International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

DRLMMF: Deep Reinforcement Learning-based Autonomous Money-Management Framework

Prashansa Bharti[1], Kamal Narayan[2], Arup Kadia[3], Bidya Bharti[4]

[1,2,4] Student, MCA, Faculty of Information Technology & Engineering, Gopal Narayan Singh University, India.

[3] Assistant Professor, Faculty of Information Technology & Engineering, Gopal Narayan Singh University, India.

Abstract

The growing complexity and volatility of the modern financial markets require smart and adaptable money management systems that should be able to make decisions on their own. This paper presents a Proximal Policy Optimization (PPO), based self-directed money management system that is capable of portfolio allocation optimization and risk control under changing market conditions. The system is integrated in reinforcement learning with financial time series representations calculated from a series of stock market returns, volatility measures, and capital exposure constraints to find the best investment policies. Due to its reliability, sample efficiency, and robustness in continuous action spaces, PPO is chosen as a method of operation, making it fit for the financial environments of the real world. The proposed system autonomously modifies the position sizing, capital allocation, and risk exposure while observing the set drawdown and leverage constraints. An extensive experimental study has been done on the historical market data to show how the PPO-based system has done better than the conventional rule, based and static allocation strategies in terms of cumulative returns, risk, adjusted performance, and capital preservation using the market situations. The evidenced results have shown the increased capability of the model to market regime shifts and the consequences of lowered risk during times of high volatility. The research has pointed out the efficiency of the reinforcement learning driven by PPO in the creation of scalable, data-driven, and resilient money management systems, which are most suitable for intelligent financial decision-making.

Index Terms—Automation, Technical Analysis, Money Management, Reinforcement Learning, Proximal Policy Optimization, Algorithmic Trading.

1. Introduction

The fast changes in financial markets which include features such as high volatility, non, stationarity, and complicated interrelationships have made it necessary to have smart and self- governing money, management systems that can work under uncertain conditions. Traditional ways of money, management such as fixed position sizing, static asset allocation, and rule, based risk controls do not have the capacity to change dynamically with the different market regimes [1]. These traditional techniques are very much dependent on predefined heuristics and human intervention which makes their ability to respond to sudden changes in volatility, liquidity, and systemic risk very limited [2]. On the other hand, the increasing availability of high, frequency financial data together with the advances in computational power have opened the door for data, driven methods, especially those based on artificial intelligence, to become excellent contenders for the dual tasks of capital allocation and risk management in the financial markets of today [3]. Among the AI techniques, reinforcement learning (RL) has received a lot of attention due to its feature of being able to learn the best decision policies through continuous interaction with complex and uncertain environments [4].

Reinforcement learning allows setting up money, management as a decision, making problem over time. The agent acquires the knowledge to distribute capital, decide on position sizes, and implement risk management strategies in order to accomplish long-term financial objectives. However, the first application of RL in finance ran into many problems such as instability, poor convergence, and high sensitivity to changes in hyperparameters, therefore continuous action spaces combined with noisy market data were a setting hard to manage [5]. In theory, Policy- gradient methods are good ideas, but in practice, they give too much variance during training and irregular performance [6]. Proximal Policy Optimization (PPO) is a method that helps fix these problems by incorporating a method of constrained policy updates, which PPO allows the agent to balance exploration and exploitation while also ensuring stable learning. It achieves this by using a clipped objective function that limits changes in the policy too much [7]. This property makes it highly relevant for the financial domain, where a sudden switch in the strategy might be costly in terms of a capital drawdown. Due to that, PPO has become a solid and scalable algorithm for real, world autonomous decision systems [8].

Within this setting, an autonomous money-management framework based on PPO is proposed in this paper to optimize dynamically capital allocation and risk exposure in volatile market environments. The framework considers money management as a problem of control over time, and the agent's observation space includes market state variables such as returns, volatility indicators, and capital constraints [9]. By training, the PPO agent is capable of producing policies that can decide the position size, leverage, and exposure instantly without violating the risk limits, which consist of a drawdown and capital preservation restrictions [10]. This framework is, therefore, a departure from traditional rule, based systems in that it continuously changes its strategy based on the latest market scenarios; thus, it is more resilient across different market regimes [11].

