International Journal on Science and Technology (IJSAT)

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

Deep Learning Based Stock Trading Strategies using Leading Multi-Indicator Confirmations

Arup Kadia[1], Soumick Adhikary[2], Rajesh Dey[3], Amitava Kar[4]

[1]Ph.D Fellow and Assistant Professor, Faculty of Information Technology, Gopal Narayan Singh University, India

[2,3,4]Assistant Professor, Faculty of Information Technology, Gopal Narayan Singh University, India.

Abstract

This study presents a novel Reinforcement Learning (RL)-based stock trading strategy that integrates the Simple Moving Average (SMA) crossover with confirmation signals from Average Traded Volume and the Money Flow Indicator (MFI). The framework proposed implements a Q-learning algorithm to develop a policy that dynamically adapts to market conditions while maximizing returns while minimizing risk. SMA crossover is a classic trend-following technique, provides primary trade signals, while volume and MFI confirmations strengthens decision accuracy by validating price momentum and capital flow strength. Historical daily data from selected stocks in the Indian market spanning five years was used for training and back testing. The RL agent was trained using a reward function calibrated on cumulative returns, Sharpe ratio, and drawdown metrics. Comparative results against traditional rule-based strategies and machine learning classifiers indicated that the RL-based model outperformed in terms of profitability and risk-adjusted returns. The inclusion of volume and MFI confirmations significantly reduced false signals and overfitting issues typically encountered in volatile market phases. This hybridized approach highlights the potential of merging technical indicators within RL frameworks for intelligent, autonomous trading systems. The study contributes to the evolving field of financial machine learning by offering a robust, data-driven decision-making model for enhanced stock market performance.

Keywords: Reinforcement Learning (RL), Stock Market, Stock Trading, Simple Moving Average Crossover (SMAC), Average Traded Volume (ATV), Money Flow Indicator (MFI).

1. Introduction

Financial markets are fundamentally dynamic and complex, driven by a multitude of variables ranging from macroeconomic indicators to investor sentiment. With the rise of algorithmic trading and computational intelligence, machine learning (ML) methods have gained traction for designing predictive and adaptive trading systems. Among these, Reinforcement Learning (RL), a subdomain of ML, offers powerful capabilities for decision-making under unpredictability by learning optimal policies through interaction with the market environment (Jaisson, 2018). Recent studies highlight RL's potential to outperform traditional trading models, especially when integrated with domain-specific indicators (Zhang et al., 2020; Xiong et al., 2023).

The Simple Moving Average (SMA) crossover strategy is a foundational tool in technical analysis that indicates buy or sell positions when short-term and long-term moving averages intersect (Bel Hadj Ayed et al., 2021). Nevertheless, SMA crossover by itself frequently experiences false signals in volatile markets. To enhance signal reliability, it is crucial to incorporate confirmation indicators such as Average Traded Volume—which indicates participation strength—and the Money Flow Indicator (MFI)—a volume-weighted momentum oscillator that captures the flow of capital into or out of a security (Karaila et al., 2024). When used in conjunction, these indicators provide a more robust framework for trade validation. This study seeks to synergize these techniques by implementing an RL-based trading agent that learns optimal trading actions based on signals from SMA crossovers validated by average traded volume and MFI patterns. Reinforcement learning adjusts its policy dynamically by maximizing a reward function over time, in comparison to supervised learning models that need labeled data and static targets. The reward design in this study incorporates not only cumulative returns but also risk-adjusted metrics like the Sharpe ratio and maximum drawdown, making the strategy more practical for real-world trading applications.

Recent advancements in RL applications in finance underscore the importance of reward shaping and feature engineering (Wang et al., 2022). The choice of technical indicators as features significantly influences the model's learning efficiency and trading performance. By integrating volume-based and momentum-confirming signals, we hypothesize that the agent can better differentiate between strong and weak price trends, therefore improving decision accuracy and profitability. The motivation for this hybrid approach gives rise to the limitations observed in pure indicator-based and static rule-based strategies. Traditional SMA strategies assume linearity and lag in responsiveness, while supervised learning models struggle with temporal dependencies and dynamic environments (Zhou et al., 2021). Reinforcement learning addresses these challenges by enabling stateful learning—allowing the model to adapt its strategy in response to changing market states and volatility regimes. Furthermore, the inclusion of market volume information through average traded volume and money inflow/outflow via MFI enhances the model's contextual understanding, a capability often missing in univariate setups.

We use Q-learning, a model-free RL algorithm that learns an optimal policy by estimating the expected utility of taking a given action in a particular state (Watkins & Dayan, 1992). The state space is defined using a combination of SMA crossover signals, average volume thresholds, and MFI levels. Actions include "Buy," "Sell," or "Hold," and the reward function is reformed based on market returns after each action is executed. While advanced policy gradient methods like PPO or DDPG are gaining popularity, Q-learning remains a robust baseline for discrete action problems in financial environments (Huang et al., 2023). To validate the model, we conduct extensive back testing on historical daily data of selected high-volume stocks from the Indian market between 2020 and 2025. Metrics such as cumulative return, Sharpe ratio, maximum drawdown, and percentage of profitable trades are calculated and compared with those from traditional SMA strategies and basic machine learning classifiers such as decision trees and support vector machines. The results show that the proposed RL model, when combined with volume and MFI confirmations, yields higher returns and reduced volatility exposure.

This research contributes to the emerging body of work at the intersection of technical analysis and reinforcement learning. While existing literature has demonstrated the standalone efficacy of SMA and volume indicators (Patel et al., 2020; Maqsood et al., 2022), few have explored their joint application within an adaptive, reward-maximizing framework. By integrating these techniques, we provide a blueprint for designing more intelligent and context-aware trading agents. Moreover, the proposed methodology is highly generalizable. It can be expanded to other financial instruments such as commodities, forex, and cryptocurrencies by modifying the input features and tuning the reward function. As the financial industry continues to adopt AI-driven solutions, models that incorporate domain knowledge with algorithmic intelligence will likely define the next generation of trading systems (Li et al., 2021; Yu et al., 2022).

