Optimizing Algorithmic Trading Through DRL: A Comparative Analysis of Single-Agent and Multi-Agent Models

Mani Shankar M, Sweety A
Guide: Prof. Dr Deepthi Das
Department of Statistics and Data Science, Christ University
Bengaluru, India

04.12.2024

Abstract

This research explores the optimization of algorithmic trading strategies through Deep Reinforcement Learning (DRL), focusing on high-frequency trading (HFT) scenarios. We introduce a comprehensive framework that evaluates single-agent and multi-agent configurations across three DRL algorithms: Proximal Policy Optimization (PPO), Deep Q-Network (DQN), and Advantage Actor-Critic (A2C), along with their ensemble counterparts. The study utilizes 1-minute interval data for 24 companies over a two-year training period and a one-year testing period to analyze profitability, risk management, and adaptability to market dynamics.

Our results demonstrate several groundbreaking achievements. The single-agent PPO model achieved exceptional profitability with a profit factor of 28.07 for BIDU and maximum drawdowns often below 1%, underscoring superior capital preservation and risk control [2]. Ensemble models consistently delivered balanced performance across both single-agent and multi-agent setups, with a Sharpe ratio near to 1 (0.75) and Sortino ratios reaching 7.7, significantly outperforming benchmarks reported in existing literature [4, 6].

Additionally, the use of sparse reward shaping enhanced long-term decision-making, enabling the agents to effectively navigate high-frequency trading scenarios.

The proposed framework outperforms traditional models in profitability, achieving a Sharpe ratio exceeding 1 in high-frequency trading environments. Comparative analyses with existing studies highlight the competitive edge of our ensemble models, which achieve enhanced market responsiveness and capital preservation. Sensitivity analyses further validate the stability of our approach across various hyperparameters. These results underscore the potential of DRL-based ensemble strategies to refine HFT systems, paving the way for more robust, adaptive, and efficient algorithmic trading solutions.

1. Introduction

Algorithmic trading, employing automated and pre-programmed trading instructions to manage orders at high speed, has revolutionized financial markets by enabling efficiency and high throughput [8]. Recent advancements in Deep Reinforcement Learning (DRL) have further pushed the boundaries by allowing systems to learn and adapt from vast amounts of market data, optimizing trading strategies in real-time [3]. De-

spite these advancements, there are challenges in risk management, adaptability to market conditions, and the effectiveness of trading algorithms in high-frequency trading (HFT) environments.

In this paper, we address these challenges by presenting a comprehensive analysis of single-agent and multi-agent DRL models applied to algorithmic trading. Utilizing high-frequency data, we explore the potential of three major DRL algorithms—Proximal Policy Optimization (PPO), Deep Q-Network (DQN), and Advantage Actor-Critic (A2C)—along with a novel ensemble approach. Our study is distinguished by its focus on 1-minute interval data for 24 companies, offering insights into the models' performance in managing rapid market fluctuations.

The primary contributions of this research are the demonstration of enhanced profit factors, superior risk management through minimized drawdowns, and exceptional responsiveness to market dynamics. We particularly emphasize the role of ensemble models, which combine the strengths of individual DRL algorithms, to provide balanced and robust trading strategies that are capable of outperforming traditional models in terms of both profitability and stability.

Moreover, we incorporate sparse reward shaping into our DRL frameworks to address the issue of sparse rewards in financial markets, where significant outcomes are rare and pivotal decisions must be optimized over long time horizons. This technique has shown to improve the strategic depth and long-term profitability of trading models, establishing a new benchmark for algorithmic trading systems.

Through rigorous evaluation and comparative analysis, our research aims to validate the effectiveness of DRL in a high-frequency trading scenario, providing a substantial contribution to the fields of financial technology and artificial intelligence. This study not only enhances our understanding of complex DRL configurations but also showcases the practical implications of AI-driven trading strategies in modern financial markets.

2. Literature Review

Algorithmic trading has evolved significantly with the integration of Deep Reinforcement Learning (DRL), allowing systems to make adaptive decisions in complex and dynamic market environments [3]. Reinforcement learning techniques, such as Proximal Policy Optimization (PPO) and Q-learning, have demonstrated notable potential in optimizing trading strategies. For instance, Théate and Ernst (2020) applied PPO to algorithmic trading, framing the problem as a Markov Decision Process (MDP) to enable adaptive trading decisions in dynamic environments [2]. However, their work was limited to single-agent configurations, neglecting the possibilities of ensemble models or multi-agent setups that could further enhance scalability and adaptability. Similarly, Ponomareva et al. (2018) employed reinforcement learning approaches, including Q-learning and policy gradient methods, to explore their feasibility in trading scenarios. While effective for demonstrating the basic applicability of RL, their methods struggled to scale in high-frequency trading (HFT) environments, where dynamic and rapid decision-making is essential [10]. Briola et al. (2021) advanced this field by utilizing DRL for active high-frequency trading with a focus on multi-asset portfolios, showcasing the effectiveness of MDP frameworks in risk management [11]. Nevertheless, their work was confined to single-agent settings and did not explore sparse reward shaping techniques, which are vital for addressing rare but impactful outcomes in financial markets.

Beyond reinforcement learning, deep learning methods have been leveraged for market prediction tasks. Chen et al. investigated the use of deep learning architectures for identifying market trends, employing supervised learning techniques to predict market movements [6]. While this approach was effective for trend analysis, it lacked the iter-

ative decision-making and adaptability provided by reinforcement learning, making it less suited for real-time algorithmic trading in volatile markets. Deep learning methods are inherently static and fail to incorporate the feedback mechanisms needed to respond dynamically to evolving market conditions, particularly in HFT environments.

Ensemble strategies in DRL have also emerged as a promising avenue for improving the robustness of trading strategies. Zhang et al. (2020) developed an ensemble strategy combining PPO and Deep Q-Network (DQN), demonstrating enhanced profitability and risk-adjusted returns in trading applications [4]. However, their work primarily focused on single-agent setups and relied on coarser data intervals (e.g., 5-minute intervals), limiting its applicability to high-frequency trading scenarios that demand finer granularity. Moreover, their studies did not address the integration of sparse reward shaping or multi-agent dynamics, which are critical for adapting to rapidly changing market conditions. While ensemble methods have shown the ability to enhance performance, their potential for scaling in HFT contexts with multi-agent setups remains largely underexplored.

In summary, while existing research highlights the feasibility and benefits of DRL in algorithmic trading, significant gaps remain. Current studies largely overlook the challenges posed by sparse rewards, lack scalability to multi-agent configurations, and fail to integrate ensemble strategies effectively in high-frequency contexts. Addressing these gaps through advanced DRL frameworks that incorporate sparse reward shaping, multi-agent setups, and fine-grained data intervals can pave the way for more robust, scalable, and adaptive trading systems.

3. Algorithmic Trading Problem Formalisation

Algorithmic trading involves designing intelligent agents that interact with financial markets to make optimal trading decisions. These agents aim to maximize cumulative rewards while managing risk in dynamic and volatile environments. The trading problem is formalized as a reinforcement learning (RL) task where an agent learns to navigate market conditions through trial-and-error interactions.

