JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 40, 745-761 (2024)
DOI: 10.6688/JISE.202407_40(4).0004

Stock Recommendation Model with Investor Risk Acceptance

HEI-CHIA WANG[1,2,3,+], YI-HSIN CHENG[1,2] AND YI-HSUAN WU[3]
[1]*Institute of Information Management*
[2]*Center for Innovative FinTech Business Models*
[3]*Department of Industrial and Information Management*
National Cheng Kung University
Tainan, 701 Taiwan
E-mail: hcwang@mail.ncku.edu.tw

In this era of high inflation and low interest rates, the public often invests in financial products to increase their passive income and thus increase their savings. Current data show that the public still considers stocks as investment targets. However, because investment is risky and each person's investment personality (risk tolerance) is different, some prefer to take risks to obtain the maximum return, and some are afraid of risks and avoid them to obtain stable returns; nevertheless, the recommendations of current research give little consideration to the risk characteristics of the investment target and the personal characteristics of the investor. However, investment is a trade-off between risk and reward, and the risk acceptance associated with an investment depends on the investor's ability to accept risk. Therefore, in the context of recommending, a consideration of investors' personal characteristics brings recommendations more in line with users' expectations.

This study proposes a method to manage investment portfolios based on investors' risk personalities. It is mainly used to classify stocks according to beta indicators (select stocks according to investors' personalities), and financial indicators and an autoencoder are used to score stocks; moreover, technical indicators, covariance matrices, deep reinforcement learning-advantage actor critic (A2C), and proximal policy optimization (PPO) are used for asset allocation with the aim of developing investment portfolios with good performance that are optimally suitable for different types of investors.

In the training results, the cumulative return rate obtained by the A2C model is as high as 49.61% for conservative investments, 82.04% for stable investments, and 99.69% for active investments. The cumulative return obtained by the PPO model is as high as 39.92% for conservative investments, 89.89% for stable investments, and 85.61% for active investments.

Keywords: financial portfolio, deep reinforcement learning, asset allocation, A2C, PPO

1. INTRODUCTION

Central banks around the world are implementing negative interest rate policies [1]. In this era of high inflation and low interest rates, people who previously saved their money are tending to invest in other products. Current data show that the public still considers stocks as investment targets. The main purpose of investing is to obtain maximum returns. If the goal is to reduce risk and obtain the maximum return, it is not optimal to invest in a single stock. Constructing an optimal investment portfolio requires the use of different asset classes and weights according to investors' risk tolerance.

Through a review of the literature on investment portfolios, two distinct types of portfolio construction methods have been identified: (1) methods that are designed to accommodate different levels of risk tolerance and (2) dynamic investment portfolio methods.

Received April 20, 2023; revised May 25 & June 13, 2023; accepted June 27, 2023.
Communicated by Mu-Yen Chen.

The former approach is typically based on investors' personal risk profiles and seeks to tailor the composition of a portfolio to the corresponding individual's unique risk tolerance level. The latter approach, on the other hand, is characterized by an ongoing process of portfolio adjustment that is designed to take advantage of changes in market conditions over time.

A fundamental aspect of portfolio construction is the classification of stock risk, which has been traditionally categorized into three types based on risk index beta values. Specifically, the three categories are conservative, stable, and active. This classification system enables investors and financial advisors to develop investment strategies that are tailored to both their specific risk preferences and their overall investment objectives [2]. The creation of an appropriate investment portfolio necessitates a consideration of not only the fluctuations in the stock market but also the idiosyncrasies of investors' personalities. A well-crafted portfolio should be aligned with the investor's risk tolerance, investment goals, and time horizon. These personal factors can significantly influence investment decisions and should not be overlooked in the portfolio construction process. As such, both financial advisors and investors must thoughtfully examine the unique attributes of individual investors when designing portfolios that are tailored to meet both their financial objectives and their psychological dispositions [3]. In choosing the right stocks for a portfolio, one can create a stock scoring mechanism to evaluate candidate stocks based on their fundamental and technical characteristics with the aim of obtaining the maximum profit [4]. Dimensionality reduction and feature selection are performed using constrained stacked autoencoders, ensuring that only the most informative abstract features are retained to reduce risk [5-8]. Therefore, in this study, we hope to combine the advantages of both, aiming to screen out low-risk stocks from a high-return list.

After we construct a portfolio from a list of stocks, we address portfolio trading as an optimization problem involving a sequential decision-making process across multiple rebalancing cycles. Deep reinforcement learning (DRL) has been widely used in the financial industry by many scholars in previous literature [7, 9-13]. Ma *et al.* [10] presented the TC-MARL algorithm, a multi-agent deep reinforcement learning approach for portfolio optimization. This approach creates two agents with different reward functions but sharing the same policy and value models. Ngo's research [13] findings illustrate the superiority of reinforcement learning models over traditional optimization models in terms of cumulative returns and the Sharpe ratio within the Vietnamese and U.S. securities markets. This empirical evidence underscores the dynamic capabilities inherent in reinforcement learning and its transformative potential in reshaping risk management practices within the financial industry. Moreover, it can be used to explore the high-dimensional and nonlinear characteristics of the stock market. In this study, we use the advantage actor critic (A2C) [14] and proximal policy optimization (PPO) methods for asset allocation. The purpose of this research is delineated by the following objectives: (1) Propose a method to effectively manage investment portfolios that align with investors' risk personalities. This approach aims to tailor investment strategies to individual risk preferences, considering the unique risk tolerance and appetite of each investor; (2) Highlight the significance of portfolio decomposition in addition to portfolio optimization. Recognizing the interconnectedness between portfolio composition and returns, this research emphasizes the importance of understanding the composition of portfolios and its direct correlation with investment returns; (3) Investigate how the returns of portfolios with different risk profiles are influenced by

distinct financial environments. By analyzing the performance of portfolios associated with varying risk types, this research aims to uncover the relationship between investment returns and specific financial circumstances, providing insights into the impact of different market conditions on portfolio outcomes.