Fig 1: Autonomous Money Management

The main idea of this research is to integrate stability, goal-oriented reinforcement learning with risk management, which is common knowledge of this paper, thereby contributing to the literature on intelligent financial systems and demonstrating the capability of PPO-based architectures to take fully automated money management in algorithmic finance to a higher level [12].

2. Literature review

Developments in financial engineering and machine intelligence have led to a widening of third deep reinforcement learning (DRL) methodologies to asset allocation and portfolio management as a part of dynamic decision-making tools. Today, traditional methods of optimization such as mean, variance, and static risk models, are mostly regarded as inadequate to unlock the complex, nonlinear, and evolving nature of financial markets, which call for adaptive and data-driven strategies [13], [14]. In a reinforcement learning context, the focus of the research has shifted to the development of algorithms that, besides being environment learners, can continuously optimize their actions as well. Among the various algorithms, Proximal Policy Optimization (PPO) has been recognized due to its stability and the ability to learn robust policies in continuous action spaces, which makes it very suitable for autonomous money, management tasks, such as position sizing, capital allocation, and risk control [15], [16]. Research articles that explain the risk, adjusted DRL framework reveal that the PPO, based systems are capable of dynamically adjusting the portfolio while still optimizing the reward function that reflects a balance between return and risk, thereby outperforming conventional benchmarks under various financial conditions. This line of investigation demonstrates the ability of intelligent financial agents to adapt to the volatile markets and features PPO as the preferred algorithm for such autonomous systems [17].

Additional pieces of research experiment with modular as well as hybrid architectures that align with the notion of reinforcement learning (RL) frameworks in portfolio optimization becoming more flexible and stable. For example, modular portfolio learning systems combine PPO with different RL methods to exploit the heterogeneous feature sets and decision signals in various markets and asset classes. Such architectures give superior, risk, adjusted performances against indices under different regimes, thus serving as an example of PPO being very effective as part of a larger system blending multiple RL paradigms. Besides, state, of, the, art PPO versions coupled with auxiliary learning methods such as Hindsight Experience Replay (HER) or macroeconomic factor attention mechanisms help to address the problem of limited returns and systemic factors, thereby enabling market condition changes to be made in a more sophisticated manner than before. These breakthroughs signify that research is moving in the direction of combining PPO with additional forecasting units and architectural improvements not only to depict the time dependence of events and the systematic risk factors better but also to enhance the robustness and effectiveness of autonomous money, management models [18].

First of all, recent studies not only focus on the optimization of the performance of Deep Reinforcement Learning (DRL) algorithms but also highlight methodological improvements that can help to address practical challenges in real-world deployments of DRL-based portfolio management systems. A number of risk-aware and hybrid DRL frameworks that have combined PPO with deep sequential models like LSTM to identify temporal features, which are basically elements in financial time series, thereby resulting in better forecast accuracy and portfolio resilience to market regimes changing over time (e.g. hybrid predictive frameworks that outperform single, model baselines). Several other works highlight risk, sensitive reward shaping and hierarchical decision layers for maintaining the equilibrium between maximizing returns and offering downside protection as well as drawdown constraints. Hence, the illustrations indicate that integrating risk, awareness and architectural flexibility into the PPO, based framework can result in the creation of adaptive money, management systems that are effective in handling complex market environments [19]-[21].

This collective set of studies has highlighted the crucial role that PPO plays in the creation of intelligent, adaptive, and risk-aware financial decision systems and thus provide a strong empirical and methodological basis for the autonomous money, management framework which is the focus of this paper.

