Cryptocurrency Asset Management with Proximal Policy Optimization

Jacob T. Cassady*
Johns Hopkins Whiting School of Engineering, Baltimore, Maryland, 21218

Since the 1990s, algorithms have been used to analyze financial asset markets and make trading decisions. Yet, unique challenges exist when trading on securities markets due to tax regulations like the "wash sale" rule. Cryptocurrencies are considered digital assets, not securities, by the Internal Revenue Service (IRS) and are taxed as property. Therefore, cryptocurrency represents a unique opportunity for experimenting with trading algorithms that would be further constrained in traditional securities markets. This paper analyzes two reinforcement learning models for cryptocurrency trading: a Buy/Sell/Hold (BSH) model and a Managed Risk model. Both models use Proximal Policy Optimization (PPO) to learn trading strategies but have different action spaces and reward functions. The performance of the models is compared to baseline investment strategies and the paper concludes with a discussion on the results and future work.

I. Introduction

Since the 1990s, algorithms have been used to analyze financial asset markets and make trading decisions [1]. Yet, unique challenges exist when trading on security markets due to tax regulations like the "wash sale" rule. The "wash sale" rule disallows a tax deduction for a security sold at a loss and repurchased at a higher price within 30 days before or after the sale [2]. Consequently, security trading algorithms must be designed to avoid wash sales which complicates algorithm design and adds constraints to the action space or a deficit on the reward function.

Cryptocurrencies are considered digital assets, not securities, by the Internal Revenue Service (IRS) and are taxed as property [3]. This means that although there is a different capital gains tax rate, security specific tax regulations such as the wash sale rule do not apply. Therefore, cryptocurrency represents a unique opportunity for experimenting with trading algorithms that would be difficult to implement in traditional securities markets.

This paper analyzes two reinforcement learning models for cryptocurrency trading: a Buy/Sell/Hold (BSH) model and a Managed Risk model. This paper begins in Section II with a description of the models designed for cryptocurrency trading including the feature space, reinforcement learning algorithm, action spaces, and reward functions. Section III details the python implementation of the models with a focus on the leveraged 3rd party libraries. Section IV presents

*Graduate Student, Robotics and Autonomous Systems the performance of the agents on historical data with comparisons to some baseline investment strategies. The paper finishes with a discussion of the results in Section V and an introduction to possible future work in Section VI.

II. Model Design

Two models were tested for cryptocurrency trading: a Buy/Sell/Hold (BSH) model and a Managed Risk model. Both models use the same feature space and reinforcement learning algorithms, but the action spaces and reward functions are different. Section II.A describes the feature space used for the models with a focus on leveraged technical analysis methods. Section II.B describes the reinforcement learning algorithm. Section II.C introduces the action space and reward function of the BSH model. Section II.D details the action space and reward function of the Managed Risk model.

A. Feature Space

Technical analysis indicators are mathematical calculations derived from a financial asset's price and volume data [4]. Traders and algorithms have incorporated technical analysis indicators into their decision making and predictions of future trends and price movements. At each time t, the technical analysis indicators described in Sections II.A.1 through II.A.10 were calculated for the target asset.

This paper assumes the target cryptocurrency's price is influenced by the trends of other large market capital cryptocurrencies. To address this, the feature space included a subset of technical analysis indicators for the other cryptocurrencies tested in this paper. Table 1 shows the technical analysis indicators calculated for target and related assets as well as the hyperparameters used in calculations. The feature space for the BSH and Managed Risk models is a combination of technical analysis indicators for a target asset, technical analysis indicators for related assets, and price data for the target asset.