The remainder of this paper is organized as follows. Section 2 offers a succinct review of the extant literature concerning portfolio management with a specific focus on investor risk personality and DRL algorithms. In Section 3, we introduce our proposed method for developing a stock recommendation model, which considers investors' acceptance of risk. The subsequent section, Section 4, outlines the experimental procedures employed to evaluate the performance of different models and presents a comparative analysis of the outcomes. Finally, Section 5 provides a comprehensive summary of the findings obtained.

2. RELATED WORK

This section provides a comprehensive review of the relevant literature on portfolio management involving investor risk personality utilizing the principles of DRL.

2.1 Construction of Investment Portfolios

Portfolio construction, which involves selecting stocks to include in a portfolio, is the initial step in the investment process. Research has been conducted on identifying and selecting stocks to optimize portfolio performance. For instance, Hajjami and Amin [15] proposed a two-dimensional framework for stock selection that considers the perspectives of investors seeking high-yield stocks and creditors prioritizing a company's maximum repayment ability. This investment strategy has been demonstrated to be feasible in practical implementation. Similarly, Yang *et al.* [16] developed a novel model for stock selection that utilizes a multifactor valuation approach incorporating financial indicators and stock forecasting techniques to capture the future trajectory of stocks.

In a related study, Lee and Woo [17] introduced the concept of cohesion in the stock market network, which is induced by the correlation of stocks, as a seminal measure. This measure indicates that the characteristics of a stock network are related to stock returns and that its dynamics can be utilized to construct a stock portfolio and predict changes in the stock markets. Li *et al.* [18] proposed a collective intelligence mechanism that leverages social media data to extract and consolidate opinions expressed over a social investing platform. This approach creates investment portfolios by analyzing the knowledge, authority, and opinions of other investors related to the investment target. Experimental results obtained from eToro.com demonstrate the superiority of this method over benchmark approaches in various financial performance aspects.

Inspired by the findings of these previous studies, we aim to utilize the approach of categorizing stocks into three groups and subsequently applying a financial statement analysis model to select appropriate stocks for inclusion in a portfolio.

2.2 Deep Learning in Investment

Predicting stock returns is a time series forecasting problem due to daily changes in stock prices. A neural network consisting of more than three layers, including inputs and outputs, can be thought of as a deep learning (DL) algorithm. DL is merely a subset of ma-

chine learning. Many researchers have conducted extensive studies on this topic [19-21]. Kuo *et al.* [22] presented a system with a genetic algorithm based on a fuzzy neural network and an artificial neural network that achieved commendable buy-sell performance. Hargreaves [23] introduced a method for stock portfolio selection using neural network and logistic regression approaches for data mining. The method was found to outperform the market index in the healthcare and financial sectors, with improvements of 18% and 1%, respectively. Vo *et al.* [24] proposed a deep responsible investment portfolio (DRIP) model containing a multivariate bidirectional long short-term memory neural network to predict stock returns for the construction of a socially responsible investment portfolio.

2.3 Deep Reinforcement Learning in Investment

Reinforcement learning (RL) is a machine learning technique that enables an agent to learn by trial and error in a changing environment and to adjust actions based on observations. Almahdi and Yang [25] used a hybrid method that combined recurrent RL and particle swarm optimization to derive a portfolio trading strategy considering real-world constraints.

Deep neural networks (DNNs) can enhance RL algorithms, but the combination of online RL algorithms with DNNs was initially considered unstable. Various approaches have been proposed to stabilize the algorithm [26-28], such as storing the agent's data in experience replay memory and randomly sampling it from different time steps to reduce nonstationarity and decorrelate updates. These methods are limited to off-policy RL algorithms. If this approach is used to aggregate over memory, the performance of RL algorithms using DNNs can be improved. DRL is a combination of deep learning and reinforcement learning.

DRL algorithms that employ experience replay have demonstrated exceptional performance in challenging domains such as Atari 2600. Nonetheless, experience replay suffers from several drawbacks. First, it entails increased memory usage and computational requirements per real interaction. Second, it necessitates the use of off-policy learning algorithms, which can update on the basis of data produced by a different policy [14]. Numerous reinforcement learning algorithms have been employed in the domain of portfolio management. Some studies have utilized deep Q-learning [9, 29], while others have employed PPO [30, 31].

Katongo *et al.* [32] sought to address the tactical asset allocation problem by using the A2C method in the context of the US stock market. Vishal *et al.* [33] used actor-critic-based RL methods such as PPO, deep deterministic policy gradient (DDPG), A2C, and twin delayed DDPG (TD3) in the Indian stock market. Gunawan *et al.* [34] chose PPO to predict buy and sell stock market prices. Both A2C and PPO belong to the family of DRL algorithms. Multiple academic studies have shown that these algorithms consistently produce results that are superior in terms of cumulative profit to those of other techniques.

3. PROPOSED APPROACH

The primary objective of this paper is to present a comprehensive approach for managing investment portfolios based on investors' risk personalities to obtain maximum returns. Fig. 1 provides an overview of our proposed approach.