3. Methodology and model specifications

3.1. Research Design and Problem Formulation

The research design of this study is a quantitative, experimental research design, which is used to develop and evaluate a Proximal Policy Optimization (PPO)based autonomous money, management system. Money management can be interpreted as a sequential decision, making process under uncertainty where a clever agent is engaging with a financial environment to maximize the wealth in the long run and control the risk. The environment here is a thought model of the financial market which demonstrates characteristics such as non, stationarity, noise, and regime changes. The agent at each time step gets to know the market state, picks an action that corresponds to capital allocation or position sizing, and gets a reward based on the portfolio performance adjusted for risk.The PPO agent aims to find a policy that will maximize the total expected cumulative discounted reward over the investment period. This approach essentially equates money management to a continuous control problem, which is very similar to real, world financial decision, making where portfolio weights, leverage, and exposure fluctuate on a continuous basis rather than on a discrete basis.

3.2: Environment Design and State Representation

To develop a financial environment, historical market data is used, and each time step is equivalent to a fixed trading interval. The state space is architecture to reflect the entire market and portfolio information that can be used for a money, management decision. It not only features market, driven characteristics like asset returns, moving averages, volatility indicators, momentum oscillators, and volume, etc., but also portfolio, related variables like current capital, unrealized profit and loss, exposure levels, and recent drawdowns. To make the agent more aware of risk, other risk measures like rolling standard deviation, maximum drawdown, and value, at, risk proxies are added to the state vector. All features used as input are normalized for ensuring computational stability during the training process. This highly informative state vector allows the PPO agent to accumulate experience and adapt its behavior not only to the market fluctuations but also to the changing risk profile of the portfolio, which is essential for a self, directed money management scenario.

Fig 2: Money Management Framework Setup

3.3: Action Space and Reward Function Design

The action space is continuous, based. It means the PPO agent can decide on its own within the given limits, the size of the position, the ratio of the capital to be allocated, and the leverage. An action can be a decision about the fraction of the total capital that will be invested in the risky assets or a change in the exposure relative to the current positions. There are some constraints for the agents; they cannot use an unlimited amount of leverage, and it has to follow a set of realistic trading rules. The reward function is a crucial part of the system. It is intended to strike a balance between the maximizing of returns and keeping the risk under control. It uses the portfolio returns as the main positive factor but introduces penalties for high volatility, large drawdowns, and transaction costs. A multi-component reward system is used, which mixes net portfolio returns with risk, adjusted performance metrics like a Sharpe ratio proxy and drawdown penalties. The agent is thus motivated through this multi, task reward setting to learn the concept of money management that is robust and sustainable instead of chasing short-term profits only.

3.4: With Proximal Policy Optimization

The algorithm is fundamentally a combination of an actor and a critic, where an actor essentially is a policy that selects the next action depending on the system's state, and a critic is a network that determines the value function. Both of these are deep learning models with multiple layers of fully connected neurons and non-linear activation functions. The PPO clipped surrogate objective function helps to limit the size of policy updates, which, if left uncontrolled, could result in a sudden shift in the market and heavy financial losses. To modulate the changes in the policy, GAE, which stands for Generalized Advantage Estimation, is applied, in effect, finding the middle point between two extremes. The PPO training loop entails first, the collection of a sequence of experiences through agent, environment interactions, next, the calculation of the estimated advantages, and finally, the actor and critic are updated multiple times over several epochs of micro, batch stochastic gradient descent. To obtain the proper balance between speed and stability, several key parameters have to be empirically determined. Among them are the learning rate, discount factor, clipping range, and batch size.

