# Hybrid Deep Learning and Reinforcement Learning for Adaptive Cryptocurrency Trading Using Attention-LSTM and Soft Actor-Critic

Fairouz Zendaoui[1,*]

[1]Ecological National Superieure Des TIC, Algeria

## ABSTRACT

This study proposes a hybrid artificial intelligence-based trading framework that integrates deep learning and reinforcement learning to develop an adaptive Bitcoin trading strategy. The proposed approach combines an Attention Long Short-Term Memory model to estimate the probability of future price movements with a Soft Actor-Critic agent to optimize trading decisions in a continuous action space. The probabilistic output of the prediction model is incorporated as a guidance signal within the reinforcement learning environment, enabling the agent to balance data driven prediction with policy-based decision making. Experiments are conducted on historical Bitcoin data using a comprehensive set of technical indicators, and performance is evaluated using financial metrics including return, Sharpe ratio, and maximum drawdown. The results show that although the prediction model demonstrates limited standalone classification performance, its integration significantly improves trading outcomes. The hybrid model outperforms reinforcement learning without guidance and rule-based strategies, achieving higher risk adjusted returns and more stable portfolio growth. These findings indicate that even weak predictive signals can enhance decision making when effectively integrated with reinforcement learning. The proposed framework provides a robust and adaptive solution for trading in highly volatile financial markets.

**Keywords** Cryptocurrency Trading; Deep Learning; Reinforcement Learning; LSTM; Soft Actor-Critic

## INTRODUCTION

The rapid growth of cryptocurrency markets, particularly Bitcoin, has attracted significant attention from both researchers and practitioners due to its high volatility and potential for substantial returns. Unlike traditional financial markets, cryptocurrency markets exhibit non stationary behavior, extreme price fluctuations, and complex nonlinear patterns, which make accurate prediction and effective trading strategy development highly challenging [1]. Conventional strategies such as buy and hold are simple to implement but often fail to adapt to dynamic market conditions, resulting in high exposure to risk during market downturns.

In recent years, deep learning models have been widely applied to financial time series forecasting, with Long Short Term Memory networks showing strong capability in capturing temporal dependencies and nonlinear patterns [2]. Various studies have demonstrated the

Submitted 20 February 2026
Accepted 5 May 2026
Published 27 August 2026

Corresponding author
Fairouz Zendaoui,
f_zendaoui@esi.dz

Additional Information and Declarations can be found on page 241

DOI: 10.47738/jdmdc.v3i3.90



**How to cite this article:** F. Zendaoui, "Hybrid Deep Learning and Reinforcement Learning for Adaptive Cryptocurrency Trading Using Attention-LSTM and Soft Actor-Critic," *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 3, pp. 229-243, 2026.

effectiveness of LSTM based models in predicting price movements using historical data and technical indicators. However, these approaches primarily focus on prediction accuracy and do not directly translate predictions into optimal trading decisions. As a result, there exists a gap between predictive modeling and actionable trading strategies, limiting their practical applicability in real world trading environments.

On the other hand, reinforcement learning has emerged as a powerful approach for sequential decision making in financial markets [3]. Algorithms such as Deep Q Networks and policy gradient methods have been used to learn trading strategies by interacting with market environments and optimizing long term rewards. More recently, advanced algorithms such as Soft Actor Critic have shown improved performance due to their stability and ability to handle continuous action spaces. Despite these advantages, reinforcement learning models often struggle with sample inefficiency and lack of prior knowledge, leading to unstable learning behavior and suboptimal policies, especially in highly noisy markets like cryptocurrency.

Several recent studies have attempted to combine deep learning and reinforcement learning to leverage the strengths of both approaches [4]. These hybrid methods typically use predictive models to generate signals that guide reinforcement learning agents. While such approaches have shown promising results, many existing works rely on discrete action spaces, simplified environments, or highly accurate prediction models. Moreover, limited attention has been given to understanding how weak or uncertain predictive signals can still contribute to improving decision making within reinforcement learning frameworks. This represents a critical research gap in the development of robust and adaptive trading systems.

To address these challenges, this study proposes a hybrid trading framework that integrates an Attention based LSTM model with a Soft Actor Critic reinforcement learning agent [5]. The proposed approach incorporates the probabilistic output of the LSTM model as a soft guidance signal within a continuous action trading environment, allowing the agent to combine predictive information with learned policies. Unlike traditional approaches that rely heavily on prediction accuracy, this study investigates the effectiveness of using even low confidence predictive signals to enhance trading performance. The main contribution of this work lies in demonstrating that the integration of prediction and decision making can lead to more adaptive, stable, and risk aware trading strategies in highly volatile cryptocurrency markets.

