A Comprehensive Investigation of Reinforcement Learning-based Financial Quantitative Analysis: Taking Stock Trading and Risk Control as Examples

Yifei Zeng*

Mathematics and Financial Accounting, Queen Mary University of London, E1 4NS London, England

Abstract. Reinforcement learning (RL) has emerged as a transformative approach for addressing complex decision-making challenges, particularly in the financial sector, where its application has garnered substantial interest. This paper offers a comprehensive review of the foundational concepts and classical methods of RL, while providing an in-depth exploration of its advanced applications in crucial domains: stock prediction & quantitative trading, and risk estimation. By analyzing recent advancements over the past decade, the study underscores the expanding role of RL in optimizing financial strategies, improving decision-making processes, and driving innovation in quantitative finance. In addition to reviewing key developments, the paper discusses persistent challenges related to risk management, such as the trade-off between risk and reward, data scarcity, and the need for algorithms to adapt to dynamic and volatile market conditions. Through these insights, this research aims to provide a roadmap for future studies, addressing the limitations. Meanwhile, this study is contributed to guiding the continued evolution of RL applications in finance, ensuring they remain robust, adaptive, and effective in a rapidly changing economic landscape.

1 Introduction

Predicting financial markets remains a highly challenging task due to their inherent uncertainty and complexity [1]. Market behavior is shaped not only by the dynamics of supply and demand but also by external influences, such as global economic conditions, policy shifts, news events, and even social media commentary [1]. The non-linearity, noise, and sudden, unpredictable changes substantially increase market risks.

In this context, quantitative techniques have emerged as a promising solution by employing mathematical models and algorithms to systematically analyze financial data. These approaches aim to extract valuable insights from historical data, allowing for market trend forecasting, investment optimization, and risk reduction [2].

* Corresponding author: Ah22434@qmul.ac.uk

Traditional models, such as linear regression, have been widely used to capture historical price trends based on relatively stable and linear market changes. However, these models struggle wit the complexities and non-linearities of real-world market behavior. Time series methods, like Auto Regression Integrated Moving Average (ARIMA) [3], address some of these limitations by handling time-dependent data, but they still face challenges regarding to sudden events or capturing long-term dependencies effectively. Over the past years, deep learning (DL) techniques have dominated fnancial analysis, with models such as Convolutional Neural Networks (CNNs) [4,5], Recurrent Neural Networks (RNNs) [6], and Transformers [7,8] showing signifcant advantages. These models are particularly well-suited for fnancial data analysis due to their ability to extract complex features and recognize intricate patterns in large datasets.

A unique branch of machine learning, Reinforcement Learning (RL), has shown its potential in fnancial markets. Unlike traditional methods, RL models are designed to self-learn by optimizing decisions within dynamic environments, continuously adjusting strategies to maximize rewards. This makes RL particularly suitable for developing adaptive trading strategies in fast-changing market conditions. Furthermore, the integration of Deep Reinforcement Learning (DRL) has further advanced algorithmic trading by combining RL's optimization with DL's powerful feature extraction capabilities. This combination has paved the way for the development of fully automated trading systems.

Despite the promising capabilities of RL in fnancial applications [9], several challenges remain. Financial markets are characterized by signifcant random fluctuations in price sequences, which can disrupt model training. Moreover, the rapidly evolving and complex nature of markets often requires RL algorithms to handle high-dimensional state and action spaces, leading to huge model complexity and computational demands [10]. High-frequency trading, for example, demamds vast amounts of real-time data, placing considerable pressure on computational resources. Additionally, factors such as market liquidity and trading frequency can influence RL performance, with constraints potentially leading to low accuracy in decision [11].

This study focuses on the application of RL in fnancial quantitative analysis, with particular emphasis on stock prediction and quantitative trading, and risk estimation. This study aim to provide an overview of the development of RL in this domain, summarize the current research landscape, and outline the key challenges that remain.

2 Reinforcement learning

RL has emerged as a unique machine learning technique, demonstrating remarkable advancements in various fields in recent years. Its application in finance shows significant potential. By simulating the interactions between an agent and the environment, RL aims to learn an optimal policy that maximizes long-term rewards.

