Ekonomi, Politika & Finans Araştırmaları Dergisi, 2026, 11(1): 106-119
Journal of Research in Economics, Politics & Finance, 2026, 11(1): 106-119
Araştırma Makalesi / Research Article, https://doi.org/10.30784/epfad.1811319

DYNAMIC PORTFOLIO OPTIMIZATION WITH DEEP REINFORCEMENT LEARNING: EVIDENCE FROM BORSA ISTANBUL*

Derin Pekiştirmeli Öğrenme ile Dinamik Portföy Optimizasyonu: Borsa İstanbul Örneği

Hidayet BEYHAN, Erhan ERGİN**** & Binali Selman EREN*****

Keywords:
Reinforcement Learning, Portfolio Optimization, Borsa Istanbul, DDPG, Markowitz Model, Emerging Markets.

JEL Codes:
C45, C61, C63, G11, G15, O16.

Abstract

In this study, portfolio optimization has been conducted using the reinforcement learning approach, one of the artificial intelligence algorithms. The data is considered for constituents of the BIST30 index, which is the blue-chip index of Borsa Istanbul. The performance of Deep Deterministic Policy Gradient (DDPG), a deep learning algorithm of reinforcement learning, has been tested against the Markowitz mean-variance and equal-weighted portfolios as benchmark models; the BIST30 index itself has also been taken as a benchmark portfolio. This study contributes to the relevant literature in terms of Türkiye as an example of a developing country and the method employed. The study demonstrates the potential of RL approaches that are becoming widespread for portfolio optimization. The obtained results reveal that the portfolio formed with the DDPG approach shows a superior Sharpe ratio portfolio over portfolios obtained with other classical approaches. These findings, while highlighting the potential of RL approaches in practice, emerge as an alternative option for fund managers, especially in a volatile market environment.

Anahtar Kelimeler:
Pekiştirmeli Öğrenme, Portföy Optimizasyonu, Borsa İstanbul, DDPG, Markowitz Modeli, Gelişmekte Olan Piyasalar.

JEL Kodları:
C45, C61, C63, G11, G15, O16.

Öz

Bu çalışmada, yapay zekâ algoritmalarından biri olan pekiştirmeli öğrenme yaklaşımı kullanılarak portföy optimizasyonu gerçekleştirilmiştir. Veriler, Borsa İstanbul endeksi olan BIST30 endeksinin bileşenleri için ele alınmıştır. Pekiştirmeli öğrenmenin derin öğrenme algoritmalarından biri olan Derin Deterministik Politika Gradyanının (DDPG) performansı, kıyaslama modeli olarak Markowitz ortalama-varyans ve eşit ağırlıklı portföylere karşı test edilmiş, ayrıca BIST30 endeksinin kendisi de kıyaslama portföyü olarak alınmıştır. Bu çalışma, gelişmekte olan bir ülke örneği olarak Türkiye ve kullanılan yöntem açısından ilgili literatüre katkıda bulunmaktadır. Çalışma, giderek yaygınlaşan RL yaklaşımlarının portföy optimizasyonu için potansiyelini göstermektedir. Elde edilen sonuçlar, DDPG yaklaşımıyla oluşturulan portföyün, diğer klasik yaklaşımlarla elde edilen portföylere göre daha üstün bir Sharpe oranına sahip portföy gösterdiğini ortaya koymaktadır. Bu bulgular, RL yaklaşımlarının pratikteki potansiyelini vurgularken, özellikle dalgalı piyasa ortamında fon yöneticileri için alternatif bir seçenek olarak ortaya çıkmaktadır.

* This study is an extended version of the abstract previously presented at a conference (Beyhan, H., Ergin, E., & Eren Binali, S. (2025, May 8–10). AI-driven portfolio optimization: A reinforcement learning approach for dynamic asset allocation in Borsa Istanbul. 9th IERFM Economic Research and Financial Markets Congress with International Participation.

** Asst. Prof. Dr., Bitlis Eren University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Türkiye, hbeyhan@beu.edu.tr (Corresponding Author)
*** Asst. Prof. Dr., Bitlis Eren University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Türkiye, eergin@beu.edu.tr
**** Asst. Prof. Dr., Bitlis Eren University, Tatvan Vocational School, Department of Accounting and Taxation, Türkiye, bseren@beu.edu.tr

Received Date (Makale Geliş Tarihi): 27.10.2025 *Accepted Date (Makale Kabul Tarihi): 17.03.2026*

1. Introduction

Decision-making in financial markets represents a critical phase under conditions fraught with uncertainty and rapid change. One of the most crucial of these decisions is determining a portfolio with ideal weights, which plays a crucial role in competitive markets. In making this decision, not only the portfolio's return but also its resilience to varying levels of risk is crucial. To support such decisions, a number of approaches have been developed that prioritize low risk over high returns (Markowitz, 1952). While Markowitz's (1952) pioneering contribution constituted a significant turning point in financial decision-making, evolving market conditions, changing buying and selling habits due to technology, and particularly the pace of trading, have rendered the model's fundamental assumptions extremely naive, especially in volatile market environments (Michaud, 1989). Emerging market stock markets are more volatile than those in developed economies and are more susceptible to shocks such as macroeconomic disruptions. As a result, the inherent volatility of these economies further increases the difficulty of decision-making in such markets (Rjoub et al., 2009; Erdoğan et al., 2022).