Technical Analysis Indicator	Hyperparameters	Target Asset	Related Asset(s)
Simple Moving Average	$k = 30$	✓	
Simple Moving Average	$k = 60$	✓	
Relative Strength Index	$k = 14$	✓	✓
Consumer Commodity Index	$k = 30$	✓	
Average Directional index	$k = 30$	✓	
Moving Average Convergence/Divergence		✓	✓
Bollinger Bands	$k = 20, m = 2$	✓	
Average True Range	$k = 14$	✓	
Rate of Change	$k = 10$	✓	
On-Balance Volume	$k = 30$	✓	
Stochastic Oscillator	$k = 14$	✓	

Table 1 Technical Analysis Indicator Hyperparameters

1. Simple Moving Average

The simple moving average (SMA) is an arithmetic average of the last *n* prices of an asset. The equation for simple moving average is shown in equation 1 where k is the number of periods used to calculate the average.

$$SMA_{t,k} = \frac{1}{k}\sum_{i=0}^{k-1} Close_{t-i} \tag{1}$$

2. Relative Strength Index

The relative strength index (RSI) is a momentum oscillator that is designed to indicate the strength of a financial asset market [5]. RSI is calculated using equations 2, 3, and 4 where k is the number of periods used to calculate the average gain (G_t) and average loss (L_t).

$$G_{t,k} = \begin{cases} \frac{1}{k}\sum_{i=0}^{k-1} \max\left(\text{Close}_{t-i+1} - \text{Close}_{t-i}, 0\right), & \text{if } t < k, \\ \\ \frac{(k-1)G_{t-1} + \max(\text{Close}_{t-i+1} - \text{Close}_{t-i}, 0)}{k}, & \text{if } t \geq k. \end{cases} \tag{2}$$

$$L_{t,k} = \begin{cases} \frac{1}{k}\sum_{i=0}^{k-1} \max\left(\text{Close}_{t-i} - \text{Close}_{t-i+1}, 0\right), & \text{if } t < k, \\ \\ \frac{(k-1)L_{t-1} + \max(\text{Close}_{t-i} - \text{Close}_{t-i+1}, 0)}{k}, & \text{if } t \geq k. \end{cases} \tag{3}$$

$$RSI_{t,k} = \begin{cases} 100 - \frac{100}{1+\frac{G_{t,k}}{L_{t,k}}}, & \text{if } t < k, \\ \\ 100 - \frac{100}{1+\frac{(G_{t-1,k}\cdot k)+G_{t,k}}{(L_{t-1,k}\cdot k)+L_{t,k}}}, & \text{if } t \geq k. \end{cases} \tag{4}$$

3. Commodity Channel Index

The commodity channel index (CCI) compares the current price of an asset to an average price over a period [6]. CCI is calculated using equations 5, 6, and7 where k is the number of periods used to calculate the typical price and mean deviation.

$$\text{Typical Price}_t = \frac{\text{High}_t + \text{Low}_t + \text{Close}_t}{3} \tag{5}$$

$$\text{Mean Deviation}_{t,k} = \frac{1}{k}\sum_{i=0}^{k-1} \left|\text{Typical Price}_{t-i} - \text{SMA}_{t,k}\right| \tag{6}$$

$$CCI_{t,k} = \frac{\text{Typical Price}_t - \text{SMA}_{t,k}}{0.015 \cdot \text{Mean Deviation}_{t,k}} \tag{7}$$

4. Average Directional Index

The average directional index (ADX) is a momentum oscillator designed to indicate financial asset market trend strength like RSI [5]. ADX is calculated using equation 8 where k is the number of periods.

$$ADX_t = \frac{1}{k} \sum_{i=0}^{k-1} \frac{\left|\text{High}_{t-i} - \text{Low}_{t-i}\right|}{\text{High}_{t-i}} \tag{8}$$

5. Moving Average Convergence/Divergence

The moving average convergence/divergence (MACD) technical indicator made up of two exponential moving averages (EMA) [7]. The equation for EMA is shown in equation 9, where k is the number of periods used to calculate the average. MACD is the difference between what are called “fast” (k=12) and “slow” (k=26) averages at the same time as shown in equation 10. Historically traders have analyzed when MACD$_t$ crosses Signal Line$_t$, shown in equation 11, as a buy or sell signal.