This section provides a brief introduction to the fundamental principles of RL. This study will begin by exploring essential concepts, including the agent, environment, state, and reward. Subsequently, following sections will classify and discuss various RL methods, distinguishing between On-Policy and Off-Policy approaches, as well as model-based and modelfree techniques.

2.1 Basic concepts in reinforcement learning

RL centers around achieving specific goals through interactive learning. In this framework, the learner, commonly referred to as the agent, interacts with everything external to it, collectively known as the environment. This interaction unfolds through a series of actions and reactions: the agent selects actions, while the environment responds by presenting new

states and providing a corresponding reward. The agent's primary objective is to maximize this reward over time. An illustration of a typical reinforcement learning schema is shown in Fig. 1.

More specifically, the agent engages with the environment at discrete time steps, denoted as $t = 0, 1, 2, ...$ At each time step t, the agent observes the current state of the environment, represented as $s_t \in S$, and selects an action $a_t \in A(s_t)$ from the available actions in that state. After executing the action, the agent receives a reward $r_{t+1} \in R$ and transitions to a new state.

At each time step, the agent determines the probability of selecting each action based on the current state, which is characterized by a policy denoted as $\pi_t(s, a)$. Over time, RL algorithms provide feedbacks for the agent to update this policy based on accumulated experience, with the ultimate goal of maximizing total long-term rewards. The boundary between the agent and the environment is not static. In contrast, it can be adjusted depending on the specific task. In more intricate scenarios, such as robotic control, multiple agents may interact, each with distinct boundaries.

In real-world applications, once states, actions, and rewards are defined, the process and the boundaries become clear. The RL framework offers an abstract representation for goal-oriented learning problems, positing that any such problem can be modeled using three types of signals exchanged between the agent and the environment: the action taken by the agent (a), the current state of the environment (s), and the reward received by the agent (r). Although this framework may not cover every decision-learning problem, its broad applicability has been demonstrated across various fields.

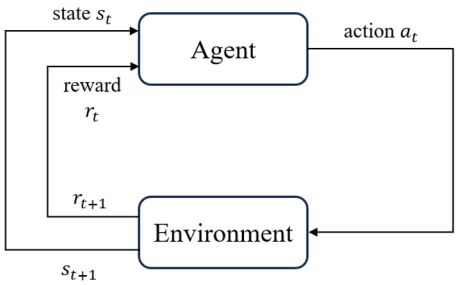

Fig. 1. The agent-environment interaction in reinforcement learning (Photo/Picture credit: Original).

2.2 Markov decision process and bellman equation

The Markov Decision Process (MDP) is a mathematical framework for modeling the interaction between an agent and its environment, commonly used for sequential optimization problems. An MDP consists of four key éléments: state space S, action space A, state transition function P, and reward function R :

$$(S, A, P, R) \tag{1}$$

where : 1) S denotes the state space, encompassing all possible states. Each individual state $s \in S$ can be either a discrete or continuous variable. 2) A is the action space, which includes all possible actions for the agent to take in each state s. Similarly, an action $a \in A$ can also be discrete or continuous. 3) P represents the state transition function, describing the probability of transitioning to the next state s' given the current state s and action a. This is defined as :

$$P_a(s, s') = \Pr(s_{t+1} = s' | s_t = s, a_t = a) \tag{2}$$

R refers to the reward function, which quantifies the reward the agent receives when transitioning from state s to s' as a result of taking action a:

$$R_a(s, s') = \mathbb{E}[r_{t+1}|s_t = s, a_t = a, s_{t+1} = s'] \tag{3}$$

The goal of an MDP is to maximize expected discounted cumulative rewards:

$$\mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k R_{a_{t+k}}(s_{t+k}, s_{t+k+1}|s_t = s, a_t = a\right] \tag{4}$$

2.3 Categories of reinforcement learning

Table 1. Classical reinforcement learning methods and their corresponding category in terms of different rules.

