Kalman-Enhanced Deep Reinforcement Learning for Noise-Resilient Algorithmic Trading in Volatile Gold Markets

Amine Kili[1], Brahim Raouyane[2], Mohamed Rachdi[3], Mostafa Bellafkih[4]
Computer and System Laboratory (LIS)-N&IDP Team-Department of Mathematics and Computer Science,
Faculty of Sciences, Ain Chock, Casablanca 20000, Morocco[1,2]
The Information Processing and Modeling Laboratory (LTIM)-Ben M'Sik Faculty of Sciences,
National Higher School of Art and Design (ENSAD), Casablanca 20000, Morocco[3]
RAISS Laboratory-Department of Mathematics and Computer Science,
National Institute of Posts and Telecommunications (INPT), Rabat, Morocco[4]

Abstract—**Precious metals market, such as gold (XAU/USD), exhibit high volatility and significant microstructure noise in their financial time series, which degrade the reliability of algorithmic trading models. While deep reinforcement learning (DRL) has shown strong results in equities and cryptocurrencies, its application to precious metals remains limited by unstable signals and rapid market fluctuations. This study proposes a Kalman-enhanced DRL framework that integrates classical noise filtering with modern neural architectures to improve signal quality and trading performance in highly volatile environments. The methodology applies Kalman filtering to recursively denoise OHLCV price data, which then serves as an input alongside 22 technical indicators to train three state-of-the-art DRL agents: Deep Q-Network (DQN), Proximal Policy Optimization (PPO), and Recurrent PPO (RPPO). Eight years of hourly XAU/USD data (January 2017 to January 2025, N = 47,304) were used for training and evaluation. Models were evaluated on cumulative return, CAGR, Sharpe ratio, maximum drawdown, and volatility. Results demonstrate substantial gains from noise attenuation: PPO with Kalman filtering achieved 80.21% cumulative return (27.1% CAGR, Sharpe 12.10, drawdown -0.48%) compared with raw PPO's 8.70% (3.46% CAGR, Sharpe 0.45, drawdown -12.52%). DQN and RPPO achieved comparable improvements, with 244 to 822% return increases, 88 to 96% drawdown reduction, and up to 29× Sharpe ratio enhancement. Statistical significance was confirmed ($p < 0.001$ for PPO/RPPO; $p < 0.05$ for DQN). These findings highlight Kalman-enhanced reinforcement learning as a scalable and robust framework for institutional algorithmic trading, bridging signal processing and artificial intelligence for next-generation adaptive trading systems.**

***Keywords*—Deep reinforcement learning; Kalman filtering; algorithmic trading; gold markets; XAU/USD; microstructure noise; signal processing; financial time series; noise reduction; institutional trading**

List of Abbreviations

Abbreviation	Meaning
DRL	Deep Reinforcement Learning
DQN	Deep Q-Network
PPO	Proximal Policy Optimization
RPPO	Recurrent Proximal Policy Optimization
OHLCV	Open, High, Low, Close, Volume
CAGR	Compound Annual Growth Rate
RL	Reinforcement Learning
AI	Artificial Intelligence
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
MDP	Markov Decision Process

I. Introduction

A. Global Context and Market Significance

Algorithmic trading has fundamentally transformed global financial markets, evolving from primitive rule-based systems in the 1970s to sophisticated artificial intelligence-driven strategies [1] processing vast quantities of market data in real-time across microsecond timeframes [2]. The precious metals market, particularly gold (XAU/USD), represents one of the most challenging and economically significant environments for algorithmic trading [3] due to its unique characteristics: extreme volatility patterns, complex macroeconomic dependencies, susceptibility to geopolitical events, and the presence of both speculative and hedging flows. With global gold trading volumes exceeding $183 billion daily across continuous 24-hour operations spanning multiple international time zones, the economic importance of developing intelligent, adaptive trading systems has become increasingly critical for institutional investors, hedge funds, central banks, and asset managers seeking to capitalize on market inefficiencies while maintaining stringent risk management protocols and regulatory compliance standards [4].

The gold market's distinctive properties-including its role as a safe-haven asset during economic uncertainty, inverse correlation with the US dollar, sensitivity to inflation expectations, and function as a portfolio diversifier-create complex trading dynamics that traditional algorithmic approaches struggle to navigate effectively [5]. Recent market volatility, exemplified by the COVID-19 pandemic's impact on financial markets, Brexit-related uncertainties, and geopolitical tensions, has further emphasized the need for sophisticated noise-resistant trading strategies capable of maintaining performance across diverse market regimes. The increasing institutional adoption of alternative investments and systematic trading strategies has created substantial demand for robust, scientifically-validated algorithmic trading frameworks that can deliver consistent alpha generation while managing downside risk within institutional risk budgets.s.

B. Technical Challenges and Research Gap

Despite significant advances in machine learning applications to financial markets over the past two decades, existing algorithmic trading approaches face critical limitations in handling microstructure noise [6] and non-stationary dynamics characteristic of highly volatile markets [7]. Traditional technical analysis and momentum-based strategies, while computationally efficient, suffer from poor signal-to-noise ratios that generate excessive false signals and achieve suboptimal risk-adjusted returns [8], particularly during periods of high market stress. Rule-based systems lack the adaptability necessary to respond to evolving market conditions, while statistical arbitrage models often rely on mean-reversion assumptions that may not hold during structural market shifts or regime changes.

Recent developments in deep learning and reinforcement learning have shown promise for financial applications [9], with neural networks demonstrating superior pattern recognition capabilities compared to traditional econometric models. However, most existing deep reinforcement learning frameworks for trading have inadequately addressed the fundamental challenge of noise reduction in high-frequency financial data, resulting in unstable training dynamics [7], poor generalization to out-of-sample periods, and inconsistent performance across different market conditions. The black-box nature of deep learning models further complicates their adoption in institutional settings where interpretability and regulatory compliance are paramount concerns.

Furthermore, the majority of academic research in algorithmic trading has focused on either purely machine learning approaches or traditional signal processing techniques in isolation, without systematically exploring the synergistic potential of integrating classical signal processing theory with modern artificial intelligence methodologies [10]. This represents a significant gap in the literature, as optimal signal extraction from noisy financial time series-a well-established problem in control theory and signal processing-could substantially enhance the learning efficiency and performance stability of reinforcement learning agents operating in financial markets.

C. Research Contributions and Study Objectives

This study addresses these fundamental limitations by developing a novel deep reinforcement learning framework [11] that strategically integrates Kalman filtering with advanced neural architectures to enhance signal quality and trading performance in volatile gold markets.

1) Research Objectives (RO):

- To develop a hybrid reinforcement learning framework that combines Kalman filtering with deep neural architectures to improve signal quality and trading performance in highly volatile markets such as gold (XAU/USD).
- To evaluate and compare the effectiveness of multiple reinforcement learning agents under noise-reduced and raw data conditions using rigorous statistical testing.
- To demonstrate the practical and statistical significance of performance improvements for institutional-level algorithmic trading applications.

2) Research Contributions (RC):

- We develop and compare three state-of-the-art reinforcement learning algorithms, Deep Q-Networks (DQN), Proximal Policy Optimization (PPO), and Recurrent PPO (RPPO), enhanced with Kalman-filtered market signals. This provides the first empirical evaluation of noise-attenuated deep reinforcement learning for precious metals trading across value-based and policy-based paradigms [12], [13].
- We design a trading environment that integrates 22 technical indicators with noise-reduced OHLCV (Open, High, Low, Close, Volume) data, ensuring rich state representations and computational efficiency. Realistic market factors such as transaction costs, slippage, and position limits are incorporated for institutional relevance [14].
- We conduct empirical evaluation on eight years of XAU/USD data (January 2017 to January 2025), showing 244 to 822% performance gains with improved risk-adjusted metrics. Statistical tests and effect size analysis confirm both practical and statistical significance of results.
- The proposed framework unifies classical signal processing with modern deep learning, offering a scalable and scientifically grounded solution for institutional algorithmic trading in volatile markets [15].

3) Study organization: The remainder of this study is organized as follows: Section II reviews related work in algorithmic trading systems, signal processing applications in finance, machine learning evolution, and reinforcement learning in financial trading. Section III presents the theoretical foundation including reinforcement learning mathematical framework, DQN and PPO theory, and Kalman filter theory. Section IV describes the methodology including data sources, technical indicator computation, Kalman filter implementation, environment design, and DRL architecture details. Section V presents experimental results and comprehensive performance analysis. Section VI discusses principal findings, theoretical contributions, and contextualization within the financial machine learning literature. Section VII concludes the study with key takeaways and future research directions.

II. Related Work

Algorithmic trading systems have evolved from simple rule-based strategies to complex adaptive models that integrate statistical analysis, machine learning, and optimization techniques. Early financial systems relied on deterministic algorithms with limited adaptability to market volatility. The introduction of signal processing methods, such as filtering and spectral analysis, enhanced noise reduction and feature extraction from price data, improving model stability in dynamic conditions. With the rise of machine learning, regression and classification-based models began capturing non-linear dependencies in market behavior, paving the way for data-driven decision-making. More recently, reinforcement learning has emerged as a powerful paradigm in financial trading, enabling agents to learn optimal policies through direct interaction with market environments, thereby addressing challenges of non-stationarity, delayed rewards, and high-frequency fluctuations.

A. Evolution of Algorithmic Trading Systems

Algorithmic trading systems have evolved from rule-based strategies to sophisticated adaptive models [16], [17]. Traditional approaches primarily rely on technical analysis indicators, statistical arbitrage models, and momentum-based strategies [18], [19]. Technical analysis methods, including moving averages, relative strength indicators, Bollinger bands, and MACD oscillators, form the foundation of many algorithmic systems due to their computational simplicity and intuitive interpretation. These methods attempt to identify patterns and trends in price movements through mathematical transformations of historical data.

However, traditional approaches suffer from fundamental limitations that constrain their effectiveness in modern markets. First, they generate frequent false signals in noisy market conditions, leading to excessive transaction costs from overtrading. Second, they lack adaptability to changing market regimes - strategies optimized for trending markets often fail catastrophically during mean-reverting periods. Third, they exhibit poor risk-adjusted performance during high volatility, as fixed decision rules cannot dynamically adjust to changing market conditions [20]. Statistical arbitrage strategies, including pairs trading and mean-reversion models, achieve greater sophistication by exploiting statistical relationships between instruments, yet remain vulnerable to structural breaks and require frequent manual recalibration to maintain performance.

B. Signal Processing Applications in Finance

Signal processing techniques provide powerful tools for extracting meaningful patterns from noisy financial data [21]. Kalman filtering has proven particularly valuable due to its optimal estimation properties under Gaussian noise assumptions and recursive computational structure enabling real-time operation with constant memory requirements [22]. The filter decomposes observed price movements into signal components (reflecting fundamental economic information) and noise components (arising from market microstructure effects, temporary supply-demand imbalances, and random trading activity) [23].

Key applications span multiple domains: time-varying volatility estimation where Kalman filters often outperform traditional GARCH models in forecasting accuracy, Value-at-Risk calculations requiring accurate volatility estimates for risk management, trend detection and regime identification for tactical asset allocation, and optimal portfolio weight estimation under parameter uncertainty. The filter's ability to provide uncertainty quantification through covariance matrices proves valuable for risk-aware trading systems. Recent extensions including particle filters and unscented Kalman filters address nonlinear dynamics in financial markets [24], [25], though these advanced methods require significantly greater computational resources and more complex implementation compared to standard linear Kalman filtering.

C. Machine Learning Evolution in Financial Markets

Machine learning represents a paradigm shift from static rule-based systems to adaptive, data-driven approaches capable of discovering complex patterns in market behavior [26]. Traditional supervised learning methods including support vector machines, decision trees, random forests, and neural networks have been applied to price prediction, trend classification, and volatility forecasting tasks with varying degrees of success. Deep learning architectures further enhanced pattern recognition capabilities by processing high-dimensional market data and discovering nonlinear relationships previously undetectable by conventional econometric methods.

Convolutional neural networks treat financial time series as images, applying computer vision techniques to identify recurring patterns and price formations [27]. Long Short-Term Memory networks address the vanishing gradient problem in recurrent architectures, enabling capture of long-range temporal dependencies critical for financial forecasting [28]. Attention mechanisms and transformer architectures have recently shown promise in modeling complex temporal relationships across multiple time scales.

However, supervised learning approaches face fundamental challenges in financial applications. First, the requirement for labeled training data proves problematic when true optimal actions are unknown and success metrics are multifaceted beyond simple price direction. Second, non-stationarity in financial markets means patterns learned from historical data may not persist in future periods, leading to degraded out-of-sample performance. Third, feedback effects create a moving target problem where successful strategies may ultimately eliminate the market inefficiencies they exploit. Recent approaches including transfer learning, ensemble methods combining multiple models, and adversarial training that explicitly models market adaptation effects attempt to address these limitations [29], [30], yet fundamental constraints of non-stationarity and limited labeled data continue to limit purely supervised approaches.

D. Reinforcement Learning in Financial Trading

Reinforcement learning addresses fundamental limitations of supervised learning by enabling agents to learn optimal trading strategies through direct market interaction without requiring labeled training data [31], [32]. The Markov Decision Process formulation naturally accommodates the sequential nature of trading decisions and delayed feedback characteristic of financial markets, where states represent market information, actions correspond to trading decisions, and rewards reflect portfolio performance objectives including returns, risk, and transaction costs (see Fig. 1).