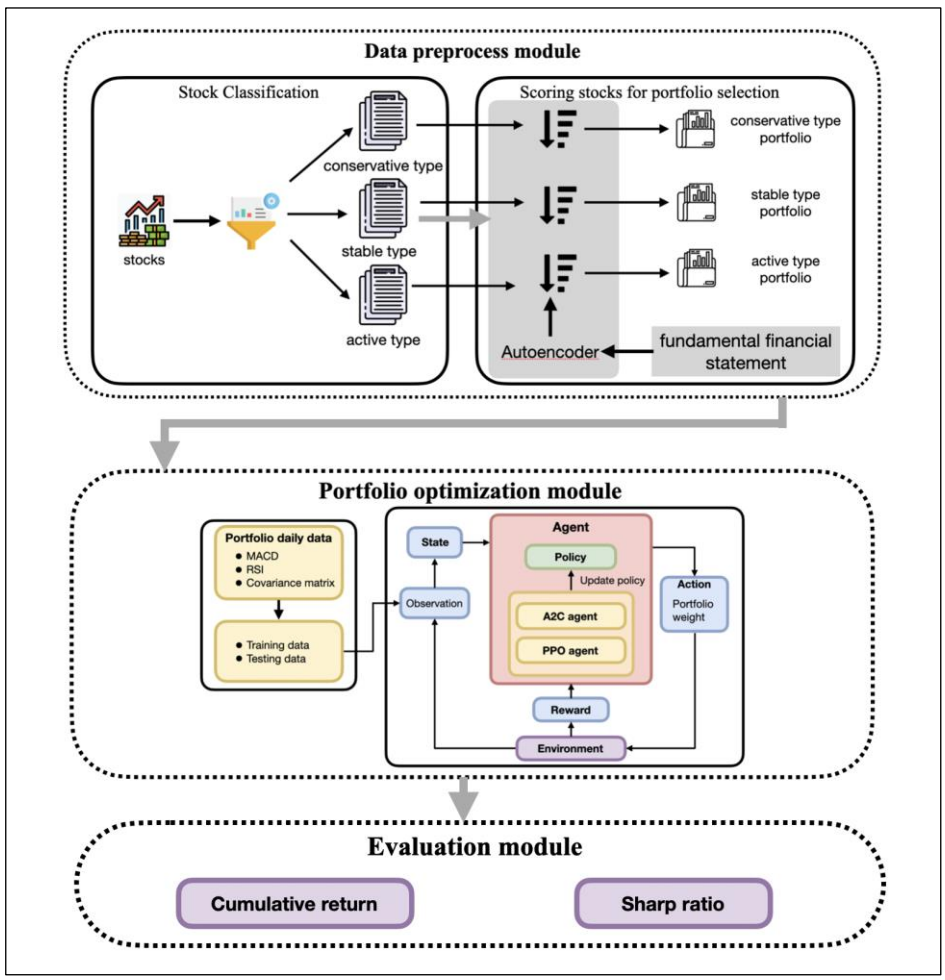

Fig. 1. An overview of our proposed approach.

The data pre-processing phase in this study encompasses two primary components. Firstly, all stocks are classified into three distinct groups based on the calculation of their beta values. Subsequently, fundamental financial data are utilized along with an autoencoder to assign scores to the stocks and identify those that meet the specified investment criteria, thereby selecting them for inclusion in the portfolio. Following the construction of the portfolio, a reinforcement learning algorithm known as Deep Reinforcement Learning (DRL) is employed as a key component of the stock recommendation model, facilitating portfolio optimization.

3.1 Stock Classification

To this end, the level of stock risk is first categorized into three distinct groups, which are determined based on the beta values of the risk indexes and are detailed in Table 1. The beta value of each stock is calculated according to Eq. (1), which is an essential aspect of

this approach. By incorporating investors' risk personalities into the investment management process, this methodology seeks to enhance the efficiency and effectiveness of portfolio management and ultimately yield superior returns.

Table 1. Performance of three types of portfolios.

Stock Classification	Rule
conservative type	Beta < 0.75
stable type	1.25<= Beta <=0.75
active type	Beta > 1.25

$$Beta_i = \frac{Covariance(r_i, r_m)}{Varience(r_m)} \tag{1}$$

The beta values are calculated as the covariance between the return r_i of the stock and the return r_m of the market index divided by the variance of the market index (over a period of three years) [18]. To illustrate this, let's consider the calculation of the beta of Apple (AAPL) compared to the SPDR S&P 500 ETF Trust (SPY). Given a covariance of 0.022 between AAPL and SPY, along with a variance of SPY amounting to 0.017, we can apply these values to the appropriate formula. By plugging them into the formula, we can determine the beta of AAPL relative to SPY, providing insight into the stock's volatility compared to the broader market. The computed beta value for AAPL is determined to be 1.294. Based on this finding, it can be inferred that AAPL falls within the category of an active stock type.

3.2 Scoring Stocks for Portfolio Selection

As the first step, in the previous section, we categorized the stocks into conservative, stable, and active groups; here, we utilize the candidate lists for each category. Once the stock classification is obtained, we use fundamental financial data and an autoencoder to score the stocks and select the ones that meet our investment criteria for inclusion in our portfolio. To evaluate the stocks, we use various financial ratios such as return on assets, the debt-to-asset ratio, the cash flow ratio, the inventory turnover ratio, and the receivables turnover ratio [15, 16, 18]. We score and rank the stocks based on their financial performance, which aids in identifying the top-performing stocks within each category. The closing price of stock data is reconstructed by passing it through the autoencoder, and then the reconstruction error of each stock is calculated. We select stocks with lower reconstruction errors for our portfolio [5]. Both the input layer and the output layer are the closing price of the stock. The structure of the autoencoder is shown in Fig. 2.

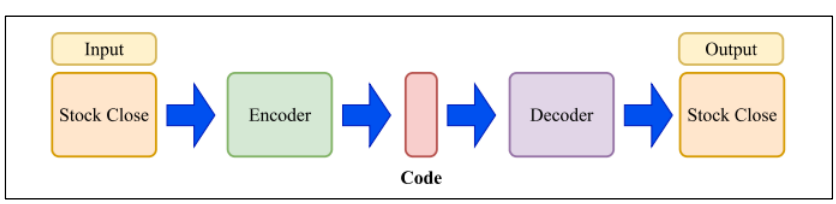

Fig. 2. The structure of autoencoder.

