*International Journal of Innovative Computing, Information and Control*
Volume **20**, Number **2**, April **2024**

ICIC International ©2024 ISSN 1349-4198
pp. 643–652

# A STOCK TRADING STRATEGY BASED ON DEEP REINFORCEMENT LEARNING AND HONG-KONG CAPITAL POSITION

Jianming Li[1], Cui Zhu[1], Xiaojun Chen[1] and Xiangpei Hu[2]

[1]School of Computer Science
[2]School of Management and Economics
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116023, P. R. China
{ lijm; drhxp }@dlut.edu.cn; zhucuii@163.com; 877540927@qq.com

Received July 2023; revised October 2023

Abstract. *As processing power and artificial intelligence improve, more and more investors are turning to quantitative trading to build more effective trading techniques for the turbulent stock market. Hong-Kong funded institutions have attracted a lot of attention due to their stable investment style, and research on their trading behavior is abundant. When investing in stock markets, investors meet a couple of difficulties: 1) environmental unpredictability: the stock market is volatile and unstable, raw financial data is noisy, and it is difficult to accurately predict the stock price; 2) data heterogeneity: information obtained by investors is diversified and incomplete, as well as noisy and difficult to use, such as Hong-Kong capital position data and financial reports. To deal with the above issues, we develop a simulation trading environment based on the Chinese A-share trading market and train a trading agent for autonomous trading employing deep reinforcement learning and Hong-Kong shareholder's position data in A-share market. Experimental results show that the trained agent is capable of making profits in the volatile stock market and outperforms the baseline methods under most conditions.*
**Keywords:** Deep reinforcement learning, Stock trading, Hong-Kong capital, Auto trading

1. **Introduction.** Professional investment institutions with sufficient funds can devote a large amount of manpower to statistical analysis in order to comprehend and even influence market, while experienced traders can develop their own unique trading strategies using technical factor analysis or fundamental analysis. However, information barriers, a lack of investment experience, and a proclivity to be swayed by emotions can all make it difficult for investors to survive in the stock market. As a result, investors must consider how to profit in the complex and volatile stock market, as well as how to profit in the long term.

Reinforcement Learning (RL) is a popular machine learning method. Its principle is to build an agent to interact with the environment to acquire experience, as opposed to supervised learning, which requires prior knowledge, and unsupervised learning, which requires a large amount of data. The agent receives feedback from the environment while interacting with it and adjusts its strategy, eventually obtaining high or even excessive rewards. Reinforcement learning is applicable to control and decision problems and has been successfully applied in a variety of fields, including automatic driving [1], Unmanned Aerial Vehicle (UAV) tracking [2], chess [3], video games [4,5] and even neural network parameter

DOI: 10.24507/ijicic.20.02.643

search [6]. Financial applications of reinforcement learning include profit management, automated trading, quantitative trading, high-frequency trading, futures trading, etc.

Stock market is complicated, dynamic, noisy, and unstable. People attempt to derive useful information from stock transaction data using various statistical methods. While these methods sometimes work, they often fail in the short term because of market volatility. Moody et al. [7] introduced the Recurrent Reinforcement Learning (RRL) to investigate investing strategies in 1998, with raw stock data as input and trading strategies as output. However, the value function of traditional reinforcement learning is too simple to deal with the high-dimensional state space and action space, so it cannot adapt to the dynamic and complex stock market. Previous studies have shown that Artificial Neural Networks (ANN) [8] can effectively extract features, and Deep Reinforcement Learning (DRL) is a combination of ANN and RL. Therefore, DRL can not only represent high-dimensional features, but also enable agents to learn optimal strategies in interaction with the environment.

China's securities market has made remarkable achievements in a few decades as an arising securities market with a rapid expansion. It has gradually become the main securities market in the world and attracted many foreign investors to invest in China's security market. Hong-Kong funding is a significant representative among them. According to [9,10], "The funding from Hong-Kong" in mainland China has played a positive role in the economic development of the entire nation, especially for attracting foreign capital, improving job opportunities, and promoting technological progress. This paper attempts to construct factors with Hong-Kong capital data from various perspectives as additional input for agents with the goal to reduce stock market noise and improve agent's trading performance. Particularly, we manufactured a trading simulation environment and a trading system based on the rules of China's A-share trading rules.