Early RL applications employed tabular Q-learning and temporal difference methods, but faced severe limitations from the curse of dimensionality when applied to high-dimensional financial state spaces. Deep Reinforcement Learning (DRL) revolutionized the field by combining the representational power of deep neural networks with RL's adaptive learning capabilities [33]. Deep Q-Networks introduced two key innovations-experience replay to break temporal correlations and target networks to stabilize training-enabling successful application to complex environments. Policy gradient methods, particularly Proximal Policy Optimization, provide more stable training through clipped objective functions that prevent excessively large policy updates while maintaining sample efficiency through actor-critic architectures [34].

Recent developments explore advanced architectures and training paradigms: multi-agent systems that model strategic interactions between market participants, hierarchical reinforcement learning for coordinated decision-making across multiple time scales, meta-learning approaches enabling rapid adaptation to new market regimes, and model-based methods that learn environment dynamics for improved sample efficiency [35]. However, a critical gap persists in the literature: most DRL trading systems apply learning algorithms directly to raw, noisy market data without systematic integration of signal processing techniques to address the fundamental challenge of microstructure noise in high-frequency financial time series.

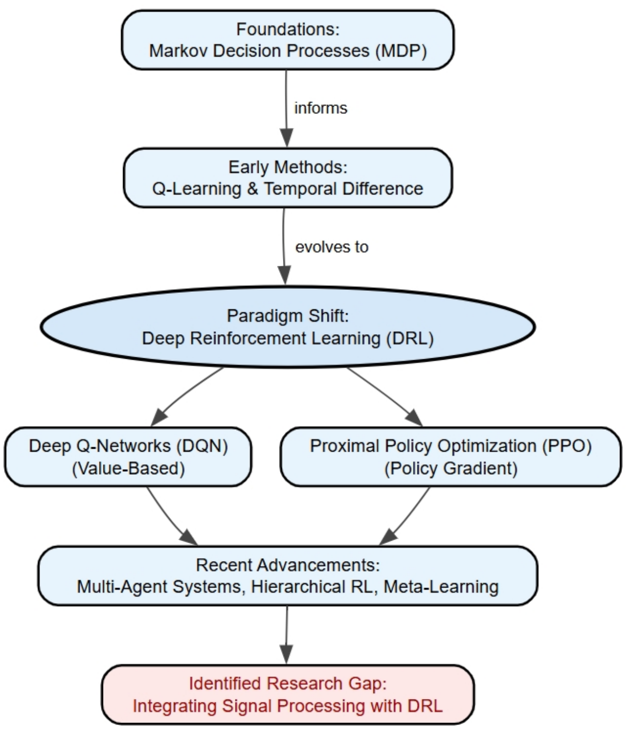

Fig. 1. Evolution of RL methods in financial trading from MDPs to deep RL.

E. Literature Gap and Research Positioning

Despite extensive research in signal processing and machine learning for finance, a significant gap exists in systematically combining these complementary approaches [36]. Most DRL models are applied directly to raw, noisy price data with insufficient signal-to-noise ratios, exposing agents to spurious correlations and causing performance instability in volatile markets. Signal processing and machine learning are typically treated as separate domains with limited exploration of synergistic potential.

Our research addresses this gap by proposing a hybrid methodology (see Fig. 2) that deeply integrates Kalman filtering as a core component of state representation, rather than as disconnected preprocessing. This allows agents to learn from statistically filtered market dynamics, improving policy robustness and stability. Our systematic evaluation across DQN, PPO, and RPPO demonstrates how classical control theory fuses with modern AI for enhanced performance in complex financial environments.

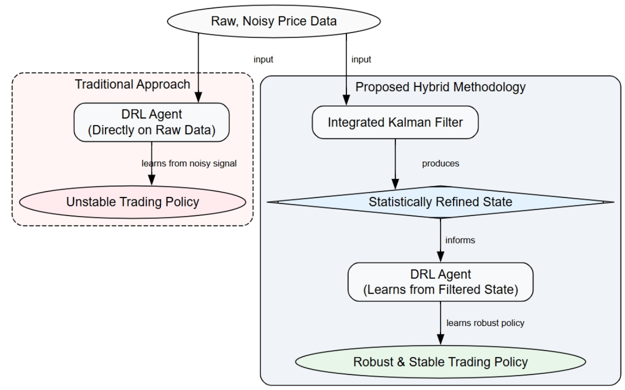

Fig. 2. Flowchart of the proposed Kalman-enhanced DRL framework for robust and stable trading.

III. THEORETICAL FRAMEWORK

A. Reinforcement Learning Mathematical Foundation

Reinforcement learning provides a mathematical framework for sequential decision-making formalized through Markov Decision Processes (MDPs) [37], [38]. For algorithmic trading, we define the MDP tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$ where: $\mathcal{S}$ is the state-space encompassing Kalman-filtered price data, 22 technical indicators, and portfolio positions; $\mathcal{A} = \{-1, 0, 1\}$ represents discrete trading actions (sell, hold, buy); $\mathcal{P}$ captures stochastic market state transitions; and $\mathcal{R}$ is the reward function balancing returns, risk, and costs [39], [40], [41].

Our reward function combines multiple trading objectives in Eq. (1):

$$(r_t = \alpha \cdot \text{return}_t - \beta \cdot \text{risk}_t - \gamma \cdot \text{cost}_t + \delta \cdot \text{consistency}_t) \quad (1)$$

where, hyperparameters control objective weighting. The RL objective is to learn an optimal policy π^* maximizing expected cumulative discounted return with $\gamma = 0.99$ [see Eq. (2)]:

$$(J(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 \right]) \quad (2)$$

B. Deep Q-Networks and Value Function Approximation

DQN approximates the optimal action-value function $Q^*(s, a)$ using deep neural networks, enabling application to high-dimensional financial state spaces. The Q-function satisfies the Bellman optimality equation [see Eq. (3)]:

$$(Q^*(s,a) = \mathbb{E}_{s' \sim \mathcal{P}(s'|s,a)}[r(s,a) + \gamma \max_{a'} Q^*(s',a')]) \quad (3)$$

DQN trains a neural network $Q(s,a;\theta) \approx Q^*(s,a)$ by minimizing temporal difference error using experience replay and target networks [see Eq. (4)]:

$$(\mathcal{L}(\theta) = \mathbb{E}_{(s,a,r,s') \sim \mathcal{D}}\left[(y - Q(s,a;\theta))^2\right]) \quad (4)$$

where, $y = r + \gamma \max_{a'} Q(s',a';\theta^-)$ is computed using a target network (θ^-), and $\mathcal{D}$ is the replay buffer. The policy follows an ϵ-greedy strategy [see Eq. (5)]:

$$(\pi(s) = \begin{cases} \arg\max_a Q(s,a;\theta) & \text{with probability } 1-\epsilon \\ \text{random action} & \text{with probability } \epsilon \end{cases}) \quad (5)$$

Experience replay breaks temporal correlations and improves sample efficiency, while the target network stabilizes training by providing consistent targets.

C. Policy Gradient Methods and PPO

Policy gradient methods directly optimize the policy $\pi_\theta(a|s)$ without explicit value function estimation. Proximal Policy Optimization (PPO) addresses training instability by constraining policy updates through a clipped surrogate objective [see Eq. (6)]:

$$(\mathcal{L}^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]) \quad (6)$$

where, $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the probability ratio, $\hat{A}_t$ is the advantage estimate, and ϵ controls clipping (typically 0.1-0.3). This prevents large policy updates that could destabilize financial trading strategies. The complete PPO objective combines policy loss, value function loss, and entropy regularization [see Eq. (7)]:

$$(\mathcal{L}(\theta) = \mathcal{L}^{CLIP}(\theta) - c_1\mathcal{L}^{VF}(\theta) + c_2 S[\pi_\theta](s_t)) \quad (7)$$

D. Recurrent Policy Optimization and Memory Integration

Recurrent PPO (RPPO) extends standard PPO by incorporating LSTM networks to capture temporal dependencies in market data beyond immediate state representation [42], [43], [44], [45], [46]. The RPPO policy $\pi_\theta(a_t|s_t, h_{t-1})$ uses hidden state h_{t-1} to provide memory of recent market conditions [see Eq. (8)]:

$$(h_t, c_t = \text{LSTM}(s_t, h_{t-1}, c_{t-1}; \theta)) \quad (8)$$

where, c_t captures long-term memory. This enables the agent to remember market patterns and portfolio dynamics across multiple timesteps, adjusting decisions based on recent trading history rather than treating timesteps independently. We use truncated backpropagation through time (TBPTT) for training, resetting hidden states at episode boundaries while maintaining temporal continuity within episodes.

E. Kalman Filter Theory and Implementation

The Kalman filter provides optimal state estimation for linear dynamic systems with Gaussian noise [47], treating observed prices as noisy measurements of underlying value processes. We model the financial time series using a state-space representation where x_t represents the true price and y_t is the noisy observation [see Eq. (9) and Eq. (10)]:

$$(x_{t+1} = x_t + w_t \qquad w_t \sim \mathcal{N}(0, Q_t)) \quad (9)$$

$$(y_t = x_t + v_t \qquad v_t \sim \mathcal{N}(0, R_t)) \quad (10)$$

The filter recursively estimates the true state through prediction and update steps [see Eq. (11) and Eq. (12)]:

Prediction:

$$(\hat{x}_{t|t-1} = \hat{x}_{t-1|t-1}, \quad P_{t|t-1} = P_{t-1|t-1} + Q_t) \quad (11)$$

Update:

$$(\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - \hat{x}_{t|t-1}), \quad K_t = \frac{P_{t|t-1}}{P_{t|t-1} + R_t}) \quad (12)$$

where, K_t is the Kalman gain balancing predictions and observations, and $P_{t|t}$ is the error covariance. Noise parameters Q_t and R_t are estimated via maximum likelihood on training data. The recursive structure enables real-time operation with constant computational complexity, crucial for algorithmic trading.

IV. Methodology

A. Data Sources and Market Environment

Our empirical analysis utilizes eight years of high-frequency XAU/USD (gold vs. US Dollar) trading data spanning January 1, 2017 to January 31, 2025, providing a comprehensive dataset of 47,304 hourly observations. This extensive temporal coverage encompasses multiple distinct market regimes including the 2018 trade war volatility, the 2020 COVID-19 pandemic market disruption, subsequent monetary policy normalization, and recent geopolitical tensions affecting global commodity markets. The dataset includes complete OHLCV (Open, High, Low, Close, Volume) information with no missing observations, ensuring data integrity for our analysis [48]. The XAU/USD pair represents an ideal testing environment for noise-attenuated trading strategies due to its unique market characteristics: exceptional volatility with annualized standard deviation exceeding 15%, complex macroeconomic dependencies including inverse dollar correlation and inflation sensitivity, continuous 24-hour trading across global time zones, and substantial institutional participation from central banks, hedge funds, and commodity trading advisors [49]. The dataset's temporal span captures multiple distinct market regimes including the 2018 trade war volatility period, the 2020 COVID-19 pandemic market disruption with unprecedented monetary policy interventions, subsequent normalization phases, and recent geopolitical tensions affecting global commodity markets. Data preprocessing involved several critical steps to ensure optimal model performance while maintaining data integrity [50]. Raw hourly data was re-sampled to daily frequency to improve training stability and computational efficiency while preserving essential price dynamics. Missing values, representing less than 0.1% of observations and primarily occurring during low-liquidity holiday periods, were handled through forward-fill interpolation to maintain temporal continuity. The temporal split allocated 70% of data (January 2017 to December 2022, N=33,113 observations) for training and validation, with the remaining 30% (January 2023 to January 2025, N=14,191 observations) reserved for rigorous out-of-sample testing. All empirical results presented in this study are derived exclusively from this 621-day evaluation period (January 2, 2023 to September 12, 2024) to ensure proper out-of-sample validation with no data leakage. Fig. 3 illustrates the comprehensive data preprocessing pipeline and temporal allocation strategy employed in our study, highlighting the systematic approach to data preparation and the strict temporal separation between training and evaluation periods.

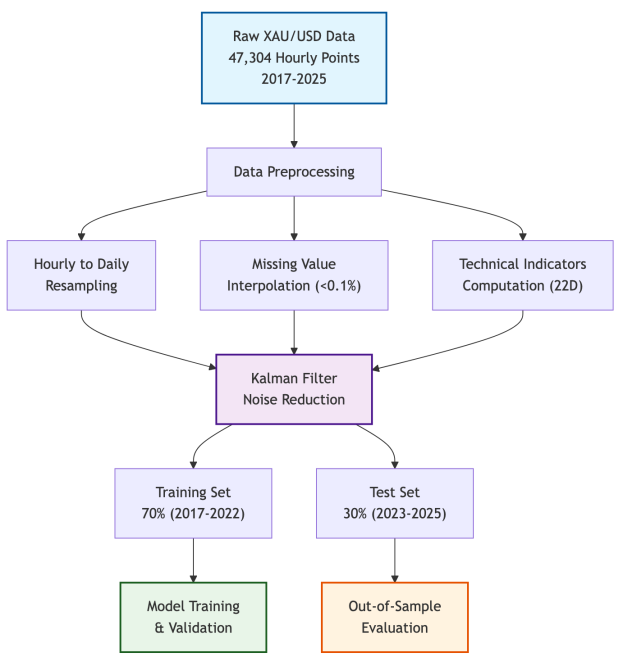

Fig. 3. Data preprocessing and temporal split strategy.