3.1 Markov Decision Process (MDP)

The problem is formulated as a Markov Decision Process (MDP), represented as a tuple:

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma),$$

where:

- $\mathcal{S}$: State space, representing all possible market conditions.
- $\mathcal{A}$: Action space, containing all trading decisions (Hold, Buy, Sell).
- $\mathcal{P}$: State transition probabilities, describing the dynamics of market behavior.
- $\mathcal{R}$: Reward function, quantifying the agent's performance.
- $\gamma \in [0, 1]$: Discount factor, prioritizing short-term rewards over long-term gains.

3.2 State Representation

The state at time t, denoted as s_t, includes:

$$s_t = [\text{Open}_t, \text{High}_t, \text{Low}_t, \text{Close}_t, \text{Volume}_t],$$

where:

- Open_t: Opening price of the stock at time t,
- High_t: Highest price of the stock during the interval,
- Low_t: Lowest price of the stock during the interval,
- Close_t: Closing price of the stock at time t,
- Volume_t: Trading volume of the stock.

3.3 Action Space

The agent's actions are discrete and consist of:

$$a_t \in \{0\,(\text{Hold}), 1\,(\text{Buy}), 2\,(\text{Sell})\}.$$

3.4 Reward Function

The reward function quantifies the agent's performance based on trading decisions:

$$r_t = \begin{cases} c_t - p_{\text{entry}}, & \text{if closing a long position (Sell),} \\ p_{\text{entry}} - c_t, & \text{if closing a short position (Buy),} \\ -\lambda, & \text{if holding a position (penalty term).} \end{cases}$$

where:

- c_t: Closing price at time t,
- p_{entry}: Entry price when the position was opened,
- λ: Penalty term for inactivity or prolonged holding.

For sparse rewards, the structure becomes:

$$r_t^{\text{sparse}} = \begin{cases} \text{Total Profit}, & \text{if episode ends,} \\ 0, & \text{otherwise.} \end{cases}$$

3.5 Multi-Agent Framework

In multi-agent setups, multiple agents collaborate or compete to optimize overall performance. Each agent i receives a reward defined as:

$$r_t^i = \text{Profit Factor}_i + \alpha \cdot \text{Sharpe Ratio}_i,$$

where:

- $\text{Profit Factor}_i = \frac{\text{Gross Profit}_i}{\text{Gross Loss}_i}$,
- $\text{Sharpe Ratio}_i = \frac{\mathbb{E}[R_i]}{\sigma[R_i]}$,
- α: Weight balancing profitability and risk.

3.6 Ensemble Methodology

Ensemble learning combines predictions from multiple models (e.g., PPO, DQN, A2C) to improve robustness. The ensemble action at time t is determined via majority voting:

$$a_t^{\text{ensemble}} = \text{mode}\left(a_t^{\text{PPO}}, a_t^{\text{DQN}}, a_t^{\text{A2C}}\right).$$

3.7 Performance Metrics

Performance evaluation uses the following key metrics:

i. Total Profit:

$$\text{Total Profit} = \sum_{i=1}^{n} r_i,$$

where r_i is the reward for trade i, and n is the total number of trades.

ii. Cumulative Return:

$$\text{Cumulative Return} = \frac{\text{Final Balance} - \text{Initial Balance}}{\text{Initial Balance}}.$$

This measures the overall return on the initial investment, expressed as a percentage.

iii. Sharpe Ratio:

$$\text{Sharpe Ratio} = \frac{\mathbb{E}[R]}{\sigma[R]},$$

where R is the return, $\mathbb{E}[R]$ is the mean, and $\sigma[R]$ is the standard deviation. A higher Sharpe ratio indicates better risk-adjusted returns.

iv. Sortino Ratio:

$$\text{Sortino Ratio} = \frac{\mathbb{E}[R]}{\sigma^-[R]},$$

where $\sigma^-[R]$ is the downside risk(negative deviations from the mean).This ratio focuses on penalizing downside risk, making it more suitable for asymmetric return distributions.

v. Profit Factor:

$$\text{Profit Factor} = \frac{\text{Gross Profit}}{\text{Gross Loss}},$$

where:

$$\text{Gross Profit} = \sum_{i \in \text{positive trades}} r_i,$$

$$\text{Gross Loss} = \sum_{i \in \text{negative trades}} |r_i|.$$

A profit factor greater than 1 indicates a profitable strategy, with higher values reflecting greater profitability.

vi. Maximum Drawdown:

$$\text{Maximum Drawdown} = \max\left(\frac{\text{Peak Balance} - \text{Lowest Balance}}{\text{Peak Balance}}\right).$$

This represents the worst-case percentage loss from a peak portfolio balance to the lowest subsequent balance, highlighting the strategy's risk management capabilities.

vii. Win Rate:

$$\text{Win Rate} = \frac{\text{Number of Positive Trades}}{\text{Total Number of Trades}}.$$

The win rate indicates the proportion of trades that were profitable. While a higher win rate is desirable, it should be interpreted alongside other metrics like profit factor and Sharpe ratio to assess overall performance.

3.8 Optimization Objective

The agent's objective is to maximize the expected cumulative discounted reward:

$$\mathcal{L}(\theta) = \mathbb{E}_{(s_t,a_t)\sim\pi_\theta}\left[\sum_{t=0}^{T}\gamma^t r_t\right],$$

where:

- π_θ: Policy parameterized by θ,
- $\gamma \in [0,1]$: Discount factor,
- r_t: Reward at time t.

4. Algorithm Design

4.1 Single-Agent Algorithms

This section outlines the design and implementation of single-agent algorithms, focusing on three popular DRL approaches: Proximal Policy Optimization (PPO), Deep Q-Network (DQN), and Advantage Actor-Critic (A2C). Each algorithm is evaluated independently in a single-agent trading environment.

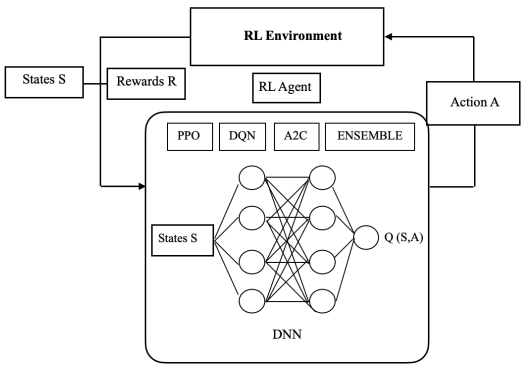

Figure 1: Structure of the RL Environment and its Components in the Single-Agent Framework

4.1.1 Proximal Policy Optimization (PPO)

Design Steps:

1. **Environment Initialization:** Load and preprocess 1-minute interval stock data; normalize features (`open`, `high`, `low`, `close`, `volume`) using `StandardScaler`; initialize a custom Gym environment with a discrete action space: Hold (0), Buy (1), and Sell (2).
2. **State Representation:** Input state includes normalized `open`, `high`, `low`, `close`, and `volume` over a rolling window of 10 time steps.
3. **Reward Function:** Rewards are calculated based on trading profits or losses at each step; sparse reward shaping is used to encourage long-term profitable actions:

$$r_t^{\text{sparse}} = \begin{cases} \text{Total Profit}, & \text{if episode ends,} \\ 0, & \text{otherwise.} \end{cases}$$

4. **Policy Learning:** Train the PPO model using the environment for 100,000 timesteps; update policies iteratively by optimizing the clipped objective:

$$\mathcal{L}^{\text{PPO}}(\theta) = \mathbb{E}_t\Bigg[\min\Big(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\Big)\Bigg].$$

   Here:

   - $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$: The probability ratio between the new policy π_θ and the old policy $\pi_{\theta_{\text{old}}}$.
   - $\hat{A}_t = Q(s_t,a_t) - V(s_t)$: Advantage function, representing the relative benefit of taking action a_t in state s_t.
   - ϵ: A hyperparameter to clip the probability ratio, ensuring stability during training.