The organizational structure of this paper is as follows. The second part introduces the development of deep reinforcement learning and quantitative trading. In the third part, a trading system based on raw financial data is introduced, where the state is augmented by the Hong-Kong position data. The fourth part presents the experiments, and finally the conclusions are drawn in the fifth part.

## 2. **Related Work.**

### 2.1. **Deep reinforcement learning.**

DRL uses deep neural networks to approximate the representation of strategies or value functions so that it can handle high-dimensional, nonlinear, complex state spaces and action spaces. It can be classified into two categories: value-based reinforcement learning and policy-based reinforcement learning. The value-based method learns Q values that represent the value of actions and then chooses the action with the largest Q value when making a decision. Examples of such methods include Q-learning [11], SARSA, and DQN [12]. The Policy Gradients [13] aims to maximize the expected cumulative rewards by optimizing the policy directly. The most basic policy gradient algorithm is REINFORCE. Building upon REINFORCE, the Trust Region Policy Optimization (TRPO) [14] algorithm introduces constraints to ensure that the policy does not change too much with each update, thus ensuring policy stability. However, the high computational complexity and inefficient sampling of the TRPO algorithm limits its range of applications. Proximal Policy Optimization (PPO) [15] solves the problem of sample efficiency by utilizing surrogate objectives to avoid the new policy changing too far from the old policy. The parameters of the policy network are adjusted using back-propagation to make the current policy closer to the optimal policy. In this paper, PPO is used to train trading agent because of its efficient performance.

2.2. **Quantitative trading and Hong-Kong capital position.** Quantitative Trading (QT) is a way to make trading decisions by using mathematical models and computer technology. It can effectively reduce the impact of human emotions and subjective judgments. Traditional QT strategies such as momentum [16] and mean reversion [17] methods discover trading opportunities based on heuristic rules. Finance expert knowledge is incorporated to capture the underlying pattern of the financial market. However, rule-based methods exhibit poor generalization ability and only perform well in certain market conditions. With the rise of neural networks, people use ANN to predict stock prices. Qiu et al. [18] combined dimensionality reduction techniques and neural networks to predict daily stock market quote. Wu et al. [19] proposed a Long Short-Term Memory (LSTM) neural network to automatically sense the dynamic stock market, as well as some carefully chosen indicators to reduce the effect of market noise. Zhou et al. [20] proposed an optimized multi-factor neural network prediction model with Principal Component Analysis-LSTM (PCA-LSTM) by analyzing historical trading data of Hong-Kong-funded institutions, and then using the model for dynamic trading prediction. Sulistio and Suhartono [21] forecast stock prices using Convolutional Neural Network-LSTM (CNN-LSTM) for technical factor analysis and BERT for sentiment factor analysis using social media data.

Obviously, the high volatility and noisy nature of the financial market make it extremely hard to predict future prices accurately. Recently, DRL has become an appealing approach in QT owing to not only its stellar performance but also the attractive property of learning meaningful representations from scratch. Bajpai [22] applied the DQN, as well as the Double DQN and Dueling DQN, to the Indian stock market to automate trading and maximize profits. Yang et al. [23] developed an automated trading system using ensemble strategies. They used three algorithms, PPO, A2C, and DDPG, to shift within a time window, and whichever algorithm performed better used that algorithm's strategy in the following step, and the results showed that the ensemble strategy outperformed any of the independent algorithms. Liu et al. [24] introduced a DRL library FinRL that facilitates beginners to expose themselves to quantitative finance and to develop their own stock trading strategies.

On November 17, 2014, Shanghai-Hong Kong Stock Connect was officially opened, which had a significant impact on Shanghai Stock Exchange (SSE) and Hong Kong Stock Exchange (HKEX), especially the A-share market. By studying the stock interconnection mechanism, Hu and Yao [25] found that the introduction of Shanghai-Hong Kong Stock Connect Program has improved the efficiency of stock pricing in a certain extent, and the degree of information response has improved a lot. We try to introduce Hong-Kong capital position to further improve the performance of DRL in QT.