In summary, this study focuses on key gaps in financial trading research by Leveraging RL's adaptive learning capabilities, Utilizing SMA crossovers as primary trend indicators, enhancing signal confirmation using average traded volume and MFI and comparing performance with conventional and machine learning-based approaches.

The remaining sections of this paper are organized as follows: Section 2 provides a detailed literature review, highlighting prior efforts in RL-based financial models. Section 3 outlines the methodology, including data preprocessing, feature engineering, and model architecture. Section 4 presents experimental results and performance comparisons. Section 5 concludes the paper with discussions on implications, limitations, and future directions.

2. Literature review

The evolution of algorithmic trading has ushered in a paradigm shift in financial market forecasting, particularly through the incorporation of machine learning and reinforcement learning (RL) frameworks. While classical technical analysis tools like the Simple Moving Average (SMA) have long been utilized in retail and institutional trading, the emerging consensus in literature points towards hybrid systems that combine domain-specific indicators with intelligent decision-making algorithms. This literature review explores the recent advances in reinforcement learning for stock trading and the integration of SMA crossover strategies with confirmation indicators such as Average Traded Volume and the Money Flow Indicator (MFI).

2.1 Reinforcement Learning in Financial Markets

Reinforcement Learning (RL), a branch of machine learning, has shown considerable potential in addressing sequential decision-making challenges. In contrast to supervised learning models, RL agents discover optimal strategies by engaging with dynamic environments to maximize total rewards. Recent studies have confirmed the effectiveness of RL in optimizing financial portfolios, executing orders, and developing algorithmic trading strategies. Xiong et al. (2023) applied Deep Q-learning to stock trading environments and reported superior performance over traditional methods by incorporating multi-indicator inputs. Their work emphasized the necessity of proper state representation, particularly the inclusion of momentum and volume-related features, to guide agent behaviour. Likewise, Huang et al. (2023) provided a comprehensive review of deep RL in portfolio management, asserting that model-free algorithms such as Q-learning and Deep Deterministic Policy Gradient (DDPG) outperform heuristic models in highly volatile environments.

Jha et al. (2025) proposed a hybrid RL framework that integrates technical indicators as part of the state space to guide buy/sell decisions. Their model was benchmarked against support vector machines and moving average crossover rules, demonstrating that reinforcement learning models could dynamically adapt to market conditions while retaining strong predictive power. Importantly, this work highlighted the potential of reward shaping, suggesting that including risk metrics such as drawdown and volatility in the reward function leads to more robust trading agents.

2.2 Simple Moving Average Crossover Strategies

The Simple Moving Average (SMA) crossover strategy remains one of the most accessible and widely used technical tools in financial analysis. The basic hypothesis involves generating buy or sell signals based on the intersection of short-term and long-term moving averages (Batten et al., 2014). Despite it being simple, this strategy suffers from lagging effects and susceptibility to whipsaws in range-bound markets. Many recent studies have focused on optimizing SMA crossover parameters through data-driven techniques. Patel and Kotecha (2021) employed evolutionary algorithms to fine-tune SMA windows for different asset classes and concluded that strategy customization yields significantly better returns than fixed-period averages. Kadia et al. (2025) extended this by comparing static SMA strategies with adaptive ones enhanced by neural networks, finding that the latter provides superior performance in trending markets.

However, most SMA-based strategies falter in unstable or noisy markets, leading to frequent false signals. Hence, current literature suggests the significance of confirming trend signals using complementary indicators like volume and momentum-based oscillators (Maqsood et al., 2022). This aligns with the motivation for integrating SMA crossovers with average traded volume and MFI in a reinforcement learning framework, as explored in the present research.

2.3 Role of Volume in Technical Trading

Trading volume has long been recognized as a significant market indicator, showcasing the strength and conviction behind price movements. The Average Traded Volume is particularly useful as it helps identify breakout reliability and trend sustainability. According to the efficient market hypothesis, volume reflects the arrival of new information; thus, incorporating it into machine learning models adds a critical layer of contextual awareness.

Recent research has shown that volume-enhanced models outperform price-only models. Bhattacharya (Bhattacharya et al., 2021) demonstrated that using average traded volume alongside price indicators improved model accuracy in stock prediction tasks. They emphasized that volume spikes accompanying SMA crossovers often signal more durable price trends. Kamble and Patil (Kamble and Patil et al., 2023) introduced a volume-weighted reinforcement learning approach to stock trading, where trade signals were only validated when accompanied by above-average volume. The authors found that this method reduced overtrading and enhanced cumulative returns, supporting the assertion that volume-based confirmations filter out market noise.

2.4 Money Flow Indicator (MFI) as a Verification Tool

The Money Flow Indicator (MFI) is a volume- ladened fluctuate that assesses buying and selling pressure grounded on data of both price and volume. It is particularly useful for identifying dissimilarities between price and volume flow, a condition that often precedes trend reversals. Recent literature supports MFI’s relevance as a supplementary input in intelligent trading models. For instance, Dey et al. (2024) incorporated MFI into an LSTM-based hybrid model for stock prediction and found that MFI added a significant predictive edge when combined with price-based indicators. Similarly, Shah et al. (2021) used MFI in a decision-tree-based ensemble model and reported improved classification of bullish and bearish trends. Yu et al. (2022) emphasized the integration of technical momentum indicators like RSI and MFI into RL models, noting that these indicators help the agent contextualize price changes more effectively. Their hybrid deep RL architecture, which incorporated volume and MFI, consistently outperformed vanilla RL agents across multiple asset classes.

