A2C Reinforcement Learning for Cryptocurrency Trading and Asset Management

Abstract—**Unlike the traditional stock markets, the 24/7 nature of the cryptocurrency market poses unique challenges and opportunities, particularly in asset trading and management. These dynamic market conditions have accelerated the development of sophisticated trading strategies, increasingly leveraging the power of Artificial Intelligence (AI). Among these, AI-driven trading bots have become a prominent tool, offering enhanced decision-making capabilities over conventional methods. This paper proposes the application of the Advantage Actor-Critic (A2C) model, a reinforcement learning technique ideally suited for the unpredictable nature of the cryptocurrency market. Our research aims to optimize asset allocation within a diverse portfolio, including both high-volatility cryptocurrencies and the more stable US Dollar. The proposed A2C model strategically leverages current and predicted price data of cryptocurrencies with current asset allocation to make new asset allocation decisions. Our experiments demonstrate the A2C model's efficacy in managing asset allocations under varying market conditions. We particularly focus on how the model responds to alterations in the loss penalty factor within its reward function, which enables a shift between aggressive and conservative investment strategies. The model effectively balances risk and return, showing promising potential in achieving stable asset growth in rising markets while mitigating losses during market downturns.**

Index Terms—**Cryptocurrency, Trading, Asset Management, A2C Reinforcement Learning**

I. Introduction

The emergence of cryptocurrency has reformed the financial landscape, introducing a digital asset class characterized by high volatility and the potential for significant returns. Unlike traditional stock markets, the cryptocurrency market operates 24 hours a day, presenting unique challenges and opportunities for traders and asset managers. This new market environment has motivated the development of sophisticated trading strategies, especially leveraging Artificial Intelligence (AI) to navigate its dynamic conditions. Among these innovations, AI-driven trading bots have evolved, promising enhanced decision-making capabilities compared to human traders [1].

Despite the progress in applying AI within the cryptocurrency domain, challenges remain, particularly in adapting to the market's rapid volatility and in effective asset management. Machine Learning (ML) has proven integral in predicting cryptocurrency price movements by processing vast amounts of market data to uncover complex patterns. However, the application of these insights into real-time asset management is still nascent. While several previous studies have utilized deep neural network models for optimizing asset allocation ratios [2], [3], our research introduces the novel application of the Advantage Actor-Critic (A2C) model [4] in this context. A2C, a reinforcement learning technique, is uniquely suited for the dynamic and unpredictable nature of the cryptocurrency market.

Our research aims to bridge this gap by proposing the use of an A2C reinforcement learning model, customized specifically for cryptocurrency trading and asset management. This approach is designed to optimize asset allocation ratios, including cryptocurrencies and USD, by leveraging both current and predicted cryptocurrency price data. We validate the efficacy of our A2C model and show how variations in its reward function parameters impact its behavior and adaptability to evolving market dynamics. Experiment results demonstrated that although the model showed somewhat extreme variations between profitability and loss prevention based on the factor that imposes higher penalties for losses in the portfolio, it effectively reallocated assets according to market changes.

The remainder of the paper is organized as follows. Section II covers related work in cryptocurrency price prediction, trading, and asset management. Our proposed methodology is presented in Section III, followed by the validation of the proposed model in Section IV. Section V provides some discussions, and finally, we conclude the paper in Section VI.

II. Related Work

A. *Cryptocurrency Price Prediction*

The field of cryptocurrency price prediction has seen a proliferation of research, with a significant focus on Bitcoin [5]. Researchers have predominantly targeted Bitcoin due to its prominence and volatility, applying various machine learning techniques for price prediction. Notable among these studies, Kim [6] utilized LSTM [7] to predict Bitcoin prices, leveraging its ability to handle nonlinearity and long-term memory characteristics of the market. This approach not only predicted prices but also analyzed the profitability of investment strategies based on these predictions. Similarly, another study by Li et al. [8] introduced a hybrid data decomposition-based bidirectional deep-learning model, combining variational mode decomposition and BiLSTM [9], to enhance Bitcoin market prediction accuracy. This model effectively decomposed the Bitcoin price data into various intrinsic modes, improving its predictability remarkably. These studies reflect a trend in the field where advanced machine learning and deep-learning techniques are being increasingly employed to address the challenges posed by the volatile nature of cryptocurrency prices, especially Bitcoin.