Following previous studies, alternative approaches that consider investor behavior have been developed, adding a new dimension to studies aimed at capturing price behavior in stock markets (Tversky and Kahneman, 1974). While this approach makes similar assumptions to Markowitz (1952), they fall short in addressing extreme scenarios. Reinforcement learning (RL), a relatively newer alternative, is emerging as a promising paradigm characterized by its ability to dynamically adapt to evolving market conditions (Jiang et al., 2017). Considering this, RL approaches are worth testing for trading in stock markets. However, RL models have not been sufficiently implemented in emerging markets such as Borsa Istanbul; they have been more commonly used for portfolio optimization in developed markets. Given the growing role of these markets in the global financial landscape and their share of approximately 10% of global stock market capitalization, such an innovative approach therefore becomes even more meaningful (Deng et al., 2016). This research aims to fill this gap and present different investment opportunities and challenges.

This study utilizes RL, specifically the DDPG algorithm, to evaluate its applicability for portfolio optimization with the use of the 30 biggest blue-chip stocks listed in Borsa Istanbul, represented by the BIST30 index. The DDPG model's performance is compared against the Markowitz mean-variance model, an equal-weight portfolio, and the BIST30 index as a market benchmark, over the test period from April 2022 to December 2024.

The aim is to assess whether RL can achieve competitive risk-adjusted returns in a volatile emerging market context. This study has two main contributions: it advances the understanding of computational decision-making in finance by applying RL to an understudied market; it offers guidance for practitioners and scholars seeking dynamic strategies to navigate emerging market volatility.

The paper is organized as follows: Section 2 reviews the literature on portfolio optimization, RL in finance, and emerging markets. Section 3 details the methodology, including data, RL environment, DDPG model, and benchmarks. Section 4 presents the results, and Section 5 concludes with recommendations for future research and practice.

2. Literature Review

The approaches developed under static market assumptions are comprehensively examined in early attempts of portfolio optimization approaches and the application of methods that adapt to changing market conditions has increased significantly in recent years (Fabozzi et al., 2007). RL, which emerged from advances in artificial intelligence, is methods that promises to offer a powerful alternative to classical portfolio optimization approaches. Financial market asset prices have non-linear structure and, by taking this into account, and RL approaches have the ability to adapt to changing environments (Sutton and Barto, 2018). One of its most important features is its assumption-free structure, making this method particularly suitable for uncertain market conditions, as it specifically targets asset return behavior (Meng and Khushi, 2019).

The use of RL in portfolio optimization is not a new idea; however, it has gained increasing attention in recent years. In the early 2000s, RL-based methods began to be used to develop trading strategies (Moody and Saffell, 2001). Over time, RL algorithms developed for portfolio optimization have fallen into three main groups: model-free methods, online learning approaches, and actor-critic-based models (Sutton and Barto, 2018). The DDPG algorithm, which belongs to this last group, has been used to dynamically determine stock weights in portfolios and has yielded quite successful results in developed markets (Jiang et al., 2017). These studies clearly demonstrate the adaptability of RL to volatile and complex market conditions. However, most of this research focuses on developed markets with high liquidity and stable regulations; the generalizability of the findings remains limited. Emerging markets like Borsa Istanbul, on the other hand, present a more challenging environment with high volatility and limited liquidity. Because such markets are more sensitive to macroeconomic and geopolitical fluctuations, they require special attention in terms of portfolio optimization (Bekaert and Harvey, 2003). While machine learning applications in emerging market finance have increased in recent years, it is noteworthy that RL-based studies are still limited (Ozbayoglu et al., 2020; Bai et al., 2025). This highlights the need for new research, especially in understudied market contexts like Borsa Istanbul.

Table 1. Overview of Key Portfolio Optimization Methods and RL Studies, Summarizing Their Approaches, Strengths, and Contributions.

Reference	Methodology	Strength	Limitations	Key Contributions
Markowitz (1952)	Mean-variance optimization	Pioneered quantitative portfolio theory	Static; single-period	Introduced risk-return tradeoff framework
Samuelson (1975)	Dynamic stochastic programming	Addresses intertemporal choices	Relies on subjective utility	Multi-period asset allocation model
Merton (1973)	Continuous-time optimization	Integrates consumption-investment	High complexity	Continuous-time asset allocation
Black and Litterman (1990); Black and Litterman (1992)	Black-Litterman model	Balances views with priors	Requires confidence estimates	Unified subjective and market expectations
Rockafellar and Uryasev (2000)	CVaR optimization	Tail-risk focus; coherent	Ignores time structure	Risk measure for extreme losses
Qian (2011); De Prado (2016)	Risk parity models	Balanced risk allocation	Static weights	Alternative to mean-variance allocation
Sutton et al. (1999); Williams (1992)	Policy Gradient RL	Foundation for RL control	Simple environments	Base for RL in finance
Moody (1999) et al.	RL (PG, Q-learning)	Early RL use in finance	Single-asset focus	RL applied to trading decisions
Jiang et al. (2017); Liang et al. (2018)	DRL (DDPG)	Handles costs; model-free	Sample inefficiency	DRL with transaction cost integration
Almahdi and Yang (2017); Aboussalah and Lee (2020)	RRL with risk ratios	Reward-aware RL	Narrow reward types	RRL using Sharpe/Calmar
Lim et al. (2022)	RL with NAV reward	Adaptive rebalancing	Reward-specific	NAV-based RL reward design
Kochliaridis et al. (2023)	DRL + technical indicators	Hybrid strategy	Limited generalizability	Merged TA with DRL trading
Wang and Zhou (2020)	Continuous-time RL	Continuous decision setting	Implementation complexity	RL in continuous-time finance
Jang and Seong (2023)	DRL with MPT	Theory-driven design	Market-specific	Linked MPT with DRL frameworks
Yue et al. (2023)	DRL (DDPG)	Autonomous management	Static backtesting	DDPG for portfolio tasks
Gort et al. (2022)	DRL + hypothesis testing + cross-validation	Rejects overfitted agents; improves robustness	Crypto-specific setup; complex tuning	Combines hypothesis testing with DRL to improve portfolio performance
Li and Hai (2023)	Advanced DRL models	Multi-module learning	High complexity	DRL with expert systems
Sun et al. (2024)	DRL + GNN / BL	Models' relations, views	Poor convergence	DRL with GNN and BL integration
Yu et al. (2019)	Model-based DRL	Predictive augmentation	Data-heavy	Combined prediction and cloning