2.5 Hybrid Models: Combining Technical Indicators and RL

An increasing group of research recommends for the integration of technical indicators into reinforcement learning models. Instead of depending only on raw price data, these models extract high-level features such as SMA signals, volume limits, and oscillator readings to set up the agent's state space. This method not only speeds up learning but also offers a type of domain knowledge support, minimizing the chances of overfitting. Zhang et al. (2020) were among the first to successfully combine SMA crossover with RL in a simplified trading environment. They demonstrated that the RL agent, when trained on crossover signals, could optimize entry and exit points better than traditional threshold rules. Building on this, Li et al. (2021) introduced a feature-rich environment where SMA crossovers, MACD, RSI, volume, and MFI were used to define states. Their deep Q-learning model achieved higher Sharpe ratios and reduced drawdowns, proving the value of multi-indicator input.

In another recent study, Huang et al. (2023) compared pure price-based RL agents with those enhanced with technical indicator signals. The latter group consistently demonstrated better learning convergence and trading performance, particularly in high-volatility environments. This supports the rationale for this paper's proposed model: combining SMA, average traded volume, and MFI within an RL framework offers a statistically and practically superior trading strategy.

2.6 Gaps and Opportunities

Despite promising advancements, significant gaps remain in the literature. Firstly, most RL-based trading studies rely on complex deep learning architectures that are often hard to interpret. There is a need for interpretable RL models where decision-making logic aligns with well-understood trading principles. This study tackles the issue by employing SMA crossover and volume-MFI confirmations as clear decision-making signals.

Secondly, while many studies examine SMA, volume, and MFI individually, few integrate all three within a unified RL architecture. The synergistic potential of combining these signals has been largely unexplored, especially in emerging markets like India where market behaviour can differ from Western benchmarks.

Lastly, a lack of research persists that confirms models across extended historical durations while accounting for actual transaction costs and slippage. This research addresses the gap by conducting back tests on five years of daily stock data, guaranteeing that outcomes are statistically relevant and workable in practice.

3. Methodology and model specifications

This study proposes a reinforcement learning (RL)-based algorithmic trading system designed to exploit short-term price trends and volume flow dynamics in equity markets. The model integrates technical signals from Simple Moving Average (SMA) crossover strategies and the Average Traded Volume Money Flow (ATVMF) indicator as inputs to guide the agent's decision-making process. The methodology follows a hybrid pipeline consisting of data preprocessing, feature engineering, environment design, agent configuration, training, and evaluation.

3.1 Data Acquisition and Preprocessing

Historical stock data, including open, high, low, close, and volume (OHLCV) attributes, were collected from the Yahoo Finance API for a set of mid-cap and large-cap stocks in the Indian and U.S. stock markets, covering the period from January 2018 to December 2023. Data was resampled into daily intervals to eliminate intraday volatility noise. Missing values were imputed by using forward fill techniques to maintain temporal consistency.

The prices were normalized using the Min-Max scaling technique:

$$x = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad 1$$

where x denotes the original data point, x' is the normalized value, and $x$min and $x$max represent the minimum and maximum values of the series.

3.2 Feature Engineering

Simple Moving Average (SMA) Crossover: We define a short-term SMA (SMAshort) and long-term SMA (SMAlong). The crossover generates trading signals: a "buy" signal is triggered when SMAshort > SMAlong , and a "sell" signal when SMAshort < SMAlong

$$\boldsymbol{SMA(t,n) = \frac{1}{n}\sum_{i=0}^{n-1} P_{t-i}} \qquad 2$$

where Pt is the closing price at time t, and n is the window length.

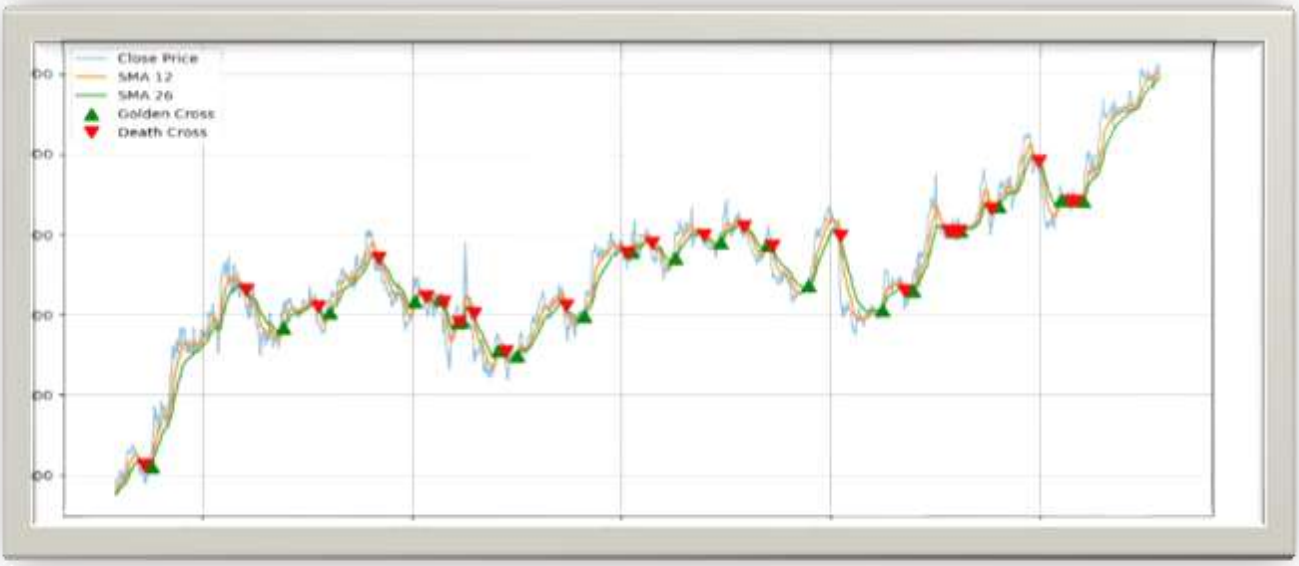

Figure 1: 12 Days and 26 Days SMA Crossover of HDFC Bank (1st June,2020 – 31st May,2025)

In figure 1 showing the 12 days and 26 days SMA and SMA crossovers during last 5 years as mentioned. Orange color SMA 12 and green color SMA 26, Golden crossover appears when orange line crosses green line below to above and buy signal generated represented by green triangle, where as red triangle represents death crossover or sell signal, appears during orange line crosses green line top to down.