Ortu et al. [10] explored the use of deep learning algorithms for cryptocurrency price classification, focusing on Bitcoin and Ethereum [11]. They investigated the impact of combining technical, trading, and social media indicators on the predictability of price movements. The study demonstrated that the inclusion of social media indicators, such as sentiments and emotions extracted from platforms like GitHub [12] and Reddit [13], considerably improved the accuracy of price predictions. This research emphasized the importance of considering a broad range of factors, including social interactions and sentiment, in the volatile cryptocurrency market.

These studies collectively highlight a trend towards the utilization of advanced machine learning and deep learning techniques in cryptocurrency price prediction. They demonstrate the effectiveness of these models in handling the complex and nonlinear nature of cryptocurrency price movements, emphasizing the importance of incorporating diverse data sources, including social media sentiment, for enhanced predictive accuracy.

B. Cryptocurrency Trading and Asset Management

In the area of cryptocurrency trading and asset management, most approaches [14]–[16] have focused on investments to a single asset. Unlike them, Jiang and Liang [3] attempted to optimize portfolio management using deep learning methods. They presented a model-less convolutional neural network that took historic prices of crypto assets to output portfolio weights, focusing on maximizing accumulative returns. Their method was unique as it applied a full machine-learning approach without relying on pre-existing financial market models, directly capturing market trends and patterns. The results from their study indicated that their CNN [17] strategy notably outperformed other portfolio selection algorithms.

However, there are key areas that could be improved for more robust analysis. Firstly, the duration of the test set used to validate their model's performance was relatively short, approximately two months. This time frame may not be sufficient to capture the wide array of market conditions, especially considering the high volatility and rapid changes characteristic of cryptocurrency markets. Secondly, while they mentioned utilizing a portfolio of 12 cryptocurrencies including Bitcoin, the exact list of these assets was not specified. Including a stable asset like USDT [18] or US Dollar in the portfolio is crucial, given that many cryptocurrencies are heavily influenced by Bitcoin's price trends. The inclusion of a stable asset can provide a hedge against market volatility and should be a key consideration in any portfolio management strategy. Lastly, their model did not include the varying asset allocation ratios at each timestamp as an input. This omission means that the objective function used for training their model does not account for transaction fees that would arise from changes in asset allocation. Transaction fees can significantly impact the net performance of trading strategies, especially in a high-frequency trading context like cryptocurrency markets. The ability to factor in these costs is important for practical portfolio management strategies. Ignoring these fees could result in an overly optimistic assessment of the model's performance and may not accurately reflect the strategy's effectiveness in a real-world trading environment.

III. METHODOLOGY

This section presents our approach using the Advantage Actor-Critic (A2C) reinforcement learning model to distribute assets within a portfolio comprising various cryptocurrencies and US Dollar, aiming to enhance returns and mitigate losses.

We have included Bitcoin (BTC), Bitcoin Cash (BCH) [19], Ethereum (ETH), Ethereum Classic (ETC) [20], and US Dollar (USD) in our asset portfolio. USD can be replaced by USDT, a stablecoin pegged to the USD, under certain conditions. These cryptocurrencies were chosen due to their high market capitalization and the availability of extensive historical data.

A. Price Prediction

We first obtained historical price data of our portfolio's cryptocurrencies from Binance [21], spanning from the end of 2019 to January 2024. As the focus of this paper is not on predicting cryptocurrency prices with high accuracy, we used a simple Long Short-Term Memory (LSTM) model [7] suitable for time series data prediction. Our LSTM model consisted of a LSTM layer with 10 units and a Dense layer with a single unit. We pre-processed the downloaded price data into daily timestamps, with the model predicting the next day's price based on the previous 20 days' data. The data was split into a 70:30 ratio for training and testing.