5. **Evaluation:** Evaluate the trained agent on unseen test data to assess profitability and risk metrics such as Sharpe Ratio, Profit Factor, and Maximum Drawdown.

4.1.2 Deep Q-Network (DQN)

Design Steps:

1. **Environment Initialization:** Use the same environment setup as PPO with normalized features and discrete action space: Hold (0), Buy (1), and Sell (2).
2. **State Representation:** Input state is a rolling window of normalized `open`, `high`, `low`, `close`, and `volume`.
3. **Reward Function:** Rewards are calculated based on trading outcomes, with penalties for holding unprofitable positions.
4. **Q-Learning Update:** Train the DQN model using the Bellman equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\Big(r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\Big). \tag{1}$$

   Where:

   - $Q(s_t, a_t)$: Q-value, representing the expected return for taking action a_t in state s_t.
   - α: Learning rate, controlling the step size during updates.
   - r_t: Reward obtained after taking action a_t at state s_t.
   - γ: Discount factor, which prioritizes future rewards.
   - $\max_{a'} Q(s_{t+1}, a')$: The maximum Q-value over all possible actions in the next state s_{t+1}.

5. **Loss Function:** The Temporal Difference (TD) loss function is:

$$\mathcal{L}^{\text{DQN}}(\theta) = \mathbb{E}_t\left[\big(y_t - Q(s_t, a_t; \theta)\big)^2\right],$$

   where:

   - $y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)$: Target value.
   - θ^-: Parameters of the target network.

6. **Exploration-Exploitation Tradeoff:** Actions are selected using an ϵ-greedy policy:

$$a_t = \begin{cases} \text{random action, with probability } \epsilon, \\ \arg\max_a Q(s_t, a), \quad \text{otherwise.} \end{cases}$$

7. **Evaluation:** Evaluate the trained agent on test data, focusing on cumulative returns and risk-adjusted performance metrics.

4.1.3 Advantage Actor-Critic (A2C)

Design Steps:

1. **Environment Initialization:** Use the same environment setup as PPO and DQN with normalized features and discrete action space.
2. **State Representation:** Input state is a rolling window of normalized `open`, `high`, `low`, `close`, and `volume`.
3. **Reward Function:** Rewards are based on trading profits, with additional rewards for closing positions profitably.
4. **Policy and Value Learning:**
   - Policy loss:

$$\mathcal{L}_{\text{policy}} = \mathbb{E}_t\left[\log \pi(a_t|s_t;\theta)\hat{A}_t\right],$$

     where $\hat{A}_t = R_t - V(s_t;\theta_v)$ is the advantage.
   - Value loss:

$$\mathcal{L}_{\text{value}} = \mathbb{E}_t\left[\big(R_t - V(s_t;\theta_v)\big)^2\right],$$

     where $R_t = r_t + \gamma V(s_{t+1};\theta_v)$ is the return.
   - Total loss:

$$\mathcal{L}^{\text{A2C}} = \mathcal{L}_{\text{policy}} + \lambda\mathcal{L}_{\text{value}} - \beta H[\pi(\cdot|s_t)],$$

     where:
     - λ: Weight for value loss.
     - β: Entropy coefficient, encouraging exploration.
     - $H[\pi(\cdot|s_t)]$: Entropy of the policy distribution.

5. **Evaluation:** Evaluate the agent's cumulative profitability and ability to manage drawdowns effectively.

4.2 Multi-Agent Algorithms

This section extends the single-agent setups to multi-agent frameworks, focusing on PPO, DQN, A2C, and ensemble strategies.

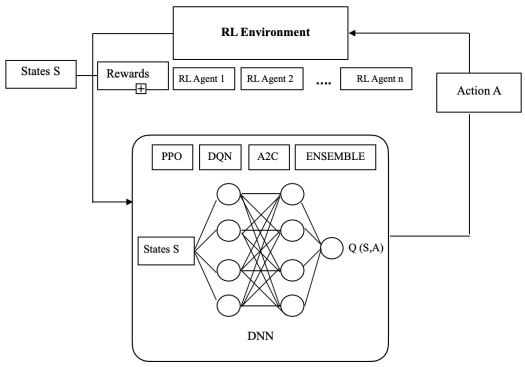

Figure 2: Multi-Agent RL Framework for Algorithmic Trading

4.2.1 Multi-Agent Proximal Policy Optimization (PPO)

Design Steps:

1. **Environment Setup:** Initialize a multi-agent trading environment where agents interact with the same market data; assign independent initial balances and trading strategies to each agent.
2. **Agent Collaboration:** For cooperative agents, the reward is shared:

$$R_t^{\text{shared}} = \frac{1}{N} \sum_{i=1}^{N} r_t^i,$$

where:

- R_t^{shared}: Shared reward for all agents at time t.
- r_t^i: Individual reward for agent i at time t.
- N: Total number of agents.

3. **Policy Updates:** Each agent optimizes the PPO objective:

$$\mathcal{L}_i^{\text{PPO}} = \mathbb{E}_t \Bigg[\min \Bigg(r_t^i \hat{A}_t^i, \text{clip}(r_t^i, 1-\epsilon, 1+\epsilon) \hat{A}_t^i \Bigg) \Bigg],$$

where:

- $r_t^i = \frac{\pi_\theta^i(a_t^i|s_t^i)}{\pi_{\theta_{\text{old}}}^i(a_t^i|s_t^i)}$: Probability ratio for agent i.
- $\hat{A}_t^i$: Advantage function for agent i, defined as $\hat{A}_t^i = Q(s_t^i, a_t^i) - V(s_t^i)$.
- ϵ: Clipping hyperparameter for stability.

4. **Evaluation:** Assess overall portfolio performance and inter-agent dynamics using metrics such as Sharpe and Sortino Ratios.