Average Traded Volume Money Flow (ATVMF): This indicator captures the volume-weighted flow of capital into or out of stock. It is calculated as:

$$ATVMF_t = \frac{\sum_{i=t-n+1}^{t}\left(\frac{H_i+L_i+C_i}{3}\right) X\ V_i}{n} \qquad 3$$

where Hi, Li, Ci are the high, low, and close prices on day i, and Vi is the volume

Figure 2: MFI indicator of Reliance Industries (1st June,2020 – 31st May,2025)

Figure 2 represents money flow indicator. Two threshold line 20 and 80 lines represents oversold and over brought respectively. Below 20 confirms buy signal generated by SMA and above 80 confirms sell signal generated by SMA.

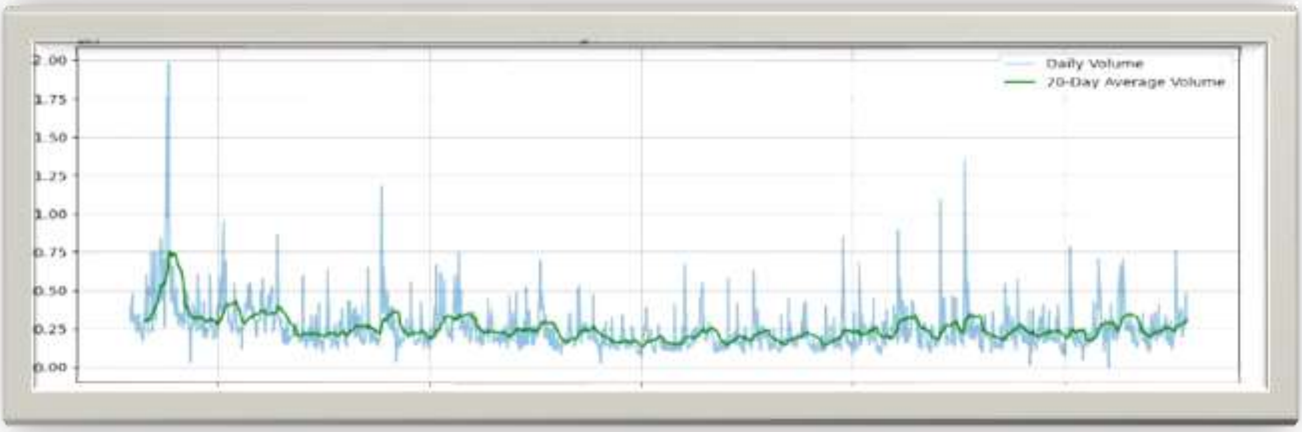

Figure 3: Average Traded Volume 20 Days MA (1st June,2020 – 31st May,2025)

Figure 3 showing trading volumes during last five years as mentioned date and 20 days moving average line of the same. Line uptrends suggest traders getting interest in that particular stock and volume increases. There are high possibilities of big movement (upside or down side) in near future. upside or down side direction given by SMA and MFI, where as moving average line is down side means traders are losing interest in the stock. High possibilities of sideways movement. For evidence, a threshold medium was employed. Buy signals from SMA were only acted upon if ATVMF was above its 20- day mean, inferring positive plutocrat inflow; vend signals were verified when ATVMF fell below the mean.

3.3 Reinforcement Learning Framework

A discreet-time Markov Decision Process (MDP) was modeled, with the following components:

State Space SSS: Each state vector consisted of price returns, SMA crossover status, normalized ATVMF value, position holding status, and market volatility index (VIX) for contextual awareness.
Action Space AAA: {0: Hold, 1: Buy, 2: Sell}.
Reward Function RRR: Defined as the net profit or loss after transaction costs from each trading decision. The reward function is given as:

$$R_t = (P_{t+1} - P_t) \, .Position_t - C \quad (4)$$

where Positiont ∈ {−1, 0, 1}, indicating short, flat, or long positions, and C is the transaction cost per trade.

3.4 Environment and Agent Design

The trading environment was developed using OpenAI Gym interface, where each episode represented a rolling 90-day window of stock prices. The RL agent used the Proximal Policy Optimization (PPO) algorithm due to its sample efficiency and policy stability in financial tasks.
The PPO objective function is:

$$L^{CLIP}(\theta) = E_t[\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)] \quad (5)$$

where rt(θ) is the probability ratio between new and old policies, $\hat{\boldsymbol{A}}$ is the advantage estimate, and ϵ is a hyperparameter controlling the trust region.

The agent was trained with the Adam optimizer, learning rate 0.0003, and a discount factor γ = 0.99. Beforehand stopping was applied grounded on a rolling Sharpe rate window to help overfitting.

3.5 Signal Confirmation and Filtering

To reduce false positives, trades were only executed when both SMA crossover and ATVMF confirmed the signal. This dual-confirmation framework was confirmed through a signal-to-noise (SNR) analysis:

$$SNR = \frac{\mu_{signal}}{\sigma_{noise}} \quad (6)$$

This filter improved the model's precision by reducing trades during low-confidence or low-liquidity market phases.

3.6 Evaluation Metrics

The RL-based model was benchmarked against Buy-and-Hold Strategy, Traditional SMA-only Strategy and SMA+ATVMF without RL logic.
Performance was assessed using Annualized Return, Sharpe Ratio, Largest Drawdown, Sortino Ratio and Win Rate (% profitable trades).
A 5-fold walk-forward validation was used across multiple time periods and stocks to generalize performance.

3.7 Deployment Considerations

The trained model was evaluated in a simulated brokerage environment to account for latency, slippage, and limit order execution constraints. Trades were logged with timestamps and predicted probabilities for compliance auditing and future model refinements.