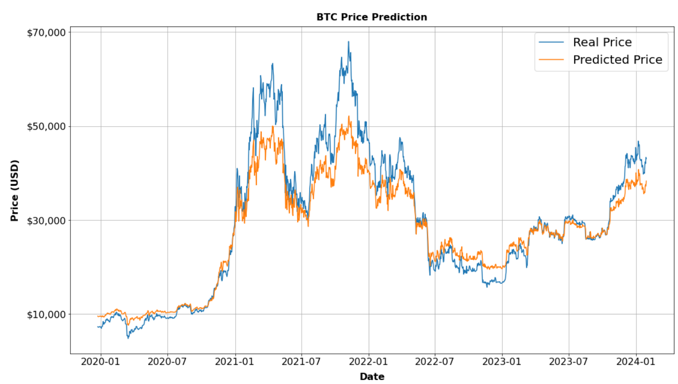


Fig. 1: The Actual and Predicted BTC Prices

Despite its simplicity, the model effectively follows the actual price trends, as shown in Fig. 1. In our subsequent A2C model, we use this predicted price data, accepting some noise, to test our model's effectiveness under less-than-perfect conditions.

B. A2C Model for Crypto Trading and Asset Management

The A2C model that consists of two main components, the Actor and the Critic, is a type of reinforcement learning algorithm that integrates the advantages of both value-based and policy-based methods (Fig. 2). The Actor is in charge of determining the actions, essentially dictating the policy of how an agent should behave in a given state. On the other hand, the Critic scores the action taken by the Actor by computing the value function, which is an estimate of the future rewards.

One of the key features of A2C is its ability to balance between exploring new actions (exploration) and leveraging the knowledge already acquired (exploitation). This balance is crucial in complex environments where the agent must make decisions under uncertainty. A2C improves upon earlier Actor-Critic methods by updating the policy and value function simultaneously, leading to more stable and efficient learning.

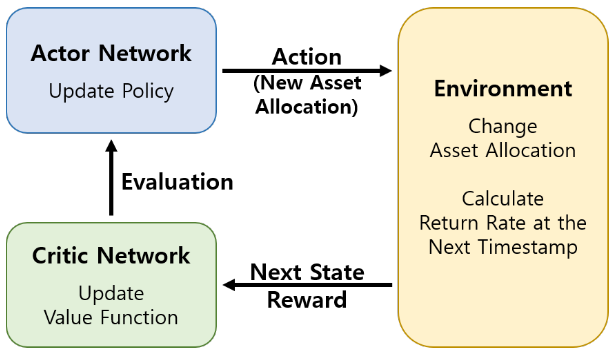


Fig. 2: A2C Model for Crypto Trading and Asset Management

In the context of crypto trading and asset management, A2C's capability to handle continuous action spaces and its robustness in dealing with complex, dynamic environments make it a suitable choice. The model can learn optimal strategies over time, making it ideal for finding optimal asset allocation ratios in our portfolio. The design of our model's state, action, and reward is as follows:

- *State*: Includes cryptocurrency prices for the current and previous 10 timestamps, current asset allocation ratios, and predicted prices for the next timestamp.
- *Action*: New asset allocation ratios for each cryptocurrency and USD, summing to 1.
- *Reward*: The reward function is designed to balance between the portfolio's profitability and market volatility. The reward is calculated as follows:

$$Reward^{(t)}(s^{(t)}, a^{(t)}) = (1 + \alpha^{(t)}) * \gamma^{(t)} - \delta^{(t)} \tag{1}$$

where

$$\begin{aligned} t &= \text{Timestamp} \ , \\ s^{(t)}, a^{(t)} &= \text{State and Action at } t \ , \\ \gamma^{(t)} &= \text{Portfolio's return rate} - \text{trading fees at } t \ , \\ \delta^{(t)} &= \text{Portfolio's avg. price change rate at } t \ , \\ \alpha &= \text{Loss penalty factor} \ , \\ \alpha^{(t)} &= \begin{cases} 0, & \text{if } \gamma^{(t)} \geq 0 \text{ or } \sum_{t-2}^{t} \gamma^{(t)} \geq 0 \ , \\ \alpha_{\text{given}}, & \text{otherwise} \end{cases} \end{aligned}$$