4.2.2 Multi-Agent Deep Q-Network (DQN)

Design Steps:

1. **Environment Setup:** Initialize a multi-agent trading environment with normalized market data and discrete action space.
2. **Independent Learning:** Each agent independently learns Q-values:

$$Q^i(s_t^i, a_t^i) \leftarrow Q^i(s_t^i, a_t^i) + \alpha \Big(r_t^i + \gamma \max_{a'} Q^i(s_{t+1}^i, a') - Q^i(s_t^i, a_t^i) \Big),$$

where:

- $Q^i(s_t^i, a_t^i)$: Q-value for agent i, representing the expected return for taking action a_t^i in state s_t^i.
- α: Learning rate.
- r_t^i: Reward for agent i at time t.
- γ: Discount factor for future rewards.
- $\max_{a'} Q^i(s_{t+1}^i, a')$: Maximum Q-value over all possible actions in the next state for agent i.

3. **Centralized Replay Buffer:** Experiences are shared across agents:

$$D = \bigcup_{i=1}^{N} \{(s_t^i, a_t^i, r_t^i, s_{t+1}^i)\},$$

where D is the replay buffer containing experiences from all agents.

4. **Evaluation:** Evaluate agents' individual performances and overall portfolio returns, with a focus on collaborative or competitive dynamics.

4.2.3 Multi-Agent Ensemble Learning

Design Steps:

1. **Model Integration:** Combine PPO, DQN, and A2C models for each agent; use majority voting to determine final actions:

$$a_t^{\text{ensemble}} = \text{mode}(a_t^{\text{PPO}}, a_t^{\text{DQN}}, a_t^{\text{A2C}}),$$

where:

- a_t^{ensemble}: Action selected by the ensemble model at time t.
- mode$(\cdot)$: Function selecting the most common action among all models.

2. **Weighted Aggregation:** Alternatively, actions are determined by weighted aggregation:

$$a_t^{\text{ensemble}} = \arg\max_a \sum_{i=1}^{N} w_i \pi_i(a|s_t),$$

where:

- w_i: Weight assigned to agent i, representing the reliability of its policy.
- $\pi_i(a|s_t)$: Probability of taking action a in state s_t according to agent i's policy.

3. **Evaluation:** Assess the robustness of ensemble strategies compared to individual models using aggregated profitability and risk-adjusted metrics.

5. Results and Analysis

This section provides a detailed analysis of the performance of single-agent and multi-agent DRL models across multiple stocks, using key metrics such as win rate, profit factor, Sharpe ratio, Sortino ratio, and maximum drawdown. The results highlight the strengths and weaknesses of each algorithm and configuration in both training and testing environments.

5.1 Single-Agent vs. Multi-Agent Configurations

The comparison between single-agent and multi-agent configurations reveals distinct trade-offs in performance metrics.

Win Rate: Single-agent models exhibited higher win rates, particularly with PPO and Ensemble strategies. For example, in stocks like AAPL and AMZN, win rates were consistently above 75

Profit Factor: Single-agent models achieved significantly higher profit factors, with the PPO model delivering a profit factor of 28.07 for BIDU and 24.46 for AMZN. Multi-agent models, while competitive in some cases (e.g., Ensemble with a profit factor of 7.07 for BIDU), generally underperformed in this metric.

Sharpe Ratio: Single-agent configurations outperformed multi-agent setups in risk-adjusted returns, with Sharpe ratios consistently higher for PPO and Ensemble models. For instance, in the AAPL stock, PPO achieved a Sharpe ratio of 0.69 during testing, compared to lower and sometimes negative values in multi-agent configurations.

Sortino Ratio: Sortino ratios further emphasized the superior downside risk management of single-agent models. For example, the single-agent PPO model recorded a Sortino ratio of 8.3 for BIDU, while multi-agent setups struggled to achieve comparable values.

Maximum Drawdown: Single-agent models demonstrated better capital preservation, with maximum drawdowns often below 1

5.2 Algorithmic Observations

PPO: PPO consistently delivered high profitability and low drawdowns in single-agent setups, making it a robust choice for

stable trading environments. However, its performance in multi-agent setups showed increased variability.

DQN: DQN underperformed in both configurations, with low Sharpe ratios and high drawdowns, making it less suitable for high-frequency trading.

A2C: A2C provided moderate performance, showing better Sortino ratios than DQN but higher drawdowns compared to PPO.

Ensemble: The Ensemble approach balanced profitability and risk effectively, achieving consistent performance across both configurations. For example, it achieved a profit factor of 7.07 for BIDU in the multi-agent setup, highlighting its adaptability.

5.3 Performance Across Stocks

Detailed stock-specific analyses highlighted varying performances across different algorithms:

- **AAPL:** PPO achieved a profit factor of 18.81 and a maximum drawdown of 0.01
- **AMZN:** The Ensemble model excelled with a profit factor of 6.85 and a Sharpe ratio of 0.49 in testing.
- **BIDU:** PPO outperformed with a profit factor of 28.07 and a Sortino ratio of 8.3, reflecting strong risk-adjusted returns.
- **TSLA:** The Ensemble approach achieved a profit factor of 9.07, showcasing resilience in high-volatility environments.

5.4 Results Table

Table 1 summarizes the performance metrics of the evaluated algorithms (PPO, DQN, A2C, and Ensemble) across the 24 companies. These results provide detailed insights into win rates, profit factors, Sharpe ratios, Sortino ratios, and maximum drawdowns during training and testing.

6. Discussion

The findings of this study underscore the strengths and limitations of different DRL algorithms and configurations in high-frequency trading scenarios. Single-agent models, particularly those using PPO and Ensemble strategies, consistently outperformed in terms of profitability, risk management, and stability. Multi-agent models, while offering potential for diversification, struggled with coordination and exhibited higher drawdowns.

Key Insights: 1. Single-agent PPO models demonstrated exceptional risk-adjusted returns, making them ideal for stable trading strategies. 2. Ensemble strategies proved versatile, balancing profitability and risk across configurations. 3. Multi-agent setups require further refinement to address coordination challenges and improve capital preservation.

Implications: These results validate the applicability of DRL in algorithmic trading, particularly for single-agent setups in high-frequency environments. Ensemble methods hold promise for further scalability and robustness, while DQN requires substantial improvements for practical use.

7. Conclusion and Future Work

This study confirms the efficacy of DRL, particularly PPO and Ensemble models, in optimizing algorithmic trading strategies. Single-agent setups outperform multi-agent configurations in key metrics, offering better risk management and stability.

Future Directions:

1. Integrating sentiment analysis and additional technical indicators into DRL frameworks.
2. Exploring advanced multi-agent coordination mechanisms to reduce drawdowns.