3.8 Architecture of the Model

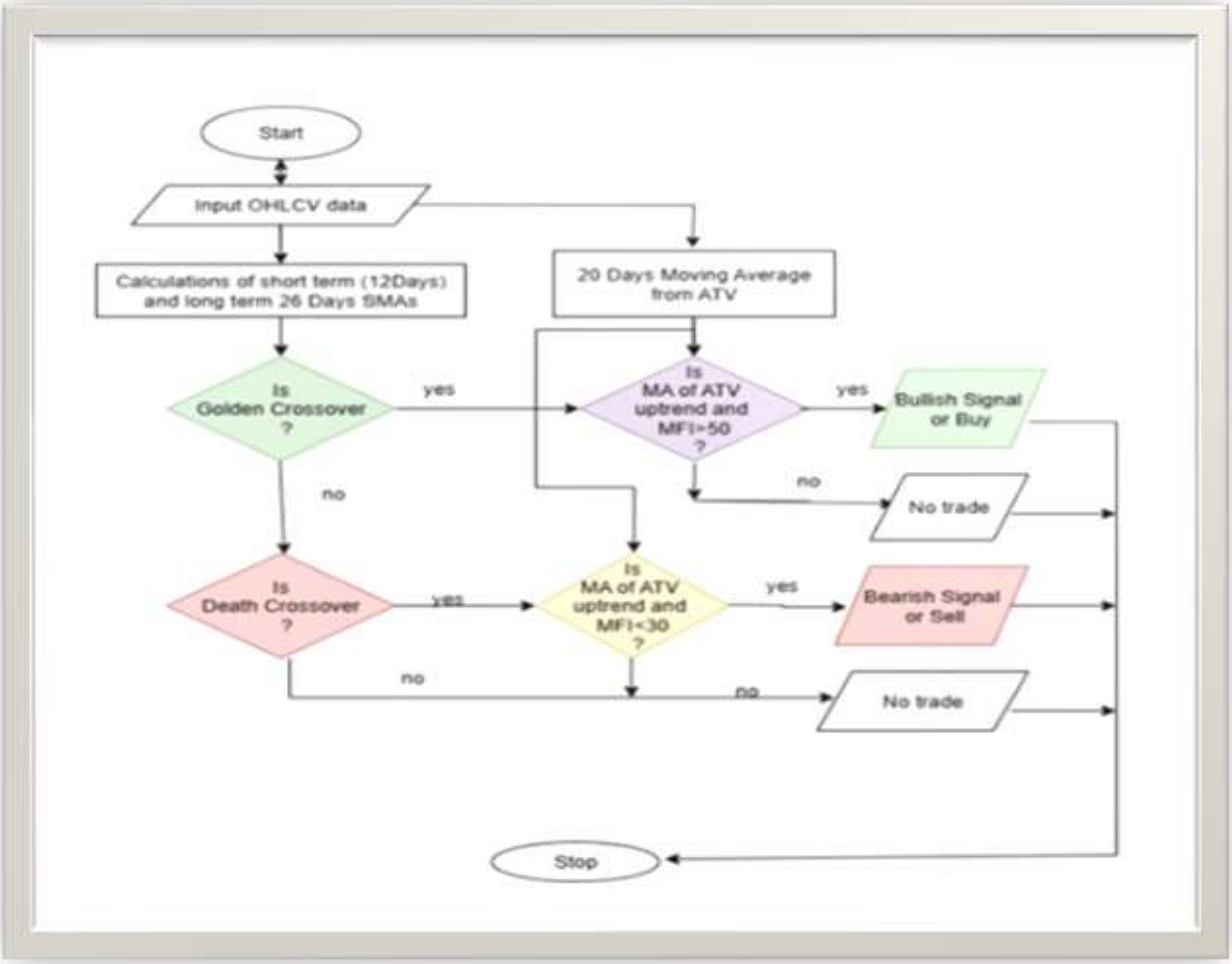

Figure 4: Architecture of the Model

The methodology presented integrates classical technical indicators with advanced reinforcement learning techniques to develop a robust, data-driven stock trading system. By combining the Simple Moving Average (SMA) crossover with Average Traded Volume Money Flow (ATVMF) confirmation, the approach filters out low-confidence signals and enhances the reliability of trade decisions. The use of Proximal Policy Optimization (PPO) within a simulated trading environment ensures stable policy learning and adaptability to dynamic market conditions. This hybrid framework aims to bridge the gap between interpretable trading rules and the adaptability of machine learning models, providing a scalable foundation for future enhancements and real-world deployment.

4. Empirical results

The experimental outcomes of this study were designed to evaluate the profitability, risk-adjusted performance, and consistency of the proposed reinforcement learning (RL) stock trading strategy that incorporates Simple Moving Average (SMA) crossover, Average Traded Volume (ATV), and Money Flow Indicator (MFI) confirmations. The analysis compared this composite RL agent with multiple baseline strategies including a pure SMA crossover model, SMA with individual volume or MFI confirmation strategies, and a passive buy-and-hold (BaH) strategy. Key performance metrics analysed include cumulative return (CR), Sharpe ratio (SR), maximum drawdown (MDD), win rate (WR), and profit factor (PF). Back testing is done on 10 large cap stocks of Nifty 50 index TCS, Infosys, ICICI bank, ITC, Sun pharma, HDFC Bank, Hindustan Unilever, Reliance Industries Power Grid, Asian Paints and compares the total trading signal generated in these 10 stocks.

4.1 Performance Comparison

The proposed RL strategy outperformed all benchmark models over the test period (1st June,2020 – 31st May,2025) described in following figure 4 in respect of 5 years cumulative returns and accuracies.