The loss penalty factor is dynamically adjusted based on the performance of the portfolio. It is set to 0 if the portfolio's return rate is non-negative or if the average return rate over the last three timestamps is non-negative. Otherwise, the loss penalty factor is set to a predefined value α. This mechanism is designed to penalize negative returns, encouraging the model to avoid strategies that consistently result in losses. The reason for considering the average return rate over the last three timestamps in the selection of α is to mitigate sensitivity to short-term losses and to smooth the volatility impact.

The inclusion of the average price change rate of the portfolio is also important to assess the portfolio's return rate considering the average market movement. This metric allows for a more nuanced evaluation of the portfolio's performance by contextualizing its returns against the backdrop of market-wide price fluctuations. By using this measure, the model is better equipped to differentiate between genuine portfolio gains and those merely reflective of the overall market trend.

In the calculation of the reward function's γ, we accounted for the occurrence of trading fees. The trading fees were applied to each transaction amount at a predetermined fee rate. This is to prevent the model from making excessively frequent and substantial changes to the asset allocation ratios without considering the trading fees.

The design of this reward function, with its components and dynamic adjustment, aligns with the primary goal of our methodology: optimizing asset allocation to improve returns while managing risk effectively in the volatile environment of cryptocurrency trading.

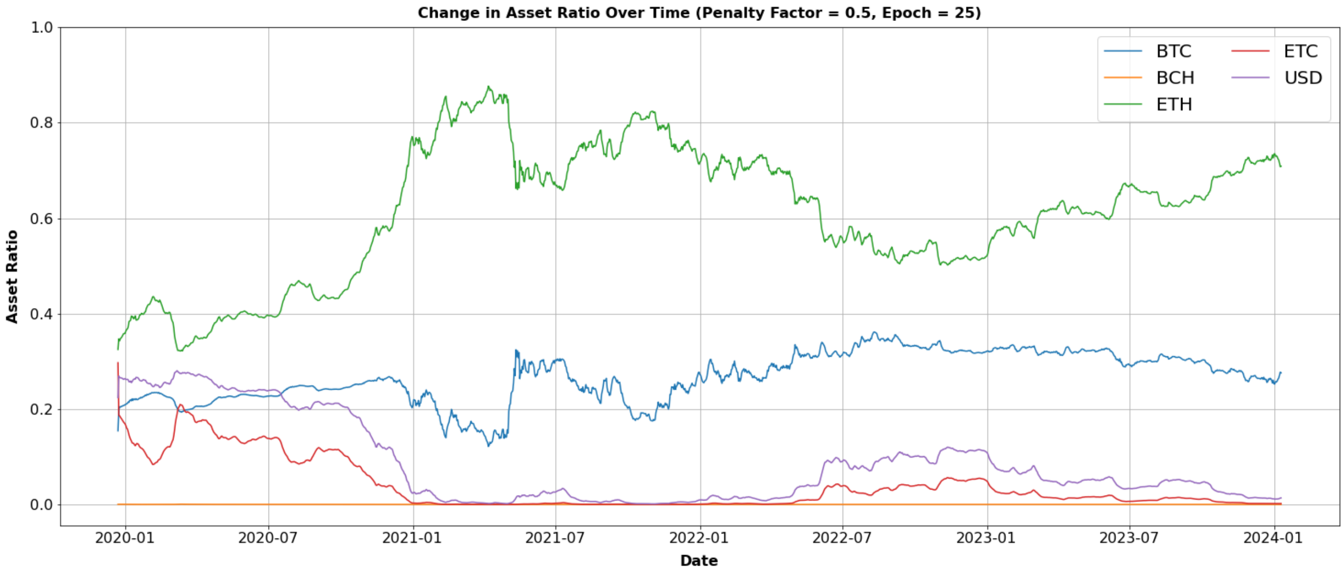
IV. Experiment and Results

In this section, we conducted a comprehensive evaluation of the A2C model that we previously described, using the pre-processed dataset. The focus was to observe how varying the loss penalty factor α in the reward function influenced the behavior of the trained model.

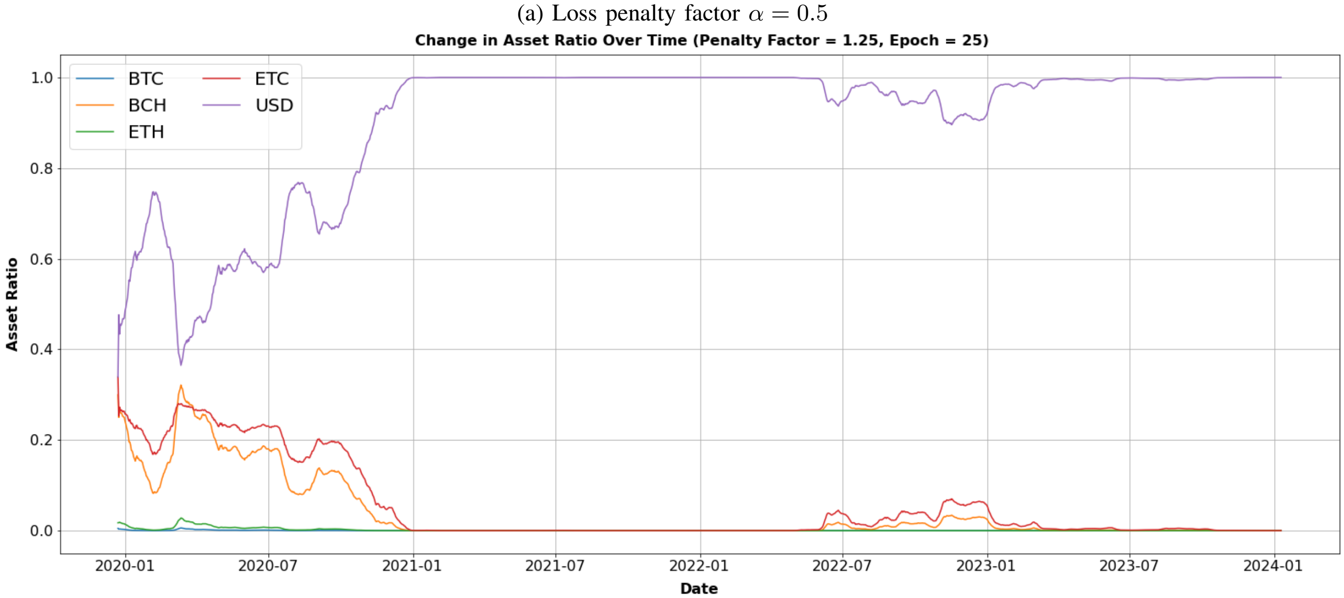
For this experiment, both the Actor and Critic networks in our A2C model were designed with three hidden layers, each containing 256 units. We employed the ReLU [22] activation function for these layers, following standard practices for enhancing non-linear learning capabilities. In the Actor network, we integrated an output layer with 5 units, corresponding to the number of assets (BTC, BCH, ETH, ETC, USD) in our portfolio. This output layer used the softmax activation function, suitable for generating a probability distribution across our portfolio for asset allocation. Conversely, the Critic network was equipped with a single-unit output layer to estimate the value function.

