	601888	China Tourism Group Duty Free	Consumer Retail

Data Processing The raw input is 5-minute K-line data for the ten stocks from January 5, 2015, to December 31, 2025. Each record contains open, high, low, close, and volume. Since the agent makes one allocation decision per trading day, I convert the intraday data into a daily decision series by taking the close of the final 5-minute bar as the daily closing price:

$$P_{i,t}^{\text{close}} = \text{Close}_{i,t}^{\text{last bar}}$$

Daily returns are then computed from the aligned close-price series. To reduce the influence of abnormal records and to remain consistent with the price-limit environment of A-shares, returns are clipped at the pre-specified bounds used in the experiment. After requiring complete observations for all ten stocks, the final panel contains 2,265 valid trading days.

Dataset Splitting The split is strictly chronological. The training period covers 1,537 trading days from January 5, 2015, to December 31, 2022. The 2023 sample contains 242 trading days and is used only for validation and model selection. The final evaluation uses 484 trading days from January 2024 to December 2025. No observation from the test period is used to update the PPO network or to choose the checkpoint.

The main experimental parameters are eta = 0.2% for two-way transaction costs, lambda = 0.3 for risk aversion, gamma = 0.99 for discounting, and Delta = 0.25 for the maximum deviation in the action mapping. PPO hyperparameters are summarized in Table 2.

DOI: 10.54254/2755-2721/2026.34592

Table 2. Main hyperparameters of the PPO algorithm

Hyperparameter	Symbol	Value
Policy network hidden layer dimension	d_h	64
Learning rate	α	3×10^{-4}
Weight decay (L2 regularization)	λ_{wd}	1×10^{-4}
PPO clip parameter	ε_{clip}	0.2
Policy updates per epoch	K	8
GAE parameter	λ_{GAE}	0.95
Training epochs	-	800

Model Selection Strategy To reduce overfitting, I do not select the last training epoch automatically. Every 50 epochs, the policy is evaluated on the validation set using the deterministic mean action. The checkpoint with the highest validation Sharpe ratio is retained for the 2024-2025 test. This procedure separates training reward improvement from out-of-sample model selection and avoids choosing a checkpoint based on test results.

3.2. Evaluation metrics and benchmark strategies

Performance is evaluated from the net asset value series generated in the test period. I report three standard measures: annualized return, Sharpe ratio, and maximum drawdown. Together they describe profitability, return per unit of volatility, and the worst peak-to-trough loss experienced by the strategy.

• **Annualized Return (AR)**: Measures the absolute profitability of the strategy:

$$\mathrm{AR} = \left(\prod_{t=1}^{T} (1 + r_{p,t}) \right)^{252/T} - 1$$

• **Sharpe Ratio (SR)**: Reflects the excess return generated per unit of total risk, defined as $\mathrm{SR} = \bar{r}_p / \widehat{\sigma}_p \cdot \sqrt{252}$, where $\bar{r}_p$ and $\widehat{\sigma}_p$ are the sample mean and standard deviation of daily returns, respectively.

• **Maximum Drawdown (MDD)**: Measures the extreme downside risk the portfolio may face during the out-of-sample period, defined as $\mathrm{MDD} = \max_{t \in [1,T]} \frac{\max_{\tau \le t} V_\tau - V_t}{\max_{\tau \le t} V_\tau}$.

Two benchmarks are used so that PPO is compared with both a robust rule and an estimation-based optimizer.

(1) Equal Weight (1/N, EW) Strategy: Each stock receives weight 1/M on each trading day. This strategy is intentionally simple. Because it does not estimate expected returns or covariances, it is less exposed to parameter instability and is a useful benchmark for judging whether a more complex learning method adds practical value [3].

(2) Rolling Mean-Variance (MV) Maximum Sharpe Strategy: This benchmark estimates the expected-return vector and covariance matrix using the most recent 252 trading days and then solves a maximum-Sharpe portfolio problem:

DOI: 10.54254/2755-2721/2026.34592

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top \widehat{\boldsymbol{\mu}}}{\sqrt{\mathbf{w}^\top \widehat{\boldsymbol{\Sigma}} \mathbf{w}}}, \quad \text{s.t.} \quad \mathbf{w}^\top 1 = 1,\ \mathbf{w} \geq 0$$

The MV portfolio is rebalanced every 20 trading days. This rule is closer to the classical mean-variance logic than the equal-weight rule, but it is also more vulnerable to estimation error when the recent historical window is not representative of the next market regime [1].

3.3. Out-of-sample performance and results analysis

Table 3 summarizes the out-of-sample results for the 2024-2025 test period. The table is interpreted as a genuine post-training evaluation because the model checkpoint was fixed using the 2023 validation set before the test sample was examined.

Table 3. Summary of out-of-sample performance (test period: Jan 2024 – Dec 2025, 484 trading days)

Strategy	Annualized Return (%) ↑	Sharpe Ratio ↑	Maximum Drawdown (%) ↓
PPO (Proposed)	10.24	0.6164	18.67
Equal Weight (1/N)	12.62	0.7399	18.56
Mean-Variance (MV)	1.98	0.1976	16.85

Note: Arrows indicate the preferred direction of each metric; bold numbers mark the best value within a column.