3. Methodology

The DDPG algorithm, a model-free, off-policy actor-critic method, is employed to determine optimal daily portfolio weights for the selected stocks. DDPG is an extension of the Deterministic Policy Gradient (DPG) algorithm proposed by Silver et al. (2014), which was further developed by Lillicrap et al. (2015) to operate effectively in continuous action spaces using deep neural networks. By maximizing the expected discounted cumulative reward, the model aims to achieve high risk-adjusted returns. The sample consists of stocks listed as components of Borsa Istanbul's BIST30 index. The performance of the resulting portfolio is compared against three benchmark portfolios: the BIST30 index, the Markowitz mean-variance portfolio, and an equal-weighted portfolio. The RL environment, reward function, algorithm configuration, and evaluation metrics are designed to ensure robustness and reproducibility, addressing the limitations of traditional methods in volatile markets like Borsa Istanbul (Jiang et al., 2017). The reward function is specifically designed to balance return and risk: it combines the 30-day rolling Sharpe ratio, for risk-adjusted returns, with an entropy-based diversification term ($\lambda = 0.05$) that penalizes concentrated allocations and reduces volatility.

Daily adjusted closing prices for BIST30 stocks are employed, and the BIST30 index is obtained using the Python yfinance library, covering the period from January 1, 2015, to December 31, 2024. After excluding stocks with more than 15% missing data, daily returns are calculated as:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}} \times 100 \tag{1}$$

where P_t is the adjusted closing price on day t. The training period is from January 1, 2015, to April 16, 2022, for all approaches; and the testing period for both is from April 16, 2022, to December 31, 2024.

The DDPG algorithm is implemented using Stable-Baselines3 (Brockman et al., 2016; Raffin et al., 2021). The environment is implemented as a custom PortfolioEnv class using OpenAI Gymnasium. The environmental architecture is structured as follows:

Action Space: The agent outputs a weight vector w subject to

$$\sum_{i=1}^{n} w_i = 1 \ , w_i \in [0,1] \tag{2}$$

Observation Space: The observation space includes a rolling 60-day window of standardized daily returns, normalized using a rolling StandardScaler.

Reward Function: Designed to balance risk-adjusted returns with diversification

$$R_t = \frac{R_P - r_f}{\sigma_p} - \lambda \sum_{i=1}^{n} w_i \ln\left(w_i + \epsilon\right) \tag{3}$$

where $\lambda = 0.05$ is the entropy regularization coefficient and r_f is the annualized risk-free rate. The model is optimized using a replay buffer of size 100,000 to store transitions.

To ensure full reproducibility, the specific hyperparameters and training configurations used in this study are. In the DDPG algorithm, both the actor and critic networks are multi-layer

perceptron with two hidden layers of 256 neurons each. Key hyper parameters include: learning rate: 0.0001, replay buffer size: 100000, batch size: 64, start training after: 5000-time steps, soft update parameter: τ = 0.001, discount factor: γ = 0.99. To encourage exploration, the agent employs Ornstein-Uhlenbeck noise with parameters: $\theta = 0.15, \sigma = 0.05, dt = 0.01$. The model is trained 50 times over 20000 timesteps each to employ different random seeds that ensures the robustness of the model and the trained policy is used for out-of-sample testing.

To ensure the generalizability of the results and account for the inherent stochasticity in neural network initialization and RL exploration, the DDPG model was trained and tested over 50 independent simulations. Each simulation utilized a different random seed for weight initialization. The performance metrics are compared with DDPG's performance, a one-sample t-test was conducted to determine if the mean Sharpe ratio of the DDPG agent was significantly different from those of the static measures of the Markowitz and Equal-Weight benchmarks at a 5% significance level ($p < 0.05$).

To quantify the diversification of the portfolio weights at each time step t, we employ Shannon Entropy (H_t), defined as:

$$H_t = -\sum_{i=1}^{n} w_{i,t} \ln (w_{i,t} + \epsilon) \tag{4}$$

where $w_{i,t}$ is the weight of asset i and ϵ is a small to ensure numerical stability. Higher entropy values indicate a more evenly distributed (diversified) portfolio, whereas lower values signify asset concentration.