Algorithm	Year	Model	Policy	Action	State	Operator
DQN [12]	2013	Model-Free	Off-Policy	Discrete	Discrete or Continuous	Q-value
DDPG [13]	2015	Model-Free	Off-policy	Continuous	Discrete or Continuous	Q-value
A3C [14]	2016	Model-Free	On-Policy	Continuous	Discrete or Continuous	Advantage
TRPO [15]	2017	Model-Free	On-Policy	Discrete or Continuous	Discrete or Continuous	Advantage
PPO [16]	2017	Model-Free	On-Policy	Discrete or Continuous	Discrete or Continuous	Advantage
TD3 [17]	2018	Model-Free	Off-policy	Continuous	Continuous	Q-value
MBPO [18]	2017	Model-Based	Off-policy	Continuous	Continuous	Q-value
MPC [19]	1997	Model-Based	On-Policy	Continuous	Continuous	Cost function

Based on different criteria, RL methods can be classified in to various subclasses, with the most common being On-Policy vs. Off-Policy, and model-based vs. model-free. Table 1 presents a classification of some representative methods according to these two criteria.

2.3.1 On-policy and Off-policy

In On-policy methods, the agent selects actions based on the current observed state using a specific policy, and the outcomes of these actions are gathered to iteratively improve the same policy's parameters. This approach involves both the behavior policy and the target policy to explore the state and action spaces, optimizing the learning objective using the collected data. Onpolicy algorithms generally incorporate some degree of action randomness to balance exploration and exploitation. This allows the agent to explore uncertain actions when needed, while still choosing those that maximize expected rewards in other situations.

Off-policy methods use two distinct policies: the behavior policy, responsible for generating actions, and the target policy, which is trained based on the resulting data. This separation allows for decoupling data collection from policy training, enabling the agent to learn an optimal target policy without relying on the behavior policy's exploration.

The advantage of On-policy methods is their ability to directly improve the policy being followed, which can simplify learning in certain environments. However, because exploration and exploitation are linked, on-policy methods may struggle to fully explore the state-action space, especially in environments where randomness alone cannot ensure adequate exploration.

In contrast, Off-policy approaches like Q-learning offer more flexibility by separating exploration from the optimization process. This allows for more efficient learning and exploration. The downside is that they can be more complex to implement and may require more data to converge, especially when learning from large or complex environments.

2.3.2 Model-based and model-free

Model-based methods construct an internal model of the environment's dynamics through interactions, estimating transition probabilities and reward functions. This allows the agent to predict outcomes and actions to maximize future rewards. A prime example of this approach is dynamic programming. However, accurately modeling the environment remains a significant challenge, particularly in uncertain domains like financial markets.

Consequently, inaccurate models can lead to suboptimal decisions, making the effectiveness of model-based methods heavily reliant on model precision and available computational resources. Conversely, model-free methods operate without relying on the environment's transition probability distribution or reward function. This means constructing a model of the Markov Decision Process (MDP) is unnecessary. Instead, these methods enhance policy gradually through direct interactions with the environment, leveraging feedback from trial-and-error learning. Common model-free algorithms include Monte Carlo (MC) estimation and Temporal Difference (TD) methods. The Monte Carlo method evaluates the policy by averaging cumulative rewards after episode completion, while TD methods perform incremental updates using the estimated values of the current state, without waiting for the episode to end.

Notably, recent advancements have merged deep learning with model-free algorithms, resulting in techniques such as Deep Q-Networks (DQN) [12], Asynchronous Advantage Actor-Critic (A3C) [14], Proximal Policy Optimization (PPO) [16], and Deep Deterministic Policy Gradient (DDPG) [13].

The primary advantage of model-free approaches lies in their independence, which renders them particularly suitable for complex tasks like financial quantitative problems. Nevertheless, they exhibit low sample efficiency, often necessitating substantial amounts of data for effective training. Moreover, they may display instability during learning, especially in scenarios involving high-variance policy gradients. In comparison to model-based methods, model-free approaches generally demonstrate lower exploration efficiency due to their lack of predictive capabilities inherent in an internal model of the environment.

2.4 Trend

Fig. 2. Performance comparison of different RL methods (Photo/Picture credit: Original).