3. **A Stock Trading System Based on DRL and Hong-Kong Capital.** To solve the above problems, we built a stock trading system based on DRL and Hong-Kong capital data, as shown in Figure 1. In the trading system, we build a simulation environment based on the China A-share rules and use DRL to train the trading agent to automate trading. To narrow the gap with the real environment, we also take account of trading constraints such as transaction cost, market liquidity, and the investor's degree of risk aversion. In order to make the agent perceive more effective information, we fuse the basic stock information, technical index, and factors designed after the analysis of Hong-Kong capital through the encoder to enhance the state representation of the agent. The function of the encoder is to concatenate the heterogeneous data into a vector as the input. After DRL training, the agent outputs decision actions, including buy, sell, and hold.

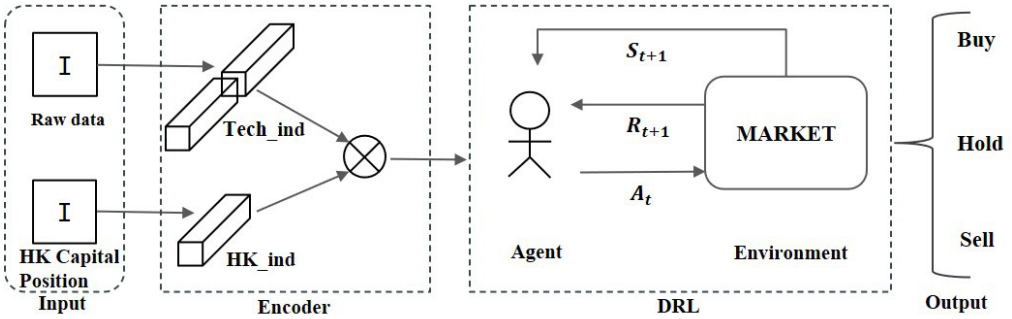


FIGURE 1. The trading system with raw financial data and HK capital position

3.1. **Hong-Kong capital related factors.** In finance, a factor refers to a set of key variables that affect the performance of an asset or portfolio. Examples of common technical factors are Moving Averages (MACD), Relative Strength Indicators (RSI), excess inflows or outflows, etc. In fact, factors can be understood as further processing and information extraction of basic stock data, which can be used to predict stock price changes or directly guide investment behavior, and technical factor analysis is a typical quantitative trading method. This paper obtains the historical data of A-share trading market of Hong-Kong-funded institutions in recent years on the website of HKEX news. Since Hong-Kong-funded institutions have entered the Chinese stock market for a relatively short period of time, the earliest historical data we can obtain is from 2016. In order to further extract valid information about the trading behavior of Hong-Kong-funded institutions, the following factors related to Hong-Kong capital are designed in this paper.

(a) Hong-Kong capital position change

This factor is expressed as a discrete value to record the change of Hong-Kong capital's position on a trading day compared with the previous trading day, where the signal of position increase is 1, the signal of position decrease is $-1$, and the signal of position immobility is 0.

$$vol\_signal = sign(volume_{t+1} - volume_t) \tag{1}$$

(b) Hong-Kong capital flow

For simplicity, we use the close price on the trading day as the stock price to calculate the market value of stocks held by Hong-Kong capital, and use the difference in market value as the change in Hong-Kong capital flows.

$$value\_chg = volume_{t+1} * price_{t+1} - volume_t * price_t \tag{2}$$

(c) Profit-loss ratio

The break-even ratio is a measure of risk and reward that can be used to assess the profitability of a Hong-Kong-funded institution. The P/L ratio represents the ratio between the average profit and the average loss obtained in a trade. Normally, a P/L ratio greater than 1 means that the average profit is greater than the average loss, indicating that the trading strategy has a certain level of profitability. Formula (3) is the most representative calculation method for the profit and loss ratio factor, which is the ratio of the trading profit and loss to the position cost.

$$profit\_loss\_ratio = (close - price_{cost})/price_{cost} * 100\% \tag{3}$$

(d) Excess inflows or outflows

Sudden and continuous net buying or selling is often an important signal for the start of a short-term rally in A-shares. Research data shows that when the buying and selling volume of a single Hong-Kong-funded institution or multiple anchor institutions reaches

5% of the total stock volume, it will have a certain impact on the stock price, and when it reaches 20%, it can almost determine the price trend.