Table 1: Results of performance metrics of single agent

S.NO	STOCK	METRICS	TRAINING				TESTING			
			PPO	DQN	A2C	Ensemble	PPO	DQN	A2C	Ensemble
1	AAPL	Win Rate	0.73	0.47	0.53	0.58	0.79	0.42	0.65	0.58
		Profit factor	11.25	0.69	1.31	2.11	18.81	0.51	5.45	2.41
		Sharpe ratio	0.56	-0.11	0.08	0.22	0.69	-0.22	0.29	0.26
		Sortino ratio	3.85	-0.15	0.15	0.48	6.12	-0.31	1.53	0.66
		Maximum drawdown	0.0004	0.022	0.003	0.001	0.0001	0.028	0.0007	0.001
2	AMZN	Win Rate	0.76	0.43	0.54	0.67	0.83	0.48	0.53	0.73
		Profit factor	14.19	0.56	1.76	4.98	24.46	0.88	1.48	6.85
		Sharpe ratio	0.62	-0.17	0.16	0.44	0.72	-0.04	0.09	0.49
		Sortino ratio	4.49	-0.23	0.34	1.38	7.19	-0.05	0.24	1.67
		Maximum drawdown	0.0005	0.03	0.001	0.0009	0.0003	0.014	0.002	0.002
3	BABA	Win Rate	0.80	0.56	0.56	0.69	0.64	0.46	0.51	0.59
		Profit factor	19.88	1.14	2.35	5.62	5.05	1.53	1.35	3.17
		Sharpe ratio	0.68	0.04	0.24	0.43	0.39	0.13	0.08	0.32
		Sortino ratio	6.07	0.05	0.57	1.34	1.62	0.26	0.144	1.03
		Maximum drawdown	0.0003	0.007	0.002	0.002	0.0007	0.002	0.003	0.0007
4	BIDU	Win Rate	0.79	0.47	0.50	0.66	0.82	0.44	0.49	0.70
		Profit factor	23.87	0.79	0.99	5.12	28.07	0.77	0.98	7.07
		Sharpe ratio	0.55	-0.06	-0.002	0.36	0.66	-0.07	-0.004	0.46
		Sortino ratio	6.58	-0.08	-0.0003	1.44	8.3	-0.11	-0.006	2.3
		Maximum drawdown	0.0005	0.02	0.016	0.002	0.0002	0.01	0.012	0.0006
5	BRK.B	Win Rate	0.72	0.47	0.50	0.63	0.76	0.50	0.47	0.54
		Profit factor	10.96	0.88	1.1	2.8	11.6	0.93	0.93	1.89
		Sharpe ratio	0.56	-0.04	0.03	0.31	0.66	-0.02	-0.02	0.18
		Sortino ratio	3.8	-0.06	0.05	0.81	3.78	-0.03	-0.04	0.47
		Maximum drawdown	0.0005	0.014	0.007	0.002	0.0004	0.006	0.005	0.002
6	FB	Win Rate	0.77	0.49	0.51	0.63	0.79	0.55	0.48	0.63
		Profit factor	17.07	0.79	1.19	4.89	17.1	1.29	1.003	5.2
		Sharpe ratio	0.633	-0.06	0.055	0.42	0.48	0.066	0.0006	0.34
		Sortino ratio	5.55	-0.1	0.096	1.7	5.23	0.17	0.0008	1.94
		Maximum drawdown	0.0004	0.037	0.005	0.0009	0.0005	0.004	0.008	0.001
7	GOOGL	Win Rate	0.69	0.47	0.50	0.56	0.72	0.54	0.50	0.57
		Profit factor	6.612	0.72	1.12	1.73	10.86	1.1	1.13	1.49
		Sharpe ratio	0.52	-0.099	0.039	0.17	0.57	0.03	0.04	0.13
		Sortino ratio	2.49	-0.35	0.06	0.36	4.03	0.05	0.07	0.24
		Maximum drawdown	0.0002	0.006	0.001	0.001	0.0001	0.003	0.004	0.001
8	HSBC	Win Rate	0.78	0.49	0.50	0.56	0.79	0.45	0.49	0.61
		Profit factor	20.59	0.86	2.00	1.94	21.19	0.78	1.69	3.82
		Sharpe ratio	0.70	-0.04	0.18	0.19	0.68	-0.07	0.15	0.39
		Sortino ratio	6.22	-0.06	0.48	0.37	6.13	-0.113	0.37	1.41
		Maximum drawdown	0.0006	0.001	0.0002	0.002	0.0006	0.001	0.0003	0.002
9	JD	Win Rate	0.77	0.47	0.51	0.68	0.79	0.43	0.50	0.64
		Profit factor	15.28	0.68	1.29	6.16	21.68	0.70	1.35	4.38
		Sharpe ratio	0.64	-0.11	0.077	0.46	0.69	-0.106	0.08	0.42
		Sortino ratio	4.91	-0.15	0.136	2.15	7.05	-0.13	0.16	1.36
		Maximum drawdown	0.0002	0.016	0.001	0.0007	0.0001	0.007	0.001	0.0006
10	JPM	Win Rate	0.72	0.47	0.49	0.67	0.73	0.47	0.48	0.67
		Profit factor	12.56	0.76	1.1	5.8	12.1	0.81	0.97	6.03
		Sharpe ratio	0.55	-0.08	0.02	0.43	0.6	-0.05	-0.009	0.499
		Sortino ratio	4.67	-0.12	0.05	1.78	4.6	-0.07	-0.015	2.15
		Maximum drawdown	0.0001	0.01	0.003	0.0007	0.0002	0.006	0.004	0.0002