Combining Table 1 and Fig. 2, it becomes evident that the rise of deep learning has led to an increasing number of recent studies favoring model-free methods. This trend can be attributed to several factors. On one hand, a growing array of practical applications

necessitates modeling through reinforcement learning (RL), yet these scenarios often pose challenges in accurately capturing the environment's dynamics. On the other hand, model-free methods provide greater flexibility by enabling the direct learning of optimal policies through interactions with the environment, without requiring a predefined explicit model. This flexibility proves especially vital in complex and dynamic environments such as financial markets, robotic control, and intricate games, where the characteristics are difficult to fully capture or predict.

Moreover, as computational resources and data have increased, the costs associated with model-free methods have become relatively manageable. Particularly when integrated with deep neural networks, these methods excel in high-dimensional state spaces. Algorithms such as DQN, PPO, and SAC demonstrate the capability to adapt to various complex environments and tasks through end-to-end training, thereby further promoting the widespread application of model-free approaches.

3 Reinforcement learning-based quantitative methods in financial market

In recent years, RL has attracted significant attention in financial. The daily-increasing complexity in financial markets, along with the large amounts of data, together lead to an increasing need for RL-based quantitative methods.

This section will focus on two key application areas of RL in financial markets: stock prediction and quantitative trading, as well as risk estimation and control. These tasks are significant for both institutional and individual investors.

3.1 Stock prediction and quantitative trading

There has been widespread attention on the application of DRL in stock prediction and quantitative trading. Existing literature indicates that DRL demonstrates significant potential in automating trading.

For instance, Li [20] provided a systematic review of DRL in financial applications, validating its effectiveness in the stock market. Through comprehensive analysis, the study compared performance differences among several classic DRL models regarding stock returns. It is reported that DRL models can optimize strategies through continuous interactions with the market environment. Based on this, Li [21] proposed a novel DRL model aimed at generating trading signals and maximizing returns. This research highlighted the superior performance of DRL models in capturing complex signals compared to traditional methods, confirming their effectiveness in intelligent mechanisms.

Additionally, Li [22] introduced a Cooperative Multi-Agent Deep Reinforcement Learning model (CMPS) for stock portfolio management. A three-agent system was developed, leveraging DQN and self-attention mechanisms to capture market features. Furthermore, Li also proposed the CMPS-Risk Free model, which utilizes a risk-free asset strategy to mitigate market risks, significantly improving the strategy's stability and yield.

Beyond these independent RL-based models, increasing research explores the integration of RL with other deep learning methods. For example, Fu [23] proposed a hybrid model combining Long Short-Term Memory (LSTM) networks with the Deep Deterministic Policy Gradient (DDPG). By incorporating RL in the early stages of supervised learning, the model accelerated convergence and improved trading strategy accuracy. Similarly, Shin [24] developed a deep multimodal RL system that integrates CNN and LSTM, effectively extracting multidimensional market features through various chart inputs and time series data.

The results demonstrated stability under different market conditions, showcasing its adaptability and potential returns in financial markets.

Recently, Transformer architectures have garnered significant interest. Gao [10] proposed StockFormer, utilizing Transformers to capture global market dynamics while integrating strategy optimization for trading decisions. This model enables the joint training of reinforcement learning agents, allowing for investment decisions across state space.

Altuner [25] introduced Graph Neural Networks (GNN) to analyze relationships among users, illustrating the potential of using social knowledge graphs and sentiment analysis for stock market prediction. By building social network relationships and integrating deep Q-learning networks, he successfully improved the accuracy of stock price predictions. This GNN-based strategy expands the research direction of intelligent quantitative trading by capturing complex relationships among individuals in the stock market.

A search on DBLP was conducted using the keywords *stock*, *quantitative*, and *RL* for research published after 2019, categorizing the findings into five groups: RL Solely, DRL, RL+RNN, RL+CNN, and RL+Transformer. It is found that an increasing number of studies are beginning to combine reinforcement learning with deep learning methods. This trend can be summarized to several reasons.

Firstly, it has always been challenging for solely RL methods to capture the latent relationships among various market factors due to their complexity. LSTM excels at processing time series data, enabling models to better capture the long-term dependencies of historical trends. Meanwhile, GNN effectively models the relational networks among individuals in the stock market, thereby enhancing the model's ability to capture structural information. Transformers, with their self-attention mechanisms, improve the model's understanding of both long-term and shortterm market dynamics, further optimizing trading strategies. Each of these techniques finds its unique position in financial quantitative analysis.