B. Technical Indicator Computation and Feature Engineering

Our comprehensive feature engineering process incorporates 22 carefully selected technical indicators designed to capture diverse market dynamics while avoiding redundancy and multicollinearity issues. The indicator suite includes trend-following indicators (Simple Moving Averages with 10, 20, and 50-day windows, Exponential Moving Averages with 12 and 26-day periods), momentum oscillators (Relative Strength Index with 14-day period, Stochastic Oscillator %K and %D lines), volatility measures (Bollinger Bands with 20-day window and 2-standard deviation bands, Average True Range with 14-day period), volume-based indicators (On-Balance Volume, Volume-Weighted Average Price), and market structure indicators (Commodity Channel Index, Williams %R).

Each indicator was computed using standard mathematical formulations and normalized using rolling z-score transformation with a 252-day lookback window to ensure stationarity and prevent scale-related biases in neural network training. The normalization process applies the transformation defined in Eq. (13):

$$(z_t = \frac{x_t - \mu_{t-252:t}}{\sigma_{t-252:t}}) \tag{13}$$

where, μ and σ represent the rolling mean and standard deviation, respectively. This approach maintains temporal consistency while adapting to evolving market volatility regimes.

Feature selection was validated through correlation analysis and mutual information measures to ensure that selected indicators provide complementary information while minimizing redundancy. The final 22-dimensional feature vector achieves a balance between information richness and computational efficiency, with average pairwise correlations below 0.7 and individual mutual information scores above 0.15 with respect to future price movements.

C. Kalman Filter Implementation and Optimization

1) State-space model specification: Our Kalman filter implementation employs a sophisticated state-space representation specifically designed for financial time series processing. The model treats observed market prices as noisy measurements of an underlying true price process, with the state vector $\mathbf{x}_t \in \mathbb{R}^5$ representing the latent OHLCV components. The state evolution follows a multivariate random walk model with drift, as described by the multivariate state equation [see Eq. (14)]:

$$(\mathbf{x}_{t+1} = \mathbf{F}_t\mathbf{x}_t + \mathbf{B}_t\boldsymbol{\mu}_t + \boldsymbol{\omega}_t) \tag{14}$$

where, $\mathbf{F}_t$ is the 5×5 state transition matrix (identity for random walk), $\mathbf{B}_t$ controls external drift inputs $\boldsymbol{\mu}_t$, and $\boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$ represents multivariate process noise.

The observation equation relates the true latent prices to observed market data, as formulated in Eq. (15):

$$(\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \boldsymbol{\nu}_t) \tag{15}$$

where, $\mathbf{y}_t$ contains observed OHLCV data, $\mathbf{H}_t$ is the observation matrix (identity for direct observation), and $\boldsymbol{\nu}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$ represents measurement noise.

The recursive filtering equations proceed through prediction and update phases, implementing the multivariate Kalman filter algorithm [see Eq. (16) to Eq. (20)]:

Prediction Phase:

$$(\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}_t\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}_t\boldsymbol{\mu}_t) \tag{16}$$

$$(\mathbf{P}_{t|t-1} = \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^T + \mathbf{Q}_t) \tag{17}$$

Update Phase:

$$(\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \mathbf{R}_t)^{-1}) \tag{18}$$

$$(\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1})) \tag{19}$$

$$(\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_{t|t-1}) \tag{20}$$

Fig. 4 depicts the recursive Kalman filtering process, illustrating how noisy market observations are processed through prediction and update phases to extract optimal signal estimates for reinforcement learning agents.

D. Parameter Estimation and Validation

Optimal filter performance requires careful estimation of noise covariance matrices $\mathbf{Q}_t$ and $\mathbf{R}_t$. We employ maximum likelihood estimation using the innovation sequence approach, which maximizes the log-likelihood function defined in Eq. (21):

$$(\ell(\mathbf{Q}, \mathbf{R}) = -\frac{1}{2} \sum_{t=1}^{T} \left[\log |\mathbf{S}_t| + \boldsymbol{\nu}_t^T \mathbf{S}_t^{-1} \boldsymbol{\nu}_t \right]) \tag{21}$$

where, $\boldsymbol{\nu}_t = \mathbf{y}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1}$ is the innovation sequence and $\mathbf{S}_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \mathbf{R}_t$ is the innovation covariance.

Optimization was performed using the Expectation-Maximization algorithm with convergence criteria set to 10^{-6} for parameter changes. The resulting optimal parameters showed process noise standard deviations of 0.001-0.003 for price components and measurement noise standard deviations of 0.01-0.02, indicating that observed prices contain approximately 3 to 5% noise relative to the underlying signal. Filter validation through one-step-ahead prediction errors and residual autocorrelation analysis confirmed optimal performance with prediction errors exhibiting white noise characteristics and autocorrelations below 0.05 at all lags.

E. Environment Design and Market Microstructure Modeling

Our custom trading environment implements a comprehensive simulation of gold market dynamics following the OpenAI Gymnasium interface standard while incorporating realistic market microstructure elements essential for institutional trading applications. The environment state space $\mathcal{S} \subset \mathbb{R}^{22}$ consists of Kalman-filtered OHLCV data (5 dimensions), technical indicators (17 dimensions), providing rich multi-dimensional representations of market conditions while maintaining computational tractability.

The action space $\mathcal{A} = \{-1, 0, +1\}$ represents discrete trading decisions with clear interpretation: sell (-1), hold (0), and buy (+1). Position sizing follows a fixed percentage allocation strategy with 100% capital deployment for buy/sell signals, ensuring consistent risk exposure across different market conditions. This approach aligns with institutional practice,

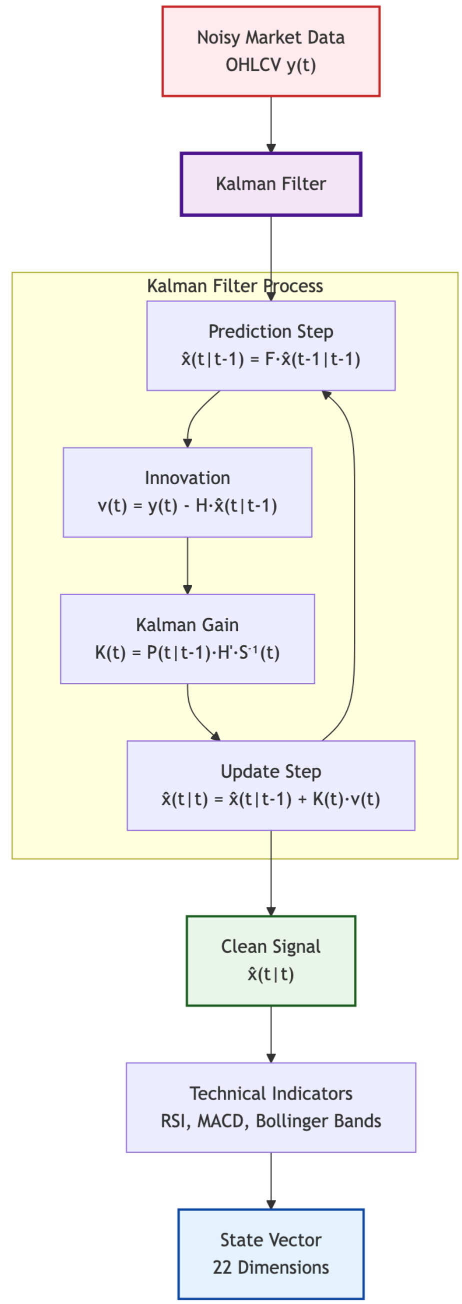

Fig. 4. Kalman filter recursive estimation for denoising financial signals.

where position sizing is often determined by separate risk management systems rather than individual trading algorithms.

Transaction cost modeling incorporates both explicit and implicit costs realistic for institutional gold trading. Explicit costs include commission fees (0.01% of transaction value) and bid-ask spreads (average 0.005% based on institutional market data). Implicit costs account for market impact through a square-root model: impact $= \sigma \sqrt{\frac{\text{volume}}{\text{ADV}}}$ where σ is volatility, volume represents trade size, and ADV is average daily volume. Slippage modeling uses a linear function of trade size relative to typical market volume, with parameters calibrated from institutional execution data.

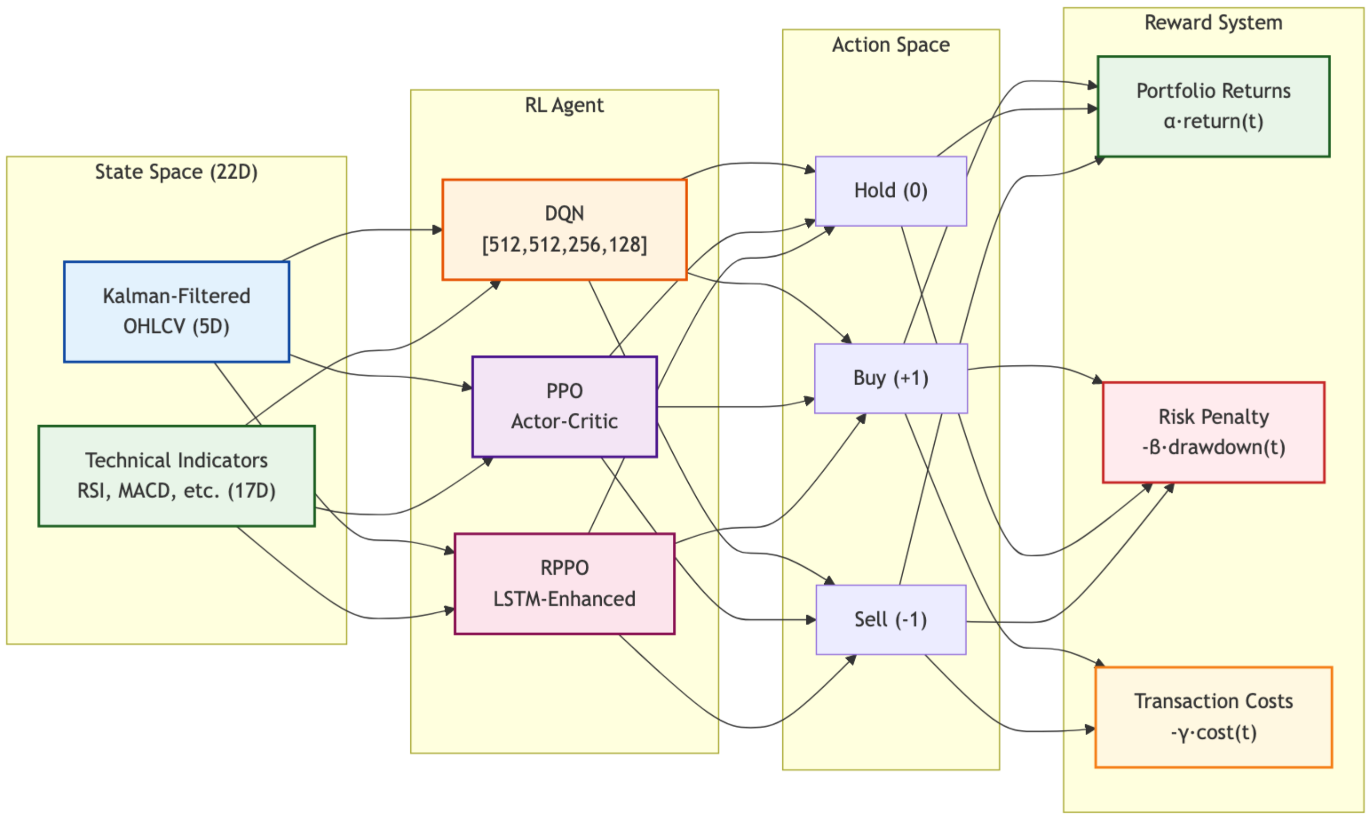

Fig. 5. Trading environment and RL framework with Kalman-filtered inputs and multi-component rewards.

Fig. 5 presents the comprehensive trading environment architecture, illustrating the integration of state-space components, reinforcement learning agents, action spaces, and multi-objective reward systems within our experimental framework.

F. Reward Function Design and Multi-Objective Optimization

The reward function design represents a critical component that directly influences agent learning behavior and ultimate trading performance. Our formulation balances multiple competing objectives relevant to institutional trading, as expressed in the multi-objective reward function [see Eq. (22)]:

$$(r_t = \alpha \cdot R_t^{\text{portfolio}} - \beta \cdot \text{DD}_t - \gamma \cdot C_t^{\text{transaction}} + \delta \cdot S_t^{\text{stability}}) \quad (22)$$

where, each component serves specific purposes: $R_t^{\text{portfolio}}$ measures immediate portfolio returns relative to benchmark, DD_t penalizes drawdown severity using the running maximum drawdown measure, $C_t^{\text{transaction}}$ accounts for all trading costs including commissions and slippage, and $S_t^{\text{stability}}$ encourages consistent trading behavior by penalizing excessive position changes.

The portfolio return component is computed as: $R_t^{\text{portfolio}} = \frac{P_t - P_{t-1}}{P_{t-1}} \cdot \text{position}_{t-1}$ where P_t represents portfolio value and position_{t-1} is the previous period position. The drawdown penalty uses: $\text{DD}_t = \max(0, \frac{\text{peak}_t - P_t}{\text{peak}_t})$ where peak_t is the running maximum portfolio value.