$$EMA_{t,k}(x_t) = \left(x_t \cdot \frac{2}{k+1}\right) + \text{EMA}_{t-1,k}\left(1 - \frac{2}{k+1}\right) \tag{9}$$

$$MACD_t = EMA_{t,12}(\text{Close}_t) - EMA_{t,26}(\text{Close}_t) \tag{10}$$

$$\text{Signal Line}_t = EMA_{t,9}\left(\text{MACD}_t\right) \tag{11}$$

6. Bollinger Bands

Bollinger Bands are meant to provide a metric for the volatility of a financial asset [8]. The upper and lower bands represent an m number of standard deviations from the simple moving average as shown in equations 12 and 13 where k is the number of periods used to calculate the average and standard deviation.

$$\text{Upper Band}_{t,k} = SMA_{t,k} + m \cdot \sigma_{t,k} \tag{12}$$

$$\text{Lower Band}_{t,k} = SMA_{t,k} - m \cdot \sigma_{t,k} \tag{13}$$

7. Average True Range

The average true range (ATR) technical analysis indicator is designed to measure volatility in a financial asset markets [5]. ATR is calculated by taking an average of the true range (TR) which captures the largest range of price movement in a single period. The equations for TR and ATR are shown in equations 14 and 15 respectively where k is the number of periods.

$$\mathrm{TR}_t = \max\left(\mathrm{High}_t - \mathrm{Low}_t, \left|\mathrm{High}_t - \mathrm{Close}_{t-1}\right|, \left|\mathrm{Low}_t - \mathrm{Close}_{t-1}\right|\right) \tag{14}$$

$$\mathrm{ATR}_t = \frac{1}{k}\sum_{i=0}^{k-1} \mathrm{TR}_{t-i} \tag{15}$$

8. Rate of Change

The rate of change (ROC) measures the percentage change in price from one period to the next [9]. The equation for ROC is shown in equation 16 where k is the number of periods.

$$\mathrm{ROC}_{t,k} = \left(\frac{\mathrm{Close}_t - \mathrm{Close}_{t-k}}{\mathrm{Close}_{t-k}}\right) \times 100 \tag{16}$$

9. On-Balance Volume

On-Balance Volume (OBV) is a momentum indicator that combines volume and price changes to predict buying and selling pressures in financial asset markets [9]. OBV is calculated using equation 17.

$$\mathrm{OBV}_t = \mathrm{OBV}_{t-1} + \mathrm{Volume}_t \cdot \mathrm{Sign}\left(p_t - p_{t-1}\right) \tag{17}$$

10. Stochastic Oscillator

The Stochastic Oscillator compares the closing price to the range of prices over a period of k [4]. The equation for Stochastic Oscillator is shown in equation 18.

$$\text{Stochastic Oscillator}_{t,k} = \left(\frac{\mathrm{Close}_t - \mathrm{Low}_k}{\mathrm{High}_k - \mathrm{Low}_k}\right) \tag{18}$$

B. Algorithm Selection

Algorithm selection for BSH and Managed Risk models done by leveraging the work of Mohammadshafie, et. al which tested the performance of deep learning reinforcement strategies on stock market trading [10]. Mohammadshafie's team tested the performance of Soft Actor-Critic (SAC) [11], Asynchronous Advantage Actor-Critic (A2C) [12], Deep

Deterministic Policy Gradient [13], Twin Delayed Deep Deterministic Policy Gradient (TD3) [14], and Proximal Policy Optimization (PPO) [15] on a set of stocks from a variety of sectors. The results of their study are highlighted in figure 1 and show superior results from PPO, TD3, and A2C. Mohammadshafie, et al. notes the volume of stock trades overtime were higher for PPO than A2C and for this reason PPO was chosen as the algorithm for the BSH and Managed Risk models. Although Mohammadshafie, et al. ignores the tax implications of the wash sale rule, the results of their study are very much relevant to cryptocurrency trading.