Moreover, intelligent trading systems must possess greater adaptive capabilities due to the high risks and volatility in financial markets. RL models integrating deep learning methods can make more precise decisions in the face of unknown market conditions compared to those based solely on RL strategies. This trend illustrates the collaborative evolution of deep learning and reinforcement learning in financial markets.

Lastly, the rapid advancement of computational resources over the past decade has propelled the swift development of the AI framework and industrial ecosystem. This has given rise to fast model-building frameworks like PyTorch and TensorFlow, excellent deployment protocols such as ONNX and TensorRT that ensure compatibility across different platforms, and integrated frameworks like OpenMMLab that encompass everything from training to deployment. Thanks to advancements in computational capabilities, complex deep learning methods are no longer difficult to implement, allowing RL to better utilize multidimensional data for more efficient strategy optimization. The combination of these new technologies not only improves model performance but also enhances flexibility and robustness against varying market conditions.

3.2 Risk estimation and control

Traditional risk modeling primarily relies on conventional RL methods, as proposed by researchers like Mihatsch [26]. These methods aim to assist agents in effectively handling risk in sequential decision-making problems. Mihatsch [26] introduced a series of risk-sensitive RL algorithms that capture latent risk characteristics in human behavior by applying functions to the temporal difference error. These algorithms reveal differing risk preferences regarding gains and losses, providing a theoretical foundation for subsequent research. Additionally, He further developed a risk-sensitive Q-learning algorithm, validating its

convergence in the presence of unknown transition probabilities, thereby offering a more stable theoretical support for risk management.

Building on this, Jaimungal [27] explored more robust risk-related RL methods, proposing a ranking-dependent expected utility (RDEU) approach to optimize the value of different strategies. While these traditional methods have provided important insights for understanding and managing risks in financial markets, they face challenges due to the increasing complexity of the financial environment.

In recent years, researchers have increasingly tended to integrate DL methods into RL to enhance risk modeling capabilities [28-30]. For instance, Shin [28] introduced a DRL-based autonomous trading agent that adjusts its greediness parameter to favor low-risk actions. In experiments conducted in the cryptocurrency market, this model achieved an impressive return of 1800% while maintaining low risk.

Furthermore, additional studies have combined Generative Adversarial Networks (GAN) with RL to tackle issues of missing and imbalanced financial data [29], proposing an innovative RL framework. This approach optimizes performance through customizable reward functions tailored to the needs of various financial institutions, effectively mitigating risks associated with misclassification. This flexibility allows financial decision-making to better adapt to real-world challenges.

Meanwhile, Coache [30] introduced a dynamic spectral risk measurement framework based on deep neural networks to enhance the precision and reliability of risk management. This framework not only expands the theoretical basis for risk measurement but also supports risk decision-making in practical operations. Overall, the combination of traditional RL methods with modern DL techniques is gradually transforming the landscape of financial risk modeling, providing new solutions for addressing decision-making challenges in complex financial environments.

Table 2. Summary of Relevant Literature.

Method	Year	Schema	Aim & Solved Problems
[27]	2002	RL	TD error; sequential decision-making in uncertainty.
[28]	2019	DRL	Target policy; low-risk actions with profit maximization.
[29]	2021	DRL; GAN	GANs; risk reduction; missing and imbalanced data;
[30]	2023	DRL	Dynamic risk measures; deep neural networks; effective in arbitrage.
[26]	2024	RL; XGBoost	Q-table optimization; XGBoost for financial risk.

Table 2 summarizes relevant literature on risk quantification in the financial domain. Over the past decade, there has been a notable shift towards integrating RL modeling of risk with deep learning techniques. This trend reflects a growing recognition among researchers of deep learning's advantages in feature extraction and complex pattern recognition, which, when combined with RL, can significantly enhance model performance in dynamic financial environments.

In the future, continuous advancements in computational power and algorithm optimization are expected to enable risk-sensitive RL models to adapt more effectively to volatile market conditions. This evolution will facilitate greater intelligence and automation in financial decision-making processes, making them more efficient and accurate. Ultimately, this transition will not only improve the efficiency of risk management but also provide

stronger decision support for financial institutions navigating complex market dynamics. Such advancements are crucial for reducing potential economic losses and minimizing risk exposure, thereby promoting overall stability in the financial sector.