3.3 Deep Reinforcement Learning

In this paper, we employ a DRL algorithm as a component of our stock recommendation model to perform portfolio optimization. Our research architecture for deep reinforcement learning in portfolio strategy is shown in Fig. 3. We add the covariance matrix and technical indicators of each stock as feature selections to the DRL algorithm. Referring to previous research [12, 32, 35], we choose the technical indicators and related parameters shown in Table 2.

Table 2. Summary of technical indicators.

Technical indicator	Parameters
Moving Average Convergence Divergence (MACD)	Fast MA Period: 12 Slow MA Period: 26
Relative Strength Index (RSI)	Period: 10

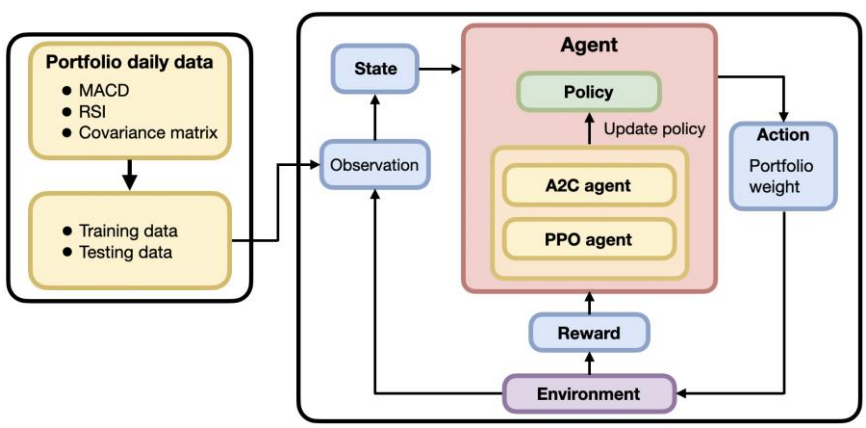

Fig. 3. A Research architecture for deep reinforcement learning in portfolio strategy.

There are four integral components that are crucial for the effective utilization of DRL; (1) Environment: The simulation environment is a critical component in the context of this study. Specifically, the Gym framework provided by OpenAI has been utilized for this purpose. Markov decision process (MDP) modeling has been utilized to design simulation environments; (2) State space: The state space is expanded to encompass technical indicators and the covariance matrix. The purpose of this expansion is to incorporate the intrinsic informational characteristics of the stock market into the state space; (3) Action Space: the action space is formulated as the portfolio weights of each stock. The action space is constrained to the range of 0 to 1, and the sum of all actions is set to 1; (4) Reward function: The reward function in this study is based on the return rate (R_p) of the portfolio that is computed at each time step. In other words, when the agent observes a state (S_t), takes an action (A_t), and transitions to a new state (S_{t+1}), the return rate of the portfolio is used to compute the reward. The details of the portfolio management process using DRL are shown by the pseudocode in Algorithm 1.

Algorithm 1: Portfolio Management Using Deep Reinforcement Learning

Input: *s*, state space includes technical indicators and covariance matrix for stocks.
Output: Portfolio Value

1. Initialize P_0 = \$1,000,000, $W_0 = (\frac{1}{m}, ..., \frac{1}{m})$, P_0 is the initial portfolio value, W_0 is the initial portfolio weight, m is number of stocks in the portfolio, *Unit* represents the quantity of shares purchased in a single lot, *Cost* is the transaction cost;
2. **for** $t = 1, \ldots$ **do**
   - 2.1 Agent Observe a state as s and outputs a portfolio weights vector W_t as actions;
   - 2.2 Normalize The weight W_t to sum to 1, $\sum W_t = 1$;
   - 2.3 AM_t is the quantity that can be purchased by weight,
     $Close_t$ *represents the closing price per share of the stock at time t*
     $AM_t = (P_0 * W_t) / (Clos_t * Unit)$;
   - 2.4 Calculate portfolio period return – R_p
     $R_p = \sum((Close_t - Close_{t-k}) * AM_{t-k} * Unit)$
     $- (Close_{t-k} * AM_{t-k} * Unit * Cost)$
     $- (Close_t * AM_{t-k} * Unit * Cost)$
     , k is trading day;
   - 2.5 Update portfolio value $P_t = P_{t-1} + R_p$;

end for

4. EXPERIMENT AND EVALUATION

In this section, we verify the accuracy and reliability of the proposed stock recommendation model considering investor risk acceptance through a simulation and comparison of its performance with several experimental schemes.

4.1 Simulation Environment

To evaluate the effectiveness of our stock recommendation model considering investor risk acceptance for financial portfolio management, we conduct experiments on a real-world stock market dataset. The dataset includes the historical stock prices of various assets, as well as economic indicators and news sentiment data that may affect stock market performance.

We used Python, NumPy, pandas, and the scikit-learn machine learning library to implement and train the A2C and PPO models on this dataset. To accelerate the computation process, we used a PC with an Intel four-core CPU 2.7G, DDR4 16G RAM, and Ubuntu Desktop 20.04.5 LTS operating system. The CPU was manufactured by Intel Corporation (located in Santa Clara, California, USA) and the RAM was manufactured by ASUS (located in Taiwan). The experimental environment is further described in Table 3. In the data preprocessing phase, we need to construct three different types of portfolios. The process of obtaining three different risk types of portfolios from the stock market, using a dataset consisting of 911 stocks and one year of related data, requires approximately 5 mins. This timeframe accounts for the necessary computations and analyses involved in categorizing the stocks based on their risk profiles and generating the corresponding port-folios. During the training stage, which utilizes a dataset spanning ten years of data, the A2C agent takes

around 0.5 minutes to train, while the PPO agent requires approximately 1.5 minutes. In the prediction stage, both agents perform similarly, taking approximately 3 seconds to generate results. For our study, the daily stock price data are sourced from Yahoo Finance, a well-known and widely used platform for accessing financial data. Yahoo Finance provides historical stock price information, including open, high, low, and closing prices, as well as trading volumes, for a wide range of publicly traded companies. On the other hand, the financial statement data utilized in our research is obtained from the Taiwan Market Observatory System. The Taiwan Market Observatory System is a reputable and authoritative source for financial information in Taiwan. It provides comprehensive and reliable financial statements of companies listed on the Taiwanese stock exchange, including income statements, balance sheets, and cash flow statements.