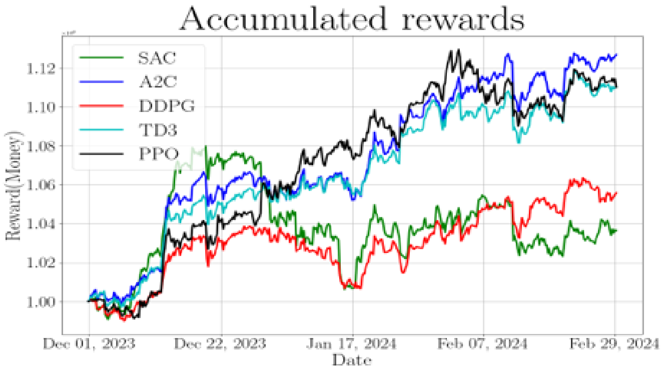

Fig. 1 Performance of Reinforcement Learning Strategies on Financial Asset Trading [10]

C. Buy/Sell/Hold Agent

The Buy/Sell/Hold (BSH) agent is designed to be the simplest trading agent possible. The action space, described in Section II.C.1, is binary. The reward function, described in Section II.C.2, is based on the position of the agent at time *t* and the change in closing price.

1. BSH Action Space

The BSH agent has a discrete action space as shown in table 2. The agent can choose to invest in cash or the asset at each time step. Figure 3 shows an example of the BSH action space. Rising edges on the graph are buy signals and falling edges are sell signals.

Action	Description
0	Invested in cash.
1	Invested in asset.

Table 2 Buy/Sell/Hold Agent Action Space

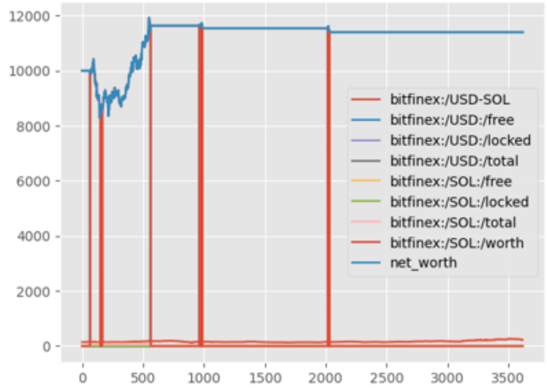

Fig. 2 BSH Action Plot Example

2. *Position-Based Reward Function*

The reward function for the BSH agent is based on the position of the agent (x_t) at time t and the change in closing price ($Close_t$). The reward function is shown in equation 19 where x_t is the position of the agent at time t.

$$R_t = (\text{Close}_t - \text{Close}_{t-1}) \cdot x_t \tag{19}$$

D. Managed Risk Agent

The Managed Risk agent is designed to use limit orders to manage risk and maximize returns. The action space, described in Section II.D.1, is a combination of **stop loss**, **take profit**, and trade size values. The reward function, described in Section II.D.2, is based on the Sortino ratio which measures volatility as a function of downside risk.

1. *Limit Order Action Space*

The Managed Risk agent has a discrete action space as shown in table 3. Stop loss and take profit values are percentages at time t relative to the change in closing price from the time a limit order was placed. Stop loss values represent the maximum loss before selling and take profit values represent the maximum profit before selling. Trade size values are fractions of the total account value. Figure 3 shows an example of the Managed Risk action space.