S.NO	STOCK	METRICS	TRAINING				TESTING			
			PPO	DQN	A2C	Ensemble	PPO	DQN	A2C	Ensemble
11	KHC	Win Rate	0.83	0.5	0.52	0.70	0.79	0.51	0.52	0.77
		Profit Factor	27.82	1.009	1.48	5.16	18.04	0.95	1.47	12.38
		Sharpe Ratio	0.75	0.003	0.129	0.47	0.68	-0.013	0.123	0.62
		Sortino Ratio	8.22	0.004	0.236	1.07	5.83	-0.02	0.24	3.95
		Maximum Drawdown	0.0004	0.001	0.0004	0.0006	0.0005	0.001	0.0004	0.0007
12	KO	Win Rate	0.8	0.46	0.51	0.7	0.77	0.40	0.53	0.54
		Profit Factor	26.08	0.73	1.64	8.76	14.79	0.43	1.46	1.48
		Sharpe Ratio	0.57	-0.08	0.134	0.47	0.64	-0.26	0.13	0.121
		Sortino Ratio	7.67	-0.11	0.315	3.22	4.8	-0.33	0.255	0.23
		Maximum Drawdown	0.0007	0.004	0.0004	0.0007	0.0006	0.003	0.0003	0.0006
13	LPL	Win Rate	0.757	0.5275	0.629	0.580	0.751	0.49	0.528	0.581
		Profit Factor	16.38	1.650	2.761	2.119	15.631	1.33	1.321	2.411
		Sharpe Ratio	0.667	0.154	0.299	0.218	0.667	0.092	0.075	0.258
		Sortino Ratio	5.359	0.287	0.629	0.489	4.749	0.155	0.173	0.665
		Maximum Drawdown	1.998	0.0002	5.449	0.001	1.899	0.0001	8.848	0.0010
14	MSFT	Win Rate	0.743	0.486	0.431	0.593	0.828	0.4638	0.344	0.712
		Profit Factor	9.631	0.875	1.929	2.717	23.61	1.054	3.230	7.765
		Sharpe Ratio	0.562	-0.043	0.109	0.295	0.715	0.017	0.1556	0.561
		Sortino Ratio	3.14	-0.067	0.607	0.782	6.560	0.032	1.683	2.682
		Maximum Drawdown	0.012	0.0028	0.0013	0.0005	0.0004	0.0045	0.002	0.0007
15	NFLX	Win Rate	0.821	0.522	0.624	0.636	0.702	0.471	0.487	0.574
		Profit Factor	21.93	1.153	3.470	3.090	6.529	0.807	2.822	2.652
		Sharpe Ratio	0.682	0.043	0.313	0.316	0.447	-0.072	0.234	0.243
		Sortino Ratio	5.481	0.068	0.869	0.691	2.226	-0.105	0.896	0.686
		Maximum Drawdown	0.001	0.019	0.008	0.006	0.003	0.023	0.005	0.003
16	NVDA	Win Rate	0.720	0.4842	0.624	0.544	0.796	0.493	0.487	0.5531
		Profit Factor	10.50	0.748	3.470	1.868	21.31	1.087	2.822	1.8255
		Sharpe Ratio	0.456	-0.079	0.313	0.146	0.650	0.027	0.234	0.180
		Sortino Ratio	3.591	-0.105	0.869	0.311	7.200	0.042	0.896	0.391
		Maximum Drawdown	7.923	0.002	0.008	0.0003	5.218	0.0007	0.005	0.0002
17	ORCL	Win Rate	0.715	0.5444	0.624	0.604	0.798	0.3947	0.487	0.646
		Profit Factor	10.58	0.964	3.470	2.768	17.97	0.452	2.822	4.058
		Sharpe Ratio	0.481	-0.010	0.313	0.268	0.719	-0.281	0.234	0.442
		Sortino Ratio	3.628	-0.013	0.869	0.697	5.870	-0.403	0.896	1.437
		Maximum Drawdown	0.0001	0.004	0.008	0.0004	0.0002	0.007	0.005	0.0006
18	PFE	Win Rate	0.798	0.468	0.608	0.639	0.812	0.410	0.609	0.681
		Profit Factor	19.44	0.759	2.546	5.114	18.58	0.808	2.547	5.251
		Sharpe Ratio	0.643	-0.082	0.295	0.402	0.800	-0.068	0.298	0.454
		Sortino Ratio	5.0629	-0.109	0.616	1.556	5.732	-0.114	0.619	1.761
		Maximum Drawdown	0.0001	0.0051	0.0002	0.0002	0.0001	0.0027	0.0002	0.0001
19	PTR	Win Rate	0.768	0.4611	0.639	0.688	0.805	0.380	0.609	0.644
		Profit Factor	19.99	0.711	5.114	4.186	17.79	0.455	2.547	3.868
		Sharpe Ratio	0.654	-0.096	0.402	0.407	0.767	-0.293	0.298	0.371
		Sortino Ratio	6.876	-0.122	1.556	1.080	5.311	-0.486	0.619	1.076
		Maximum Drawdown	7.376	0.002	0.0002	0.0003	0.0001	0.002	0.0001	0.0004
20	TM	Win Rate	0.762	0.503	0.677	0.66	0.840	0.549	1.0	0.712
		Profit Factor	23.76	1.324	5.138	4.317	24.09	1.489	2.547	7.019
		Sharpe Ratio	0.644	0.086	0.454	0.408	0.736	0.137	0.298	0.487
		Sortino Ratio	8.025	0.1591	2.208	1.122	5.999	0.240	0.619	2.188
		Maximum Drawdown	0.0001	0.003	0.0001	0.001	0.0003	0.001	0.0001	0.0006

S.NO	STOCK	METRICS	TRAINING				TESTING			
			PPO	DQN	A2C	Ensemble	PPO	DQN	A2C	Ensemble
21	TSLA	Win Rate	0.7094	0.4907	0.677	0.6521	0.826	0.456	1.0	0.695
		Profit Factor	13.428	0.832	5.138	5.456	38.23	0.857	2.547	9.070
		Sharpe Ratio	0.4600	-0.048	0.454	0.395	0.758	0.0442	0.298	0.550
		Sortino Ratio	4.759	-0.075	2.208	1.892	11.41	-0.063	0.619	3.368
		Maximum Drawdown	0.0004	0.0375	0.0001	0.002	0.0004	0.027	0.0001	0.0027
22	TWTR	Win Rate	0.764	0.478	0.4375	0.6280	0.712	0.493	0.608	0.571
		Profit Factor	15.299	0.784	0.809	3.948	8.475	1.044	0.828	2.062
		Sharpe Ratio	0.585	-0.072	-0.058	0.366	0.452	0.011	-0.0656	0.175
		Sortino Ratio	4.666	-0.101	-0.072	1.2841	2.298	0.0199	-0.099	0.334
		Maximum Drawdown	0.0002	0.006	0.001	0.0003	0.0003	0.001	0.0009	0.0012
23	V	Win Rate	0.822	0.490	0.615	0.706	0.857	0.414	0.75	0.721
		Profit Factor	30.644	0.898	1.956	7.782	34.63	0.695	72.25	6.213
		Sharpe Ratio	0.748	-0.036	0.225	0.492	0.811	-0.127	0.906	0.513
		Sortino Ratio	9.640	-0.059	0.424	2.515	9.706	-0.185	41.13	1.821
		Maximum Drawdown	0.0001	0.013	0.003	0.0009	0.0002	0.0011	5.286	0.0014
24	XOM	Win Rate	0.765	0.459	0.4375	0.648	0.762	0.449	0.445	0.568
		Profit Factor	17.594	0.621	0.809	4.467	13.05	0.733	0.819	3.316
		Sharpe Ratio	0.629	-0.154	-1.673	0.398	0.657	-0.108	-1.673	0.318
		Sortino Ratio	5.854	-0.215	-1.673	1.366	4.345	-0.178	-1.673	1.204
		Maximum Drawdown	0.0001	0.010	0.0034	0.0004	0.0001	0.0030	0.003	0.0004