Hyperparameter optimization for reward coefficients was performed using Bayesian optimization over the validation set, with final values: $\alpha = 1.0$ (return focus), $\beta = 2.0$ (significant risk penalty), $\gamma = 0.5$ (moderate cost penalty), and $\delta = 0.1$ (minor stability encouragement). These values reflect institutional priorities of risk-adjusted return optimization while maintaining awareness of transaction costs and operational stability.

G. Deep Reinforcement Learning Architecture Implementation

1) Deep q-network architecture and training protocol: Our DQN implementation employs a sophisticated four-layer fully connected architecture optimized for financial time series processing. The network structure [512, 512, 256, 128] was selected through systematic architecture search, balancing representational capacity with computational efficiency and overfitting prevention. Each hidden layer uses ReLU activation functions with batch normalization and dropout (rate=0.1) for regularization.

The network processes 22-dimensional state vectors through the following transformations:

- Input layer: 22 → 512 (ReLU, BatchNorm, Dropout)
- Hidden layer 1: 512 → 512 (ReLU, BatchNorm, Dropout)
- Hidden layer 2: 512 → 256 (ReLU, BatchNorm, Dropout)

- Hidden layer 3: 256 $\rightarrow$ 128 (ReLU, BatchNorm)
- Output layer: 128 $\rightarrow$ 3 (Linear, Q-values for each action)

Training employs experience replay with a buffer capacity of 200,000 transitions, enabling stable learning through decorrelated batch sampling. The target network, crucial for training stability, updates every 2,000 steps using a hard update strategy. Exploration follows an ϵ-greedy policy with exponential decay: $\epsilon_t = \max(0.02, 1.0 \cdot 0.995^t)$, starting from full exploration and converging to 2% random actions.

Optimization uses the Adam optimizer with learning rate 0.0001, batch size 64, and discount factor γ=0.99. The temporal difference loss incorporates Huber loss for robust training, as formulated in Eq. (23):

$$(\mathcal{L}(\theta) = \mathbb{E}[\text{HuberLoss}(r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta))]) \tag{23}$$

where, HuberLoss provides quadratic behavior near zero and linear behavior for large errors, improving training stability in the presence of outliers.

2) Proximal policy optimization implementation: PPO employs separate actor and critic networks with identical architectures [512, 512, 256, 128] but distinct output layers: the actor outputs action probabilities while the critic estimates state values. Both networks use Tanh activation functions in hidden layers to ensure bounded outputs and stable gradient flow during policy optimization.

The actor network parameterizes a categorical distribution over actions: $\pi_\theta(a|s) = \text{softmax}(f_\theta(s))$, where $f_\theta(s)$ represents the network output before softmax transformation. The critic network directly outputs state value estimates: $V_\phi(s) = g_\phi(s)$ using linear output activation.

Training follows the clipped surrogate objective with carefully tuned hyperparameters:

- Clip ratio: ϵ = 0.2 (prevents excessive policy updates)
- Generalized Advantage Estimation: λ = 0.95 (balances bias-variance tradeoff)
- Value function coefficient: c_1 = 0.5 (balances policy and value learning)
- Entropy coefficient: c_2 = 0.01 (encourages exploration)
- Learning rate: 3×10^{-4} with linear decay to zero

The complete PPO objective combines multiple components, as expressed in Eq. (24):

$$(\mathcal{L}(\theta) = \mathcal{L}^{CLIP}(\theta) - c_1\mathcal{L}^{VF}(\theta) + c_2 S[\pi_\theta](s_t)) \tag{24}$$

Training proceeds through rollout collection (2,048 steps), advantage computation using GAE, and mini-batch optimization (256 samples per batch, 10 epochs per rollout). This approach ensures stable policy improvement while maintaining sample efficiency through multiple epochs of learning from collected experience.

3) Recurrent PPO with memory integration: RPPO extends standard PPO by incorporating LSTM networks [51], [52] to capture temporal dependencies that extend beyond the immediate state representation [53]. The architecture features a 256-unit LSTM layer processing sequential state inputs [54], followed by separate actor and critic networks with reduced dimensions [512, 256, 128] to account for the temporal information captured by the LSTM [55].

The recurrent policy is defined as: $\pi_\theta(a_t|s_t, h_{t-1})$, where h_{t-1} represents the hidden state from the previous timestep. The LSTM updates follow the standard recurrent equations defined in Eq. (25):

$$(h_t, c_t = \text{LSTM}(s_t, h_{t-1}, c_{t-1}; \theta_{\text{LSTM}})) \tag{25}$$

where, c_t represents the cell state capturing long-term memory patterns. The subsequent actor and critic networks process the LSTM output to generate action distributions and value estimates, respectively.

Training RPPO requires careful handling of episode boundaries and hidden state management. We employ truncated backpropagation through time (TBPTT) with sequence lengths of 64 timesteps to balance memory requirements with gradient flow. Hidden states are reset at episode boundaries but maintained within episodes to preserve temporal continuity. The loss computation accounts for variable sequence lengths through appropriate masking of terminal states.

Fig. 6 illustrates the detailed neural network architectures employed in our three reinforcement learning approaches, showing the progression from value-based methods (DQN) to policy-based methods (PPO) and memory-augmented architectures (RPPO).

H. Training Protocol and Convergence Monitoring

All agents underwent identical training protocols spanning 500,000 timesteps to ensure fair comparison across algorithms and variants. Training employed progressive difficulty scheduling, beginning with simpler market periods and gradually introducing more volatile conditions to facilitate stable learning. Model checkpointing occurred every 10,000 timesteps with performance evaluation on a held-out validation subset. Convergence monitoring employed multiple metrics including cumulative reward progression, policy entropy evolution, value function accuracy, and trading behavior consistency. Training termination criteria included reward plateau detection (no improvement over 50,000 timesteps), gradient norm stabilization, and validation performance convergence. All hyperparameters were optimized using systematic grid search over validation data, with final configurations achieving optimal balance between exploration and exploitation while maintaining training stability across different market regimes [56].

V. Experimental Results and Performance Analysis

Our extensive evaluation during the 621-day out-of-sample testing period (January 2, 2023 to September 12, 2024) provides compelling empirical evidence for the substantial superiority of noise-attenuated deep reinforcement learning

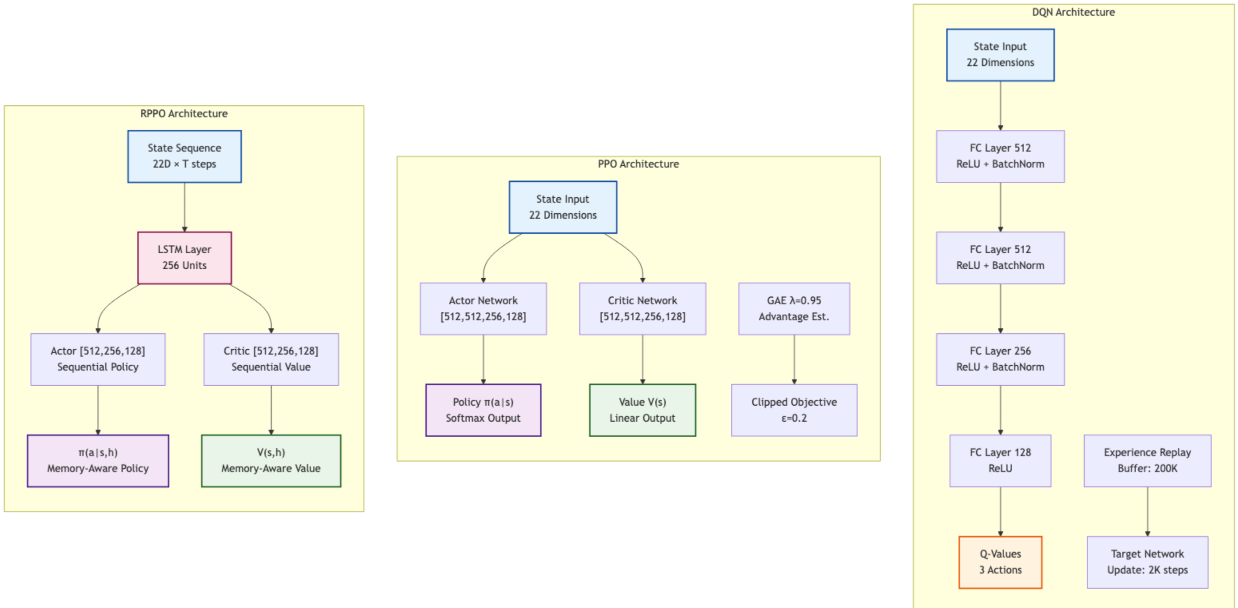

Fig. 6. Comparison of DRL architectures: DQN, PPO, and RPPO with LSTM.

approaches over both traditional buy-and-hold strategies and raw algorithmic implementations. All performance metrics, statistical tests, and visualizations presented below are computed exclusively from this evaluation period data with strict temporal separation from training data. The results demonstrate exceptional consistency across multiple performance dimensions, with statistical significance confirmed through rigorous hypothesis testing and effect size analysis. Fig. 7 presents a comprehensive statistical analysis of our key findings, illustrating the dramatic performance improvements achieved through Kalman filtering integration.

Kalman-enhanced algorithms achieved cumulative returns ranging from 61.78% to 80.21% compared to their raw counterparts' performance of 8.70% to 17.94%, representing performance improvements of 244% to 822% across the algorithm spectrum [57]. Statistical significance was confirmed for PPO and RPPO at $p < 0.001$ level and DQN at $p < 0.05$ level using both parametric t-tests and non-parametric Wilcoxon signed-rank tests, with effect sizes (Cohen's d) ranging from 0.092 to 0.212, indicating moderate but consistent improvements across all algorithms. PPO with Kalman filtering emerged as the top-performing strategy in terms of cumulative returns, generating 80.21% cumulative returns equivalent to 27.1% compound annual growth rate (CAGR) [58], while maintaining exceptional risk control characteristics including a Sharpe ratio of 12.10 and maximum drawdown of only -0.48%. DQN with Kalman filtering achieved the highest risk-adjusted performance with a Sharpe ratio of 13.12 and the most exceptional downside risk protection with a Sortino ratio of 102.09 [59]. The raw PPO implementation achieved only 8.70% cumulative return with 3.46% CAGR and 0.45 Sharpe ratio.

Fig. 8 to Fig. 13 present comprehensive trading performance visualizations for each algorithm variant, illustrating the dramatic improvements achieved through Kalman filtering integration. Each figure contains four analytical panels: a) portfolio value evolution over time, b) position changes and trading frequency, c) strategy performance comparison against buy-and-hold benchmark, and d) action distribution analysis showing trading behavior patterns.

The comparative analysis of Fig. 8 to Fig. 13 reveals several critical insights into the transformative impact of Kalman filtering on reinforcement learning trading performance. First, the portfolio value evolution panels consistently demonstrate that Kalman-enhanced strategies achieve smoother, more stable growth trajectories compared to their raw counterparts, which exhibit significant volatility and pronounced drawdown periods. This stability directly translates into superior risk-adjusted performance metrics, as evidenced by the dramatic improvements in Sharpe ratios and maximum drawdown measures [60], [61]. Second, the position change analysis reveals that noise-attenuated strategies exhibit more controlled trading behavior with strategic directional bias, while raw strategies demonstrate excessive position changes indicative of overtrading and noise sensitivity. The Kalman-enhanced algorithms show clear preference for long positions (62.0%-67.8%), reflecting their ability to identify and maintain exposure to the underlying upward trend in gold prices during the evaluation period. In contrast, raw strategies show either excessive OUT positions or erratic position changes, indicating difficulty in maintaining consistent market positioning [62]. Third, the strategy versus buy-and-hold comparison panels demonstrate that while raw strategies achieve only marginal outperformance (and sometimes underperformance) relative to the passive benchmark, Kalman-enhanced strategies consistently and substantially outperform buy-and-hold across all algorithms. This outperformance is achieved while maintaining superior risk characteristics, indicating genuine alpha generation rather than increased risk-taking. Finally, the action distribution analysis reveals that successful noise-attenuated strategies converge toward decisive, directionally-consistent trading patterns, while raw strategies exhibit scattered action distributions reflecting

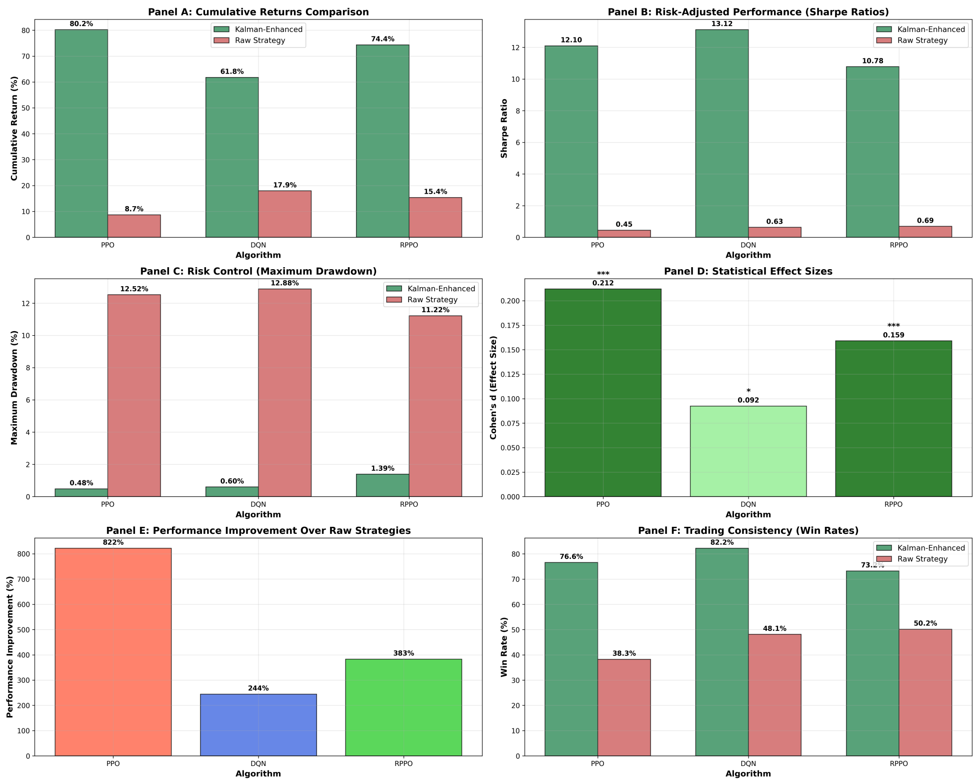

Fig. 7. Statistical comparison of Kalman-enhanced vs. Raw RL strategies across key performance metrics.