Table 3. Simulation environment.

Numerical and Machine Learning Package	
Python 3.9.2	scikit-learn
	Stable-baseline3
	NumPy
	SciPy
	Tensorflow
	Keras
	pandas

4.2 Simulation Parameters

This paper follows the DRL parameter settings proposed by Katongo *et al.* [32]. During the experiment, we tried to reduce the total time step size of the A2C and PPO agents and found that the results did not get worse when the value was reduced to 10,000. The final parameter setting is depicted in Table 4.

Table 4. Summary of parameter settings.

	Parameters
A2C Agent	ent_coef: 0.005
	learning rate: 0.0002
	total time steps: 10,000
PPO Agent	ent_coef: 0. 005
	learning_rate: 0.0001
	batch_size: 128
	total time steps: 10,000
AutoEncoder	batch_size = 300
	Epochs = 5
	hidden_layers = 5

4.3 Evaluation Metrics

To test our proposed approach, we use cumulative returns and the Sharpe ratio to perform the evaluation. As the calculation of the annualized rate of return over a single year is equivalent to that of the cumulative rate of return over the same period, the an-

nualized rate of return is not appropriate for evaluating performance when returns from multiple years are combined. Therefore, in this study, we employ the cumulative rate of return and the Sharpe ratio to compare research results and assess performance. The cumulative rate of return represents the total profits over a specified time period. Moreover, the Sharpe ratio measures the amount of excess return generated per unit of risk accepted by investors.

4.4 Experimental Results

Experiment 1: Different stock quantities in portfolio

This experiment explores the performance of portfolios with 5 and 10 stocks and is validated with 10 years of training and 1 year of back testing. The experimental results in Table 5 show that the 5-year average cumulative return on the portfolio of 5 assets was higher than that on the portfolio of 10 assets.

Table 5. Portfolio results with different stock quantities.

A2C Agent	portfolio – 5 stocks		
	Conservative	Stable	Active
Cumul. return	20.65%	8.21%	28.46%
Sharp ratio	0.2758	0.1946	0.1813
A2C Agent	portfolio – 10 stocks		
	Conservative	Stable	Active
Cumul. return	12.73%	7.39%	12.27%
Sharp ratio	0.265	0.2356	0.1239

Experiment 2: Portfolios can maximize returns with a few days trading strategy

An analysis of the experimental results in Table 6 shows that in these three risk categories, trades occur every 15 days (buying on the first day and selling on the 15th day), and the cumulative return obtained is better than that obtained over other numbers of trading days. Therefore, we calculate the cumulative return every 15 days.

Table 6. Cumulative return with different trading strategies.

type	days\year	2016	2017	2018	2019	2020	avg
Conservative	5 days	–4.60%	39.55%	0.68%	16.68%	10.56%	12.57%
	10 days	5.27%	45.98%	–3.46%	22.13%	6.92%	15.37%
	15 days	7.67%	49.61%	6.40%	22.74%	16.82%	20.65%
	monthly	6.69%	59.39%	–2.68%	20.50%	14.54%	19.69%
Stable	5 days	6.93%	6.39%	19.15%	7.34%	–0.24%	7.92%
	10 days	11.57%	11.30%	12.32%	7.79%	–5.29%	7.54%
	15 days	10.03%	8.38%	7.27%	10.63%	4.73%	8.21%
	monthly	13.95%	7.44%	7.75%	8.16%	0.95%	7.65%
Active	5 days	20.47%	17.30%	–17.70%	6.21%	30.04%	11.26%
	10 days	38.90%	35.42%	–18.41%	16.00%	33.25%	21.03%
	15 days	32.29%	44.18%	–21.63%	24.56%	62.98%	28.47%
	monthly	49.11%	33.41%	–18.58%	21.41%	56.95%	28.46%

Experiment 3: Portfolio performance comparison between the A2C agent and the PPO agent.

In this experimental scheme, we use different training periods for the A2C agent and the PPO agent and calculate the returns.

Table 7 presents a comprehensive summary of the profits generated by both the A2C and PPO agents in the conservative portfolio, and it shows the 5-year average cumulative return and the Sharpe ratio obtained after 1 to 10 years of training. The return of the A2C agent is better than that of the PPO agent for both 5-year and 10-year training. However, in the other training periods, the PPO agent performs better than the A2C agent, as shown in Fig. 4. In terms of overall performance, the PPO agent performs better than the A2C agent.

Table 7. Profit summary table of conservative type portfolio.

Training period	A2C Agent		PPO Agent	
	Cumul. return	Sharp ratio	Cumul. return	Sharp ratio
1 year	8.75%	0.1341	9.48%	0.1489
2 years	8.66%	0.1933	9.92%	0.2170
3 years	9.93%	0.2208	10.51%	0.2250
4 years	7.87%	0.2294	8.95%	0.2900
5 years	**8.46%**	**0.1723**	5.12%	0.1156
6 years	12.60%	0.1982	15.17%	0.2420
7 years	19.60%	0.2443	23.50%	0.3083
8 years	16.71%	0.2123	19.25%	0.2994
9 years	14.12%	0.1809	17.90%	0.2102
10 years	**20.65%**	**0.2759**	15.54%	0.2771

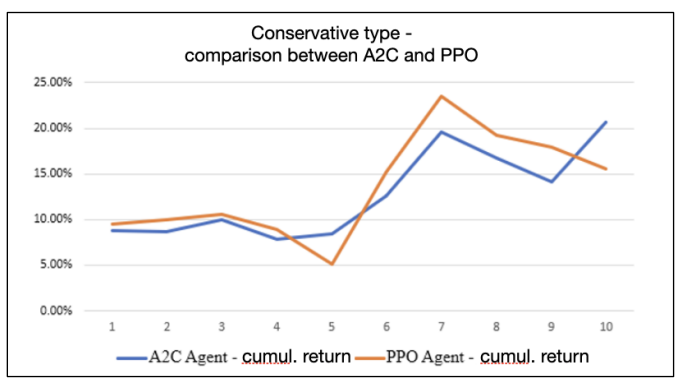

Fig. 4. Conservative type with different training period.