Three benchmark portfolios are employed for comparison: the Markowitz Mean-Variance Portfolio, equal- weight portfolio, and BIST30 index.

The Markowitz Mean-Variance Portfolio optimizes the Sharpe ratio with constraints and regularization:

$$\max_{w} \frac{w^T \mu - r_f}{\sqrt{w^T (\sum + \epsilon I) w}} \quad subject\ to: w \geq 0, \qquad \sum_{i=1}^{n} w_i = 1 \tag{5}$$

where $\epsilon = 0.000001$ ensures numerical stability. The optimization is solved using Sequential Least Squares Programming (SLSQP).

A naive benchmark, the equal-weight portfolio distributes the weight equally to the number stocks:

$$w_i = \frac{1}{n}, \quad \forall i = 1, \ldots, n \tag{6}$$

where w_i is the weight of i_{th} stock and n is the number of stocks.

The third benchmark is the BIST30 index representing the Turkish equity market.

Performance is assessed using the following metrics: the first metric is the Sharpe Ratio (SP) defined as:

$$SR = \frac{R_P - r_f}{\sigma_p} \tag{7}$$

where R_P, r_f and σ_p are the annualized portfolio return, risk-free rate, and volatility, assuming 252 trading days. Sharpe ratios are computed using a constant stylized risk-free rate of 2% for consistency across all strategies. As this is a comparative evaluation, alternative realistic r_f values would produce a uniform shift in all ratios without changing the observed performance ordering. The second metric is Maximum Drawdown (MDD), which is defined as:

$$MDD = \max_t \left(\frac{max_{s\leq t}V_s - V_t}{max_{s\leq t}V_s}\right) \quad (8)$$

where V_s is the portfolio value at time s, V_t is the portfolio value at time t, and the maximum is taken over all time points t in the test period. This metric captures the worst-case loss an investor might experience. The third metric is cumulative return over the test period, initialized at $V_0 =$ \$1.

The study is implemented in Python 3.12.12 using a central processing unit (CPU) architecture, and the libraries employed are NumPy 2.0.2, Pandas 2.2.2, SciPy 1.16.3, Stable-Baselines3 2.7.1, Matplotlib 3.10.0, and Gymnasium 1.2.3. Training is performed on the training dataset, with out-of-sample evaluations conducted on the test dataset.

4. Results

This study evaluates the effectiveness of the DDPG algorithm in optimizing portfolios of BIST30 stocks. After cleaning data, 28 stocks were included in the analysis, achieved by excluding those with more than 15% missing data. The training period for all models is from January 2, 2015, to April 16, 2022, while the test period spans 625 trading days, from April 16, 2022, to December 31, 2024. These periods were selected to reflect a broad range of market conditions in a volatile market, allowing the DDPG to showcase its adaptive capabilities.

Performance of all approaches is compared based on the final value of the portfolio, the Sharpe ratio, and the maximum drawdown, and they are given in Table 2. To be comparable, all portfolios start with \$1, and the final values of the portfolios are compared. The portfolio value for the average of 50 simulated DDPG portfolios increases to \$5.25, delivering a relatively high final portfolio value and the strongest Sharpe ratio of 2.12. On the other hand, the BIST30 index had the lowest drawdown at 21.96% while DDPG recorded the third-lowest maximum drawdown at 26.88%.

Table 2. Portfolio Performance for Test Period

Portfolio	Final Value	Sharpe Ratio	Max Drawdown (%)
Mean DDPG	5.25	2.12	26.88
Equal-Weight	5.32	2.01	26.37
Markowitz	2.54	1.07	33.31
BIST30	3.86	1.68	21.96

While the Markowitz mean-variance optimization is theoretically optimal under quadratic utility and perfectly estimated inputs, its out-of-sample performance often deteriorates substantially due to estimation errors in expected returns and covariance. In such cases, simpler strategies like the equal-weight portfolio may outperform Markowitz allocations because they completely avoid parameter estimation risk and prevent extreme weight concentrations arising

from noisy input estimates. In contrast, the DDPG-based approach demonstrates superior performance by leveraging its dynamic, multi-period optimization framework and its ability to learn nonlinear relationships in the data without relying on distributional assumptions or explicit moment estimates. Unlike static quadratic optimization, DDPG directly learns a policy that maximizes realized performance over time, enabling it to better mitigate estimation errors and adapt to evolving market conditions.

These findings show the DDPG model's ability to generate competitive risk-adjusted returns and effectively manage downside risk in a volatile, emerging market context. Across 50 independent simulations, the DDPG strategy achieved a significantly higher final portfolio value than the Markowitz model, the BIST 30, and the Equal-Weight portfolio; it has a significantly higher Sharpe ratio relative to all three benchmarks. Additionally, DDPG exhibited significantly lower maximum drawdown than the Markowitz and Equal-Weight portfolios but significantly higher maximum drawdown than BIST 30, indicating superior return and risk-adjusted performance overall, with downside risk exceeding only that of the index benchmark. Therefore, the success of the DDPG portfolio underscores its potential when compared to other alternatives. In Figure 1, the portfolio value over time is shown for each approach—DDPG, Equal-Weight, Markowitz, and BIST 30—where the DDPG curve includes a 95% confidence interval, as it represents the mean performance across multiple simulated DDPG runs.