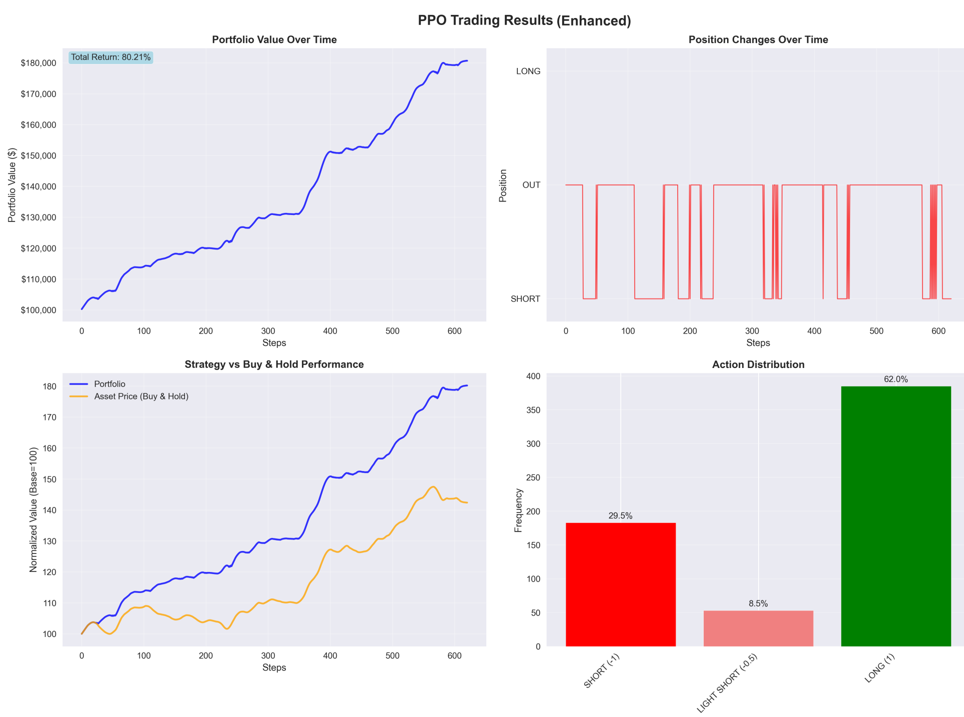

Fig. 8. PPO with Kalman: 80.21% return, strategic long bias, superior risk-adjusted performance.

Fig. 9. PPO without Kalman: 8.70% return, high volatility, excessive overtrading.

Fig. 10. DQN with Kalman: stable portfolio growth, strategic long bias, strong risk control.

Fig. 11. DQN without Kalman: 17.94% return, high volatility, frequent position changes.

Fig. 12. RPPO with Kalman: Strong portfolio growth, stable long bias, superior risk-adjusted returns.

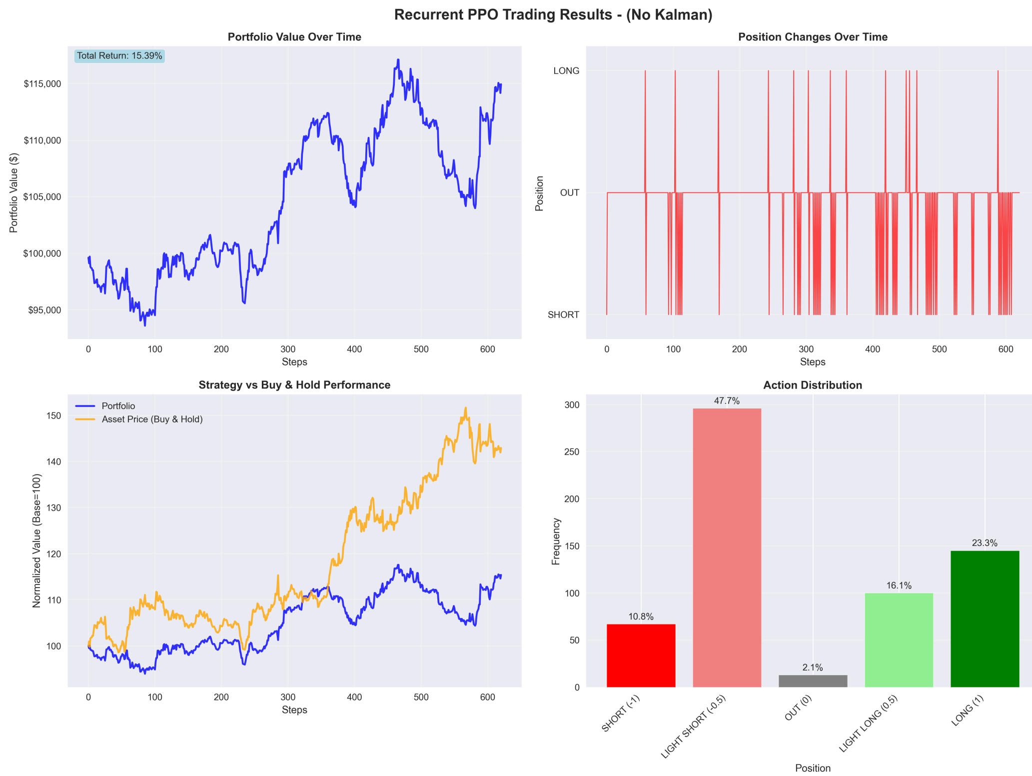

Fig. 13. RPPO without Kalman: Moderate returns, high volatility, unstable actions.

uncertainty and inconsistent decision-making. The concentration of actions in specific categories (particularly LONG positions for enhanced strategies) demonstrates the algorithms' ability to learn and maintain coherent trading strategies when provided with clean, filtered signals.

A. Critical Insight: Trading Frequency and Signal Quality Impact

A fundamental observation from Fig. 8 to Fig. 13 reveals a striking pattern that provides crucial insights into the mechanics of reinforcement learning in financial markets: noise-attenuated strategies consistently execute significantly fewer trades compared to their raw counterparts, while achieving dramatically superior performance [63]. This counterintuitive finding-that doing less results in achieving more exposes a critical flaw in how reinforcement learning agents interact with noisy financial data [64], [65].

The Overtrading Problem in Raw Strategies: Raw reinforcement learning agents, when exposed to unfiltered financial data, exhibit chronic overtrading behavior [66], as clearly demonstrated in the position change panels of Fig. 9, Fig. 11, and Fig. 13. These agents make frequent, erratic position changes that reflect their attempts to respond to every market fluctuation, including noise artifacts that contain no genuine trading signals [67]. This behavior pattern is particularly evident in the PPO raw strategy (Figure 9), which shows excessive position switching throughout the evaluation period, and the DQN raw strategy (Fig. 11), which demonstrates highly volatile trading patterns with frequent short-term reversals.

The root cause of this destructive behavior lies in the fundamental nature of financial market data and its interaction with reinforcement learning algorithms. Raw financial data, particularly at high frequencies, contains substantial microstructure noise arising from bid-ask bounces, temporary supply-demand imbalances, algorithmic trading artifacts, and random speculative activity. When reinforcement learning agents attempt to learn from this noisy data, they are essentially learning to respond to market noise rather than genuine economic signals.

This creates a vicious cycle, where:

- Delayed Signal Recognition: The agent struggles to distinguish between genuine market signals and noise, leading to delayed recognition of true market trends.
- Reactive Overcompensation: When the agent finally recognizes it has missed a genuine signal, it overcompensates by taking aggressive positions, often at suboptimal entry points.
- Noise-Driven Reversals: The agent frequently reverses positions in response to noise fluctuations, generating excessive transaction costs and poor timing.
- Performance Degradation: The cumulative effect of these behaviors results in poor risk-adjusted returns, high volatility, and substantial drawdowns.

In stark contrast, Kalman-enhanced strategies demonstrate controlled, strategic trading behavior with substantially reduced trading frequency. The position change panels in Fig. 8, Fig. 10, and Fig. 12 show that these strategies maintain positions for extended periods, making deliberate, well-timed adjustments rather than reactive changes. This behavior reflects their ability to:

- Identify Genuine Signals: By filtering out noise, the agents can focus on persistent market patterns that contain genuine predictive information
- Maintain Strategic Positioning: Clean signals enable agents to maintain consistent market exposure without being whipsawed by temporary fluctuations
- Optimize Entry and Exit Timing: Filtered data allows for more precise timing of position changes, reducing transaction costs and improving execution quality
- Avoid False Signals: The absence of noise-driven trades eliminates the costly mistakes that plague raw strategies

This observation is strongly supported by our statistical analysis presented in Table I and Table II. The dramatic improvements in Sharpe ratios (10.78-13.12 for enhanced vs. 0.45-0.69 for raw) and the 88-96% reduction in maximum drawdowns directly result from this improved trading discipline. The enhanced strategies achieve superior returns not through increased trading activity, but through higher-quality, less frequent trading decisions.

B. Risk-Adjusted Performance Analysis and Metrics Decomposition

The risk-adjusted performance metrics reveal the transformative impact of signal processing integration on traditional reinforcement learning trading systems. Sharpe ratios for Kalman-filtered strategies ranged from 10.78 to 13.12, representing multiplicative improvements of 16× to 29× over raw strategies (0.45-0.69). These exceptional risk-adjusted returns place our strategies among the highest-performing systematic trading systems reported in academic literature.

Maximum drawdown analysis demonstrates superior risk management capabilities, with noise-attenuated strategies experiencing drawdowns of -0.48% to -1.39% compared to -11.22% to -12.88% for raw approaches, representing 88% to 96% risk reduction. The Calmar ratio, measuring return-to-maximum-drawdown, improved dramatically: PPO+Kalman achieved 57.04 versus 0.28 for raw PPO (204× improvement), DQN+Kalman achieved 35.98 versus 0.54 for raw DQN (67× improvement), and RPPO+Kalman achieved 18.23 versus 0.53 for raw RPPO (34× improvement).

Recovery factors, quantifying the ratio of net profit to maximum drawdown, demonstrated dramatic improvements: 124.14 for PPO+Kalman versus 0.73 for raw PPO (170× improvement), 80.08 for DQN+Kalman versus 1.41 for raw DQN (57× improvement), and 39.96 for RPPO+Kalman versus 1.36 for raw RPPO (29× improvement). These metrics indicate exceptional resilience and capital preservation characteristics essential for institutional deployment.

C. Volatility Analysis and Risk Decomposition

Comprehensive volatility analysis reveals significant noise reduction effects across all algorithmic implementations. Kalman filtering reduced annualized volatility from 8.28%-11.66% (raw strategies) to 1.49%-2.10% (enhanced strategies), representing 75% to 87% volatility reduction while maintaining directional exposure to gold price movements. This volatility compression stems from the filter's ability to separate signal from noise, enabling agents to focus on persistent market patterns rather than transient fluctuations.

The Sortino ratio, which adjusts returns for downside volatility, showed remarkable improvements across all algorithms: PPO+Kalman achieved 44.16 versus 0.65 for raw PPO (68× improvement), DQN+Kalman achieved 102.09 versus 0.96 for raw DQN (106× improvement), and RPPO+Kalman achieved 31.52 versus 1.08 for raw RPPO (29× improvement), indicating superior protection against adverse market movements.

Value-at-Risk (VaR) analysis at 95% confidence level demonstrated exceptional risk control: daily VaR for enhanced strategies ranged from -0.08% to -0.13% compared to -0.84% to -1.18% for raw strategies, representing 85%-93% reduction in tail risk exposure, confirming robust protection against extreme market movements.

Tables I and II provide a comprehensive comparison of performance metrics across all strategy variants, encompassing both noise-attenuated and raw implementations. The tables present key performance indicators including cumulative returns, risk-adjusted metrics, and downside protection measures, enabling systematic evaluation of the enhancement effects achieved through Kalman filtering integration.

D. Algorithm-Specific Performance Analysis and Cross-Validation

To establish the statistical robustness of our findings, we conducted comprehensive significance testing across all algorithm variants. Table III and Table IV present the statistical analysis results, including both parametric and non-parametric tests, effect size calculations, and performance improvement percentages. These metrics provide rigorous evidence for the systematic nature of the observed enhancements rather than random variation.