3.2. **Trading agent and environment.** Considering the interactivity and complexity of the stock market, we model the trading behavior as a Markov model. In the Markov model, the state transfer of the trading agent is Markovian in nature. The trading agent gets the current market state and the historical state of the past period at each time step. The agent makes the corresponding decision based on the state, and the environment gives feedback based on the agent action. After a certain number of iterations, the agent learns how to make the right decision to get a higher reward. The Markov model contains the following key elements.

**State $s$**: The state is a one-dimensional vector shaped as $[p, h, b, tech_{ind}, hk_{ind}]$, which contains the historical stock price $p$, the agent's position $h$, the account balance $b$, the technical factor $tech_{ind}$, and the Hong-Kong capital-related factor $hk_{ind}$. The state is all the information that the agent can perceive, and if the information content is too low, then the decision making of the agent is more random and uninformed. On the contrary, if the information content is too high, then the data noise will also increase, which will make the agent lose the direction of optimization and thus difficult to converge.

**Action $a$**: An action is the set of effective behaviors that the specified agent can take. We use a continuous action space, i.e., the output action is a decimal within $[-1, 1]$, indicating the buy or sell $h_{\max} * a$ number of shares, where $h_{\max}$ is the maximum number of shares in a single operation.

**Reward $r$**: The reward function represents the reward that an agent can get for taking action $a$ in state $s$ and reaching the next state $s_{t+1}$, we design the reward function as the difference Sharpe ratio of two transactions.

$$r(s_t, s_{t+1}, a) = sharpe_{t+1} - sharpe_t \tag{4}$$

The objective is to maximize the cumulative return while minimizing the risk taken by the agent to obtain the return.

**Policy $\pi$**: The policy represents the probability distribution of the action taken by the agent when it is in state $s$. During the training process, the agent selects the action with the maximum probability in the distribution with a certain probability of $1 - \epsilon$ and a certain probability of $\epsilon$ selecting a random action, which aims to increase the randomness of exploration and avoid falling into a local optimum. And in the real trading process, the agent will directly choose the action with the highest probability of output.

**Environment**: Environment is one of the core elements of reinforcement learning. In this paper, we simulate the environment of the Chinese A-share trading market and let the agent train in a realistic stock market as much as possible to minimize the deviation from the real scenario. The Chinese A-share market implements a $T + 1$ trading system, where stocks bought on the same day cannot be sold on the same day, and the agent must strictly comply with this rule. Considering the complex transaction fees in the stock market, we charge a flat fee of one thousandth of the transaction amount. Suppose that the set of stocks that the agent needs to operate is $D = \{d_1, d_2, \ldots, d_{|D|}\}$, and on trading day $t$, the agent performs action $a_i$ on stock $d_i$, then on the next trading day $t + 1$, the agent's position in stock $d_i$ is

$$hold_i^{t+1} = hold_i^t + a_i * h_{\max} \tag{5}$$

Meanwhile, the account balance is updated to

$$b_{t+1} = b_t + \sum_{i=1}^{|D|} a_i * h_{\max} * price_t^i - cost_t \tag{6}$$

where $price_t^i$ represents the closing price of stock $i$ on trading day $t$, and $cost_t$ represents the transaction fee generated on trading day $t$. The corresponding change in net worth is the sum of the account balance and the value of the stocks, i.e.,

$$profit_t = b_t + \sum_{i=1}^{|D|} hold_i^t * price_i^t \tag{7}$$

Our goal is to get as big a net worth as possible at the end of the transaction.

3.3. **Training algorithms for trading agents.** In this paper, we use a deep reinforcement learning algorithm to train a trading agent. PPO is a stable and efficient deep reinforcement learning algorithm, and the main idea of PPO is to limit the magnitude of the policy change when updating the policy, so as to ensure the stability and convergence of the policy. In PPO, we use an optimization objective function that is composed of two parts, the first part is the ratio between the old and new versions of the policy, i.e., the ratio of the probability of comparing the output actions of the new policy and the old policy in the same state; the second part is a shearing function used to limit the magnitude of the policy change, which controls the difference between the old and new policy during the update process.