Action Parameter	Description	Values used
Stop Loss	Maximum loss before selling	[2%, 5%, 10%]
Take Profit	Maximum profit before selling.	[1%, 5%, 10%, 15%]
Trade Size	Size of next trade.	$[\frac{1}{16}, \frac{2}{16}, \ldots, 1]$

Table 3 Limit Order Agent Action Space

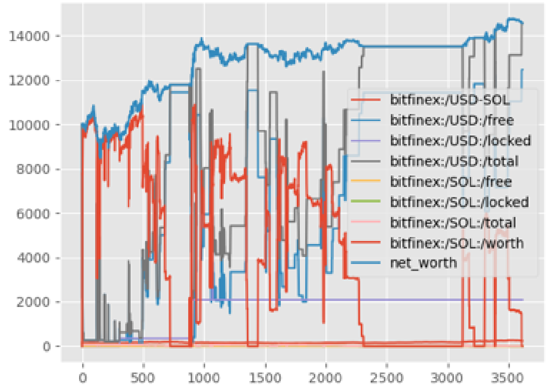

Fig. 3 Limit Order Action Plot Example

2. Sortino Ratio Reward Function

The reward function for the Managed Risk agent is based on the Sortino ratio which measures volatility as a function of downside risk [16]. The Sortino ratio is calculated using equation 21 where x_t is the position of the agent at time t, r_f is the risk-free rate of return, and σ_d, shown in equation 20, is the standard deviation of all changes in price that are less than the risk free rate.

$$\sigma_d = \sqrt{\frac{1}{k}\sum_{i=0}^{k-1} \min\left(0, \text{Close}_{t-i} - \text{Close}_{t-i-1} - r_f\right)^2} \tag{20}$$

$$R_t = \frac{\frac{x_t - x_{t-1}}{x_{t-1}} - r_f}{\sigma_d} \tag{21}$$

III. Implementation

The models described above were implemented in Python by leveraging third party libraries shown in table 4. NumPy and pandas were used for data manipulation. pandas_ta was leveraged for technical analysis calculations across pandas dataframes. TensorTrade provides models for wallets, exchanges, assets, OpenAi Gym Environments, Reward Schemes, and Action Schemes as well as an API for accessing historical data from cryptocurrency exchanges. Stable-Baseline3 provides a suite of reinforcement learning algorithms for training and testing agents.

3rd Party Libraries	Usage
NumPy [17]	Data manipulation and computations.
pandas [18]	Data storage and manipulation.
pandas_ta [19]	Technical Analysis computation.
TensorTrade [20]	Models for exchanges, assets, OpenAi Gym Environments.
stable_baselines3 [21]	Reinforcement Learning algorithm implementations.

Table 4 3rd Party Libraries

IV. Analysis

To test the implementations, a python script was written following the implementation details above to train and test 100 BSH and Managed Risk models on 6 different cryptocurrencies and compare the results of the best models to baseline investment strategies. The models were tested on Bitcoin (BTC) [22], Ethereum (ETH) [23], Cardano (ADA) [24], Solana (SOL) [25], Litecoin (LTC) [26], and Tron (TRX) [27] cryptocurrency markets with data from the Bitfinex exchange.

Some currencies had larger datasets than others due to the age of the cryptocurrency. The models were split into train and test sets with the most recent 5 months of data used for testing and the remainder of the data held for training. Agents were given $\$10,000.00$ to start for training and testing. Table 5 shows the sizes of the train and test sets for each cryptocurrency.

Cryptocurrency	Training Set Size	Test Set Size
BTC	53611	3673
ETH	53603	3673
ADA	18431	3673
SOL	18460	3673
LTC	52612	3673
TRX	52231	3618

Table 5 Train and Test Sizes

The number of training steps was randomly selected from the list [500, 1000, 5000, 10000, 15000]. Each model's window size for PPO was randomly selected for the list [96, 72, 48, 24, 12]. Due to the lack of hard research providing standards for these values, random selection was used to allow for a variety of model results. The best model performance on each cryptocurrency is shown in Figure 4. Baseline investment strategies for investing in a savings account at 4% interest or holding the asset for the entire period were plotted next to model results for comparison. The number of training steps and window size for the best models are shown in Table 6.