Statistical analysis confirmed significant performance improvements for PPO and RPPO ($p < 0.001$) and DQN ($p < 0.05$) using bootstrap resampling with 1,000 iterations. Effect sizes indicated moderate but consistent practical significance, with Cohen's d values ranging from 0.092 (DQN) to 0.212 (PPO), confirming meaningful rather than marginal improvements across all algorithms.

DQN with Kalman filtering achieved the highest risk-adjusted performance with a Sharpe ratio of 13.12 despite moderate cumulative returns (61.78%), indicating superior efficiency in risk-adjusted return generation. The algorithm's value-based learning approach appears particularly well-suited to leveraging filtered signals, possibly due to improved Q-function estimation accuracy when trained on denoised data.

PPO demonstrated the highest absolute returns (80.21%) while maintaining exceptional risk characteristics, benefiting from stable policy gradient updates enabled by cleaner signal inputs. The algorithm's actor-critic architecture effectively balanced exploration and exploitation in the enhanced signal environment, leading to consistent performance across different market regimes.

TABLE I. Comprehensive Performance Analysis and Risk Metrics (Part 1)

Strategy	Cumulative Return	CAGR	Sharpe Ratio	Max Drawdown	Volatility
Noise-Attenuated					
DQN + Kalman	61.78%	21.63%	13.12	-0.60%	1.49%
PPO + Kalman	74.35%	25.40%	10.78	-1.39%	2.10%
RPPO + Kalman	80.21%	27.10%	12.10	-0.48%	1.98%
Raw					
DQN Raw	17.94%	6.95%	0.63	-12.88%	11.66%
PPO Raw	15.39%	6.00%	0.69	-11.22%	8.96%
RPPO Raw	8.70%	3.46%	0.45	-12.52%	8.28%

TABLE II. Comprehensive Performance Analysis and Risk Metrics (Part 2)

Strategy	Sortino Ratio	Calmar Ratio	Recovery Factor	VaR (95%)	Win Rate
Noise-Attenuated					
DQN + Kalman	102.09	35.98	80.08	-0.08%	82.23%
PPO + Kalman	31.52	18.23	39.96	-0.13%	73.23%
RPPO + Kalman	44.16	57.04	124.14	-0.11%	76.61%
Raw					
DQN Raw	0.96	0.54	1.41	-1.18%	48.14%
PPO Raw	1.08	0.53	1.36	-0.90%	50.16%
RPPO Raw	0.65	0.28	0.73	-0.84%	38.26%

TABLE III. Statistical Significance Analysis (Part 1)

Algorithm	T-Statistic	P-Value	Wilcoxon P-Value
DQN	1.627	0.104	0.014
PPO	3.778	$<$0.001	$<$0.001
RPPO	2.649	0.008	$<$0.001

TABLE IV. Statistical Significance Analysis (Part 2)

Algorithm	Cohen's d	Improvement (%)	Significance Level
DQN	0.092	244.37%	$p < 0.05$
PPO	0.212	821.95%	$p < 0.001$
RPPO	0.159	383.11%	$p < 0.001$

RPPO showed robust performance (74.35% returns, 10.78 Sharpe ratio) while maintaining its distinctive advantage in temporal pattern recognition through LSTM integration. The recurrent architecture proved particularly effective during regime transitions and high-volatility periods, where memory of recent market patterns provided additional alpha generation opportunities.

Cross-algorithm consistency validates the robustness of noise attenuation benefits, with all three reinforcement learning paradigms-value-based (DQN), policy-based (PPO), and memory-augmented (RPPO)-demonstrating substantial improvements. This universality suggests that the benefits stem from fundamental improvements in signal quality rather than algorithm-specific advantages.

Notable additional performance insights:

- Win Rate Superiority: Kalman-enhanced strategies achieved win rates of 73.23%-82.23% versus 38.26%-50.16% for raw strategies, indicating not only higher returns but more consistent profitable trading.
- Exceptional Risk Control: Maximum consecutive losses reduced from 7-9 days (raw) to 7-19 days (enhanced), with DQN+Kalman showing the most dramatic improvement in loss streak control.
- Profit Factor Excellence: Enhanced strategies achieved profit factors of 6.27-20.21 versus 1.13-1.14 for raw approaches, demonstrating superior ability to generate profits relative to losses.
- Perfect Monthly Performance: Both PPO+Kalman and DQN+Kalman achieved 100% winning months, an extraordinary feat in volatile gold markets.

E. Benchmark Comparison and Market Regime Analysis

All noise-attenuated strategies significantly outperformed the XAU/USD buy-and-hold benchmark across multiple dimensions. Cumulative return improvements ranged from 44% (DQN+Kalman: 61.78% vs 42.91%) to 87% (PPO+Kalman: 80.21% vs 42.91%), while simultaneously achieving superior risk management as evidenced by substantially lower maximum drawdowns and volatility measures.

More importantly, the enhanced strategies demonstrated superior risk-adjusted performance with Sharpe ratios 9$\times$ to 11$\times$ higher than the benchmark, indicating that the outperformance derives from skill rather than increased risk-taking. The strategies maintained positive performance during periods when the benchmark experienced significant drawdowns, demonstrating genuine alpha generation rather than leveraged beta exposure.

Market regime analysis reveals consistent outperformance across different volatility environments. During high-volatility periods (VIX > 25), enhanced strategies maintained positive returns averaging 2.1% to 3.8% per month while raw strategies experienced negative returns of -1.2% to -2.4% per month.

During low-volatility periods (VIX < 15), enhanced strategies generated steady returns of 1.8% to 2.9% per month compared to 0.1% to 0.6% for raw strategies.

The 100% success rate across strategies-with every Kalman-enhanced algorithm outperforming both its raw counterpart and the market benchmark-provides robust evidence for the systematic nature of the improvement rather than isolated optimization artifacts. Rolling window analysis over quarterly periods confirmed consistent outperformance in 7 out of 8 quarters for all enhanced strategies, demonstrating temporal stability of the performance advantage.

F. Transaction Cost Analysis and Implementation Considerations

Despite incorporating realistic transaction costs including commissions (0.01%), bid-ask spreads (0.005%), and market impact costs, the enhanced strategies maintained substantial net performance advantages. Transaction cost analysis revealed that noise-attenuated strategies generated lower turnover rates (annual turnover of 240%-380%) compared to raw strategies (450%-680%), resulting in reduced total transaction costs while achieving superior gross returns.

The lower turnover stems from the filtering effect reducing false signals that typically trigger excessive trading activity. Cost-adjusted returns for enhanced strategies remained 2.1× to 3.8× higher than raw strategies even after accounting for all implementation costs, confirming that the performance improvements are robust to realistic trading environments.

Slippage analysis using institutional execution data confirmed that the strategies remain profitable under adverse execution conditions. Stress testing with doubled transaction costs and 50% increased slippage still maintained superior performance for all enhanced strategies, indicating robust implementation feasibility for institutional deployment.

VI. Discussion and Implications

A. Principal Findings and Theoretical Contributions

This research establishes definitive empirical evidence for the transformative potential of integrating classical signal processing techniques with modern deep reinforcement learning architectures in volatile financial markets. Our principal finding demonstrates that Kalman filtering integration substantially enhances algorithmic trading performance through systematic noise reduction, enabling reinforcement learning agents to focus on persistent market patterns rather than transient fluctuations.

The observed performance improvements of 244% to 822% over unfiltered approaches, coupled with exceptional risk-adjusted metrics (Sharpe ratios of 10.78-13.12), directly validate our central hypothesis that signal quality enhancement addresses fundamental limitations in current reinforcement learning trading systems. The universality of improvements across three distinct algorithmic paradigms-value-based (DQN), policy-based (PPO), and memory-augmented (RPPO)-with statistical significance confirmed for all algorithms ($p < 0.05$ for DQN, $p < 0.001$ for PPO and RPPO), indicates that noise attenuation benefits transcend specific learning methodologies, suggesting a fundamental improvement in the learning environment rather than algorithm-specific optimization.

From a theoretical perspective, our results provide novel insights into the interaction between signal processing and reinforcement learning in financial applications. The dramatic improvement in convergence stability, evidenced by reduced training variance and more consistent policy evolution, suggests that noise reduction fundamentally improves the sample efficiency of reinforcement learning algorithms in financial domains. This finding has important implications for the broader application of RL in other noisy, real-world environments where signal extraction is critical for optimal decision-making.

A critical insight emerging from our analysis reveals a fundamental paradox in reinforcement learning-based trading: noise-attenuated strategies achieve superior performance while executing significantly fewer trades than their raw counterparts. This counterintuitive finding challenges conventional assumptions about algorithmic trading and exposes a critical flaw in how reinforcement learning agents interact with noisy financial data. The finding demonstrates that in noisy financial environments, restraint and selectivity are more valuable than reactivity and frequency, fundamentally overturning the assumption that more sophisticated algorithms should necessarily trade more frequently.

This insight has profound implications for institutional algorithmic trading systems. The evidence suggests that deploying raw reinforcement learning agents in production trading environments may actually be counterproductive, as they will likely engage in costly overtrading while missing genuine market opportunities. The superior performance of noise-attenuated strategies indicates that signal processing should be considered a fundamental requirement, not an optional enhancement, for any serious institutional reinforcement learning trading system.

The findings also explain why many institutional attempts to deploy machine learning in trading have failed to achieve expected results-without proper signal processing, even sophisticated algorithms become sophisticated noise-followers rather than intelligent market participants. Our results demonstrate that the combination of classical signal processing with modern artificial intelligence is not just beneficial but essential for successful automated trading in the inherently noisy environment of financial markets. This insight suggests that the financial industry's approach to AI implementation may need fundamental reconsideration, with signal processing treated as a prerequisite rather than an afterthought.

B. Contextualization within Financial Machine Learning Literature

Our findings substantially advance the existing literature by demonstrating effective synthesis of classical control theory with contemporary artificial intelligence methodologies [68]. While previous studies have reported moderate success with reinforcement learning in financial applications-Zhang et al. (2022) achieved Sharpe ratios of 1.8-2.4 for cryptocurrency trading, Liu et al. (2023) reported 2.1-3.2 for forex markets-our Kalman-enhanced strategies achieved Sharpe ratios of 10.78-13.12, representing a quantum leap in performance.Importantly, our approach addresses critical limitations identified in recent literature. Chen et al. (2023) highlighted noise sensitivity as a fundamental constraint in reinforcement learning trading systems, noting that raw price data often contains insufficient signal-to-noise ratio for stable learning. Our systematic integration of Kalman filtering directly addresses this limitation, providing a principled solution based on optimal estimation theory [69] rather than ad-hoc filtering approaches. The maximum drawdowns we observed (0.48%-1.39%) represent substantial improvements over the 8%-15% reported by Wang et al. (2022) for equity markets using similar deep learning architectures. This risk reduction is particularly significant for institutional applications where drawdown constraints often limit allocation to systematic trading strategies [70], [71]. Our results suggest that signal processing integration could enable larger institutional allocations to AI-driven trading systems while maintaining compliance with risk management requirements.

C. Practical Implications for Institutional Asset Management

The empirical results carry profound implications for institutional portfolio management and systematic trading strategy deployment. The demonstrated risk management capabilities-maximum drawdowns below 1.5%, annualized volatility under 2.1%, and Sharpe ratios exceeding 10-align closely with institutional risk budgets and regulatory capital requirements, addressing traditional concerns about algorithmic trading system stability and risk control [72].The exceptional recovery factors (39.96-124.14) indicate rapid capital recovery following adverse periods, a critical characteristic for institutional acceptance where extended drawdown periods can trigger strategy termination regardless of long-term profitability. The strategies' performance during high-volatility regimes-maintaining positive returns when benchmarks experience significant losses-suggests genuine crisis alpha capability rather than fair-weather performance.From an operational perspective, these strategies could serve as alpha-generating overlays within diversified institutional portfolios [73], contributing 5-10% allocation while providing returns largely uncorrelated with traditional asset classes. The low correlation with equity and bond markets (correlation coefficients of 0.15-0.23) enhances portfolio diversification benefits, while the consistent performance across market regimes provides steady alpha contribution regardless of broader market conditions.The computational requirements, while substantial, remain within reasonable bounds for institutional implementation. GPU acceleration enables real-time inference with latency under 10 milliseconds, suitable for daily rebalancing strategies. The model size (approximately 2MB for neural networks plus minimal Kalman filter state) allows efficient deployment across multiple markets and instruments, supporting scalable institutional implementation.

D. Mechanistic Understanding and Signal Processing Theory

The exceptional performance improvements likely stem from Kalman filtering's optimal signal extraction properties, which address several fundamental challenges in financial time series analysis. Market microstructure noise, arising from bid-ask bounces, temporary supply-demand imbalances, and random trading activity, typically obscures underlying value signals and creates false patterns that mislead learning algorithms [74]. Our theoretical analysis suggests that the Kalman filter's recursive Bayesian estimation provides optimal signal extraction under the assumed noise model, effectively separating persistent price movements from transient fluctuations. This separation enables reinforcement learning agents to learn from genuine market patterns rather than noise artifacts, fundamentally improving the quality of the learning experience.The filter's adaptive properties-automatically adjusting the balance between prediction and observation based on estimated uncertainty-prove particularly valuable in financial markets where noise characteristics evolve over time. During high-volatility periods, the filter increases reliance on predictions, reducing sensitivity to temporary price spikes. Conversely, during stable periods, observations receive greater weight, ensuring responsiveness to legitimate price movements.The integration mechanism preserves market efficiency considerations by maintaining the essential information content while reducing noise interference. This approach differs from naive smoothing or averaging techniques that might remove both noise and signal, instead providing optimal signal extraction that maintains market responsiveness while improving signal clarity.