For training purposes, we treated the first 70% of the total data's time period as a training set and iteratively trained the model through multiple episodes (epochs), each being a subset of the training set. The loss penalty factor α was varied across experiments with values of 0.1, 0.5, 0.75, 1, and 1.25. We assumed that the market possesses sufficient liquidity to accommodate the volume of trades the model wish to execute and that the transaction fee rate is set at 0.1%.

As a result, we found that with lower α values, the model engaged in more aggressive investment strategies in pursuit of higher returns. However, one key finding was that during overall market downturns, the model maintained an optimistic outlook, continuing to allocate assets to cryptocurrencies even



(a) Loss penalty factor $\alpha = 0.5$



(b) Loss penalty factor $\alpha = 1.25$

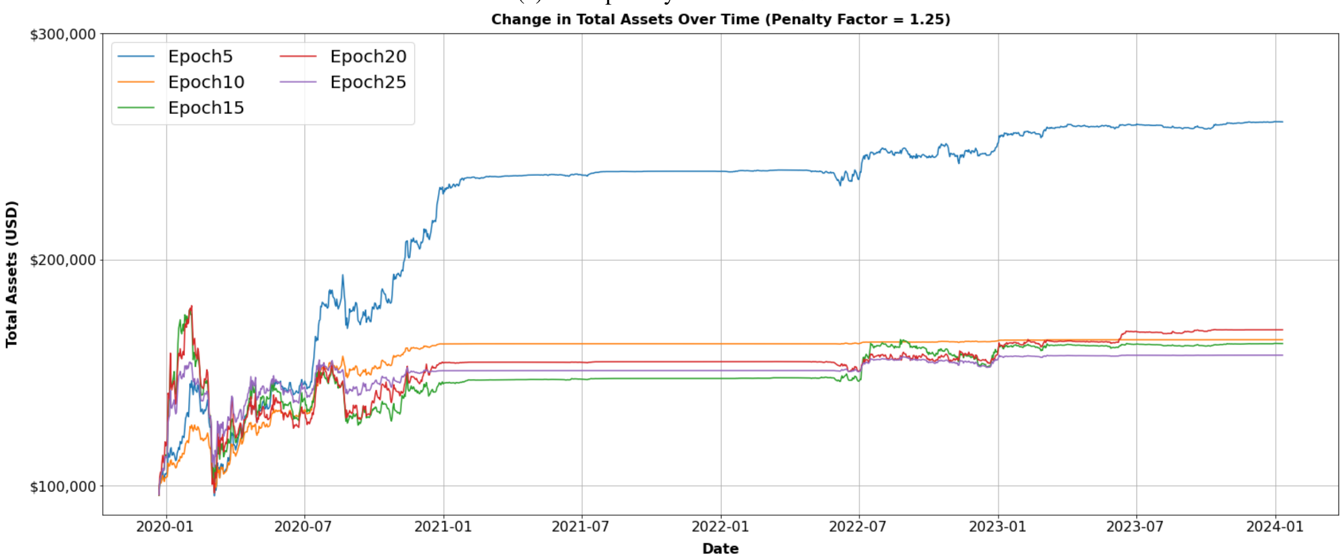
Fig. 3: Asset allocation ratio at each timestamp for the models after being trained through 25 epochs (episodes) with an α of 0.5 and 1.25.

as their prices were falling. This behavior indicates a tendency of the model to remain invested in depreciating assets under certain market conditions, suggesting a potential area for further refinement. Fig. 3(a) illustrates the asset allocation ratio changes at each timestamp when the model with an α of 0.5 was employed. Here, we observed that the model predominantly allocated assets to volatile cryptocurrencies, with minimal allocation to USD. This tendency indicates the model's inclination to pursue potentially higher returns by taking on more risk associated with these cryptocurrencies.