4 Challenges

This section furhter analyzes several challenges faced by current RL-based quantitative methods in the finance area. These challenges will be discussed at three levels: Market-level, Model-level, and External-level.

4.1 Market-level challenges

Market-level challenges primarily rise from the complexity and uncertainty of real-world markets. The design of RL strategies must adapt to different market conditions and risk preferences to maintain effectiveness in constantly changing environments.

For instance, in market-making, inventory risk and execution risk are key challenges driven by dynamic market changes [31]. To address these challenges, RL strategies are required to continuously adjust. Similarly, in portfolio allocation and optimization, RL algorithms shall respond flexibly to fluctuations by dynamically adjusting asset allocation to balance risk and reward. Moreover, the rapid reactions of other participants to market dynamics often result in price fluctuations [32]. To avoid negative impacts from market volatility during trade execution, RL algorithms must strike a balance between executing trades quickly and securing optimal prices.

Another critical consideration is the reward function, a core component in RL. It directly impacts the convergence speed and ultimate performance of the strategy. To ensure the maximization of long-term returns, the reward function are supposed to follow changes in the market. However, in complex market conditions, designing such a function is highly challenging. This information asymmetry increases the difficulty of optimizing RL strategies, particularly in large action spaces where agents must identify the optimal policy from numerous choices.

4.2 Model-level challenges

Model-level challenges are those within the inherent limitations of RL algorithms and the design and optimization of related models.

One of the biggest challenges lies in the limitations of RL methods within complex financial environments. RL algorithms often struggle to efficiently explore the policy space when dealing with high-dimensional state spaces [33], leading to sub-optimal strategy optimization. This is particularly evident in dynamic and uncertain market conditions. Although model-based RL methods show potential, many researchers have turned to model-free approaches due to the complexities in modeling the market environment. However, this shift further constrains the learning efficiency of the algorithms in complex environments.

Secondly, data scarcity directly impacts the effectiveness of the models. In financial markets, historical data is limited and highly noisy, especially under extreme market conditions [34]. The lack of sufficient and reliable data during such periods prevents models from undergoing effective training.

Furthermore, the precision of market simulation is another significant challenge. High-fidelity market simulation is critical for the success of DRL, but existing market simulation tools fail to adequately capture real market factors such as liquidity and market impact. This limitation hampers the training effectiveness of RL algorithms in complex market

environments. Future research are supposed to focus on developing more accurate and high-fidelity market simulation tools to support RL models in training within more realistic market conditions.

4.3 External-level challenges

Finally, uncontrollable external factors have a significant impact on financial markets. Key external challenges include political policies [35], military conflicts [36], and natural disasters [37].

Policy changes often have a direct impact on global finance, causing sudden and dramatic market fluctuations. Military conflicts can lead to abrupt shifts in market sentiment, resulting in global market instability. Natural disasters, such as earthquakes, floods, and hurricanes, can also lead to unforeseen market shocks, as they may disrupt supply chains and cause significant market turbulence. Particularly, shifts in the global macroeconomic environment, such as economic crises or pandemics, also have profound effects on financial markets. For instance, the COVID19 [38] triggered widespread panic in global markets, with government interventions and policy adjustments creating extreme market uncertainty.

It is unrealistic to expect RL models to predict future changes in external factor, given the extreme uncertainty and suddenness of these events. Such occurrences are beyond the predictive capabilities of any current algorithm. Therefore, rather than aiming to predict these factors, the focus here shifts to ensuring that RL models can respond immediately to changes in external conditions, adjusting their strategies to maximize returns. This requires models not only to have adaptability but also to incorporate risk-buffering strategies.