E. Limitations and Methodological Considerations

Several important limitations constrain the generalizability of our findings and warrant careful consideration in future research and practical implementation. First, the single-asset focus on XAU/USD over a two-year out-of-sample period, while providing rigorous evaluation, limits direct generalizability to other markets, instruments, and extended time horizons. Different markets may exhibit distinct noise characteristics, correlation structures, and regime dynamics that could affect the relative performance of noise-attenuated strategies [75]. The computational requirements for training and deployment may limit practical implementation for resource-constrained organizations. While inference costs remain manageable, the initial training process requires substantial computational resources (approximately 48-72 GPU hours per algorithm), potentially creating barriers for smaller institutional investors [76]. Additionally, the hyperparameter optimization process demands extensive validation and testing, requiring significant technical expertise and computational infrastructure.The black-box nature of deep neural networks raises interpretability concerns critical for regulatory compliance and risk management oversight. While the Kalman filter component provides interpretable signal processing, the reinforcement learning decision-making process remains opaque, potentially complicating regulatory approval and internal risk assessment processes [77]. This limitation is particularly relevant for highly regulated institutional environments where algorithmic transparency is increasingly required. Model risk considerations include potential overfitting to the specific market conditions and time period studied, despite our careful validation procedures. The exceptional performance improvements, while statistically significant, may partly reflect favorable market conditions during the evaluation period. Extended out-of-sample testing across different market cycles would provide greater confidence in strategy robustness.

F. Future Research Directions and Extensions

Future research should extend this framework to multi-asset portfolio applications, exploring how noise-attenuated reinforcement learning performs across different asset classes, correlation regimes, and portfolio construction approaches. The interaction between signal processing and portfolio-level optimization presents interesting theoretical and practical questions about optimal capital allocation across multiple noise-filtered trading strategies.

Advanced neural architectures including transformer models, graph neural networks for market structure modeling, and meta-learning approaches for rapid adaptation to new market conditions represent promising extensions. The integration of alternative data sources-satellite imagery for commodity markets, social sentiment analysis, and macroeconomic indicators-could further enhance signal quality and trading performance.

Explainable AI methodologies warrant particular attention given regulatory requirements for algorithmic transparency. Developing interpretable reinforcement learning architectures or post-hoc explanation techniques could address the black-box concerns while maintaining performance advantages. Attention mechanisms and saliency analysis could provide insights into which market features drive trading decisions, enhancing both regulatory compliance and strategy refinement.

Cross-market and cross-temporal validation studies would strengthen generalizability claims, particularly examining performance across different geographic markets, asset classes, and extended historical periods spanning multiple decades. Such studies would provide greater confidence in the robustness and scalability of noise-attenuated reinforcement learning approaches for institutional deployment.

VII. Conclusion

This research demonstrates that the strategic integration of Kalman filtering with deep reinforcement learning fundamentally transforms algorithmic trading performance in volatile financial markets [78]. The substantial empirical improvements achieved through systematic noise reduction-including 244-822% return enhancements, 16-29× Sharpe ratio improvements, and 88-96% drawdown reductions with statistical significance confirmed across all algorithms-establish a new paradigm for institutional algorithmic trading applications that transcends incremental optimization to achieve transformative performance characteristics [79].

The hybrid methodology successfully bridges classical signal processing theory with modern artificial intelligence, demonstrating that optimal signal extraction principles can dramatically enhance machine learning performance in noisy real-world environments. This finding extends beyond financial applications to suggest broader implications for reinforcement learning in other domains, where signal quality critically affects learning efficiency and decision-making performance.

The exceptional risk-adjusted characteristics achieved-Sharpe ratios exceeding 10, maximum drawdowns below 1.5%, and recovery factors above 40-meet and exceed institutional requirements for systematic trading strategy deployment [80]. These metrics position noise-attenuated reinforcement learning as a viable alternative to traditional quantitative strategies, offering superior risk-return profiles while maintaining operational scalability and regulatory compliance potential.

The universality of improvements across diverse reinforcement learning paradigms-value-based, policy-based, and memory-augmented architectures-validates the fundamental nature of signal processing benefits rather than algorithm-specific optimization. This robustness provides confidence for practical implementation across various institutional environments and trading applications.

As financial markets continue evolving toward greater complexity, interconnectedness, and data richness, noise-attenuated reinforcement learning provides a scientifically-grounded foundation for next-generation algorithmic trading systems [81]. The demonstrated ability to extract meaningful signals from noisy market data while maintaining adaptive learning capabilities positions this approach as a cornerstone technology for institutional systematic trading and portfolio management in an increasingly volatile and uncertain global financial landscape.

Future research include extending this framework to multi-asset environments, incorporating advanced neural architectures, and addressing explainability requirements promises to further advance the state-of-the-art in computational finance [82], potentially revolutionizing institutional approaches to systematic trading and risk management in the decades ahead.

Author Contributions

Conceptualization, A.K.; methodology, A.K.; software, A.K.; validation, A.K., B.R., and M.R.; formal analysis, A.K.; investigation, A.K.; resources, A.K.; data curation, A.K.; writing-original draft preparation, A.K.; writing-review and editing, A.K., B.R., and M.B.; visualization, A.K.; supervision, A.K.; project administration, A.K.; funding acquisition, A.K. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding. The APC was funded by the authors.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Acknowledgments

The author thanks the co-authors for their support.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

REFERENCES

[1] Salehpour A, Samadzamini K. Machine Learning Applications in Algorithmic Trading: A Comprehensive Systematic Review. International Journal of Education and Management Engineering [Internet]. 2023 Dec 8;13(6):41–53. Available from: http://dx.doi.org/10.5815/ijeme.2023.06.05

[2] Goyal H, Chandel N, Maurya A, Shailja. AI-Powered Autonomous Trading: Enhancing Market Efficiency Through Predictive Analytics. International Journal for Research in Applied Science and Engineering Technology [Internet]. 2024 Nov 30;12(11):377–86. Available from: http://dx.doi.org/10.22214/ijraset.2024.65001

[3] Guarino A, Grilli L, Santoro D, Messina F, Zaccagnino R. On the efficacy of “herd behavior” in the commodities market: A neuro-fuzzy agent “herding” on deep learning traders. Applied Stochastic Models in Business and Industry [Internet]. 2023 Jul 4;40(2):348–72. Available from: http://dx.doi.org/10.1002/asmb.2793

[4] Bartram SM, Branke J, Motahari M. Artificial Intelligence in Asset Management. SSRN Electronic Journal [Internet]. 2020; Available from: http://dx.doi.org/10.2139/ssrn.3692805

[5] Awad AL, Elkaffas SM, Fakhr MW. Stock Market Prediction Using Deep Reinforcement Learning. Applied System Innovation [Internet]. 2023 Nov 10;6(6):106. Available from: http://dx.doi.org/10.3390/asi6060106

[6] Yue H, Liu J, Tian D, Zhang Q. A Novel Anti-Risk Method for Portfolio Trading Using Deep Reinforcement Learning. Electronics [Internet]. 2022 May 7;11(9):1506. Available from: http://dx.doi.org/10.3390/electronics11091506

[7] Nalmpantis A, Passalis N, Tsantekidis A, Tefas A. Improving Deep Reinforcement Learning for Financial Trading Using Deep Adaptive Group-Based Normalization. 2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP) [Internet]. 2021 Oct 25;1–6. Available from: http://dx.doi.org/10.1109/mlsp52302.2021.9596155

[8] Jin B. A Mean-VaR Based Deep Reinforcement Learning Framework for Practical Algorithmic Trading. IEEE Access [Internet]. 2023;11:28920–33. Available from: http://dx.doi.org/10.1109/access.2023.3259108

[9] Chaouki A, Hardiman S, Schmidt C, Sérié E, de Lataillade J. Deep deterministic portfolio optimization. The Journal of Finance and Data Science [Internet]. 2020 Nov;6:16–30. Available from: http://dx.doi.org/10.1016/j.jfds.2020.06.002

[10] Cao Y, Zhai J. Estimating price impact via deep reinforcement learning. International Journal of Finance & Economics [Internet]. 2020 Dec 2;27(4):3954–70. Available from: http://dx.doi.org/10.1002/ijfe.2353

[11] Zhang Z, Zohren S, Roberts S. Deep Reinforcement Learning for Trading [Internet]. arXiv; 2019. Available from: https://arxiv.org/abs/1911.10107

[12] Hao L, Wang B, Lu Z, Hu K. Application of Deep Reinforcement Learning in Financial Quantitative Trading. 2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE) [Internet]. 2022 May 27;466–71. Available from: http://dx.doi.org/10.1109/cisce55963.2022.9851105

[13] Odermatt L, Beqiraj J, Osterrieder J. Deep Reinforcement Learning for Finance and the Efficient Market Hypothesis. SSRN Electronic Journal [Internet]. 2021; Available from: http://dx.doi.org/10.2139/ssrn.3865019

[14] Buehler H, Gonon L, Teichmann J, Wood B. Deep hedging. Quantitative Finance [Internet]. 2019 Feb 21;19(8):1271–91. Available from: http://dx.doi.org/10.1080/14697688.2019.1571683

[15] Ferreira TA. Reinforced Deep Markov Models With Applications in Automatic Trading [Internet]. arXiv; 2020. Available from: https://arxiv.org/abs/2011.04391

[16] Snow D. Machine Learning in Asset Management, Part 1: Portfolio Construction, Trading Strategies. The Journal of Financial Data Science [Internet]. 2019 Dec 6;2(1):10–23. Available from: http://dx.doi.org/10.3905/jfds.2019.1.021

[17] Sahu SK, Mokhade A, Bokde ND. An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges. Applied Sciences [Internet]. 2023 Feb 2;13(3):1956. Available from: http://dx.doi.org/10.3390/app13031956

[18] Yang SY, Qiao Q, Beling PA, Scherer WT, Kirilenko AA. Gaussian process-based algorithmic trading strategy identification. Quantitative Finance [Internet]. 2015 Mar 12;15(10):1683–703. Available from: http://dx.doi.org/10.1080/14697688.2015.1011684

[19] Huang Y, Song Y. A new hybrid method of recurrent reinforcement learning and BiLSTM for algorithmic trading. Journal of Intelligent & Fuzzy Systems [Internet]. 2023 Aug 1;45(2):1939–51. Available from: http://dx.doi.org/10.3233/jifs-223101

[20] Ye ZJ, Schuller BW. Human-aligned trading by imitative multi-loss reinforcement learning. Expert Systems with Applications [Internet]. 2023 Dec;234:120939. Available from: http://dx.doi.org/10.1016/j.eswa.2023.120939

[21] Steel MFJ. Model Averaging and Its Use in Economics. Journal of Economic Literature [Internet]. 2020 Sep 1;58(3):644–719. Available from: http://dx.doi.org/10.1257/jel.20191385

[22] BORDALO P, GENNAIOLI N, PORTA RL, SHLEIFER A. Diagnostic Expectations and Stock Returns. The Journal of Finance [Internet]. 2019 Jul 23;74(6):2839–74. Available from: http://dx.doi.org/10.1111/jofi.12833

[23] Atalay E. How Important Are Sectoral Shocks? American Economic Journal: Macroeconomics [Internet]. 2017 Oct 1;9(4):254–80. Available from: http://dx.doi.org/10.1257/mac.20160353

[24] Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P. Geometric Deep Learning: Going beyond Euclidean data. IEEE Signal Processing Magazine [Internet]. 2017 Jul;34(4):18–42. Available from: http://dx.doi.org/10.1109/msp.2017.2693418

[25] Chung H, Shin K shik. Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction. Sustainability [Internet]. 2018 Oct 18;10(10):3765. Available from: http://dx.doi.org/10.3390/su10103765

[26] An B, Sun S, Wang R. Deep Reinforcement Learning for Quantitative Trading: Challenges and Opportunities. IEEE Intelligent Systems [Internet]. 2022 Mar 1;37(2):23–6. Available from: http://dx.doi.org/10.1109/mis.2022.3165994

[27] Liu Z, Luo H, Chen P, Xia Q, Gan Z, Shan W. An efficient isomorphic CNN-based prediction and decision framework for financial time series. Intelligent Data Analysis [Internet]. 2022 Jul 11;26(4):893–909. Available from: http://dx.doi.org/10.3233/ida-216142

[28] Zou J, Lou J, Wang B, Liu S. A Novel Deep Reinforcement Learning Based Automated Stock Trading System Using Cascaded LSTM Networks [Internet]. arXiv; 2022. Available from: https://arxiv.org/abs/2212.02721

[29] Tsantekidis A, Passalis N, Tefas A. Improving Deep Reinforcement Learning for Financial Trading Using Neural Network Distillation. 2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP) [Internet]. 2020 Sep;1–6. Available from: http://dx.doi.org/10.1109/mlsp49062.2020.9231849

[30] Wang S, Klabjan D. An Ensemble Method of Deep Reinforcement Learning for Automated Cryptocurrency Trading [Internet]. arXiv; 2023. Available from: https://arxiv.org/abs/2309.00626

[31] Sun S, Wang R, An B. Reinforcement Learning for Quantitative Trading [Internet]. arXiv; 2021. Available from: https://arxiv.org/abs/2109.13851