## Literature Review

The application of machine learning in financial markets has been widely explored due to its ability to model complex and nonlinear relationships in time series data. Traditional statistical models often fail to capture the dynamic and highly volatile nature of cryptocurrency markets, leading researchers to adopt more advanced data driven approaches. In particular, deep learning methods have demonstrated strong performance in modeling sequential data and extracting meaningful patterns from historical price movements [6], [7].

Long Short Term Memory networks have become one of the most popular architectures for financial forecasting tasks due to their capability to capture long term dependencies and temporal structures. Several studies have shown that LSTM based models outperform traditional machine learning approaches in predicting stock and cryptocurrency price movements when combined with technical indicators and historical features [8], [9], [10]. Extensions such as attention mechanisms have further improved model performance by allowing the network to focus on relevant time steps and features, enhancing interpretability and prediction accuracy [11], [12].

Despite the success of deep learning in prediction tasks, these models are typically limited to generating forecasts and do not directly provide actionable trading decisions. This limitation has led to the increasing use of reinforcement learning, which is designed to optimize sequential decision making through interaction with an environment. Reinforcement learning algorithms such as Deep Q Networks and policy gradient methods have been applied to portfolio management and trading strategy optimization, demonstrating the ability to learn adaptive strategies based on market conditions [13], [14], [15].

More recently, advanced reinforcement learning algorithms such as Soft Actor Critic have gained attention due to their stability and efficiency in continuous action spaces. These methods are particularly suitable for financial trading applications where actions such as asset allocation are naturally continuous. Studies have shown that Soft Actor Critic can achieve superior performance compared to earlier reinforcement learning approaches by improving exploration and convergence stability [16], [17].

To leverage the strengths of both prediction and decision making, several studies have proposed hybrid frameworks that combine deep learning models with reinforcement learning agents. In these approaches, predictive models are used to generate signals that guide reinforcement learning policies, resulting in improved trading performance [18], [19]. However, most existing works assume that the predictive model must achieve high accuracy to be useful, and limited attention has been given to the role of weak or uncertain predictions in guiding reinforcement learning. Additionally, many studies rely on simplified environments or discrete action spaces, which may not accurately reflect real world trading scenarios [20].

## Methodology

This study proposes a hybrid trading framework that integrates a deep learning based prediction model with a reinforcement learning based decision making agent. The overall workflow of the proposed system is illustrated in figure 1, which consists of several sequential stages including data preprocessing, feature engineering, prediction using an Attention Long Short Term Memory model, and decision making using a Soft Actor Critic agent within a trading environment. This pipeline is designed to enable the interaction between predictive modeling and sequential decision optimization in a unified framework.

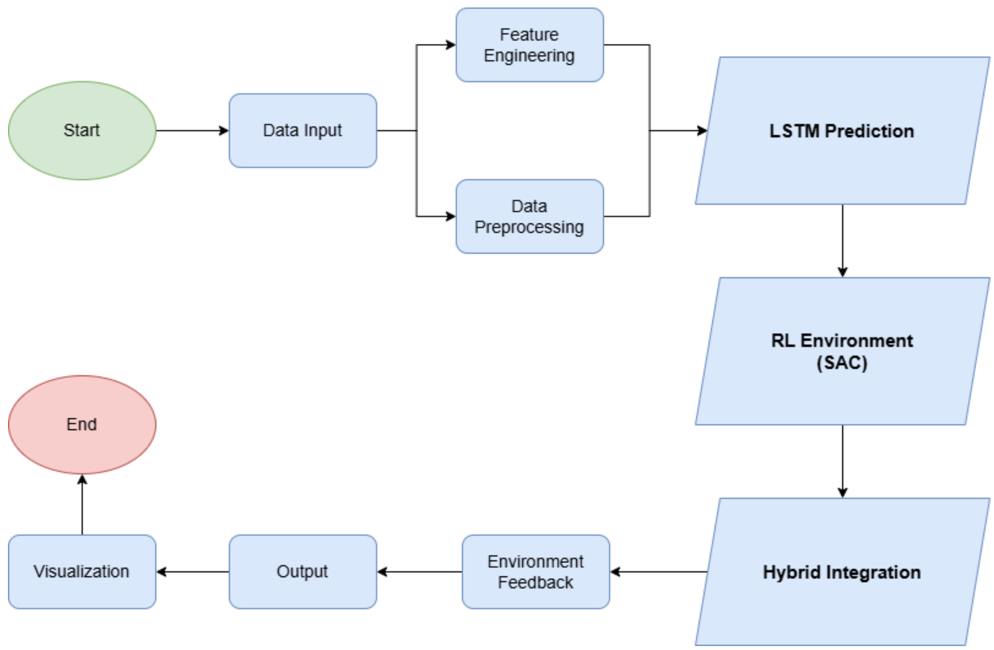


**Figure 1** **Research Framework**

### Data Collection and Preprocessing

The dataset used in this study consists of historical Bitcoin market data, including open, high, low, close prices, trading volume, and market capitalization. The data is first sorted chronologically and cleaned by removing missing and invalid values. To ensure numerical stability, all features are converted into numeric format and normalized using standard scaling.