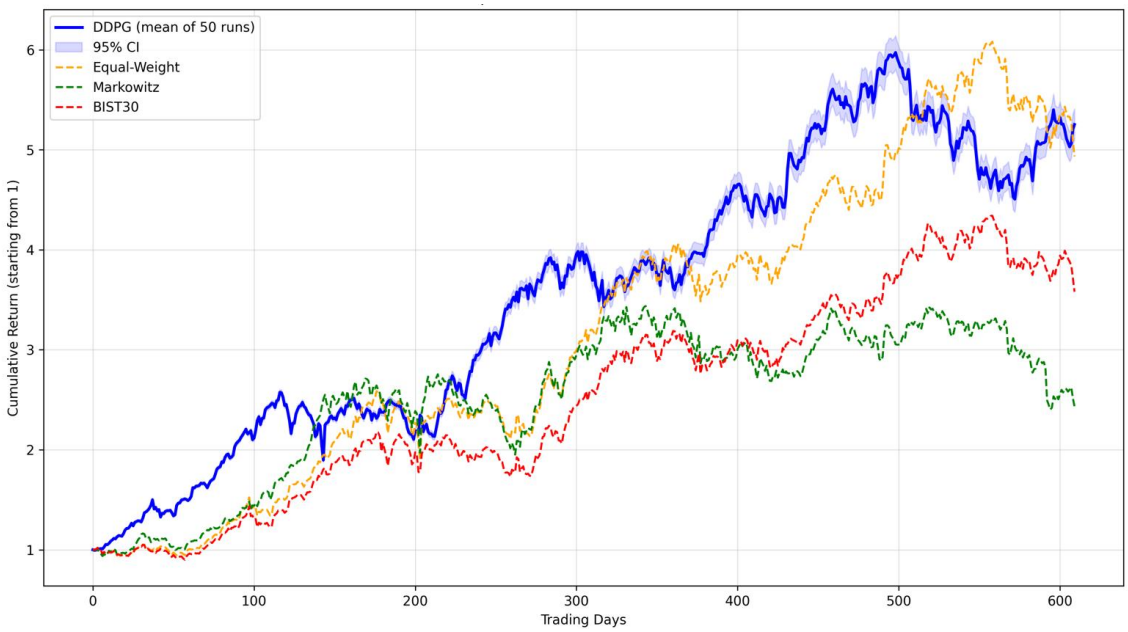

Figure 1. Portfolio Value Trajectories over Test Period for DDPG (with 95 CI), Equal-Weight, Markowitz, and BIST30 Strategies

Considering the highest total return and Sharpe ratio for the DDPG portfolio underscores DDPG's ability to balance returns and volatility more effectively, a key advantage when operating in the highly dynamic and uncertain environment of emerging markets like Borsa Istanbul.

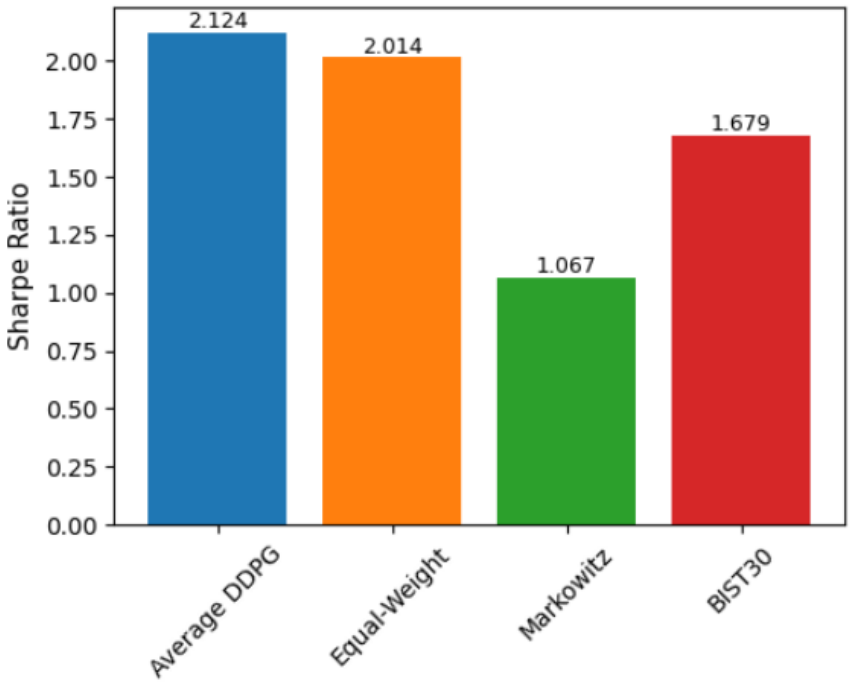

Figure 2. Sharpe Ratios of the Mean of Multiple Simulated DDPG Strategy, Equal-Weight Portfolio, Markowitz Portfolio, and the BIST30 Index During the Test Period.

To see how each portfolio allocates its wealth to each stock, the weight distribution of all portfolios is used in the test period. While the Markowitz portfolio exhibits more concentrated allocations, as determined, the equal-weight portfolio distributes weights uniformly, the DDPG strategy demonstrates a more diversified allocation pattern. This diversification is driven by the design of its reward function, which includes an entropy-based term (λ=0.05) alongside the Sharpe ratio, leading to broader exploration and sustained investment across multiple assets during its extended training period (January 1, 2015 – April 16, 2022).