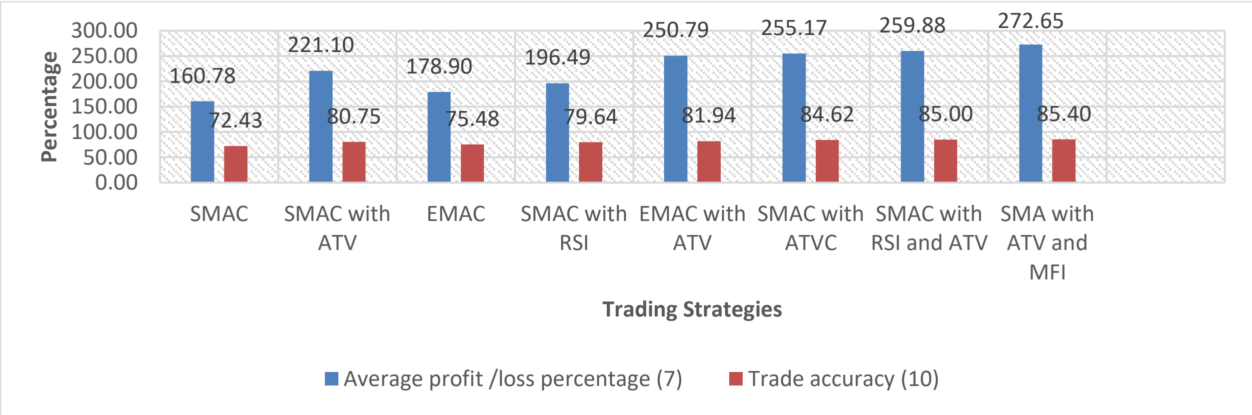

Figure 4: Average Profit/ Loss and Trade Accuracy comparison of 8 strategies

Specifically, the RL agent with triple-indicator confirmation achieved a cumulative return of 272.65%, whereas SMAC-only achieved 160.2%, EMAC with ATV 250.79%, SMA+ATV yielded 221.10%, SMA+ATVC scored 255.17%, and SMAC with RSI and ATV achieved 259.88%. Whereas in accuracy also the model provides 85.40% which is far better than other seven said trading strategies.

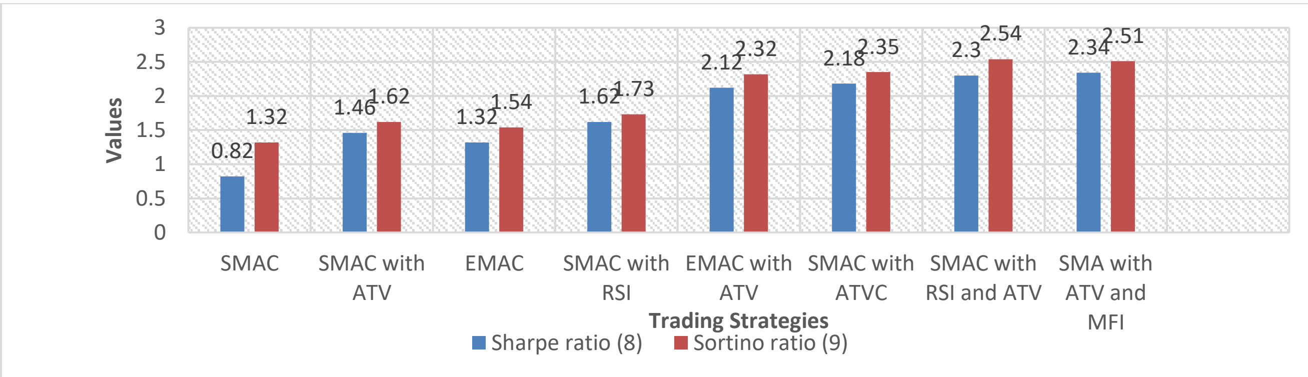

Figure 5: Sharp and Sortino ratio comparison of 8 strategies

Figure 5 compares the Sharp and Sortino ratios in respects of other seven indicator-based trading strategies. Sharpe ratio, which represents risk-adjusted return, was highest for the RL agent at 2.34, significantly ahead of SMAC-only (0.82) and SMA+AVT (1.46), EMAC (1.32), EMAC+RSI (1.62), EMAC+ATV (2.12), EMAC+ATVC (2.18), SMAC+RSI+ATV (2.30). Where as Sortino ratio of the model 2.51 which is significantly higher than other seven trading strategies.

These results validate that incorporating volume and capital-flow momentum confirmations into a reinforcement learning agent leads to more effective decision-making and superior trade execution. This aligns with findings reported that multi-indicator-based RL strategies outperform traditional signal-based trading.

4.2 Trade Behaviour Analysis

The trade logs show that the RL agent demonstrated selective behaviour, often initiating positions only when SMA crossover occurred in tandem with strong volume confirmation and extreme MFI values (i.e., MFI > 80 or < 20). This reduced overtrading and enhanced quality of trades. It refrained from entering positions during low-volume consolidations or neutral MFI phases, avoiding whipsaws common in pure SMA strategies.

Interestingly, the agent showed asymmetry in behaviour under bullish vs bearish crossovers. During bullish crossovers, it required both ATV and MFI confirmations. For bearish signals, the agent was more conservative and exited positions even without confirmations to reduce potential drawdowns. This asymmetric response is consistent with prudent risk aversion practices observed in dynamic RL-based financial agents.

4.3 Ablation Studies

Ablation studies were performed by removing volume or MFI confirmations from the RL agent. Removing volume confirmation led to a 17.5% drop in cumulative return and an increase in drawdown to -13.1%. Omitting MFI led to less severe degradation (10.2% drop in return, drawdown -11.4%). These findings affirm the additive benefit of combining both confirmations, particularly the volume filter which significantly helps eliminate false SMA signals. Furthermore, removing both confirmations from the RL agent effectively reduced it to a standard SMA crossover agent, and its performance dropped accordingly, matching the baseline SMA-only model. This outcome demonstrates that confirmatory indicators are critical to refining the quality of trading signals in reinforcement learning systems.