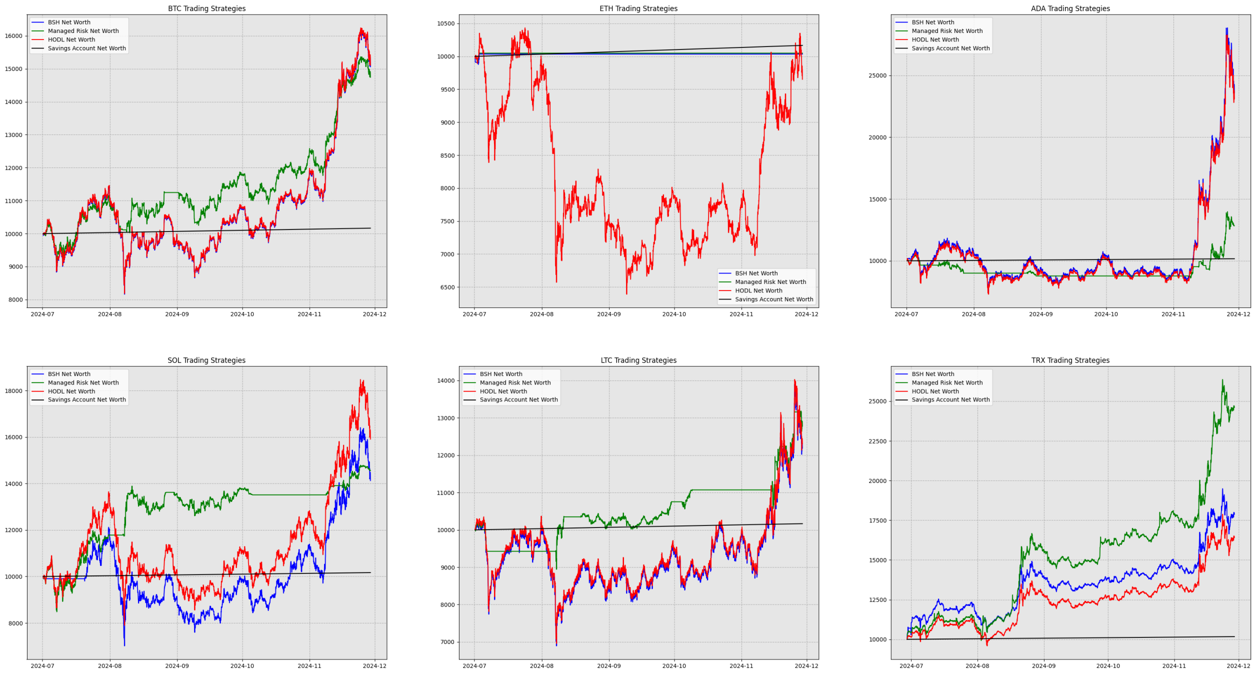

Fig. 4 Results by Cryptocurrency. The green line is the Managed Risk model's performance. The blue line is the BSH model's performance. The red line is the net worth if the agent bought the asset and held the entire time. The black line is the performance of a savings account with 4% interest.

Cryptocurrency	$\text{BSH}_{\text{training-steps}}$	$\text{BSH}_{\text{window-size}}$	$\text{Managed Risk}_{\text{training-steps}}$	$\text{Managed Risk}_{\text{window-size}}$
BTC	500	96	15000	72
ETH	50000	48	500	12
ADA	1000	48	15000	12
SOL	500	96	15000	12
LTC	1000	72	500	72
TRX	5000	48	10000	24

Table 6 Best Model Hyperparameters

To compare the performance of the models across all cryptocurrencies, the results from Figure 4 were aggregated and plotted in Figure 5.