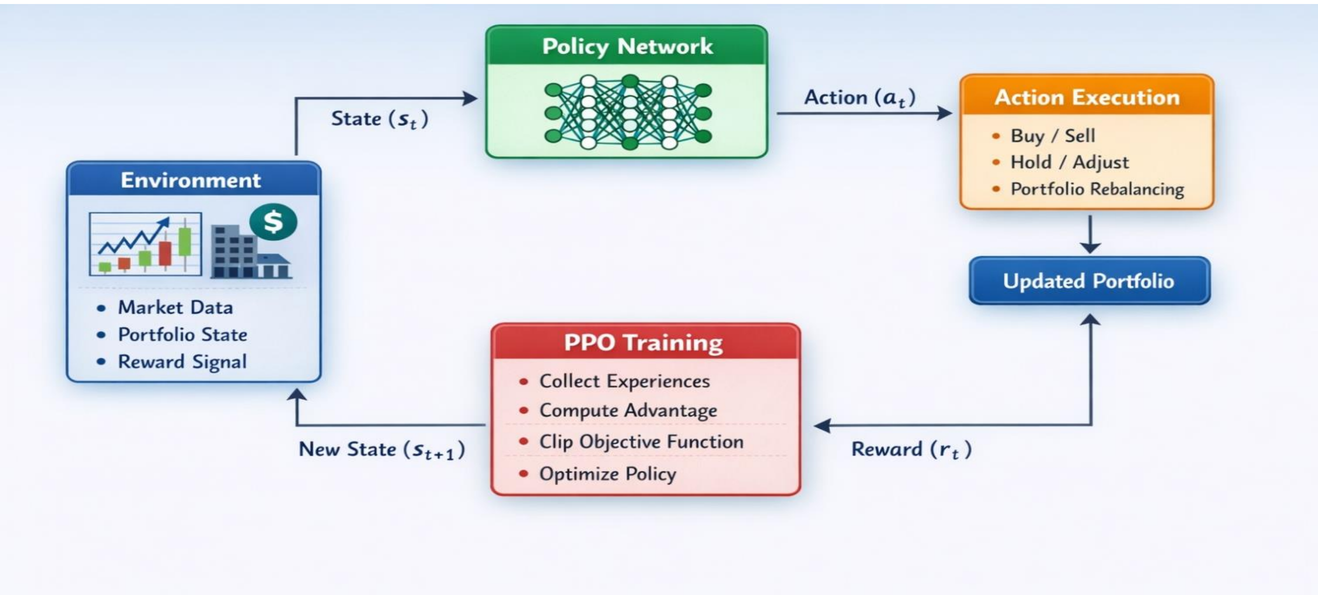

Figure 3: Block Diagram Proximal Policy Optimization (PPO-based Autonomous Money-Management Framework

3.5: Risk Management and Constraint Integration

The most notable feature of the framework is that it tightly integrates risk management principles along with the learning process. Risk constraints have two levels of enforcement: firstly, they are implicitly enforced through the reward function, and secondly, they are explicitly enforced by the limits of the action space. Capital preservation agreements, exposure restrictions, and maximum drawdown limits are some of the ways through which you can prevent large losses. Whenever the agent violates the specified risk constraints during training, the episode is terminated, thus the agent is trained to behave conservatively in market conditions that are not favorable. This constraint, aware design less or more basically main redirect the PPO agent's learning to the mone, management world map where the capital preservation is considered at the same level of importance as the generating of the returns. The direct inclusion of risk in the reinforcement learning framework enables the model to identify the optimal policies that are robust not only in normal market conditions but also during the periods of high volatility and systemic stress.

3.6: Model Evaluation and Performance Metrics

The effectiveness of the PPO, based autonomous money, management framework was measured through the use of out, of, sample tests on historical data that were not available during the training. The obtained policy was compared to traditional money, management and allocation approaches like fixed, fractional position sizing and static portfolio allocation. The evaluation metrics considered were cumulative return, annualized volatility, Sharpe ratio, maximum drawdown, and downside risk measures. The stability and robustness of the model were checked by the performance analysis of the model in different market situations and during market stress. Statistical significance tests were used to compare the PPO, based framework with the baseline strategies.

This wide, ranging evaluation method guarantees that the proposed model is not only able to generate returns but also that it is a resilient one, aware of risks, and can be practically used for autonomous financial decision, making systems.