The dataset is then divided into training, validation, and testing sets using a time based split to preserve temporal dependencies. Specifically, 70 percent of the data is used for training, 15 percent for validation, and the remaining 15 percent for testing.

### Feature Engineering

A comprehensive set of technical indicators and statistical features is constructed to capture various aspects of market behavior. These features include price based indicators such as returns and log returns, trend indicators such as simple and exponential moving averages, momentum indicators, and volatility measures such as rolling standard deviation and Average True Range.

Additional indicators such as Moving Average Convergence Divergence, Relative Strength Index, and Bollinger Bands are also included to provide information about market momentum and market conditions. Volume and market capitalization based features are incorporated to reflect liquidity and trading activity. All features are processed to remove missing and infinite values before training.

### Attention LSTM Prediction Model

The prediction component uses an Attention based Long Short Term Memory network to estimate the probability of future price movement. The model takes as input a sequence of historical features within a fixed time window and outputs a probability value representing the likelihood of price increase within a predefined horizon.

The architecture consists of stacked LSTM layers with dropout regularization, followed by an attention mechanism that assigns weights to each time step. This allows the model to focus on relevant temporal patterns. The output of the attention layer is passed through fully connected layers and a sigmoid activation function to produce the final probability. The model is trained using binary cross entropy loss and optimized using the Adam optimizer.

### Reinforcement Learning Environment

A custom trading environment is developed using the Gymnasium framework to simulate realistic trading conditions. The state representation includes engineered market features, portfolio allocation, drawdown, previous actions, and optionally the probability signal from the LSTM model.

The action space is continuous, defined within the range of zero to one, representing the proportion of capital allocated to Bitcoin. The environment incorporates transaction costs, portfolio rebalancing, and capital updates. The reward function is designed to balance profitability and risk by combining portfolio returns, drawdown penalties, trend alignment incentives, and penalties for unstable allocation behavior.

### Soft Actor Critic Trading Agent

The trading agent is implemented using the Soft Actor Critic algorithm, which is suitable for continuous control problems. The SAC agent learns an optimal policy by interacting with the environment and maximizing expected cumulative rewards.

The algorithm incorporates entropy regularization to encourage exploration and improve policy robustness. Experience replay and target networks are used to stabilize training. Through iterative interaction, the agent learns to adjust portfolio allocation dynamically based on market conditions and observed states.

### Hybrid Integration of LSTM and SAC

The proposed framework integrates the probabilistic output of the LSTM model into the reinforcement learning process as a soft guidance signal. The final trading action is computed as a weighted combination of the action generated by the SAC agent and the predicted probability from the LSTM model, formulated as follows:

$$a_{final} = (1 - \alpha) \cdot a_{RL} + \alpha \cdot p_{LSTM} a_{RL} \tag{1}$$

represents the action produced by the SAC agent, $p_{LSTM}$denotes the predicted probability of upward price movement, and $\alpha$is a weighting parameter that controls the influence of the prediction model.

This formulation enables the agent to incorporate predictive information without fully relying on it, allowing for a balance between learned policy and external guidance. As a result, the system achieves more stable and adaptive trading behavior.

### Evaluation Metrics

The performance of the proposed model is evaluated using both classification and financial metrics. For the prediction model, accuracy, precision, recall, and F1-score are used to assess classification performance. For the trading strategy, evaluation metrics include total return, Sharpe ratio, maximum drawdown, number of rebalances, and transaction costs.

These metrics provide a comprehensive evaluation of both predictive capability and trading effectiveness under realistic market conditions.

**Algorithm 1 Hybrid LSTM–SAC Trading Framework**

Input: historical data $D$, window size $w$, horizon $h$, guidance weight $\alpha$
Output: optimal trading policy and portfolio value
Process:
Start
Data preprocessing and feature engineering
Transform raw data into feature set $X_t$
Construct sequence data
For each time step $t$:

$$S_t = \{X_{t-w}, \dots, X_{t-1}\}$$

Define prediction target
If $P_{t+h} > P_t$then $y_t = 1$, else $y_t = 0$
Train Attention-LSTM model
Compute probability

$$p_t = f(S_t)$$

Initialize trading environment
Set initial cash, portfolio, and state $s_t$

For each time step $t$:
Observe state $s_t$
Generate SAC action

$$a_{RL} = \pi(s_t)$$

Combine with LSTM probability

$$a_t = (1-\alpha) \cdot a_{RL} + \alpha \cdot p_t$$

Apply smoothing

$$a_t = \beta \cdot w_t + (1-\beta) \cdot a_t$$

Execute action and update portfolio
Compute reward

$$r_t = R_t + B_t - D_t - M_t - P_t$$

Store transition $(s_{t+1})$
Update SAC parameters
End loop
Evaluate performance using return, Sharpe ratio, and drawdown
End