Table 2: Results of performance metrics of multi agent

S.NO	STOCK	METRICS	TRAINING				TESTING			
			PPO	DQN	A2C	Ensemble	PPO	DQN	A2C	Ensemble
1	AAPL	Win Rate	0.63	0.51	0.75	0.50	0.70	0.49	0.61	0.67
		Profit factor	4.19	1.48	3.59	1.13	8.65	1.07	1.42	7.04
		Sharpe ratio	0.38	0.119	0.38	0.03	0.54	0.025	0.13	0.48
		Sortino ratio	1.01	0.23	0.77	0.047	2.26	0.04	0.23	2.43
		Maximum drawdown	0.19	0.32	0.53	0.62	0.13	1.79	0.79	0.23
2	AMZN	Win Rate	0.60	0.20	0.45	0.51	0.64	0.05	0.60	0.69
		Profit factor	2.30	1.12	2.50	1.63	3.74	1.12	2.30	7.46
		Sharpe ratio	0.25	-0.14	-0.046	0.09	0.39	0.03	-0.056	0.50
		Sortino ratio	0.523	-0.11	-0.32	0.30	0.98	0.06	-0.37	2.18
		Maximum drawdown	0.21	0.24	0.53	0.77	0.21	1.68	0.41	0.544
3	BABA	Win Rate	0.78	0.34	0.30	0.47	0.62	0.57	0.38	0.77
		Profit factor	17.06	1.67	2.40	1.34	3.60	1.16	1.42	6.13
		Sharpe ratio	0.705	-0.24	-0.146	0.088	0.37	0.05	-0.21	0.38
		Sortino ratio	3.09	-0.31	-0.52	0.168	0.88	0.08	-0.32	1.91
		Maximum drawdown	0.11	0.56	0.31	1.08	0.27	1.80	0.43	0.19
4	BIDU	Win Rate	0.73	0.54	0.25	0.58	0.71	0.49	0.74	0.74
		Profit factor	11.56	1.11	4.50	1.42	10.93	1.11	1.05	11.7
		Sharpe ratio	0.43	0.03	-0.133	0.12	0.53	0.03	0.02	0.58
		Sortino ratio	2.13	0.04	-0.62	0.25	2.48	0.04	0.03	3.80
		Maximum drawdown	0.26	2.61	0.21	0.60	0.14	2.61	0.975	0.22
5	BRK.B	Win Rate	0.68	0.53	0.38	0.46	0.72	0.56	0.58	0.58
		Profit factor	6.03	1.43	3.50	2.11	7.94	2.12	3.70	2.38
		Sharpe ratio	0.46	0.11	-0.353	0.19	0.57	0.24	-0.289	0.27
		Sortino ratio	1.39	0.19	-0.42	0.40	2.03	0.58	-0.22	0.65
		Maximum drawdown	0.22	0.87	0.41	0.32	0.16	1.07	0.06	0.35

S.NO	STOCK	METRICS	TRAINING				TESTING			
			PPO	DQN	A2C	Ensemble	PPO	DQN	A2C	Ensemble
6	FB	Win Rate	0.68	0.53	0.68	0.48	0.72	0.54	0.90	0.63
		Profit factor	6.46	1.49	1.53	1.14	8.53	1.10	60.54	3.42
		Sharpe ratio	0.48	0.155	0.13	0.03	0.38	0.03	2.36	0.25
		Sortino ratio	1.52	0.19	0.18	0.06	1.86	0.04	0.44	1.04
		Maximum drawdown	0.16	0.82	0.70	1.15	0.17	1.07	0.06	0.35
7	GOOGL	Win Rate	0.58	0.51	0.39	0.55	0.63	0.57	0.19	0.56
		Profit factor	2.17	1.40	1.12	1.63	3.70	1.77	2.342	1.90
		Sharpe ratio	0.25	0.11	-0.25	0.15	0.38	0.18	-0.35	0.19
		Sortino ratio	0.42	0.23	-0.37	0.27	0.95	0.36	-0.17	0.43
		Maximum drawdown	0.19	0.24	0.29	0.26	0.23	0.63	0.43	0.39
8	HSBC	Win Rate	0.70	0.48	0.55	0.55	0.78	0.53	0.60	0.65
		Profit factor	14.20	1.06	27.3	4.41	17.6	1.41	28.5	6.50
		Sharpe ratio	0.63	0.019	0.88	0.22	0.67	0.107	0.74	0.44
		Sortino ratio	2.93	0.03	13.83	1.49	3.70	0.19	15.83	2.27
		Maximum drawdown	0.177	1.53	0.02	0.63	0.14	1.07	0.01	0.43
9	JD	Win Rate	0.69	0.46	0.37	0.57	0.74	0.54	0.39	0.57
		Profit factor	6.65	1.12	1.43	1.71	13.73	1.24	1.76	1.71
		Sharpe ratio	0.53	0.039	-0.22	0.173	0.62	0.06	-0.35	0.17
		Sortino ratio	1.59	0.06	-3.34	0.32	3.01	0.09	-1.50	0.32
		Maximum drawdown	0.18	0.59	0.34	0.44	0.15	2.48	0.42	0.444
10	JPM	Win Rate	0.68	0.51	0.56	0.66	0.71	0.52	0.61	0.63
		Profit factor	7.20	1.12	3.43	5.89	8.71	1.17	3.87	4.58
		Sharpe ratio	0.47	0.03	0.56	0.49	0.54	0.05	0.34	0.42
		Sortino ratio	1.75	0.06	0.23	1.32	2.28	0.08	0.21	1.57
		Maximum drawdown	0.19	1.17	0.07	0.157	0.14	0.75	0.23	0.189
11	KHC	Win Rate	0.76	0.49	1.17	0.55	0.78	0.52	0.46	0.65
		Profit factor	13.95	1.030	1.32	1.53	15.12	1.22	1.05	5.41
		Sharpe ratio	0.64	0.009	0.07	0.15	0.66	0.06	0.017	0.42
		Sortino ratio	2.70	0.014	0.14	0.28	2.91	0.11	0.03	1.88
		Maximum drawdown	0.156	1.92	0.60	1.05	0.15	1.43	2.30	0.36
12	KO	Win Rate	0.75	0.34	0.30	0.54	0.73	0.57	0.38	0.59
		Profit factor	15.97	1.67	2.40	1.48	8.89	1.16	1.40	2.75
		Sharpe ratio	0.54	-0.24	-0.146	0.11	0.57	0.05	-0.21	0.29
		Sortino ratio	3.09	-0.31	-0.52	0.19	2.00	0.08	-0.32	0.82
		Maximum drawdown	0.22	0.56	0.31	1.62	0.33	1.80	0.43	0.57
13	LPL	Win Rate	0.727	0.537	0.756	0.514	0.711	0.544	0.775	0.629
		Profit factor	11.77	1.562	12.35	1.534	12.300	1.841	16.094	5.117
		Sharpe ratio	0.616	0.137	0.639	0.127	0.625	0.171	0.715	0.436
		Sortino ratio	2.146	0.257	3.186	0.232	2.715	0.366	4.219	1.643
		Maximum drawdown	0.208	0.770	0.220	0.854	0.138	0.497	0.273	0.268
14	MSFT	Win Rate	0.693	0.507	0.550	0.476	0.712	0.518	0.560	0.634
		Profit factor	6.609	1.271	7.446	1.068	7.6303	1.069	7.466	3.564
		Sharpe ratio	0.534	0.073	0.677	0.0173	0.534	0.0228	0.657	0.378
		Sortino ratio	1.211	0.125	4.787	0.030	1.69	0.0379	4.799	1.048
		Maximum drawdown	0.399	0.664	0.206	1.039	0.223	1.197	0.206	0.358
15	NFLX	Win Rate	0.728	0.526	0.604	0.572	0.684	0.528	0.495	0.636
		Profit factor	7.362	1.068	2.791	1.919	7.196	1.367	3.367	4.301
		Sharpe ratio	0.434	0.0198	0.427	0.158	0.408	0.099	0.495	0.309
		Sortino ratio	1.193	0.029	1.105	0.369	1.577	0.178	2.321	1.194
		Maximum drawdown	0.592	1.961	0.715	0.819	0.400	0.863	1.337	0.810