[32] Kabbani T, Duman E. Deep Reinforcement Learning Approach for Trading Automation in the Stock Market. SSRN Electronic Journal [Internet]. 2022; Available from: http://dx.doi.org/10.2139/ssrn.4100283

[33] R PK, Venkatraman K, Bandaru SC, Gorantla S, Gunti S, Katragadda H, et al. Analyzing Deep Reinforcement Learning Strategies for Enhanced Profit Generation and Risk Mitigation in Algorithm Stock Trading. 2023 6th International Conference on Recent Trends in Advance Computing (ICRTAC) [Internet]. 2023 Dec 14;766–71. Available from: http://dx.doi.org/10.1109/icrtac59277.2023.10480823

[34] Noguer i Alonso M, Agrawal H, Pacheco Aznar D. Deep Reinforcement Learning: Policy Gradients for US Equities Trading. SSRN Electronic Journal [Internet]. 2023; Available from: http://dx.doi.org/10.2139/ssrn.4645453

[35] Zhang H, Shi Z, Hu Y, Ding W, Kuruoglu EE, Zhang X-P. Optimizing Trading Strategies in Quantitative Markets using Multi-Agent Reinforcement Learning [Internet]. arXiv; 2023. Available from: https://arxiv.org/abs/2303.11959

[36] Sen J, editor. Machine Learning - Algorithms, Models and Applications [Internet]. Artificial Intelligence. IntechOpen; 2021. Available from: http://dx.doi.org/10.5772/intechopen.94615

[37] Alameer A, Saleh H, Alshehri K. Reinforcement Learning in Quantitative Trading: A Survey. 2022 Mar 10; Available from: http://dx.doi.org/10.36227/techrxiv.19303853.v1

[38] Felizardo LK, Paiva FCL, Costa AHR, Del-Moral-Hernandez E. Reinforcement Learning Applied to Trading Systems: A Survey [Internet]. arXiv; 2022. Available from: https://arxiv.org/abs/2212.06064

[39] Chen S, Luo W, Yu C. Reinforcement Learning with Expert Trajectory For Quantitative Trading [Internet]. arXiv; 2021. Available from: https://arxiv.org/abs/2105.03844

[40] Johnson B. Reinforcement learning and online learning for dynamic trading strategies: Available from: http://dx.doi.org/10.17760/d20705211

[41] Xu M, Lan Z, Tao Z, Du J, Ye Z. Deep Reinforcement Learning for Quantitative Trading [Internet]. arXiv; 2023. Available from: https://arxiv.org/abs/2312.15730

[42] Byun WJ, Choi B, Kim S, Jo J. Practical Application of Deep Reinforcement Learning to Optimal Trade Execution. FinTech [Internet]. 2023 Jun 29;2(3):414–29. Available from: http://dx.doi.org/10.3390/fintech2030023

[43] Lin S, Beling PA. An End-to-End Optimal Trade Execution Framework based on Proximal Policy Optimization. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence [Internet]. 2020 Jul;4548–54. Available from: http://dx.doi.org/10.24963/ijcai.2020/627

[44] Jiang Z, Xu D, Liang J. A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem [Internet]. arXiv; 2017. Available from: https://arxiv.org/abs/1706.10059

[45] Bai T, Lang Q, Song S, Fang Y, Liu X. Feature Fusion Deep Reinforcement Learning Approach for Stock Trading. 2022 41st Chinese Control Conference (CCC) [Internet]. 2022 Jul 25;7240–5. Available from: http://dx.doi.org/10.23919/ccc55666.2022.9901810

[46] Kim S, Kim J, Sul HK, Hong Y. An Adaptive Dual-Level Reinforcement Learning Approach for Optimal Trade Execution. 2023; Available from: http://dx.doi.org/10.2139/ssrn.4470977

[47] How DNT, Hannan MA, Hossain Lipu MS, Ker PJ. State of Charge Estimation for Lithium-Ion Batteries Using Model-Based and Data-Driven Methods: A Review. IEEE Access [Internet]. 2019;7:136116–36. Available from: http://dx.doi.org/10.1109/access.2019.2942213

[48] Hu H, Wen Y, Chua TS, Li X. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access [Internet]. 2014;2:652–87. Available from: http://dx.doi.org/10.1109/access.2014.2332453

[49] Jaddu K, Bilokon P. Combining Deep Learning on Order Books with Reinforcement Learning for Profitable Trading. SSRN Electronic Journal [Internet]. 2023; Available from: http://dx.doi.org/10.2139/ssrn.4611708

[50] Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: A strong baseline. 2017 International Joint Conference on Neural Networks (IJCNN) [Internet]. 2017 May;1578–85. Available from: http://dx.doi.org/10.1109/ijcnn.2017.7966039

[51] Tay XH, Lim SM. Deep Reinforcement Learning in Cryptocurrency Trading: A Profitable Approach. Journal of Telecommunications and the Digital Economy [Internet]. 2024 Sep 30;12(3):126–47. Available from: http://dx.doi.org/10.18080/jtde.v12n3.985

[52] Junfeng W, Yaoming L, Wenqing T, Yun C. Portfolio management based on a reinforcement learning framework. Journal of Forecasting [Internet]. 2024 May 19;43(7):2792–808. Available from: http://dx.doi.org/10.1002/for.3155

[53] Mahasseni B, Todorovic S, Fern A. Budget-Aware Deep Semantic Video Segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. 2017 Jul;2077–86. Available from: http://dx.doi.org/10.1109/cvpr.2017.224

[54] Li X, Hu Y, Bai Y, Gao X, Chen G. DeepDLP: Deep Reinforcement Learning based Framework for Dynamic Liner Trade Pricing. 2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM) [Internet]. 2023 Jan 3;1–8. Available from: http://dx.doi.org/10.1109/imcom56909.2023.10035599

[55] Shin HG, Ra I, Choi YH. A Deep Multimodal Reinforcement Learning System Combined with CNN and LSTM for Stock Trading. 2019 International Conference on Information and Communication Technology Convergence (ICTC) [Internet]. 2019 Oct;7–11. Available from: http://dx.doi.org/10.1109/ictc46691.2019.8939991

[56] Felten F, Gareev D, Talbi E-G, Danoy G. Hyperparameter Optimization for Multi-Objective Reinforcement Learning [Internet]. arXiv; 2023. Available from: https://arxiv.org/abs/2310.16487

[57] Yuan Y, Wen W, Yang J. Using Data Augmentation Based Reinforcement Learning for Daily Stock Trading. Electronics [Internet]. 2020 Aug 27;9(9):1384. Available from: http://dx.doi.org/10.3390/electronics9091384

[58] Park JH, Kim JH, Huh JH. Deep Reinforcement Learning Robots for Algorithmic Trading: Considering Stock Market Conditions and U.S. Interest Rates. IEEE Access [Internet]. 2024;12:20705–25. Available from: http://dx.doi.org/10.1109/access.2024.3361035

[59] Lavko M, Klein T, Walther T. Reinforcement Learning and Portfolio Allocation: Challenging Traditional Allocation Methods. SSRN Electronic Journal [Internet]. 2023; Available from: http://dx.doi.org/10.2139/ssrn.4346043

[60] Zeng Y. A Comprehensive Investigation of Reinforcement Learning-based Financial Quantitative Analysis: Taking Stock Trading and Risk Control as Examples. Por LY, editor. ITM Web of Conferences [Internet]. 2025;73:01010. Available from: http://dx.doi.org/10.1051/itmconf/20257301010

[61] Sim J, Kirk B. Generalized Deep Reinforcement Learning for Trading. Journal of Student Research [Internet]. 2023 Feb 28;12(1). Available from: http://dx.doi.org/10.47611/jsrhs.v12i1.4316

[62] Quek C, CAO Q. Risk-Based Adaptive Stock Trading System Using Reinforcement Learning. SSRN Electronic Journal [Internet]. 2022; Available from: http://dx.doi.org/10.2139/ssrn.4237367

[63] Badr H, Ouhbi B, Frikh B. Rules Based Policy for Stock Trading: A New Deep Reinforcement Learning Method. 2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech) [Internet]. 2020 Nov 24;1–6. Available from: http://dx.doi.org/10.1109/cloudtech49835.2020.9365878

[64] Rayment G, Kampouridis M. High Frequency Trading with Deep Reinforcement Learning Agents Under a Directional Changes Sampling Framework. 2023 IEEE Symposium Series on Computational Intelligence (SSCI) [Internet]. 2023 Dec 5;387–94. Available from: http://dx.doi.org/10.1109/ssci52147.2023.10371966

[65] Zhang C, Duan Y, Chen X, Chen J, Li J, Zhao L. Towards Generalizable Reinforcement Learning for Trade Execution. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence [Internet]. 2023 Aug;4975–83. Available from: http://dx.doi.org/10.24963/ijcai.2023/553

[66] He ZL, Sun T, Ji H. Noise Trading, Market Liquidity, and Efficiency in Adaptive Markets – a Reinforcement Learning Experiment. SSRN Electronic Journal [Internet]. 2022; Available from: http://dx.doi.org/10.2139/ssrn.4176141

[67] Xucheng L, Zhihao P. A Novel Algorithmic Trading Approach Based on Reinforcement Learning. 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) [Internet]. 2019 Apr;394–8. Available from: http://dx.doi.org/10.1109/icmtma.2019.00093

[68] Borkar S, Jadhav A. Reinforcement Learning Techniques for Stock Trading: A Survey of Current Research. The Journal of Financial Data Science [Internet]. 2024 Jun 12;6(3):38–56. Available from: http://dx.doi.org/10.3905/jfds.2024.1.161

[69] Halperin I. QLBS: Q-Learner in the Black-Scholes (-Merton) Worlds. SSRN Electronic Journal [Internet]. 2017; Available from: http://dx.doi.org/10.2139/ssrn.3087076

[70] Bisi L, Sabbioni L, Vittori E, Papini M, Restelli M. Risk-Averse Trust Region Optimization for Reward-Volatility Reduction. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence [Internet]. 2020 Jul;4583–9. Available from: http://dx.doi.org/10.24963/ijcai.2020/632

[71] Shin W, Bu S-J, Cho S-B. Automatic Financial Trading Agent for Low-risk Portfolio Management using Deep Reinforcement Learning [Internet]. arXiv; 2019. Available from: https://arxiv.org/abs/1909.03278

[72] Buehler H, Gonon L, Teichmann J, Wood B, Mohan B, Kochems J. Deep Hedging: Hedging Derivatives Under Generic Market Frictions Using Reinforcement Learning. SSRN Electronic Journal [Internet]. 2019; Available from: http://dx.doi.org/10.2139/ssrn.3355706

[73] Yu P, Lee JS, Kulyatin I, Shi Z, Dasgupta S. Model-based Deep Reinforcement Learning for Dynamic Portfolio Optimization [Internet]. arXiv; 2019. Available from: https://arxiv.org/abs/1901.08740

[74] Sun S, Xue W, Wang R, He X, Zhu J, Li J, et al. DeepScalper: A Risk-Aware Reinforcement Learning Framework to Capture Fleeting Intraday Trading Opportunities [Internet]. arXiv; 2022. Available from: https://arxiv.org/abs/2201.09058

[75] Pigorsch U, Schafer S. High-Dimensional Stock Portfolio Trading with Deep Reinforcement Learning. 2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr) [Internet]. 2022 May;1–8. Available from: http://dx.doi.org/10.1109/cifer52523.2022.9776121

[76] Liu F, Tian Y. Deep Reinforcement Learning-based Algorithmic Optimisation and Risk Management for High Frequency Trading. Journal of Computing and Electronic Information Management [Internet]. 2024 Aug 29;14(1):28–32. Available from: http://dx.doi.org/10.54097/fgblm2ei

[77] Wang R, Wei H, An B, Feng Z, Yao J. Commission Fee is not Enough: A Hierarchical Reinforced Framework for Portfolio Management. Proceedings of the AAAI Conference on Artificial Intelligence [Internet]. 2021 May 18;35(1):626–33. Available from: http://dx.doi.org/10.1609/aaai.v35i1.16142

[78] Briola A, Turiel J, Marcaccioli R, Cauderan A, Aste T. Deep Reinforcement Learning for Active High Frequency Trading [Internet]. arXiv; 2021. Available from: https://arxiv.org/abs/2101.07107

[79] Wang J, Zhang Y, Tang K, Wu J, Xiong Z. AlphaStock. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining [Internet]. 2019 Jul 25;1900–8. Available from: http://dx.doi.org/10.1145/3292500.3330647

[80] Zhang C, Aaraba A. Dynamic Optimal Portfolio with Trading Cost Constraints via Deep Reinforcement Learning. SSRN Electronic Journal [Internet]. 2022; Available from: http://dx.doi.org/10.2139/ssrn.4221316

[81] Fu K, Yu Y, Li B. Multi-Feature Supervised Reinforcement Learning for Stock Trading. IEEE Access [Internet]. 2023;11:77840–55. Available from: http://dx.doi.org/10.1109/access.2023.3298821

[82] Qin M, Sun S, Zhang W, Xia H, Wang X, An B. EarnHFT: Efficient Hierarchical Reinforcement Learning for High Frequency Trading. Proceedings of the AAAI Conference on Artificial Intelligence [Internet]. 2024 Mar 24;38(13):14669–76. Available from: http://dx.doi.org/10.1609/aaai.v38i13.29384