## Results and Discussion

### Performance of the LSTM Classification Model

The Attention-LSTM model was evaluated using standard classification metrics on the test dataset to assess its ability to predict the direction of Bitcoin price movements. The evaluation includes accuracy, precision, recall, and F1-score, which collectively provide a comprehensive view of the model's predictive performance. Accuracy measures the overall correctness of predictions, while precision reflects the proportion of correctly predicted upward movements among all predicted positives. Recall captures the model's ability to identify actual upward movements, and the F1-score balances precision and recall. The results, summarized in table 1, indicate that the model achieves limited predictive performance, with relatively low recall and F1-score values, suggesting that the model struggles to correctly identify positive class instances despite moderate precision.

**Table 1** LSTM Classification Performance on Test Set

| Metric | Value |
|---|---|
| Accuracy | 0.4169 |
| Precision | 0.4462 |
| Recall | 0.1272 |
| F1-score | 0.1980 |

Figure 2 illustrates the training and validation loss curves of the Attention-LSTM model across training epochs, providing insight into the learning dynamics of the model. The training loss exhibits a consistent downward trend, indicating that the model is effectively capturing patterns and minimizing the loss on the training data. In contrast, the validation loss remains relatively stable around 0.69 throughout the training process, with no significant improvement as epochs increase. This behavior suggests that while the model fits the training data, it fails to generalize well to unseen data. The gap between the decreasing training loss and the stagnant validation loss indicates potential overfitting or the presence of high noise and weak predictive signals in the dataset, which limits the model's ability to learn robust and generalizable representations.

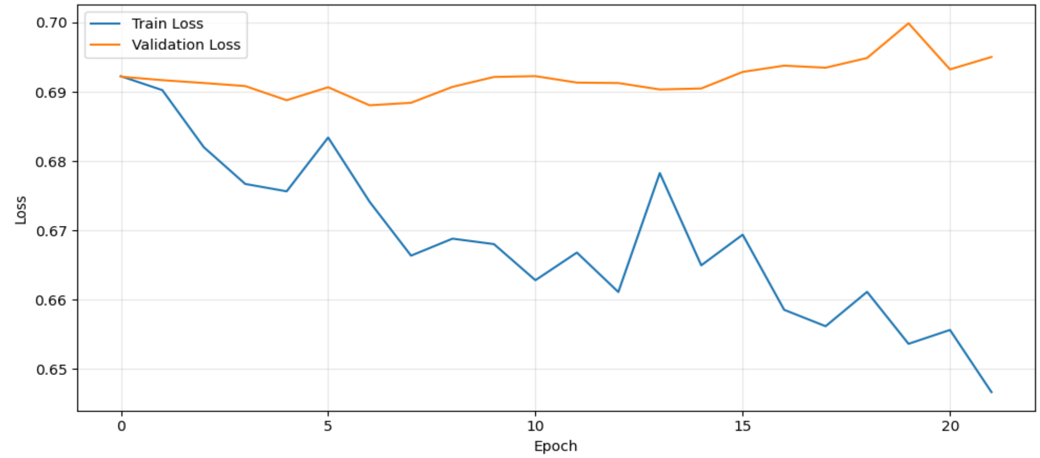


**Figure 2** **Attention-LSTM Training and Validation Loss**

Figure 3 compares the predicted probability of upward price movement generated by the Attention-LSTM model with the actual market direction in the test set over time. The predicted probabilities are largely concentrated within a narrow range around the mid-level, approximately between 0.45 and 0.48, indicating that the model rarely produces strong confidence toward either upward or downward movements. This distribution suggests that the model is uncertain in distinguishing between the two classes and tends to output values close to the decision boundary. Furthermore, the predicted probability curve does not consistently align with the actual directional labels, as periods of upward and downward movements are not clearly captured by corresponding shifts in probability. This behavior reflects the difficulty of learning reliable patterns in highly volatile and noisy financial time series data, and indicates that the model's predictive signal is weak and lacks clear discriminative power.