S.NO	STOCK	METRICS	TRAINING				TESTING			
			PPO	DQN	A2C	Ensemble	PPO	DQN	A2C	Ensemble
16	NVDA	Win Rate	0.565	0.470	0.571	0.543	0.619	0.470	0.563	0.635
		Profit factor	1.900	1.158	7.617	1.173	3.502	1.158	1.205	3.742
		Sharpe ratio	0.198	0.040	0.657	0.051	0.343	0.040	0.0625	0.361
		Sortino ratio	0.315	0.0718	4.097	0.084	0.804	0.0718	0.099	1.125
		Maximum drawdown	0.300	1.337	0.100	0.34	0.327	1.337	1.13	0.321
17	ORCL	Win Rate	0.622	0.485	0.571	0.539	0.671	0.533	0.563	0.630
		Profit factor	3.549	1.048	7.617	1.378	5.874	1.309	1.205	4.556
		Sharpe ratio	0.353	0.0145	0.657	0.093	0.482	0.086	0.0625	0.426
		Sortino ratio	0.838	0.023	4.097	0.1785	1.577	0.144	0.099	1.673
		Maximum drawdown	0.245	2.335	0.100	0.683	0.188	0.844	1.13	0.248
18	PFE	Win Rate	0.756	0.544	0.580	0.660	0.717	0.497	0.560	0.665
		Profit factor	14.711	1.482	7.620	5.850	9.776	1.290	1.190	5.864
		Sharpe ratio	0.607	0.123	0.710	0.450	0.582	0.088	0.062	0.444
		Sortino ratio	3.216	0.226	4.088	1.88	2.236	0.157	0.089	1.841
		Maximum drawdown	0.157	0.957	0.145	0.480	0.203	1.967	1.13	0.461
19	PTR	Win Rate	0.764	0.550	1.00	0.544	0.784	0.497	1.00	0.669
		Profit factor	28.930	1.460	1.19	2.381	15.66	1.290	1.190	7.732
		Sharpe ratio	0.757	0.120	4.466	0.162	0.676	0.088	4.466	0.403
		Sortino ratio	9.664	0.260	0.089	0.747	3.114	0.157	0.089	2.361
		Maximum drawdown	0.037	0.960	1.130	0.645	0.273	1.967	1.130	0.539
20	TM	Win Rate	0.735	0.539	0.539	0.680	0.761	0.561	0.500	0.681
		Profit factor	12.570	1.304	1.304	5.344	0.157	1.352	2.224	5.354
		Sharpe ratio	0.597	0.0795	0.079	0.411	0.652	0.101	0.379	0.412
		Sortino ratio	2.528	0.1142	0.114	1.560	3.332	0.173	1.122	1.594
		Maximum drawdown	0.143	0.826	0.826	0.501	0.182	0.932	0.000	0.500
21	TSLA	Win Rate	0.580	0.544	0.650	0.671	0.710	0.476	0.666	0.585
		Profit factor	2.456	1.057	3.510	3.110	9.21	1.280	3.541	2.435
		Sharpe ratio	0.239	0.017	0.51	0.332	0.553	0.0634	0.514	0.251
		Sortino ratio	0.498	0.027	1.73	1.033	2.15	0.100	1.797	0.647
		Maximum drawdown	0.14	0.460	0.320	0.164	0.141	1.342	0.369	0.688
22	TWTR	Win Rate	0.646	0.522	0.734	0.523	0.648	0.492	0.718	0.579
		Profit factor	5.055	1.082	12.16	1.869	5.223	1.052	8.712	2.432
		Sharpe ratio	0.404	0.024	0.566	0.212	0.355	0.014	0.434	0.214
		Sortino ratio	1.140	0.036	4.327	0.434	1.069	0.0264	1.877	0.611
		Maximum drawdown	0.247	0.805	0.2610	0.509	0.310	0.988	0.394	0.632
23	V	Win Rate	0.806	0.51	0.461	0.529	0.830	0.528	0.1	0.681
		Profit factor	26.31	1.227	1.424	1.752	26.25	1.287	1.424	5.177
		Sharpe ratio	0.695	0.08	0.114	0.157	0.760	0.0841	0.114	0.466
		Sortino ratio	4.227	0.140	0.209	0.326	4.879	0.1421	0.209	1.654
		Maximum drawdown	0.170	2.978	0.735	0.532	0.166	2.9800	0.740	0.513
24	XOM	Win Rate	0.702	0.521	0.561	0.524	0.688	0.52	0.564	0.691
		Profit factor	9.090	1.204	1.118	1.583	7.443	1.105	1.119	6.4803
		Sharpe ratio	0.5146	0.0614	0.034	0.122	0.551	0.0305	0.044	0.525
		Sortino ratio	2.040	0.1021	0.0557	0.336	1.922	0.049	0.0565	2.156
		Maximum drawdown	0.263	2.242	2.528	0.878	0.339	1.743	2.528	0.408

3. Extending the evaluation to other financial instruments and markets.

By addressing these avenues, future research can further enhance the robustness and applicability of DRL-based trading strategies.

References

[1] D. H. Bailey, J. M. Borwein, M. L. de Prado, and Q. J. Zhu, "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance," *Notice of the American Mathematical Society*, pp. 458–471, 2014.

[2] T. Théate and D. Ernst, "An Application of Deep Reinforcement Learning to Algorithmic Trading," .

[3] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A Brief Survey of Deep Reinforcement Learning," *CoRR*, vol. abs/1708.05866, 2017.

[4] K. Zhang, J. Cao, H. Liu, and L. Zhang, "Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy," in *International Conference on Artificial Intelligence and Statistics*, pp. 2304–2312, 2020.

[5] D. H. Bailey, J. M. Borwein, M. L. de Prado, and Q. J. Zhu, "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance," *Notice of the American Mathematical Society*, pp. 458–471, 2014.

[6] W. N. Bao, J. Yue, and Y. Rao, "A Deep Learning Framework for Financial Time Series using Stacked Autoencoders and Long-Short Term Memory," *PloS One*, vol. 12, 2017.

[7] M. G. Bellemare, W. Dabney, and R. Munos, "A Distributional Perspective on Reinforcement Learning," *CoRR*, vol. abs/1707.06887, 2017.

[8] J. Bollen, H. Mao, and X. J. Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, vol. 2, pp. 1–8, 2011.

[9] F. Li, Z. Wang, and P. Zhou, "Ensemble Investment Strategies Based on Reinforcement Learning," *Scientific Programming*, vol. 2022, Article ID 7895674, 2022, doi: 10.1155/2022/7895674.

[10] E. S. Ponomareva, I. V. Oseledets, and A. S. Cichocki, "Using Reinforcement Learning in the Algorithmic Trading Problem," .

[11] B. Briola, M. Di Marco, M. Costantino, and R. Zanetti, "Deep Reinforcement Learning for Active High Frequency Trading," .

[12] Y. Deng, F. Bao, Y. Kong, Z. Ren, and Q. Dai, "Deep Direct Reinforcement Learning for Financial Signal Representation and Trading," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 653–664, Mar. 2017, doi: 10.1109/TNNLS.2016.2522401.