In contrast, as α increased, a trend emerged where the model chose lower returns in exchange for minimizing potential losses. As shown in Fig. 3(b), we noted that most of the portfolio assets were maintained in USD throughout the entire period. This conservative strategy is even more apparent in Fig. 4(b), which shows the total asset value at each timestamp for the model with an α of 1.25. Compared to Fig. 4(a), which represents a model with a relatively lower α, Fig. 4(b) clearly shows a significantly lower return rate but almost negligible asset depreciation.

Fig. 5 shows the total asset value at each timestamp for the model with an α of 1, alongside the price change graphs of each cryptocurrency in the portfolio. Among the models trained with different α values, this particular model exhibited

(a) Loss penalty factor $\alpha = 0.5$



(b) Loss penalty factor $\alpha = 1.25$

Fig. 4: Total asset value at each timestamp for the models after being trained through different epochs (episodes) with an α of 0.5 and 1.25.

the most balanced performance. As observed in Fig. 5, the overall increase in cryptocurrency market prices led to a corresponding increase in the total asset value. Conversely, during periods of price decline, the model managed to mitigate extreme drops in asset value.

This behavior demonstrates the model's capacity to adjust its asset allocation strategies effectively in response to market conditions. In periods of market upswing, the model capitalizes on rising prices, whereas, during downturns, it employs a more cautious approach to safeguard the portfolio's value. Such a balanced approach, as evidenced in the model with an α of 1, demonstrates the model's versatility in navigating the volatile cryptocurrency market, providing a critical insight into optimizing risk-return trade-offs in asset management.

V. Discussion

Through our experiments, we discovered that the loss penalty factor α significantly influences the tendency of our model. By adjusting α, the proposed model can be customized to suit the user's objectives, whether it be chasing higher returns or focusing on loss prevention. However, a notable limitation currently exists in our model's responsiveness to changes in α. We observed that the model tends to shift abruptly between aggressive and conservative investment strategies based on the value of α. This sensitivity necessitates further refinement.

(a) BTC

(b) BCH

(c) ETH

(d) ETC

Fig. 5: (Red line) Total asset value at each timestamp for the model with an α of 1. The price graphs for each cryptocurrency (Blue line) are drawn on a different scale, intended to compare the overall trend in changes.

The implementation of AI in cryptocurrency trading and asset management offers substantial advantages over traditional algorithm-based or manual trading approaches. Though algorithm-based trading automates transactions, it still requires users to make significant decisions regarding the choice of indicators and the amount of trading, often without the benefit of continuous market monitoring. This presents a challenge for most individuals who have other commitments and cannot dedicate time to monitor a 24/7 market [23]. Furthermore, AI has the edge in generating insights from given situations, outperforming the capabilities of manual analysis. AI's ability to process and learn from vast amounts of data enables it to identify opportunities and risks that might be overlooked in manual trading. Hence, AI models like ours, which adapt and respond to market conditions with minimal user intervention, are more advantageous, particularly in the highly volatile and fast-paced cryptocurrency market.

VI. Conclusion

This paper presented an approach in the field of cryptocurrency trading and asset management using the A2C reinforcement learning model. The focus of our research was to optimize asset allocation within a portfolio that includes a mix of high-volatility cryptocurrencies and the more stable US Dollar. Our experiments demonstrated the model's effectiveness and adaptability. By adjusting the loss penalty factor α within the reward function, we observed the model's capability to switch between aggressive and conservative investment strategies in response to the user's intention. The experimental results highlighted a balance between risk and return, with the model achieving stable asset growth in rising markets and mitigating losses in declining markets.

In our future work, we plan to concentrate on two critical aspects of our A2C model for cryptocurrency trading and asset management. Firstly, we intend to analyze how the accuracy of predicted prices impacts the model's performance. If using actual future data in experiments yields better outcomes, it will highlight an opportunity for improvement in our two-step system architecture, which separates price prediction from asset allocation. Enhancing the accuracy of price predictions could boost overall system performance. Conversely, if the differences in performance are minimal, it would indicate that our model can perform comparably well with actual data, even when using predictions with some noise. Such a finding would be essential in verifying the model's applicability in real-world scenarios, where perfect price prediction is not always achievable.