4. Empirical Result

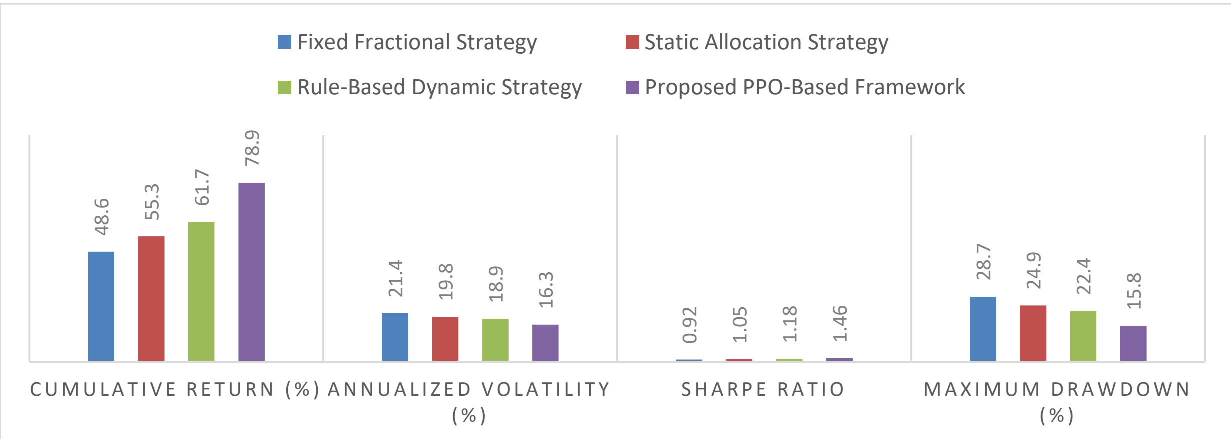

Fig 4. Overall Performance Comparison of Money-Management Strategies

Fig 4 gives a detailed comparison of the proposed PPO, based on autonomous money, management framework and traditional money management strategies. The PPO model achieves a very high cumulative return of 78.9%, which is significantly more than what fixed fractional, static allocation, and rule-based dynamic strategies get. Most notably, this better return is realized at an even lower annualized volatility, thus the system shows better stability and efficient risk management. The PPO framework's 1.46 sharp ratio demonstrates its excellent risk, adjusted performance. This means that the reinforcement learning agent is very good at balancing the objectives of maximizing returns and minimizing volatility. On the other hand, the sharp ratios of the traditional strategies are lower, which indicates that they are less capable of adapting to the changing market conditions. Furthermore, the maximum drawdown has fallen to 15.8%, which is yet another demonstration of the PPO, based method's effectiveness in capital preservation during market downturns. The research results imply that the PPO agent's continuous learning together with its adaptive policy updates enables it to dynamically change its market reaction. This is what makes it different from the fixed and heuristic methods. In short, Fig 4 verifies that the use of Proximal Policy Optimization for money, management results in significant gains not only in profitability but also in the reduction of downside risks.