Table 8 is a profit summary table in the stable portfolio generated by both the A2C and PPO agents. The return of the A2C agent is better than that of the PPO agent for 4-year to 7-year training. However, in the other training periods, the PPO agent performs better than the A2C agent, as shown in Fig. 5. In terms of overall performance, the PPO agent performs better than the A2C agent.

Table 8. Profit summary table of stable type portfolio.

Training period	A2C Agent		PPO Agent	
	Cumul. return	Sharp ratio	Cumul. return	Sharp ratio
1 year	21.74%	0.2321	23.94%	0.2498
2 years	6.71%	0.1333	6.80%	0.1542
3 years	13.14%	0.3078	13.65%	0.2755
4 years	**10.49%**	**0.1857**	10.39%	0.1579
5 years	**11.02%**	**0.2048**	10.78%	0.1996
6 years	**7.87%**	**0.1603**	6.61%	0.1338
7 years	**7.26%**	**0.1089**	5.93%	0.1155
8 years	8.53%	0.2221	10.64%	0.2381
9 years	8.87%	0.2110	9.01%	0.2195
10 years	8.21%	0.1946	9.77%	0.2459

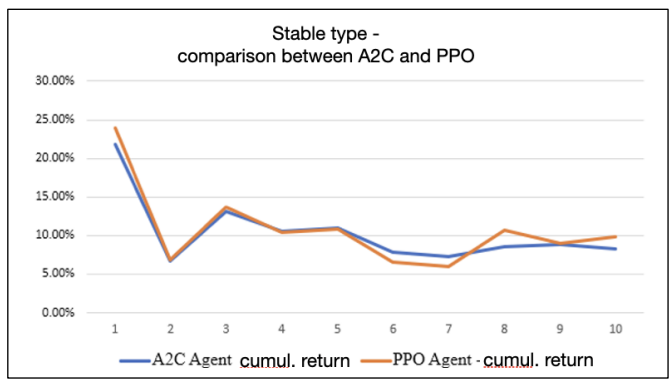

Fig. 5. Stable type with different training period.

Table 9 is a profit summary table in the active portfolio generated by both the A2C and PPO agents. The return of the A2C agent is better than that of the PPO agent for 6-year, 8-year and 9-year training. However, in the other training periods, the PPO agent performs better than the A2C agent, as shown in Fig. 6. In terms of overall performance, the PPO agent performs better than the A2C agent.

Table 9. Profit summary table of active type portfolio.

Training period	A2C Agent		PPO Agent	
	Cumul. return	Sharp ratio	Cumul. return	Sharp ratio
1 year	17.13%	0.1292	19.57%	0.1913
2 years	21.48%	0.1702	22.42%	0.1952
3 years	21.44%	0.2129	23.41%	0.2317
4 years	23.50%	0.2219	25.68%	0.2697
5 years	22.49%	0.2273	24.11%	0.2218
6 years	**15.95%**	0.1029	14.76%	0.1356
7 years	24.48%	0.2154	27.45%	0.2804
8 years	**19.14%**	0.1835	16.34%	0.1475
9 years	**45.74%**	0.2868	45.45%	0.2894
10 years	28.46%	0.1813	30.79%	0.2793

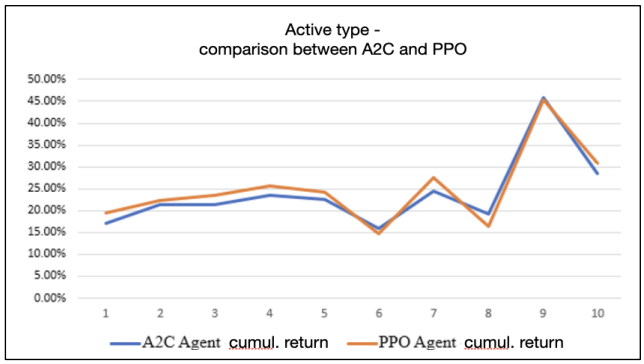

Fig. 6. active type with different training period.

Experiment 4: Portfolio performance comparison between the A2C agent and PPO agent

We select Taiwanese stocks with historical data covering 4 years for the experiment. Daily stock data from 01/01/2005 to 12/31/2019 are used for training, and data from 01/01/2016 to 12/21/2020 are used for validation. In this experiment, we want to compare the performance of the three types of portfolios using the A2C agent and PPO agent.

The Sharpe ratio specifies the relative performance of an equity investment compared to the rate of return on a risk-free investment. For this evaluation, we use not only cumulative returns but also the Sharpe ratio. The performance results are shown in Table 10.

Table 10. Performance of three types of portfolios.