To demonstrate the DDPG model's adaptive capabilities, we examined the evolution of its portfolio weights during the out-of-sample test period (April 2022 – December 2024). Unlike the Markowitz model, which remains static until a manual re-optimization, or the Equal-Weight approach, the DDPG agent re-evaluates its positions daily based on the 60-day rolling state window. As shown in the weight evolution analysis, the agent does not settle into a "buy and hold" stance. Instead, it dynamically rotates capital across BIST30 constituents (see Figure 3). During periods of high market stress in Borsa Istanbul, the entropy-based reward term (λ=0.05) successfully prevents the model from collapsing into a single-asset "winner-take-all" strategy, maintaining a mean effective number of assets higher than the concentrated Markowitz output.

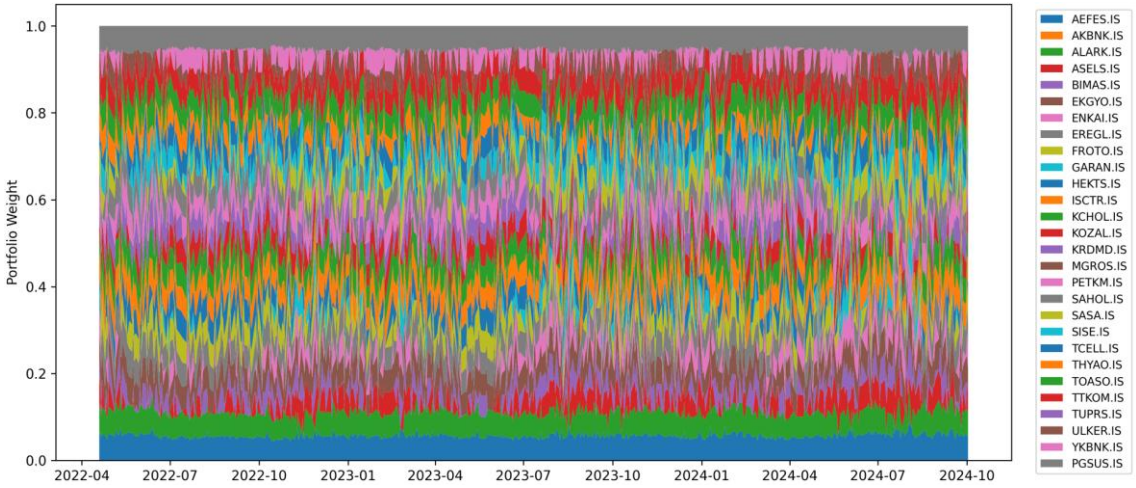

Figure 3. DDPG Dynamic Weight Allocation Evolution During the Test Period.

This Shannon Entropy (H_t) is a metric that quantifies the model's diversification consistency. As shown in Figure 4, the entropy levels remain high and stable throughout the test period, fluctuating near the theoretical maximum. This provides empirical evidence that the DDPG agent actively avoids asset concentration even while rotating capital, successfully balancing the objective of high returns with the structural requirement for a diversified risk profile.

The model exhibits high-frequency adaptability. The average daily turnover rate—defined as the sum of absolute changes in weights $\sum |w_{i,t} - w_{i,t-1}|$—was recorded at 0.44. While this reflects an active management style, the transitions remain fluid due to the actor-critic architecture, which learns to optimize the Sharpe ratio while simultaneously satisfying the entropy-based diversification constraint. This ensures that rebalancing is driven by statistical evidence of regime changes in BIST30 rather than noise-induced fluctuations.

Together, these findings underscore the DDPG's capacity to learn dynamic and diversified asset allocation strategies in the context of the Borsa Istanbul, showcasing its potential as a sophisticated tool for portfolio management in emerging markets (Jiang et al., 2017).

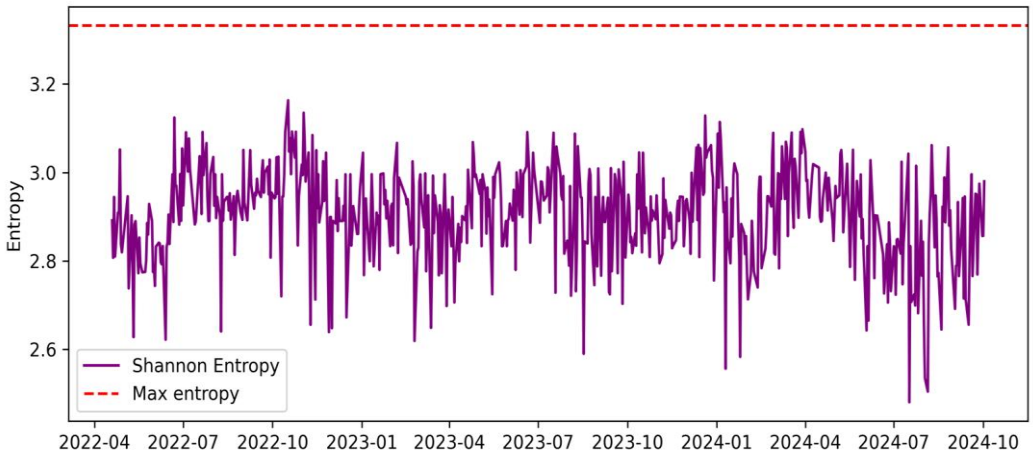

Figure 4. Portfolio Diversification Consistency (Shannon Entropy) for DDPG Algorithm

5. Conclusion

The findings indicate the superior Sharpe ratio of the RL algorithm DDPG among the different approaches developed for the BIST30 index components, they also demonstrate that the simplistic nature of other classical approaches is insufficient. This highlights the importance of approaches that can adapt to changing market conditions (Michaud, 1989).