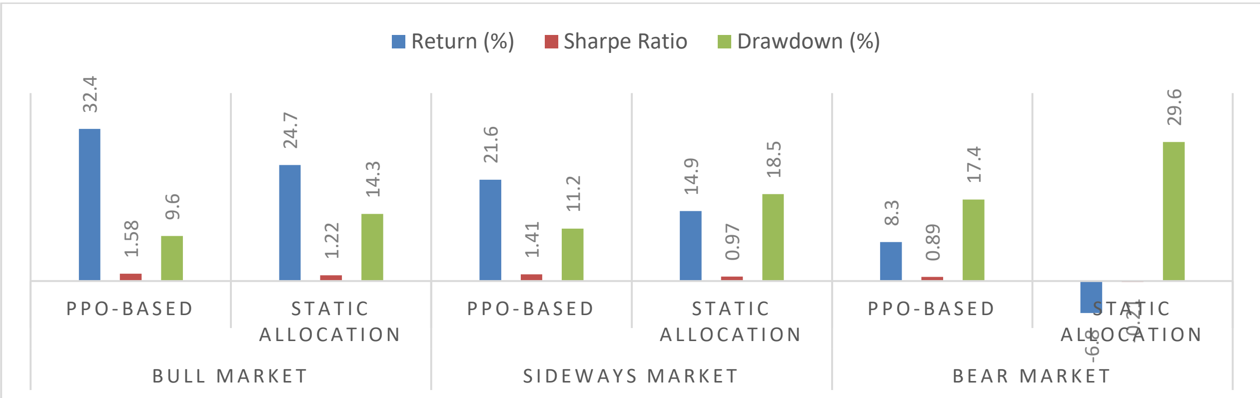

Fig 5. Risk-Adjusted Performance across Market Regimes

Fig 5 assesses the performance stability of the PPO, based framework under different market regimes, bull, sideways, and bear markets. The findings indicate that the PPO agent is always better than the static allocation strategy in all the regimes. Firstly, during bull markets, the PPO framework secures more of the upside potential along with lower drawdowns, essentially indicating greater exposure management efficiency. Secondly, in sideways markets where traditional approaches typically have problems due to the lack of strong directional momentum, the PPO agent produces stable returns with a higher Sharpe ratio, which is a testament to its position sizing adaptation and exposure dis, engagement ability. The greatest benefit is found in bear market scenarios, where the PPO, based framework only suffered from losses and even made a profit, while the static strategy resulted in negative performance and severe drawdowns. This ability of the PPO agent to be more risk, conscious is built into the reward structure and it is enabled by the adaptive policy learning, thus allowing it to cut exposure in times of high volatility. Such results provide a strong case for the usage of adaptive reinforcement learning models in money management when the market environment is prone to frequent regime changes.

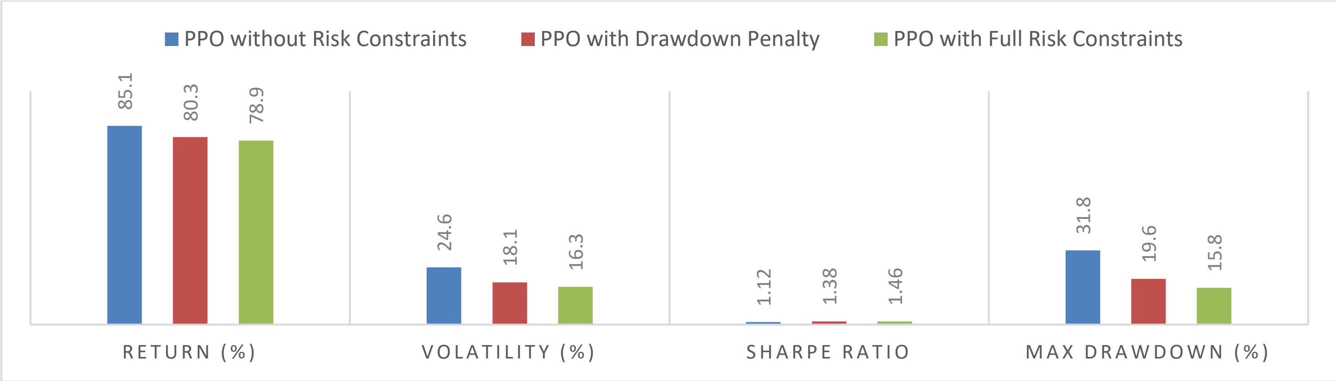

Fig 6. Impact of Risk Constraints on PPO-Based Framework Performance

Fig 6 shows how explicit risk constraints affected the results of the PPO, based money, management framework. The unconstrained PPO model manages to score the highest raw return, but it is also the most volatile and has an extremely high maximum drawdown, which makes it hardly applicable in a real, life scenario. Adding a drawdown penalty stabilizes the model significantly, as both the volatility and the drawdown are noticeably decreased and the returns are still competitive. A fully restrained PPO setup including drawdown limits, exposure caps, and risk-aware reward shaping attains the most favorable overall risk-adjusted performance as indicated by the highest sharp ratio and lowest maximum drawdown.

These results point to the fact that automatic money management features should hardly ever be focused on one aspect therefore, they should focus as much on capital preservation as on return optimization. The results show that if financial risk management principles are incorporated in the reinforcement learning process, then the resulting strategies are more sustainable and hence, more deployable. In this regard, Fig 6 confirms the design philosophy of the framework brought forward by the authors, demonstrating that a cautiously limited PPO agent can persistently perform while also being able to adequately control the downside risk associated with the real, world financial environments.