	Conservative type				Stable type				Active type			
	A2C		PPO		A2C		PPO		A2C		PPO	
year	Cumul. return	sharp ratio	Cumul. return	sharp ratio	Cumul. return	sharp ratio	Cumul. return	sharp ratio	Cumul. return	sharp ratio	Cumul. return	sharp ratio
2016	7.7%	0.095	5.4%	0.066	12.7%	0.186	8.9%	0.134	52.4%	0.406	58.9%	0.433
2017	49.6%	0.558	39.9%	0.604	12.0%	0.267	14.2%	0.317	99.7%	0.556	85.6%	0.605
2018	6.4%	0.026	0.8%	0.007	–8.8%	–0.171	–7.6%	–0.151	–22.3%	–0.282	–22.2%	–0.282
2019	22.7%	0.452	20.4%	0.524	10.7%	0.405	14.4%	0.528	62.1%	0.503	66.3%	0.466
2020	16.8%	0.246	11.2%	0.183	82.0%	0.472	89.9%	0.420	36.9%	0.249	38.6%	0.223
average	20.7%	0.275	15.5%	0.277	21.7%	0.232	23.9%	0.249	45.7%	0.286	45.5%	0.289

4.5 Discussion

In the investment process, the initial step involves selecting suitable investment targets to be included in the investment portfolio. While numerous studies [10, 12] employing DRL have focused on optimizing portfolios during the investment process, there has been limited exploration of altering the portfolio's composition. Certain papers [8, 11] emphasize the consideration of risk but primarily focus on the actions taken to control and mitigate that risk, rather than specifically addressing portfolio construction. These papers prioritize the development and implementation of strategies or measures aimed at managing and minimizing potential risks within a given context. While portfolio construction is an essential aspect of risk management, these papers concentrate on the tactical steps taken to address risks rather than the overall composition and allocation of assets within a portfolio.

By emphasizing risk control actions, these papers contribute to the broader understanding and implementation of risk management practices in specific domains or scenarios.

In Experiment 1, this study aims to investigate the efficacy of investment portfolios with varying numbers of investment targets, recognizing that different portfolio effects can arise due to the influence of initial capital. Additionally, recent research has begun incorporating transaction costs, as excessive trading can erode profits. Consequently, Experiment 2 demonstrates that employing a 15-day trading strategy can yield improved results, even when dealing with investment portfolios of varying risk types. Given that diverse risk portfolios may respond differently to varying training periods, the objective is to identify an optimal training duration that enhances portfolio performance.

Experiment 3 focused on evaluating the relative performance of two DRL agents applied to portfolios with distinct risk characteristics. The results indicated that, with few exceptions, the A2C agent exhibited subpar performance across most training sessions. In contrast, most training sessions demonstrated superior performance by the PPO agent. Notably, the performance of the PPO agent improved with longer training sessions, suggesting a positive correlation between training duration and performance outcomes.

In Experiment 4, the focus shifted to assessing the annual cumulative returns and Sharpe ratios of investment portfolios with three different risk types. The findings revealed that portfolios with varying risk profiles exhibited distinct performance characteristics in each year. For instance, in 2018, the active risk type experienced negative returns, while the conservative risk type managed to maintain a positive return. Conversely, in 2019, the active risk type achieved a cumulative return of 62%, surpassing the performance of the conservative risk type during the same year. While it is not possible to predict future trends, there appears to be a discernible business cycle pattern. Following a poor performing year, investing in the active risk type may be considered, while in years of positive performance, allocating to the conservative risk type could be favorable. However, considering the experimental data spanning five years, the overall cumulative reward was a minimum of 15%.

5. CONCLUSIONS

In this paper, we propose a stock recommendation model considering investor risk acceptance. The cumulative rate of return obtained by the A2C model in the training results is as high as 49.61% for the conservative portfolio, 82.04% for the stable portfolio, and 99.69% for the active portfolio. The cumulative return obtained by the PPO model is as high as 39.92% for the conservative portfolio, 89.89% for the stable portfolio, and 85.61% for the active portfolio. Based on the results of the research and analysis, the following conclusions can be drawn and future work proposed:

(1) The investment return in 2018 was negative. One reason could be that there are only stocks in the portfolio and no other types of assets such as bonds. Another reason is that the Taiwanese stock market is heavily influenced by the US stock market, but the model does not account for this.
(2) Different market conditions can affect the performance of portfolios with different risk types.
(3) The stock price is the final market result presenting some event. It is important to un-

derstand how the factors influencing these events are summarized and integrated into the model.

(4) In this paper, we only use A2C and PPO models. Another DRL algorithm could be integrated into our model or used in place of one of the existing models.

(5) Since the stable portfolio performs best in the one-year training set, as the training set increases, the performance will worsen. Previous work has shown that the more training data there are, the better performance becomes. Thus, we recommend that the beta value is reduced to 2 years. The expected robust portfolio can increase returns as the training set increases.

REFERENCES

1. C. P. Chandrasekhar, "Negative interest rates: Symptom of crisis or instrument for recovery," in *Economic and Political Weekly*, 2017, pp. 53-60.
2. M. Foglia, M. C. Recchioni, and G. Polinesi, "Smart beta allocation and macro-economic variables: The impact of COVID-19," *Risks*, Vol. 9, 2021, p. 34.
3. J.-R. Yu, *et al.*, "Dynamic rebalancing portfolio models with analyses of investor sentiment," *International Review of Economics & Finance*, Vol. 77, 2022, pp. 1-13.
4. L. Yu, L. Hu, and L. Tang, "Stock selection with a novel sigmoid-based mixed discrete-continuous differential evolution algorithm," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, 2016, pp. 1891-1904.
5. F. Soleymani and E. Paquet, "Financial portfolio optimization with online deep reinforcement learning and restricted stacked autoencoder – DeepBreath," *Expert Systems with Applications*, Vol. 156, 2020, p. 113456.
6. S. Y. Lin and H. Y. Lin, "Bond price prediction using technical indicators and machine learning techniques," *Journal of Information Science and Engineering*, Vol. 39, 2023, pp. 439-455.
7. J. Li, *et al.*, "Online portfolio management via deep reinforcement learning with high-frequency data," *Information Processing & Management*, Vol. 60, 2023, p. 103247.
8. J. M.-T. Wu, *et al.*, "Embedded draw-down constraint reward function for deep reinforcement learning," *Applied Soft Computing*, Vol. 125, 2022, p. 109150.
9. H. Park, M. K. Sim, and D. G. Choi, "An intelligent financial portfolio trading strategy using deep Q-learning," *Expert Systems with Applications*, Vol. 158, 2020, p. 16.
10. C. Ma, *et al.*, "Multi-agent deep reinforcement learning algorithm with trend consistency regularization for portfolio management," *Neural Computing & Applications*, Vol. 35, 2023, pp. 6589-6601.
11. C. Jiang and J. Wang, "A portfolio model with risk control policy based on deep reinforcement learning," *Mathematics*, Vol. 11, 2023, p. 19.
12. J. Jang and N. Seong, "Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory," *Expert Systems with Applications*, Vol. 218, 2023, p. 119556.
13. V. M. Ngo, H. H. Nguyen, and P. van Nguyen, "Does reinforcement learning outperform deep learning and traditional portfolio optimization models in frontier and developed financial markets?" *Research in International Business and Finance*, Vol. 65, 2023, p. 101936.
14. V. Mnih, *et al.*, "Asynchronous methods for deep reinforcement learning," in *Proceed-*