Our findings support the growing body of work that recommends the use of machine learning and, specifically, RL, to overcome these limitations (Meng and Khushi, 2019). The success of the DDPG in dynamically adjusting portfolio weights, leading to a higher Sharpe ratio than the benchmarks, is consistent with the promising outcomes reported in developed markets by Jiang et al. (2017) and Liang et al. (2018), suggesting that these deep RL methods may have broader applicability. Granular analysis of weight trajectories confirms that the DDPG agent actively rotates assets in response to market regimes, providing a level of responsiveness that static benchmarks cannot replicate. These results have important implications for professional portfolio management in emerging markets, where traditional static models often fail due to high volatility, liquidity issues, and frequent regime changes. The DDPG approach delivers superior risk-adjusted returns and better downside protection than benchmarks, providing fund managers and institutional investors with a practical, adaptive tool for real-time rebalancing. In volatile environments like Borsa Istanbul, marked by macroeconomic shocks, geopolitical risks, and limited diversification, this AI-driven strategy can enhance capital allocation efficiency, reduce drawdowns in stress periods, and improve long-term client portfolio performance. Overall, deep RL offers a pathway to greater investment resilience and market stability in under-researched emerging economies.

Moreover, the more diversified portfolio allocation learned by the DDPG approach differs from the more focused portfolios typically resulting from mean-variance optimization, likely due to the way we set the reward, a point also discussed by De Prado (2016) regarding risk diversification. While most RL research in the stock market has focused on developed markets or specific asset classes such as cryptocurrencies, this study makes a novel contribution by applying the DDPG algorithm to Borsa Istanbul, a dynamic and relatively under-researched emerging market.

A key finding in this context is the model's rebalancing frequency; the average daily turnover rate, defined as the sum of absolute changes in weights, was recorded at 0.44. This indicates that the model actively "tilts" the portfolio daily to capture momentum or hedge against emerging risks, rather than making drastic, erratic shifts. This 'smooth' but continuous rebalancing is a hallmark of the DDPG's deterministic policy gradient, allowing for stable transition between optimal states while maintaining a high level of diversification as evidenced by Shannon Entropy metrics.

These findings reinforce the growing recognition that advanced machine learning approaches can help investment decision-making in less developed financial environments, and this highlights their ability to adapt to the unique characteristics of emerging markets. It also suggests that integrating artificial intelligence methods into trading in stock markets can enhance investment efficiency and market stability in emerging markets. However, this study has limits as it only considers BIST30 stocks over a specific period and relies on a standard DDPG framework with a basic reward structure. Specifically, the current model does not account for transaction costs. Given the observed turnover rate of 0.44, the impact of brokerage fees and bid-ask spreads

in a practical application within Borsa Istanbul could influence the net profitability. Therefore, future research should move toward "cost-aware" RL by enhancing the reward function to account for transaction costs and investor risk preferences.

Declaration of Research and Publication Ethics)
This study, which does not require ethics committee approval and/or legal/specific permission, complies with the principles of research and publication ethics. This study complies with the principles of research and publication ethics.

Researcher's Contribution Rate Statement
The authors declare that they have contributed equally to this article.

Declaration of Researcher's Conflict of Interest
There are no potential conflicts of interest in this study.

Declaration of Artificial Intelligence Usage
During the preparation of this study, the author used ChatGPT / OpenAI for Language Editing. After using this tool/service, the content was reviewed and edited as necessary, and the author is solely responsible for the content of the published article.

References

Aboussalah, A.M. and Lee, C.G. (2020). Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization. *Expert Systems with Applications*, 140, 112891. https://doi.org/10.1016/j.eswa.2019.112891

Almahdi, S. and Yang, S.Y. (2017). An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications*, 87, 267–279. https://doi.org/10.1016/j.eswa.2017.06.023

Bai, Y., Gao, Y., Wan, R., Zhang, S. and Song, R. (2025). A review of reinforcement learning in financial applications. *Annual Review of Statistics and Its Application*, 12(1), 209–232. https://doi.org/10.48550/arXiv.2411.12746

Bekaert, G. and Harvey, C.R. (2003). Emerging markets finance. *Journal of Empirical Finance*, 10(1–2), 3–55. https://doi.org/10.1016/S0927-5398(02)00054-3

Black, F. and Litterman, R. (1990). Asset allocation: Combining investor views with market equilibrium. *Journal of Fixed Income,* 1(2), 7–18. https://doi.org/10.3905/jfi.1991.408013

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5), 28–43. https://doi.org/10.2469/faj.v48.n5.28

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J. and Zaremba, W. (2016). OpenAI gym. *arXiv preprint* arXiv:1606.01540. https://doi.org/10.48550/arXiv.1606.01540

De Prado, M.L. (2016). Building diversified portfolios that outperform out-of-sample. *Journal of Portfolio Management,* 42(4), 59–69. https://doi.org/10.3905/jpm.2016.42.4.059