Specifically, PPO uses the objective function:

$$L(\theta) = \widehat{E_t}\left[\min\left(r_t(\theta)\widehat{A_t}, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\widehat{A_t}\right)\right] \tag{8}$$

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \tag{9}$$

where $\theta$ is the strategy parameter, $r_t(\theta)$ is the ratio between the old and new strategies, $\widehat{A_t}$ is an estimate of the advantage of taking action $a_t$ in state $s_t$, $\epsilon$ is a hyperparameter that controls the degree of shear, and the clip function is used to restrict the value domain of the ratio $r_t(\theta)$ in the interval $[1-\epsilon, 1+\epsilon]$. The optimization process of PPO uses a stochastic gradient descent algorithm to update the policy parameters by minimizing the agent objective function. In each training iteration, we first collect a batch of empirical data, and then calculate the gradient of the agent objective function and update the policy parameters using the stochastic gradient descent algorithm.

## 4. Experiment.

4.1. **Data set and experimental setup.** We randomly selected 25 stocks listed on the SSE as the stocks pool of the agent, and also obtained all the holding data of Hong-Kong capital for these 25 stocks from the HKEX news. The data between 2016-07-01 and 2021-09-30 are used for training, and the data between 2021-10-01 and 2022-10-01 are used for test. The initial investment money is RMB 1 million, and the transaction fee is set to 1‰. Since the prices of the market contain too much noise, we introduced six of the most commonly used technical indicator to capture the main trend, including MACD (Moving Average Convergence Divergence), RSI (Relative Strength Index), CCI (Commodity Channel Index), ADX (Average Directional Movement Index), BOLL_UB (the Upper Band of the Bollinger Bands), BOLL_LB (the Lower Band of the Bollinger Bands). MACD is one of the most intuitive technical indicators around. It amplifies the value of moving averages in an elegant way to track both trend and momentum. RSI combines the average gain and average loss into a single ratio that represents price momentum. CCI uses the deviation between price and its moving average to measure price momentum. ADX is used to determine when the price is trending strongly. Bollinger bands are used to measure a market's volatility and identify "overbought" or "oversold".

4.2. **Experimental result.** In Table 1, we compare our agent with three other benchmark methods for backtesting, namely the SSE50, the buy-and-hold (Buy&Hold) strategy, and the return curve of Hong-Kong-funded institutions over the period. The backtesting results data are shown in Table 1. It can be visually seen that the annualized return of our agent is 13.63%, which is much higher than the −18.90% of the SSE50, −3.83% of the Buy&Hold method and −1.27% of the Hong-Kong capital. The SSE50 indicates a continued downturn in market sentiment during this period (probably due to the COVID-19), but our trading agent still achieved excess profits. It can also be seen that the Hong-Kong capital, despite not being profitable, also minimized its losses in the stock market with its solid investment style (Sharpe ratio of 0.0691, the only one greater than 0 among all benchmark methods). As can be seen in Figure 2, our trading agent is above the benchmark method in terms of net assets most of the time, with a Sharpe ratio of 0.6583 much higher than the other benchmark methods. The results show that our trading agent after state augmentation based on Hong-Kong capital position data is effective in discovering trading strategies, automating trades and making profits in a volatile stock market.

Table 1. Backtest metrics comparison between our agent and other baselines

| | **Our agent** | **SSE50** | **Buy&Hold** | **HK capital** |
|---|---|---|---|---|
| **Final portfolio** | **1,136,325** | 810,976 | 961,688 | 987,253 |
| **Annual return** | **13.63%** | −18.90% | −3.83% | −1.27% |
| **Cumulative return** | **13.12%** | −18.36% | −3.70% | −1.22% |
| **Annual volatility** | 23.76% | 18.57% | **17.61%** | 24.34% |
| **Sharpe ratio** | **0.6583** | −1.0390 | −0.1347 | 0.0691 |
| **Max drawdown** | −19.92% | −22.49% | **−14.28%** | −25.54% |