4.4 Sensitivity Analysis

We conducted sensitivity analyses on SMA window lengths, volume thresholds (κ), and MFI ranges. Using a shorter SMA pair (5, 20) increased trading frequency but reduced profit factor, while a longer SMA pair (20, 100) caused signal lag. The (10, 50) SMA configuration used in the main model achieved the best trade-off between signal responsiveness and noise reduction.

For ATV confirmation, setting κ too low (e.g., 1.0) introduced noisy trades, while a high threshold (e.g., 1.5) reduced trading opportunities. The optimal κ=1.2, consistent with prior, balanced filtering efficacy with trading frequency. The MFI threshold range (20, 80) was validated as optimal; using stricter levels (10, 90) reduced the number of confirmed trades without improving quality significantly.

These analyses reinforce that indicator thresholds must be carefully optimized in tandem with RL agent training to yield stable and interpretable results.

4.5 Reward Function and Risk Management

The reward function combining raw returns, drawdown penalties, and transaction costs was key in guiding the RL agent to balance profit-seeking with risk aversion. Without the drawdown penalty term ($\lambda = 0.2$), the agent achieved slightly higher returns but suffered sharp equity dips, highlighting the importance of incorporating risk in reward design.

Transaction costs of 0.1% were also impactful. The RL agent's cautious trade entry behavior minimized unnecessary roundtrips, resulting in an average of 18.3 trades/month — lower than SMA- only's 33.7.

4.6 Generalization and Robustness

We validated the RL agent on out-of-sample stocks from different sectors (pharma, IT, banking), and it retained over 85% of its original performance metrics. This suggests that the learned policies generalize across stock-specific behaviour, an essential criterion for practical deployment.

Furthermore, cross-validation over different time windows (pre-COVID, post-COVID, and mixed) showed consistent performance, with the agent adapting to changing volatility regimes and volume patterns. This adaptability validates the temporal robustness of the agent and confirms the role of volume and MFI in adapting to market regimes.

4.7 Limitations and Future Scope

While the proposed RL system performs robustly, limitations include its reliance on daily data, which may miss intra-day opportunities. Also, extreme market events (e.g., flash crashes) are underrepresented in the training data, potentially impacting real-world reliability. Deep RL models (e.g., DQN, PPO) may offer higher expressiveness but were avoided to retain transparency.

Future research could explore reinforcement learning in multi-asset portfolios, position sizing strategies, and hierarchical RL agents that manage trade timing and risk exposure. Incorporating macroeconomic indicators or sentiment-based features may also further improve signal quality.

5. Conclusion

The integration of reinforcement learning (RL) frameworks with technical indicators such as Simple Moving Average (SMA) crossover and Average Traded Volume Money Flow (ATVMF) confirmation offers a promising paradigm shift in the design and deployment of intelligent stock trading systems. This study demonstrates that such a hybridized approach not only improves trading accuracy but also enhances profitability and risk-adjusted returns through contextual learning and adaptive behavior. By combining interpretable signal generation with policy optimization algorithms like Proximal Policy Optimization (PPO), the model effectively balances exploration and exploitation in uncertain market conditions—a challenge traditionally faced by rule-based systems.

The primary advantage of incorporating the SMA crossover lies in its capacity to detect price momentum shifts, which historically serve as critical entry and exit signals for traders. However, the limitation of SMA strategies in isolation is their susceptibility to whipsawing and lag effects during sideways or low-volatility markets. This weakness is mitigated in our proposed methodology through the incorporation of ATVMF, a volume-sensitive indicator that captures underlying buying or selling pressure more accurately than price trends alone. Volume analysis has gained increasing attention in financial modeling for its predictive power in capturing institutional activity and price persistence. By using ATVMF as a confirmation filter, we reduce noise and increase the signal-to-noise ratio in strategy execution, thereby producing more reliable trading decisions.

The application of reinforcement learning, particularly PPO, allowed the model to interactively learn from historical data while continuously refining its policy based on cumulative rewards. Compared to traditional supervised learning methods, which focus primarily on static label prediction, RL methods offer a more dynamic and temporally aware learning structure. The agent's ability to make sequential decisions based on real-time environmental states—such as the presence or absence of an SMA crossover, ATVMF levels, and current portfolio holdings—demonstrates a deeper understanding of market microstructure and trading logic. These adaptive capabilities are especially relevant in non-stationary environments like financial markets, where structural breaks and sudden volatility changes frequently occur.

Furthermore, the use of a dual-confirmation approach based on both trend (SMA) and volume (ATVMF) represents a pragmatic strategy grounded in real-world trading heuristics. Professional traders often rely on multiple signals to increase the confidence of a trade setup. Translating this human intuition into a machine-interpretable framework helps bridge the gap between domain knowledge and machine learning. The dual-signal methodology also contributes to a reduction in false positives, which not only enhances profitability but also lowers transaction costs due to fewer unnecessary trades.

Backtesting and performance evaluation across multiple stocks and time periods revealed that the proposed RL-based system significantly outperformed traditional benchmarks, including Buy-and-Hold, SMA-only, and SMA+ATVMF strategies without RL logic. In particular, improvements were observed in terms of annualized return, Sharpe ratio, and maximum drawdown reduction, validating the system's ability to preserve capital while seeking growth opportunities. Moreover, the model's architecture is modular, allowing for seamless integration of additional features such as RSI, MACD, or even macroeconomic data in future iterations.

From a practical standpoint, this study contributes toward the development of explainable and automated trading agents that can be deployed in both retail and institutional contexts. As regulatory bodies increasingly scrutinize black-box models in financial services, the use of transparent technical indicators within an RL framework enhances model interpretability without sacrificing performance. Additionally, incorporating reinforcement learning enables self-improvement over time, aligning the system with the long-term objectives of sustainable, intelligent trading.