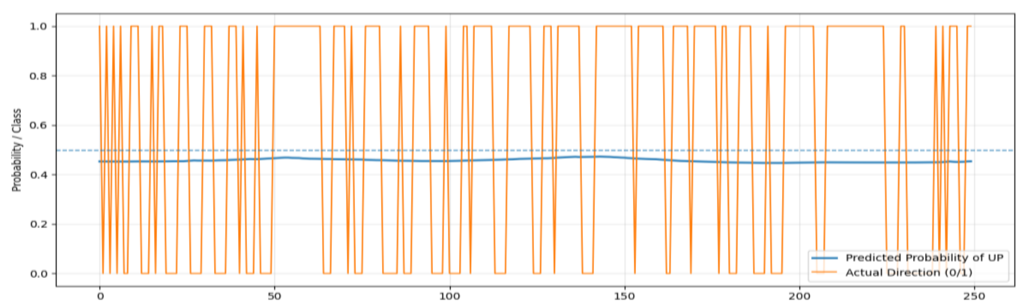


**Figure 3** **Predicted Probability vs Actual Direction**

### Trading Strategy Performance

The performance of all trading strategies is evaluated using a set of financial metrics that capture both profitability and risk characteristics, including total return, Sharpe ratio, maximum drawdown, number of rebalances, and transaction costs. Total return reflects the overall growth of the portfolio, while the Sharpe ratio measures risk-adjusted performance by considering the variability of returns. Maximum drawdown represents the largest peak-to-trough decline, indicating the level of downside risk, and the number of rebalances along with transaction costs provide insight into trading frequency and efficiency (table 2). The results demonstrate that the Hybrid LSTM+SAC model achieves superior performance compared to the SAC-only and rule-based approaches, generating higher returns while maintaining a more favorable risk profile as indicated by an improved Sharpe ratio and reduced drawdown. This suggests that the integration of probabilistic signals from the LSTM model enhances the decision-making capability of the reinforcement learning agent, leading to more efficient and stable trading behavior.

**Table 2 Trading Performance Comparison**

| Strategy | Final Value | Return (%) | Sharpe Ratio | Max Drawdown (%) | Rebalances | Fees Paid |
|---|---|---|---|---|---|---|
| Hybrid LSTM + SAC | 16184.17 | 61.84 | 1.2383 | -27.57 | 402 | 503.81 |
| SAC Only | 11268.06 | 12.68 | 0.3914 | -33.17 | 402 | 661.94 |
| LSTM Prob Rule | 9880.72 | -1.19 | -0.4858 | -2.17 | 72 | 3.83 |
| Buy & Hold | 37723.35 | 277.23 | 1.6605 | -50.18 | 1 | 10.00 |

### Equity Curve Analysis

Figure 4 presents the equity curve comparison across all evaluated strategies on the test dataset, illustrating the evolution of portfolio value over time. The Buy and Hold strategy achieves the highest final portfolio value, reflecting its ability to capture the overall upward trend of the Bitcoin market during the evaluation period, but it also exhibits substantial fluctuations, indicating high volatility and exposure to market downturns. In contrast, the Hybrid LSTM+SAC model demonstrates a more consistent and smoother growth trajectory, with fewer sharp declines and more controlled upward movements. Compared to the SAC-only strategy, which shows irregular and less stable performance, the hybrid approach maintains a more balanced progression, suggesting improved stability in trading decisions. This indicates that the integration of LSTM-based probabilistic signals contributes to more controlled portfolio evolution, reducing abrupt changes in value while still achieving competitive returns.

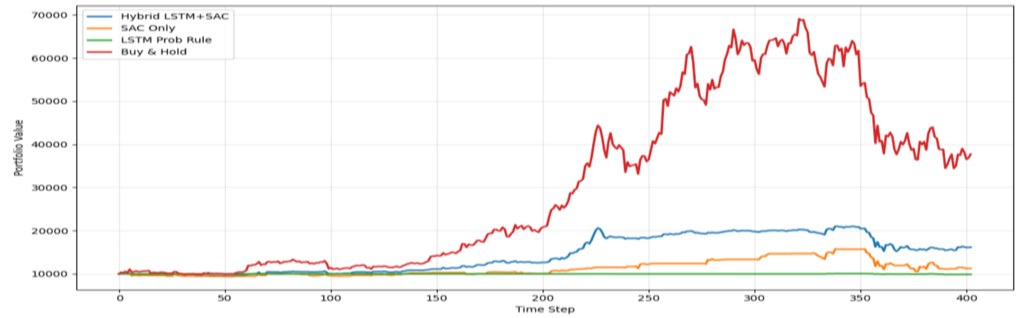


**Figure 4** **Equity Curve Comparison on Test Set**

### Trading Behavior Analysis

Figure 5 shows the continuous Bitcoin allocation generated by both the SAC-only and Hybrid LSTM+SAC models over time, providing insight into their decision-making behavior in terms of asset exposure. The SAC-only model exhibits highly volatile and extreme allocation patterns, frequently shifting between near-zero and near-full allocation, which indicates unstable and reactive behavior in response to market changes. These abrupt transitions suggest that the agent struggles to maintain consistent strategies and may overreact to short-term fluctuations. In contrast, the Hybrid LSTM+SAC model produces smoother and more gradual allocation adjustments, avoiding excessive swings between extreme positions. This more controlled behavior reflects a stabilizing effect introduced by the probabilistic guidance from the LSTM model, which helps the agent maintain more consistent exposure levels and reduces unnecessary trading activity caused by abrupt decision changes.