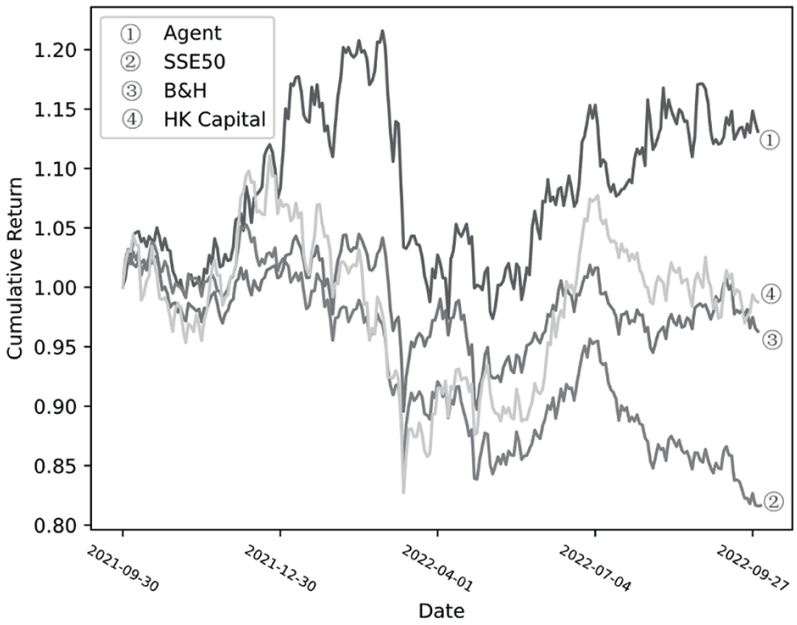


Figure 2. Cumulative return curve of our agent and other baselines

5. **Conclusion.** In this paper, we develop a deep reinforcement learning-based stock trading system, in which a Markov model is used to model the trading agent, the trading environment is simulated, and a PPO algorithm is used to train the trading agent. We

designed Hong-Kong capital related technical factors for state argumentation during the training process, which can reduce the interference of market noise to some extent while also allowing the agent to deeply explore the investment patterns of Hong-Kong capital institutions in order to profit. Finally, we back-tested the trained trading agent for nearly a year, and the results demonstrated that the trading agent could outperform even in bear market. The Sharpe Ratio indicates that the trading agent is more robust in balancing risk and return than other benchmark methods. However, the trading agent underperforms in certain areas, such as relatively high annualized volatility and high max-drawdown rates, which may present potential trading risks.

As Hong-Kong capital has a robust investment style, in the future, we can further consider using Hong-Kong capital as the core and extracting other valid information such as Hong-Kong capital related news reports and financial data of Hong-Kong institutions as heterogeneous input information for the agent to further reduce the noise of the market environment.

## REFERENCES

[1] S. Feng, H. Sun, X. Yan et al., Dense reinforcement learning for safety validation of autonomous vehicles, *Nature*, vol.615, pp.620-627, 2023.

[2] H. X. Pham, H. M. La, D. Feil-Seifer and L. V. Nguyen, Autonomous UAV navigation using reinforcement learning, *arXiv Preprint*, arXiv: 1801.05086, 2018.

[3] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot et al., Mastering the game of go with deep neural networks and tree search, *Nature*, vol.529, pp.484-489, 2016.

[4] V. Mnih, K. Kavukcuoglu, D. Silver et al., Playing atari with deep reinforcement learning, *Computer Science*, 2013.

[5] C. Berner, G. Brockman, B. Chan et al., Dota 2 with large scale deep reinforcement learning, *arXiv Preprint*, arXiv: 1912.06680, 2019.

[6] B. Zoph and Q. Le, Neural architecture search with reinforcement learning, *International Conference on Learning Representations (ICLR)*, 2017.

[7] J. Moody, L. Wu, Y. Liao et al., Performance functions and reinforcement learning for trading systems and portfolios, *Journal of Forecasting*, vol.17, nos.5-6, pp.441-470, 1998.

[8] A. K. Jain, J. Mao and K. M. Mohiuddin, Artificial neural networks: A tutorial, *Computer*, vol.29, no.3, pp.31-44, 2015.

[9] Y. C. Wang, J. J. Tsai and L. Yi, *The Influence of Shanghai-Hong Kong Stock Connect on the Mainland China and Hong Kong Stock Markets*, Social Science Electronic Publishing, 2016.