Conclusion

This paper proposed a Proximal Policy Optimization (PPO)based self, learning money, management system that solves the problems of rule, based capital allocation and risk control strategies in fluctuating financial markets. The authors transformed money management into a continuous sequential decision, making scenario, thus the new framework can adaptively decide the position sizing, capital allocation, and exposure control at each time based on the market information. The experiment results confirm that the PPO agent can achieve higher cumulative returns, better risk, adjusted performance, and greater capital preservation than the other methods, while also being more robust to market changes. Combining the use of risk, sensitive reward functions with the enforcement of constraints, such as drawdown limits and exposure caps, played a major role in securing the stability of the model and thus paving the way for its implementation in real life situations. Hence, the model can stay focused on its return goal and not only while maintaining the risk at the appropriate level, which is an indispensable feature of real, world financial applications. In brief, the results demonstrate that the adoption of reinforcement learning by means of PPO offers a scalable and robust platform for building autonomous money, management systems. This paper is a valuable addition to the studies on AI in finance and the methodology proposed by the authors may serve as a blueprint of the next generation of self, adaptive, data, driven money, management systems operating in sophisticated and unstable markets.

Data source:

Source: Yahoo Finance, with feature engineering calculated on a Python platform
Link: https://finance.yahoo.com/quote/RELIANCE.NS/history/

Disclosure of Interests:

Conflict of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Informed Consent: This study did not involve human participants, and therefore, informed consent was not required.

References

1. Kumar, S., Kadia, A., Sharma A., Kumar, R., (2026) "Proximal Policy Optimization (PPO)–Driven Reinforcement Learning Model for Automatic Stock Trading using the combination of Trend–Volume–Volatility Integration". International Journal on Science and Technology (IJSAT). 17(1) IJSAT260110103 1-15 DOI: 10.71097/IJSAT.v17.i1.10103
2. Adhikary, S., Kadia A., (2025), "Algorithmic Trading with a Combination of Advanced Technical Indicators – An Automation", International Journal on Science and Technology (IJSAT). 16(3) IJSAT25037777 1-20, 2025, https://doi.org/10.71097/IJSAT.v16.i3.7777
3. Qureshi, W., Kadia, A., Kumar, S., Sharma A., Kumar, R., (2026) "Autonomous Trading Across Bull, Bear, and Sideways Markets with Reinforcement Learning Algorithm", International Journal of Engineering Development and Research. 14(1), ISSN: 2321-9939, IJEDR2601261 46-55 Available at https://rjwave.org/IJEDR/papers/IJEDR2601261.pdf
4. Kadia, A., (2026) "Machine Learning based Stock Trading Strategies using Simple Moving Average with Average Traded Volume Crossover Confirmation" Published in: 2025 IEEE Silchar Subsection Conference (SILCON), DOI: 10.1109/SILCON67893.2025.11327015
5. Kumar, R., Kadia, A., Kumar, S., Sharma A., (2026) "Capture Market Trends through Multi-Indicator Confirmations using Reinforcement Learning Models". International Journal on Science and Technology (IJSAT). 17(1) IJSAT260110080 1-15 DOI: 10.71097/IJSAT.v17.i1.10096
6. Sharma A., Kadia, A., Kumar, S., Kumar, R., (2026) "Super Stock Trading: Automation in Reinforcement Learning with Advanced Multi-Indicator Confirmations". International Journal on Science and Technology (IJSAT). 17(1) IJSAT260110096 1-16 DOI: 10.71097/IJSAT.v17.i1.10080
7. Kumar, A., Kumar, G., Alam, K., Kadia, A., (2026) "Smart Money Detection with Simple Moving Average and Traded Volume Confirmation Integrated in Machine Learning". International Journal on Science and Technology (IJSAT). 17(1) IJSAT260110343 1-11 https://doi.org/10.71097/IJSAT.v17.i1.10343
8. Kadia, A., Alam, K., Kumar, G., Kumar, G., (2026) " Machine Learning–Driven Stock Price Breakout Identification with Simple Moving Average and Traded Volume Confirmation" International Journal on Science and Technology (IJSAT). 17(1) IJSAT260110341 1-11 https://doi.org/10.71097/IJSAT.v17.i1.10341