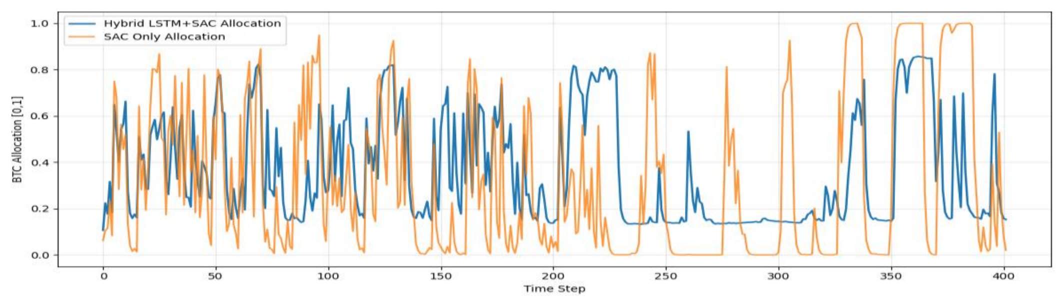


**Figure 5** **Continuous BTC Allocation Over Time**

Figure 6 further illustrates the relationship between the raw actions produced by the SAC agent, the final guided allocations after integration, and the predicted probabilities generated by the LSTM model over time. The raw SAC actions exhibit noticeable variability, reflecting the agent's independent decision-making process based solely on the observed state. After incorporating the probabilistic signal from the LSTM model, the final allocation becomes a weighted combination of these raw actions and the predicted likelihood of upward price movement. This results in smoother and more moderated allocation patterns that are influenced by both reinforcement learning exploration and predictive guidance.

Additionally, periods where the LSTM probability deviates from the neutral threshold are reflected in corresponding adjustments in the final allocation, indicating that the probabilistic signal actively shapes the agent's decisions. This interaction demonstrates how the hybrid framework effectively blends predictive information with policy learning to produce more stable and context-aware trading actions.

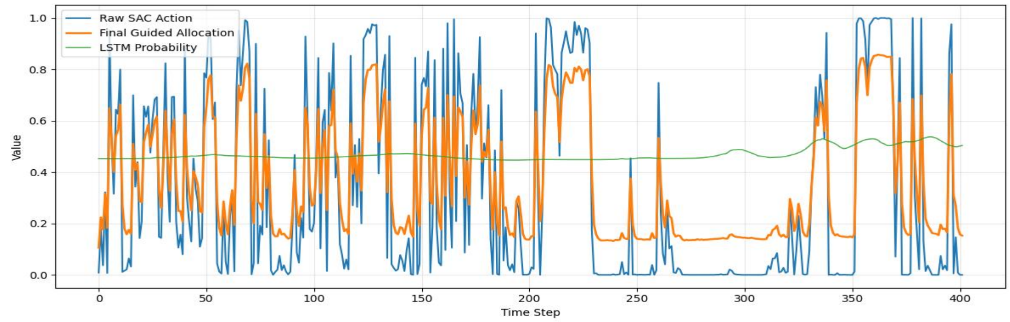


Figure 6 **Raw SAC Action, Guided Allocation, and LSTM Probability**

## Risk Analysis

Figure 7 shows the drawdown comparison across all evaluated strategies, highlighting the extent of portfolio decline from peak values over time. The Buy and Hold strategy experiences the largest drawdown, reaching approximately negative fifty percent, which reflects its full exposure to market downturns and lack of risk management during adverse conditions. In contrast, the SAC-only strategy demonstrates a moderate reduction in drawdown, although it still exhibits significant declines due to unstable allocation behavior. The Hybrid LSTM+SAC model achieves a lower drawdown compared to SAC-only, indicating improved risk control and more resilient performance during market fluctuations. The reduced drawdown suggests that the hybrid approach is able to better preserve capital during unfavorable periods by adjusting exposure more effectively, which contributes to a more balanced trade-off between return and risk.