Overall Performance Analysis The three strategies produce clearly different test-period outcomes. Equal weight performs best, with a 12.62% annualized return and a Sharpe ratio of 0.7399. PPO ranks second, with a 10.24% annualized return and a Sharpe ratio of 0.6164. The MV strategy is much weaker, with a 1.98% annualized return and a Sharpe ratio of 0.1976. Therefore, the main empirical finding is not that PPO beats every simple rule. Rather, PPO remains competitive with EW and is much more effective than the rolling mean-variance optimizer in this particular regime-shift sample.

Learning and Model-Selection Evidence During training, the PPO reward improves from -1.10 at epoch 50 to 1.01 at epoch 800, showing that the agent learns to obtain higher rewards in the training environment. However, the best validation Sharpe ratio is -1.43, which means that the validation market segment remains difficult even when the training objective improves. For this reason, the paper treats the 2024-2025 test result as an out-of-sample feasibility check rather than as proof of universal superiority. The validation-based checkpoint rule is retained to avoid selecting the final model by looking at the test period.

Why the Mean-Variance Strategy Fails The weak MV result is consistent with the known sensitivity of plug-in mean-variance optimization. The strategy estimates means and covariances from a 252-day rolling window and rebalances every 20 trading days. When the next market environment differs sharply from the estimation window, small errors in expected returns can translate into large allocation errors. In the test sample, 2024-2025 contains pronounced sector rotation and policy-driven repricing. Under such conditions, the estimated maximum-Sharpe portfolio can be pulled toward assets that looked attractive in the past but no longer offer the same return-risk trade-off. EW avoids this channel because it has no estimated parameters; PPO also

DOI: 10.54254/2755-2721/2026.34592

avoids direct mean-covariance inversion, although its learned policy is still shaped by the training sample.

Why PPO Does Not Beat Equal Weight PPO underperforms EW by 2.38 percentage points in annualized return. The gap is economically meaningful but not large relative to the MV shortfall. A likely reason is that the PPO state relies mainly on price-trend and volatility information. These signals are noisy and can lag when leadership rotates quickly across sectors. In addition, PPO pays the 0.2% two-way transaction cost whenever it changes weights, while the EW implementation in this experiment has a much lower effective turnover burden. Thus the PPO policy must overcome both signal noise and trading frictions before it can dominate a passive diversified portfolio.

Risk Control Performance PPO records a maximum drawdown of 18.67%, almost the same as EW at 18.56%. MV has the lowest drawdown at 16.85%, but this lower downside risk comes with very weak return generation. The result suggests that the rolling-variance penalty in the PPO reward helps restrain volatility but does not fully control tail losses. Future versions should consider drawdown-aware rewards or regime-state variables if downside protection is the primary objective.

4. Conclusion

This paper studies a long-only portfolio-allocation problem under market-regime shifts and implements it as an MDP solved by PPO. The model observes a small set of interpretable trend, volatility, and portfolio-state variables; maps the continuous policy output into feasible portfolio weights; and trains the policy with a reward that deducts both risk and transaction costs. The empirical design uses a chronological split so that the 2024-2025 results are purely out of sample.

The evidence is mixed but useful. PPO achieves a 10.24% annualized return and a 0.6164 Sharpe ratio over 484 test trading days, which is far better than the rolling MV benchmark. At the same time, PPO does not exceed the equal-weight strategy, whose annualized return and Sharpe ratio are 12.62% and 0.7399. The appropriate conclusion is therefore moderate: PPO is feasible and more robust than an estimation-sensitive MV optimizer in this experiment, but simple diversification remains a strong benchmark that the current PPO specification does not surpass.

Several improvements follow naturally from these findings. First, online or periodic retraining may help the policy adapt when the market regime changes after the training period. Second, the state vector could include explicit regime indicators, such as factor-rotation measures, liquidity variables, or macro-financial signals, rather than relying only on price trends and volatility spreads. Third, the reward could incorporate drawdown or downside-risk terms more directly. Finally, future work should test a broader stock universe and report performance separately across bull, bear, and high-volatility episodes so that the source of PPO's gains and failures can be identified more clearly.

References

[1] H. Markowitz, "Modern portfolio theory", Journal of Finance, vol. 7, no. 11, pp. 77–91, 1952.
[2] R. Cont, "Empirical properties of asset returns: Stylized facts and statistical issues", Quantitative finance, vol. 1, no. 2, p. 223, 2001.
[3] V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?" The review of Financial studies, vol. 22, no. 5, pp. 1915–1953, 2009.
[4] V. Mnih et al., "Human-level control through deep reinforcement learning", nature, vol. 518, no. 7540, pp. 529–533, 2015.
[5] J. Jang and N. Seong, "Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory", Expert Systems with Applications, vol. 218, p. 119556, 2023.
[6] S. Sood, K. Papasotiriou, M. Vaiciulis, and T. Balch, "Deep reinforcement learning for optimal portfolio allocation: A comparative study with mean-variance optimization", arXiv preprint arXiv: 2602.17098, 2026.

DOI: 10.54254/2755-2721/2026.34592

[7] A. Chaweewanchon and R. Chaysiri, "Markowitz mean-variance portfolio optimization with predictive stock selection using machine learning", International journal of financial studies, vol. 10, no. 3, p. 64, 2022.
[8] R. Kaniel, Z. Lin, M. Pelger, and S. Van Nieuwerburgh, "Machine-learning the skill of mutual fund managers", Journal of Financial Economics, vol. 150, no. 1, pp. 94–138, 2023.