5 Conclusion

This study explores the extensive applications of RL methods in quantitative finance sector, with a particular focus on key areas such as stock prediction & quantitative trading, and risk estimation. Firstly, this paper reviewed the fundamental concepts and classical methods of RL, including basic terms and Markov Dicision Process. These foundational frameworks provide the theoretical support needed to understand how RL learns and makes decisions in dynamic and uncertain environments. In terms of specific applications, this paper delved into the implementation of RL in stock prediction and quantitative trading. By analyzing historical market data and extracting key features, RL models can identify potential market patterns and generate efficient trading signals. The successful implementation of these models not only enhances the accuracy of trading strategies but also increases the level of automation in trading processes. Furthermore, RL's application in portfolio optimization demonstrates its flexibility, allowing for dynamic adjustments in asset allocation based on real-time market changes to achieve the optimal balance between risk and return. In risk estimation, RL offers new perspectives for financial decision-making. Through quantitative analysis of potential risks, RL models can assist investors in making more rational decisions in the face of market volatility, thereby reducing possible losses. Particularly in the realm of complex financial risk management, RL effectively evaluates risk preferences, optimizes portfolios, and enhances the robustness of overall strategies.

However, despite the significant potential of RL algorithms, practical applications still encounter multiple challenges. To name a few, data scarcity limits the training and generalization of models, while the dynamic nature of markets requires algorithms to exhibit a high degree of adaptability. Additionally, designing a suitable reward function is crucial for guiding agent learning. These challenges are particularly pronounced in risk management, where finding the right balance between maximizing returns and minimizing risks remains an urgent issue to address. Future research should focus on the integration of RL techniques

with other machine learning methods to create more efficient decision support systems and promote continuous innovation in financial practices. Through in-depth research in these areas, this study aims to provide a solid theoretical foundation and practical tools for decision optimization and risk management in the financial sector.

References

1. S. K. Sahu, A. Mokhade, N. D. Bokde, An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: recent progress and challenges. Applied Sciences, 13(3), 1956 (2023).
2. T. Théate, D. Ernst, An application of deep reinforcement learning to algorithmic trading. Expert Systems with Applications, 173, 114632 (2021).
3. A. L. S. Maia, F. de A. T. de Carvalho, Holt's exponential smoothing and neural network models for forecasting interval-valued time series. International Journal of Forecasting, 27(3), 740-759 (2011).
4. E. Hoseinzade, S. Haratizadeh, Cnnpred: CNN-based stock market prediction using a diverse set of variables. Expert Systems with Applications, 129, 273-285 (2019).
5. M. Wen, P. Li, L. Zhang, Y. Chen, Stock market trend prediction using high-order information of time series. IEEE Access, 7, 28299-28308 (2019).
6. H. Li, Y. Shen, Y. Zhu, Stock price prediction using attention-based multi-input LSTM. In Asian Conference on Machine Learning, 454-469. PMLR, 2018.
7. N. Kitaev, L. Kaiser, A. Levskaya, Reformer: The efficient transformer. arXiv preprint arXiv:2001.04451 (2020).
8. H. Wu, J. Xu, J. Wang, M. Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems, 34, 22419-22430 (2021).
9. H. Moussaoui, M. Benslimane, et al., Reinforcement learning: A review. International Journal of Computing and Digital Systems, 13(1), 1-1 (2023).
10. S. Gao, Y. Wang, X. Yang, Stockformer: Learning hybrid trading machines with predictive coding. In IJCAI, 4766-4774 (2023).
11. S. Sun, M. Qin, X. Wang, B. An, Prudex-compass: Towards systematic evaluation of reinforcement learning in financial markets. arXiv preprint arXiv:2302.00586 (2023).
12. V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. A. Riedmiller, Playing atari with deep reinforcement learning. CoRR, abs/1312.5602 (2013).
13. T. P. Lillicrap, Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971 (2015).
14. V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, K. Kavukcuoglu, Asynchronous methods for deep reinforcement learning. CoRR, abs/1602.01783 (2016).
15. J. Schulman, S. Levine, P. Moritz, M. I. Jordan, P. Abbeel, Trust region policy optimization (2017).
16. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017).
17. S. Fujimoto, H. van Hoof, D. Meger, Addressing function approximation error in actor-critic methods (2018).