ings of the 33rd International Conference on International Conference on Machine Learning, Vol. 48, 2016, pp. 1928-1937.
15. M. Hajjami and G. R. Amin, "Modelling stock selection using ordered weighted averaging operator," *International Journal of Intelligent Systems*, Vol. 33, 2018, pp. 2283-2292.
16. F. Yang, *et al.*, "A novel hybrid stock selection method with stock prediction," *Applied Soft Computing*,Vol. 80, 2019, pp. 820-831.
17. Y.-J. Lee and G. Woo, "Analyzing the dynamics of stock networks for recommending stock portfolio," *Journal of Information Science and Engineering*, Vol. 35, 2019, pp. 411-427.
18. Y.-M. Li, *et al.*, "A social investing approach for portfolio recommendation," *Information & Management*, Vol. 58, 2021, p. 103536.
19. W.-C. Chiang, *et al.*, "An adaptive stock index trading decision support system," *Expert Systems with Applications*, Vol. 59, 2016, pp. 195-207.
20. A. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, "Stock market index prediction using artificial neural network," *Journal of Economics, Finance and Administrative Science*, Vol. 21, 2016, pp. 89-93.
21. A. A. Elhag and A. M. Almarashi, "Forecasting based on some statistical and machine learning methods," *Journal of Information Science and Engineering*, Vol. 36, 2020, pp. 1167-1177.
22. R. J. Kuo, C. H. Chen, and Y. C. Hwang, "An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network," *Fuzzy Sets and Systems*, Vol. 118, 2001, pp. 21-45.
23. C. A. Hargreaves, P. Dixit, and A. Solanki, "Stock portfolio selection using data mining approach," *IOSR Journal of Engineering*, Vol. 3, 2013, pp. 42-48.
24. N. N. Y. Vo, *et al.*, "Deep learning for decision making and the optimization of socially responsible investments and portfolio," *Decision Support Systems*, Vol. 124, 2019, pp. 1-11.
25. S. Almahdi and S. Y. Yang, "An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown," *Expert Systems with Applications*, Vol. 87, 2017, pp. 267-279.
26. J. Schulman, *et al.*, "Trust region policy optimization," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 1889-1897.
27. H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 2016, pp. 2094-2100.
28. V. Mnih, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, Vol. 518, 2015, pp. 529-533.
29. G. Lucarelli and M. Borrotti, "A deep Q-learning portfolio management framework for the cryptocurrency market," *Neural Computing & Applications*, Vol. 32, 2020, pp. 17229-17244.
30. J. Du, M. Jin, P. N. Kolm, G. Ritter, Y. Wang, and B. Zhang, "Deep reinforcement learning for option replication and hedging," *The Journal of Financial Data Science*, Vol. 2, 2020, pp. 1-14.
31. S. Lin and P. Beling, "An end-to-end optimal trade execution framework based on proximal policy optimization," in *Proceedings of the 29th International Conference*

on International Joint Conferences on Artificial Intelligence, 2021, pp. 4548-4554.
32. M. Katongo and R. Bhattacharyya, "The use of deep reinforcement learning in tactical asset allocation," *SSRN*, 2021, pp. 1-18.
33. M. Vishal, Y. Satija, and B. S. Babu, "Trading agent for the Indian stock market scenario using actor-critic based reinforcement learning," in *Proceedings of IEEE International Conference on Computation System and Information Technology for Sustainable Solutions*, 2021, pp. 1-5.
34. A. A. S. Gunawan, *et al.*, "Development of stock market price application to predict purchase and sales decisions using proximal policy optimization method," in *Proceedings of the 1st International Conference on Computer Science and Artificial Intelligence*, 2021, pp. 431-437.
35. E. Fernandez, *et al.*, "A novel approach to select the best portfolio considering the preferences of the decision maker," *Swarm and Evolutionary Computation*, Vol. 46, 2019, pp. 140-153.

Hei-Chia Wang (王惠嘉) is presently working as a Professor in Institute of Information Management at National Cheng Kung University, Taiwan. His research focuses on knowledge discovery, text mining, e-learning and bioinformatics. Wang obtained both MS in Information System Engineering and Ph.D. in Informatics from the University of Manchester (UMIST), UK.

Yi-Hsin Cheng (鄭義信) received the BS and MS degrees in Industrial and Information Management from National Cheng Kung University, Taiwan. His research interests include artificial intelligence, quantitative trading, FinTech and digital finance.

Yi-Hsuan Wu (吳翌暄) received the MS degree in Industrial and Information Management from National Cheng Kung University, Taiwan. Her research interests include artificial intelligence, portfolio management and deep reinforcement learning.