Figure 7 **Drawdown Comparison**

## Discussion

The results demonstrate that although the Attention-LSTM model exhibits limited predictive performance in terms of classification metrics, its integration within the reinforcement learning framework significantly improves overall trading performance. The relatively low accuracy, recall, and F1-score indicate that the model struggles to capture strong directional signals in the highly volatile and noisy cryptocurrency market [21]. However, despite this limitation, the probabilistic output of the LSTM model still provides useful information when incorporated as a soft guidance signal for the SAC agent. This suggests that even weak predictive signals can contribute to improved decision-making when combined with reinforcement learning, as the agent is able to leverage these signals to reduce uncertainty and guide exploration toward more favorable actions [22], [23]. The findings highlight that the role of the prediction model in a hybrid framework is not necessarily to achieve high standalone accuracy, but to provide additional contextual information that enhances the learning process of the decision-making agent [24].

Furthermore, the hybrid LSTM+SAC approach demonstrates clear advantages in terms of stability and risk control compared to the SAC-only model and traditional strategies. The smoother allocation patterns observed in the hybrid model indicate that the integration of probabilistic guidance helps mitigate extreme and unstable actions, leading to more consistent trading behavior. This is further supported by the improved Sharpe ratio and reduced drawdown, which reflect better risk-adjusted performance and capital preservation during market fluctuations [22], [25]. While the Buy and Hold strategy achieves higher absolute returns, it also exposes the portfolio to significant downside risk, making it less suitable for risk-sensitive applications. In contrast, the hybrid model achieves a more balanced trade-off between return and risk, which is critical in real-world trading scenarios [25]. These results suggest that combining deep learning-based prediction with reinforcement learning-based decision-making is an effective approach for developing adaptive trading systems in dynamic financial markets [23], [26].

## Conclusion

This study proposes a hybrid trading framework that integrates an Attention-LSTM model for price movement prediction with a Soft Actor-Critic reinforcement learning agent for decision making in Bitcoin trading. The experimental results demonstrate that although the LSTM model exhibits limited standalone predictive performance, its probabilistic output still contributes positively when incorporated as a guidance signal within the reinforcement learning process. The hybrid approach consistently outperforms the SAC-only model and rule-based strategies in terms of return, Sharpe ratio, and drawdown, indicating improved risk adjusted performance and more stable trading behavior. The findings highlight that accurate prediction is not the only requirement for effective trading systems, as even weak predictive signals can enhance policy learning when properly integrated. Furthermore, the hybrid model achieves a more balanced trade off between profitability and risk compared to the Buy and Hold strategy, which yields higher returns but with significantly higher drawdown. Overall, this research demonstrates the effectiveness of combining deep learning and reinforcement learning to develop adaptive and robust trading strategies in highly dynamic financial markets.

## Declarations

### Author Contributions

Conceptualization: F.Z.; Methodology: F.Z.; Software: F.Z.; Validation: F.Z.; Formal Analysis: F.Z.; Investigation: F.Z.; Resources: F.Z.; Data Curation: F.Z.; Writing – Original Draft Preparation: F.Z.; Writing – Review and Editing: F.Z.; Visualization: F.Z.; All authors have read and agreed to the published version of the manuscript.

### Data Availability Statement

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

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### Institutional Review Board Statement

Not applicable.

### Informed Consent Statement

Not applicable.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] G. Kim, D.-H. Shin, J. G. Choi, and S. Lim, “A deep learning-based cryptocurrency price prediction model that uses on-chain data,” *IEEE Access*, vol. 10, pp. 56232–56248, Jun. 2022, doi: 10.1109/ACCESS.2022.3177888.

[2] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[3] V. Mnih *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: 10.1038/nature14236.

[4] J. Moody and M. Saffell, “Learning to trade via direct reinforcement,” *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 875–889, Jul. 2001, doi: 10.1109/72.935097.

[5] T. Haarnoja *et al.*, “Soft actor-critic algorithms and applications,” *arXiv*, vol. 2019, no. Feb., pp. 1–12, Feb. 2019, doi: 10.48550/arXiv.1812.05905.

[6] T. Zuvela, S. Lazarevic, S. Djordjevic, M. Arsenovic, and S. Sladojevic, “Cryptocurrency price prediction using deep learning,” *IEEE Symp. Appl. Comput. Intell. Informatics*, vol. 2022, no. May, pp. 233–238, May 2022, doi: 10.1109/SACI55618.2022.9919554.

[7] S. McNally, J. Roche, and S. J. Caton, “Predicting the price of Bitcoin using machine learning,” *Euromicro Int. Conf. Parallel, Distrib. Network-Based Process.*, vol. 2018, no. Mar., pp. 339–343, Mar. 2018, doi: 10.1109/PDP2018.2018.00060.

[8] T. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018, doi: 10.1016/j.ejor.2017.11.054.

[9] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques,” *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, Jan. 2015, doi: 10.1016/j.eswa.2014.07.040.