18. T. Yu, G. Thomas, L. Yu, S. Ermon, J. Y. Zou, S. Levine, C. Finn, T. Ma, Mopo: Model-based offline policy optimization. Advances in Neural Information Processing Systems, 33, 14129-14142 (2020).
19. S. J. Qin, T. A. Badgwell, An overview of industrial model predictive control technology. In AIChE Symposium Series, 93, 232-256. New York, NY: American Institute of Chemical Engineers, 1997.
20. Y. Li, P. Ni, V. Chang, An empirical research on the investment strategy of stock market based on deep reinforcement learning model. In Proceedings of the 4th International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS), 1, 52-58. SciTePress, 2019.
21. Y. Li, P. Ni, V. Chang, Application of deep reinforcement learning in stock trading strategies and stock forecasting. Computing, 102(6), 1305-1322 (2020).
22. H. Li, M. Hai, Deep reinforcement learning model for stock portfolio management based on data fusion. Neural Processing Letters, 56(2), 108 (2024).
23. K. Fu, Y. Yu, B. Li, Multi-feature supervised reinforcement learning for stock trading. IEEE Access, 2023.
24. H.-G. Shin, I. Ra, Y.-H. Choi, A deep multimodal reinforcement learning system combined with CNN and LSTM for stock trading. In 2019 International Conference on Information and Communication Technology Convergence (ICTC), 7-11. IEEE, 2019.
25. A. B. Altuner, Z. H. Kilimci, A novel deep reinforcement learning based stock price prediction using knowledge graph and community aware sentiments. Turkish Journal of Electrical Engineering and Computer Sciences, 30(4), 1506-1524 (2022).
26. O. Mihatsch, R. Neuneier, Risk-sensitive reinforcement learning. Machine Learning, 49, 267-290 (2002).
27. S. Jaimungal, S. M. Pesenti, Y. S. Wang, H. Tatsat, Robust risk-aware reinforcement learning. SIAM Journal on Financial Mathematics, 13(1), 213-226 (2022).
28. W. Shin, S.-J. Bu, S.-B. Cho, Automatic financial trading agent for low-risk portfolio management using deep reinforcement learning. arXiv preprint arXiv:1909.03278 (2019).
29. T. P. Le, C. Rho, Y. Min, S. Lee, D. Choi, A2GAN: A deep reinforcement-based learning algorithm for risk-aware in finance. IEEE Access, 9, 137165-137175 (2021).
30. A. Coache, S. Jaimungal, A. Cartea, Conditionally elicitable dynamic risk measures for deep reinforcement learning. SIAM Journal on Financial Mathematics, 14(4), 1249-1289 (2023).
31. P. Markou, D. Corsten, Financial and operational risk management: Inventory effects in the gold mining industry. Production and Operations Management, 30(12), 4635-4655 (2021).
32. C. Chiarella, R. Dieci, X.-Z. He, Heterogeneity, market mechanisms, and asset price dynamics. In Handbook of Financial Markets: Dynamics and Evolution, 277-344. Elsevier, 2009.
33. S. Zeng, M. Hong, A. Garcia, Structural estimation of Markov decision processes in high-dimensional state space with finite-time guarantees. Operations Research, 2024.
34. R. Yang, L. Yu, Y. Zhao, H. Yu, G. Xu, Y. Wu, Z. Liu, Big data analytics for financial market volatility forecast based on support vector machine. International Journal of Information Management, 50, 452-462 (2020).

35. S. Preuss, R. Königsgruber, How do corporate political connections influence financial reporting? a synthesis of the literature. Journal of Accounting and Public Policy, 40(1), 106802 (2021).
36. J. Kofroň, J. Stauber, The impact of the russo-ukrainian conflict on military expenditures of European states: security alliances or geography? Journal of Contemporary European Studies, 31(1), 151-168 (2023).
37. D. Zhang, S. Managi, Financial development, natural disasters, and economics of the Pacific small island states. Economic Analysis and Policy, 66, 168-181 (2020).
38. A. K. Khetan, S. Yusuf, P. Lopez-Jaramillo, A. Szuba, A. Orlandini, N. Mat-Nasir, A. Oguz, R. Gupta, A. Avezum, I. Rosnah, et al., Variations in the financial impact of the COVID-19 pandemic across 5 continents: a cross-sectional, individual level analysis. EClinicalMedicine, 44 (2022).