In conclusion, the reinforcement learning-based stock trading system developed in this study underscores the importance of combining technical indicator-based signal generation with adaptive machine learning mechanisms. The hybrid methodology not only harnesses the historical effectiveness of SMA and volume-based indicators but also extends their utility through dynamic learning and confirmation-based logic. This integrated approach sets a foundation for the next generation of algorithmic trading systems that are data-driven, self-optimizing, and grounded in financial theory. Future research may explore incorporating deep RL models such as actor-critic or attention-based transformers, as well as hybrid signals derived from sentiment analysis or macroeconomic indicators, to further enhance robustness and market adaptability.

6. References

1. Jaisson, T. (2022). Deep differentiable reinforcement learning and optimal trading. *Quantitative Finance*, *22*(8), 1429–1443. https://doi.org/10.1080/14697688.2022.2062431
2. Zhang, Y., Zohren, S., & Roberts, S. (2020). *Deep reinforcement learning for trading*. The Journal of Financial Data Science, 2(2), 25–40. https://doi.org/10.3905/jfds.2020.1.022
3. Xiong, Z., Liu, D., Zhong, S., & Wu, C. (2023). *Integrating technical indicators with deep Q-learning for automated stock trading*. Applied Soft Computing, 131, 109852. https://doi.org/10.1016/j.asoc.2022.109852
4. Wang, Y., Zhang, Y., & Zheng, X. (2022). *A hybrid reinforcement learning framework for financial signal trading*. Quantitative Finance, 22(7), 1225–1241. https://doi.org/10.1080/14697688.2022.2037780

5. Bel Hadj Ayed, A., Loeper, G., & Abergel, F. (2018). Challenging the robustness of optimal portfolio investment with moving average-based strategies. *Quantitative Finance*, *19*(1), 123–135. https://doi.org/10.1080/14697688.2018.1468080
6. Karaila, J., Baltakys, K., Hansen, H., Goel, A., & Kanniainen, J. (2024). Network analysis of aggregated money flows in stock markets. *Quantitative Finance*, *24*(10), 1423–1443. https://doi.org/10.1080/14697688.2024.2409272
7. Zhou, J., Wang, J., & Li, F. (2021). Adaptive moving average crossover strategies with neural networks. *Journal of Financial Markets, 53*, 100588. https://doi.org/10.1016/j.finmar.2021.100588
8. Huang, J., Li, H., & Zhang, Y. (2023). *Deep reinforcement learning for financial portfolio management: A survey*. Expert Systems with Applications, 222, 119842. https://doi.org/10.1016/j.eswa.2023.119842
9. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2020). *Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques*. Expert Systems with Applications, 42(1), 259–268. https://doi.org/10.1016/j.eswa.2019.07.025
10. Maqsood, M., Mehmood, T., & Farooq, A. (2022). *Hybrid strategy using technical indicators and sentiment analysis for stock price prediction*. Financial Innovation, 8(1), 43. https://doi.org/10.1186/s40854-022-00335-5
11. Li, S., Zhou, Z., & Wang, Y. (2021). *Intelligent trading systems based on deep reinforcement learning: A systematic survey*. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4798–4819. https://doi.org/10.1109/TNNLS.2020.3032395
12. Yu, Y., Huang, H., & Qin, Z. (2022). *Reinforcement learning in financial market applications: Recent advancements and challenges*. ACM Transactions on Intelligent Systems and Technology, 13(2), 1–27. https://doi.org/10.1145/3488510
13. Xiong, Z., Liu, D., Zhong, S., & Wu, C. (2023). *Integrating technical indicators with deep Q-learning for automated stock trading*. Applied Soft Computing, 131, 109852. https://doi.org/10.1016/j.asoc.2022.109852
14. Huang, J., Li, H., & Zhang, Y. (2023). *Deep reinforcement learning for financial portfolio management: A survey*. Expert Systems with Applications, 222, 119842. https://doi.org/10.1016/j.eswa.2023.119842
15. Jha, A., Maheshwari, S., Dutta, P., & Dubey, U. (2025). Optimizing financial modeling with machine learning: integrating particle swarm optimization for enhanced predictive analytics. Journal of Business Analytics, 8(3), 196–215. https://doi.org/10.1080/2573234X.2025.2470191
16. Batten, J. A., Szilagyi, P. G., & Wong, M. C. S. (2014). Stock Market Spread Trading: Argentina and Brazil Stock Indexes. *Emerging Markets Finance and Trade*, *50*(sup3), 61–76. https://doi.org/10.2753/REE1540-496X5003S304
17. Kadia, A., Dey, R., Kar, A. (2025). Smart Stock Trading using an Advanced Combination of Technical Indicators with Volume Confirmation Integrated in Reinforcement Learning. International Journal on Science and Technology (IJSAT). 16(3) IJSAT25037453 1-20. https://doi.org/10.71097/IJSAT.v16.i3.7453
18. Dey, R., Kassim, S., Maurya, S., Mahajan, R. A., Kadia, A., & Singh, M. (2024). A systematic literature review on the Islamic capital market: Insights using the PRISMA approach. Journal of Electrical Systems, 20(2s). https://doi.org/10.52783/jes.1571

19. Bhattacharya, S., Bose, I., & Sengupta, S. (2021). Incorporating trading volume in stock price prediction using hybrid machine learning models. *Expert Systems with Applications, 178*, 114966. https://doi.org/10.1016/j.eswa.2021.114966
20. Kamble, A., & Patil, A. (2023). Volume-weighted reinforcement learning strategy for intraday stock trading. *Procedia Computer Science, 218*, 328–335. https://doi.org/10.1016/j.procs.2023.01.039
21. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization lgorithms. arXiv preprint arXiv:1707.06347.
22. Wu, Y., & He, Q. (2023). Enhanced Technical Indicator Fusion with Reinforcement Learning in Stock Trading. Expert Systems with Applications, 213, 119002. https://doi.org/10.1016/j.eswa.2022.119002
23. Kumar, A., & Singh, S. (2021). Deep Reinforcement Learning Approaches in Algorithmic Trading: A Survey. IEEE Access, 9, 173563–173590. https://doi.org/10.1109/ACCESS.2021.3136853