Deng, Y., Bao, F., Kong, Y., Ren, Z. and Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems,* 28(3), 653–664. https://doi.org/10.1109/TNNLS.2016.2522401

Erdoğan, L., Ceylan, R. and Abdul-Rahman, M. (2022). The impact of domestic and global risk factors on Turkish stock market: Evidence from the NARDL approach. Emerging *Markets Finance and Trade,* 58(7), 1961–1974. https://doi.org/10.1080/1540496X.2021.1949282

Fabozzi, F.J., Kolm, P.N., Pachamanova, D.A. and Focardi, S.M. (2007). Robust portfolio optimization. *Journal of Portfolio Management*, 33(3), 40–48. https://doi.org/10.3905/jpm.2007.684751

Gort, B.J.D., Liu, X.Y., Sun, X., Gao, J., Chen, S. and Wang, C.D. (2022). Deep reinforcement learning for cryptocurrency trading: Practical approach to address backtest overfitting. *arXiv preprint* arXiv:2209.05559. https://doi.org/10.48550/arXiv.2209.05559

Jang, J. and Seong, N. (2023). Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory. *Expert Systems with Applications*, 218, 119556. https://doi.org/10.1016/j.eswa.2023.119556

Jiang, Z., Xu, D. and Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint* arXiv:1706.10059. https://doi.org/10.48550/arXiv.1706.10059

Kochliaridis, V., Kouloumpris, E. and Vlahavas, I. (2023). Combining deep reinforcement learning with technical analysis and trend monitoring on cryptocurrency markets. *Neural Computing and Applications,* 35(29), 21445–21462. https://doi.org/10.1007/s00521-023-08516-x

Liang, Z., Chen, H., Zhu, J., Jiang, K. and Li, Y. (2018). Adversarial deep reinforcement learning in portfolio management. *arXiv preprint* arXiv:1808.09940. https://doi.org/10.48550/arXiv.1808.09940

Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D. and Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint* arXiv:1509.02971. https://doi.org/10.48550/arXiv.1509.02971

Lim, Q.Y.E., Cao, Q. and Quek, C. (2022). Dynamic portfolio rebalancing through reinforcement learning. *Neural Computing and Applications*, 34(9), 7125–7139. https://doi.org/10.1007/s00521-021-06853-3

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91. https://doi.org/10.2307/2975974

Meng, T.L. and Khushi, M. (2019). Reinforcement learning in financial markets. *Data,* 4(3), 110. https://doi.org/10.3390/data4030110

Merton, R.C. (1973). An intertemporal capital asset pricing model. *Econometrica,* 41(5), 867–887. https://doi.org/10.2307/1913811

Michaud, R.O. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45(1), 31–42. https://doi.org/10.2469/faj.v45.n1.31

Moody, J. and Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks,* 12(4), 875–889. https://doi.org/10.1109/72.935097

Ozbayoglu, A.M., Gudelek, M.U. and Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, 93, 106384. https://doi.org/10.1016/j.asoc.2020.106384

Qian, E. (2011). Risk parity and diversification. *Journal of Investing,* 20(1), 119. Retrieved from https://www.panagora.com/

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M. and Dormann, N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268), 1–8. Retrieved from https://jmlr.org

Rjoub, H., Türsoy, T. and Günsel, N. (2009). The effects of macroeconomic factors on stock returns: Istanbul Stock Exchange. *Studies in Economics and Finance*, 26(1), 36–45. https://doi.org/10.1108/10867370910946315

Rockafellar, R.T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3), 21–41. Retrieved from https://sites.math.washington.edu

Samuelson, P.A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3), 239–246. https://doi.org/10.2307/1926559

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D. and Riedmiller, M. (2014). Deterministic policy gradient algorithms. In E. P. Xing and T. Jebara (Eds.), *Proceedings of the 31st International Conference on Machine Learning* (pp. 387–395). Retrieved from https://proceedings.mlr.press/

Sun, R., Stefanidis, A., Jiang, Z. and Su, J. (2024). Combining transformer based deep reinforcement learning with Black-Litterman model for portfolio optimization. *Neural Computing and Applications,* 36(32), 20111–20146. https://doi.org/10.1007/s00521-024-09805-9

Sutton, R.S. and Barto, A.G. (2018). *Reinforcement learning: An introduction (2nd ed.).* Cambridge: MIT Press.

Sutton, R.S., McAllester, D.A., Singh, S.P. and Mansour, Y. (1999). Policy gradient methods for reinforcement learning with function approximation. *Advances in Neural Information Processing Systems*, 12, 1057–1063. Retrieved from https://proceedings.neurips.cc/

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. Retrieved from https://www.jstor.org/

Wang, H. and Zhou, X.Y. (2020). Continuous-time mean–variance portfolio selection: A reinforcement learning framework. *Mathematical Finance*, 30(4), 1273–1308. https://doi.org/10.48550/arXiv.1904.11392

Williams, R.J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning,* 8(3), 229–256. https://doi.org/10.1007/BF00992696

Yu, P., Lee, J.S., Kulyatin, I., Shi, Z. and Dasgupta, S. (2019). Model-based deep reinforcement learning for dynamic portfolio optimization. *arXiv preprint* arXiv:1901.08740. https://doi.org/10.48550/arXiv.1901.08740