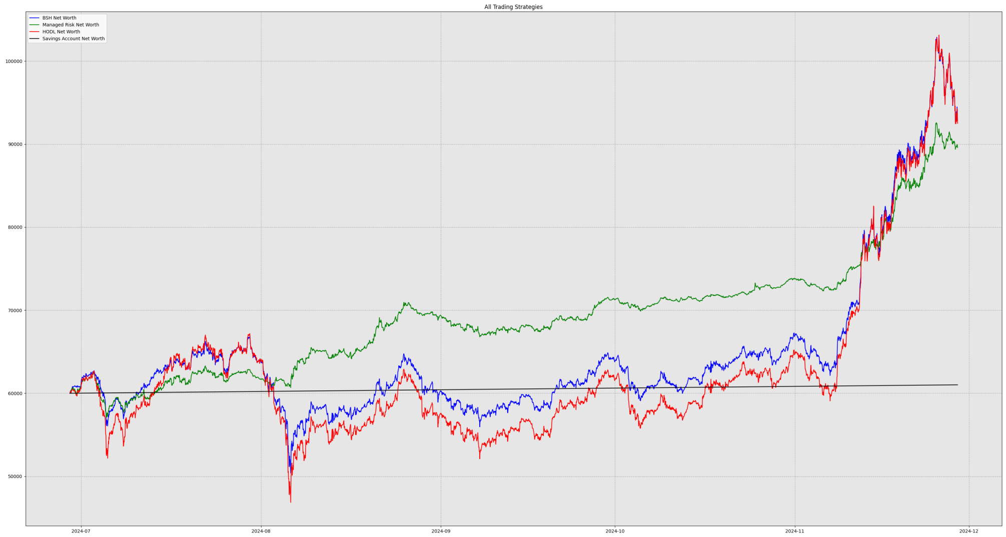

Fig. 5 Aggregate Results. The green line is the Managed Risk model's performance. The blue like is the BSH model's performance. The red line is the net worth if the agent bought the asset and held the entire time. The black line is the performance of a savings account with 4% interest.

V. Discussion

The results show that, on average, the Managed Risk model performed better than the BSH model, holding the asset, or keeping the money in a high yield savings account. The BSH model performed similarly, and slightly better, than holding the assets. Since the BSH model followed the real price signal so closely, its success is likely secondary to the recent uptrend in price and not to the well placed market orders. On the other hand, the Managed Risk model was able to protect the agent from large drops in price and take advantage of up trends in price.

VI. Future Work

The models tested in this paper have room for improvement. Currently, the feature space for the models is limited to a subset of technical analysis indicators. Including fundamental analysis indicators and sentiment analysis could improve the performance. Additionally, no feature ranking has been done to verify relevance of indicators. Dimensionality reduction techniques like Principal Component Analysis (PCA) could help identify the most relevant features or be used as a transform to reduce feature space size and therefore reduce training time [28].

The models have been tested and deployed on a centralized exchange where the US Dollar (USD) is the base currency. Another option would be to deploy on a decentralized token exchange where the base currency is a cryptocurrency such as Minswap [29]. The fees on decentralized exchanges are typically lower and the prices are more volatile. Since the Managed Risk models identify, and protecting from, large drops in price, it may perform well on one of these decentralized token exchanges.

Acknowledgements

This work was performed as a final paper for the Foundations of Reinforcement Learning class at Johns Hopkins University taught by Professor Enrique Garcia Mallada.

References

[1] Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E., and Liu, M., “Application of evolutionary computation for rule discovery in stock algorithmic trading: A literature review,” *Applied Soft Computing*, Vol. 36, 2015, pp. 534–551. https://doi.org/https://doi.org/10.1016/j.asoc.2015.07.008, URL https://www.sciencedirect.com/science/article/pii/S156849461500438X.

[2] Internal Revenue Service, “Publication 550 (2023), Investment Income and Expenses,” , 2023. URL https://www.irs.gov/publications/p550.

[3] Internal Revenue Service, “Digital assets,” , 2024. URL https://www.irs.gov/businesses/small-businesses-self-employed/digital-assets.

[4] Murphy, J. J., *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*, Penguin, 1999.