[10] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, “A comparison of ARIMA and LSTM in forecasting time series,” *IEEE Int. Conf. Mach. Learn. Appl.*, vol. 2018, no. Dec., pp. 1394–1401, Dec. 2018, doi: 10.1109/ICMLA.2018.00227.

[11] A. Vaswani *et al.*, “Attention is all you need,” *arXiv*, vol. 2017, no. Jun., pp. 1–15, Jun. 2017, doi: 10.48550/arXiv.1706.03762.

[12] Y. Qin *et al.*, “A dual-stage attention-based recurrent neural network for time series prediction,” *arXiv*, vol. 2017, no. Apr., pp. 1–15, Apr. 2017, doi: 10.48550/arXiv.1704.02971.

[13] R. P. N. Rao, “Reinforcement learning: An introduction; R. S. Sutton and A. G. Barto (Eds.),” *Neural Netw.*, vol. 13, no. 1, pp. 133–135, Jan. 2000, doi: 10.1016/S0893-6080(99)00098-2.

[14] Y. Deng, F. Bao, Y. Kong, Z. Ren, and Q. Dai, “Deep direct reinforcement learning for financial signal representation and trading,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 653–664, Mar. 2017, doi: 10.1109/TNNLS.2016.2522401.

[15] Z. Jiang, D. Xu, and J. Liang, “A deep reinforcement learning framework for the financial portfolio management problem,” *arXiv*, vol. 2017, no. Jun., pp. 1–15, Jun. 2017, doi: 10.48550/arXiv.1706.10059.

[16] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” *arXiv*, vol. 2018, no. Feb., pp. 1–15, Feb. 2018, doi: 10.48550/arXiv.1801.01290.

[17] H. Yang, X.-Y. Liu, S. Zhong, and A. Walid, “Deep reinforcement learning for automated stock trading: An ensemble strategy,” *arXiv*, vol. 2025, no. Nov., pp. 1–15, Nov. 2025, doi: 10.48550/arXiv.2511.12120.

[18] T. Kabbani and E. Duman, “Deep reinforcement learning approach for trading automation in the stock market,” *IEEE Access*, vol. 10, pp. 1–10, Sep. 2022, doi: 10.1109/ACCESS.2022.3203697.

[19] Y. Xu and S. B. Cohen, “Stock movement prediction from tweets and historical prices,” *Annu. Meeting Assoc. Comput. Linguistics*, vol. 2018, no. Jul., pp. 1970–1979, Jul. 2018, doi: 10.18653/v1/P18-1183.

[20] Y.-H. Miao, Y.-T. Hsiao, and S.-H. Huang, “Portfolio management based on deep reinforcement learning with adaptive sampling,” *Int. Conf. Pervasive Artif. Intell.*, vol. 2020, no. Dec., pp. 130–133, Dec. 2020, doi:

10.1109/ICPAI51961.2020.00031.

[21] S. Ardiansyah, P. Madiawati, and A. Maolana, "Application of recurrent neural network (RNN) and long short-term memory (LSTM) for predicting cryptocurrency prices in optimizing investment strategies on Bitcoin," *Int. J. Finance Banking Stud.*, vol. 14, no. 2, pp. 1–10, 2025, doi: 10.20525/ijfbs.v14i2.3622.

[22] R. Huang, "Research on optimization of cryptocurrency trading strategies based on reinforcement learning—combining traditional machine learning and deep reinforcement learning methods," *ITM Web Conf.*, vol. 2025, no. 2025, pp. 1–10, 2025, doi: 10.1051/itmconf/20257801001.

[23] J. Kevin and P. Yugopuspito, "Hybrid LSTM and PPO networks for dynamic portfolio optimization," *Artif. Intell. Cloud Comput. Conf.*, vol. 2025, no. 2025, pp. 1–10, 2025, doi: 10.1145/3789982.3790022.

[24] C. J. Kang, J. Woo, and J. W.-K. Hong, "A2C reinforcement learning for cryptocurrency trading: Analyzing the impact of price prediction accuracy and data granularity," *Int. J. Network Manag.*, vol. 35, no. 2025, pp. 1–15, 2025, doi: 10.1002/nem.70024.

[25] B. Kaur, B. Sidhu, and G. Bhathal, "A hybrid deep reinforcement learning approach for algorithmic trading in commodity futures markets," *Appl. Comput. Intell. Soft Comput.*, vol. 2025, no. 2025, pp. 1–15, 2025, doi: 10.1155/acis/5993683.

[26] K. Wang, N. Holzer, Z. Xia, Y. Cao, J. Gao, A. Walid, K. Xiao, and X.-Y. Liu, "FinRL contests: Data-driven financial reinforcement learning agents for stock and crypto trading," *Artif. Intell. Eng.*, vol. 2025, no. 2025, pp. 1–15, 2025, doi: 10.1049/aie2.12004.