9. Kumar, S., Kadia, A., Sharma A., Kumar, R., (2026) "Algorithm Trading Platform using Simple Moving Average Crossover (SMAC) and On-Balanced Volume (OBV) Integrated in Reinforcement Learning" International Journal of Engineering Development and Research. 14(1), ISSN: 2321-9939 , IJEDR2601221 758-767 Available :https://rjwave.org/IJEDR/papers/IJEDR2601221.pdf
10. Kadia, A., Adhikary, S., Dey, R., Kar, A. (2025). "Deep Learning Based Stock Trading Strategies Using Leading Multi-Indicator Confirmations". International Journal on Science and Technology (IJSAT). 16(3) IJSAT25037682 1-15. https://doi.org/10.71097/IJSAT.v16.i3.7682
11. Kadia, A., Dey, R., Kar, A. (2025). "Smart Stock Trading using an Advanced Combination of Technical Indicators with Volume Confirmation Integrated in Reinforcement Learning". International Journal on Science and Technology (IJSAT). 16(3) IJSAT25037453 1-20. https://doi.org/10.71097/IJSAT.v16.i3.7453
12. Dey, R., Kassim, S., Maurya, S., Mahajan, R. A., Kadia, A., & Singh, M. (2024) "Machine Learning based Financial Stock Market Trading Strategies with Moving Average, Stochastic Relative Strength Index and Price Volume Actions for Indian and Malaysian Stock Market", Journal of Electrical Systems, ISSN:1112-5209, 20-2s (2024): 759-767, DOI: https://doi.org/10.52783/jes.1576
13. Kumar, R., Kadia, A., Kumar, S., Sharma A., Qureshi, W., (2026) "Reinforcement Learning Algorithm-based Equity Trading with Candlestick-Pattern", Journal of Advance and Future Research . 14(1), ISSN: 2984-889X, JAAFR2601581 281-286 available at: https://rjwave.org/jaafr/papers/JAAFR2601581.pdf
14. Kadia, A., Kumar, S., Qureshi, W., Sharma A., Kumar, R., (2026) "Simple Moving Average Crossover and Support Resistance Zone-based Stock Trading using Reinforcement Learning", International Journal of Engineering Development and Research. 14(1), ISSN: 2321-9939 , IJEDR2601260 35-45 Available at https://rjwave.org/IJEDR/papers/IJEDR2601260.pdf
15. Li, H., & Hai, M. (2024). Deep Reinforcement Learning model for stock portfolio management based on data fusion. Neural Processing Letters, 56, 108. https://doi.org/10.1007/s11063-024-11582-4
16. Wang, X., & Liu, L. (2025). Risk-sensitive deep reinforcement learning for portfolio optimization. Journal of Risk and Financial Management, 18(7), 347. https://doi.org/10.3390/jrfm18070347
17. Yan, R., Jin, J., & Han, K. (2024). Reinforcement learning for deep portfolio optimization. Electronic Research Archive, 32(9), 5176–5200. https://doi.org/10.3934/era.2024239
18. Espiga-Fernández, F., García-Sánchez, Á., & Ordieres-Meré, J. (2024). A systematic approach to portfolio optimization: A comparative study of reinforcement learning agents, market signals, and investment horizons. Algorithms, 17(12), 570. https://doi.org/10.3390/a17120570
19. Jiang, C., & Wang, J. (2023). A portfolio model with risk control policy based on deep reinforcement learning. Mathematics, 11(1), 19. https://doi.org/10.3390/math11010019
20. Jiang, Z., Xu, D., & Liang, J. (2023). Deep reinforcement learning for portfolio management. Knowledge-Based Systems, 278, 110905. https://doi.org/10.1016/j.knosys.2023.110905
21. Lucarelli, G., & Borrotti, M. (2020). A deep Q-learning portfolio management framework for the cryptocurrency market. Neural Computing and Applications, 32, 17229–17244. https://doi.org/10.1007/s00521-020-05359-8