[5] Wilder, J., *New Concepts in Technical Trading Systems*, Trend Research, 1978. URL https://books.google.com/books?id=WesJAQAAMAAJ.

[6] Lambert, D. R., “Commodity channel index: Tool for trading cyclic trends,” *Technical Analysis of Stocks & Commodities*, Vol. 1, 1983, p. 47.

[7] Appel, G., *Technical analysis: power tools for active investors*, FT Press, 2005.

[8] Bollinger, J., *Bollinger on Bollinger bands*, McGraw-Hill New York, 2002.

[9] Granville, J., *New Key to Stock Market Profits*, Martino Publishing, 1963.

[10] Mohammadshafie, A., Mirzaeinia, A., Jumakhan, H., and Mirzaeinia, A., “Deep Reinforcement Learning Strategies in Finance: Insights into Asset Holding, Trading Behavior, and Purchase Diversity,” , June 2024. https://doi.org/10.48550/arXiv.2407.09557.

[11] Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., and Levine, S., “Soft Actor-Critic Algorithms and Applications,” , 2019. URL https://arxiv.org/abs/1812.05905.

[12] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., Silver, D., and Kavukcuoglu, K., “Asynchronous Methods for Deep Reinforcement Learning,” , 2016. URL https://arxiv.org/abs/1602.01783.

[13] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D., “Continuous control with deep reinforcement learning,” , 2019. URL https://arxiv.org/abs/1509.02971.

[14] Fujimoto, S., van Hoof, H., and Meger, D., “Addressing Function Approximation Error in Actor-Critic Methods,” , 2018. URL https://arxiv.org/abs/1802.09477.

[15] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O., “Proximal Policy Optimization Algorithms,” , 2017. URL https://arxiv.org/abs/1707.06347.

[16] Sortino, F. A., and Price, L. N., “Performance measurement in a downside risk framework,” *the Journal of Investing*, Vol. 3, No. 3, 1994, pp. 59–64.

[17] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E., “Array programming with NumPy,” *Nature*, Vol. 585, No. 7825, 2020, pp. 357–362. https://doi.org/10.1038/s41586-020-2649-2, URL https://doi.org/10.1038/s41586-020-2649-2.

[18] The pandas development team, “pandas-dev/pandas: Pandas,” , Feb. 2020. https://doi.org/10.5281/zenodo.3509134, URL https://doi.org/10.5281/zenodo.3509134.

[19] “pandas-ta,” , 2024. URL https://github.com/twopirllc/pandas-ta.

[20] “TensorTrade,” , 2024. URL https://www.tensortrade.org/en/latest/index.html.

[21] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., and Dormann, N., “Stable-Baselines3: Reliable Reinforcement Learning Implementations,” *Journal of Machine Learning Research*, Vol. 22, No. 268, 2021, pp. 1–8. URL http://jmlr.org/papers/v22/20-1364.html.

[22] Nakamoto, S., “Bitcoin: A peer-to-peer electronic cash system,” *Satoshi Nakamoto*, 2008.

[23] Buterin, V., “Ethereum Whitepaper,” , 2013. URL https://ethereum.org/en/whitepaper/, accessed: December 2024.

[24] Cardano Foundation, “Cardano: Peer-reviewed Blockchain Platform,” , 2024. URL https://cardano.org, accessed: December 2024.

[25] Solana Labs, “Solana,” , 2017. URL https://solana.com, accessed: December 2024.

[26] Lee, C., “Litecoin: Open Source Peer-to-Peer Digital Currency,” , 2011. URL https://litecoin.org, accessed: December 2024.

[27] Sun, J., “TRON Decentralized Network,” , 2017. URL https://tron.network, accessed: December 2024.

[28] Hotelling, H., “Analysis of a complex of statistical variables into principal components.” *Journal of educational psychology*, Vol. 24, No. 6, 1933, p. 417.

[29] Nguyen, L., “MIP-1 MinSwap-Multi-pool decentralized exchange on Cardano,” 2021.