Secondly, we plan to address the observed issue of the model exhibiting extreme performance variations at different alpha values. This will involve revising the reward function and engaging in more thorough hyperparameter tuning. By refining the model's sensitivity to the loss penalty factor, we aim to strike a balance, ensuring a more stable and consistent response to varying market conditions. This approach is expected to lead to a more advanced reward structure, enabling the model to effectively manage the trade-off between risk and reward.

References

[1] M. Waisi, "Advantages and disadvantages of ai-based trading and investing versus traditional methods," https://www.theseus.fi/handle/10024/347449, 2020, accessed: 2024-03-01.

[2] A. Sangadiev, R. Rivera-Castro, K. Stepanov, A. Poddubny, K. Bubenchikov, N. Bekezin, P. Pilyugina, and E. Burnaev, “Deepfolio: Convolutional neural networks for portfolios with limit order book data,” *arXiv preprint arXiv:2008.12152*, 2020.
[3] Z. Jiang and J. Liang, “Cryptocurrency portfolio management with deep reinforcement learning,” in *2017 Intelligent systems conference (IntelliSys)*. IEEE, 2017, pp. 905–913.
[4] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, “Asynchronous methods for deep reinforcement learning,” in *International conference on machine learning*. PMLR, 2016, pp. 1928–1937.
[5] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” *Decentralized Business Review*, p. 21260, 2008.
[6] S. W. Kim, “Performance analysis of bitcoin investment strategy using deep learning,” *Journal of the Korea Convergence Society*, vol. 12, no. 4, pp. 249–258, 2021.
[7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[8] Y. Li, S. Jiang, X. Li, and S. Wang, “Hybrid data decomposition-based deep learning for bitcoin prediction and algorithm trading,” *Financial Innovation*, vol. 8, no. 1, pp. 1–24, 2022.
[9] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
[10] M. Ortu, N. Uras, C. Conversano, S. Bartolucci, and G. Destefanis, “On technical trading and social media indicators for cryptocurrency price classification through deep learning,” *Expert Systems with Applications*, vol. 198, p. 116804, 2022.
[11] V. Buterin *et al.*, “Ethereum white paper,” *GitHub repository*, vol. 1, pp. 22–23, 2013.
[12] GitHub, Available at https://github.com, accessed: 2024-03-01.
[13] Reddit, Available at https://www.reddit.com, accessed: 2024-03-01.
[14] O. Sattarov, A. Muminov, C. W. Lee, H. K. Kang, R. Oh, J. Ahn, H. J. Oh, and H. S. Jeon, “Recommending cryptocurrency trading points with deep reinforcement learning approach,” *Applied Sciences*, vol. 10, no. 4, p. 1506, 2020.
[15] M. Nakano, A. Takahashi, and S. Takahashi, “Bitcoin technical trading with artificial neural network,” *Physica A: Statistical Mechanics and its Applications*, vol. 510, pp. 587–609, 2018.
[16] S. Ji, J. Kim, and H. Im, “A comparative study of bitcoin price prediction using deep learning,” *Mathematics*, vol. 7, no. 10, p. 898, 2019.
[17] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: analysis, applications, and prospects,” *IEEE transactions on neural networks and learning systems*, 2021.
[18] Tether(USDT), Available at https://tether.to, accessed: 2024-03-01.
[19] BitcoinCash, Available at https://bitcoincash.org, accessed: 2024-03-01.
[20] EthereumClassic, Available at https://ethereumclassic.org, accessed: 2024-03-01.
[21] BinanceHistoricalMarketData, Available at https://www.binance.com/en/landing/data, accessed: 2024-03-01.
[22] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
[23] H. Jazayeriy and M. Daryani, “Spa bot: Smart price-action trading bot for cryptocurrency market,” in *2021 12th International Conference on Information and Knowledge Technology (IKT)*. IEEE, 2021, pp. 178–182.