[10] R. Huo and A. D. Ahmed, Return and volatility spillovers effects: Evaluating the impact of Shanghai-Hong Kong stock connect, *Economic Modelling*, 2017.

[11] H. V. Hasselt, A. Guez and D. Silver, Deep reinforcement learning with double Q-learning, *Computer Science*, 2015.

[12] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., Human-level control through deep reinforcement learning, *Nature*, vol.518, pp.529-533, 2015.

[13] R. S. Sutton, D. McAllester, S. Singh et al., Policy gradient methods for reinforcement learning with function approximation, *Proc. of the 12th International Conference on Neural Information Processing Systems (NIPS'99)*, pp.1057-1063, 1999.

[14] J. Schulman, S. Levine, P. Abbeel et al., Trust region policy optimization, *International Conference on Machine Learning (PMLR)*, pp.1889-1897, 2015.

[15] J. Schulman, F. Wolski, P. Dhariwal, A. Radford and O. Klimov, Proximal policy optimization algorithms, *arXiv Preprint*, arXiv: 1707.06347, 2017.

[16] T. J. Moskowitz, Y. H. Ooi and L. H. Pedersen, Time series momentum, *Journal of Financial Economics*, vol.104, no.2, pp.228-250, 2012.

[17] J. M. Poterba and L. H. Summers, Mean reversion in stock prices: Evidence and implications, *Journal of Financial Economics*, vol.22, no.1, pp.27-59, 1988.

[18] M. Qiu, Y. Song and F. Akagi, Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market, *Chaos, Solitons & Fractals*, 2016.

[19] J. Wu, C. Wang, L. Xiong and H. Sun, Quantitative trading on stock market based on deep reinforcement learning, *International Joint Conference on Neural Networks (IJCNN)*, 2019.
[20] J. Li, T. Zhou and X. Hu, Prediction algorithm of stock holdings of Hong Kong-funded institutions based on optimized PCA-LSTM model, *International Journal of Innovative Computing, Information and Control*, vol.18, no.3, pp.999-1008, 2022.
[21] B. Sulistio and D. Suhartono, Utilizing BERT and CNN-LSTM in stock price prediction using data sentiment analysis and technical analysis of stock and commodity, *ICIC Express Letters*, vol.17, no.2, pp.171-179, 2023.
[22] S. Bajpai, Application of deep reinforcement learning for Indian stock trading automation, *arXiv Preprint*, arXiv: 2106.16088, 2021.
[23] H. Yang, X. Y. Liu, S. Zhong et al., Deep reinforcement learning for automated stock trading: An ensemble strategy, *Proc. of the 1st ACM International Conference on AI in Finance*, pp.1-8, 2020.
[24] X. Y. Liu, H. Yang, Q. Chen et al., FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance, *arXiv Preprint*, arXiv: 2011.09607, 2020.
[25] Z. Hu and P. Yao, The impact of Shanghai-Hong Kong, Shenzhen-Hong Kong stock connect program on the stock price efficiency – A test based on the difference-in-difference model, *Journal of Financial Development Research*, 2018.

## Author Biography

**Jianming Li** received the bachelor's degree in Ship Engineering from Dalian University of Technology, China, 1999; the M.Sc. degree in Computer Application Technology from Dalian University of Technology, China, 2002; the Ph.D. degree in Computer Application Technology from Dalian University of Technology, China, 2007.

He is currently a full-time associate professor at the Dalian University of Technology, China. His main research interests include the machine learning, classification and prediction algorithms of deep learning, software automation, and quantitative analysis and strategy research in the financial field. He has published over 50 papers in journals and conferences.

**Cui Zhu** received the bachelor's degree in Computer Science and Technology from Dalian University of Technology, China, 2021.

He is studying for a master's degree in Computer Science at Dalian University of Technology. His research interests include deep reinforcement learning, multi-agent reinforcement learning, behavioral cloning, as well as quantitative trading and automated trading in the financial domain.

**Xiaojun Chen** obtained bachelor's degree in Engineering and Literature, majoring in Computer Science and Technology, from September 2015 to June 2020, Dalian University of Technology, China.

She is currently pursuing a master's degree at Dalian University of Technology. Her main research areas include machine learning, classification and prediction algorithms in deep learning, and quantitative analysis and strategy research in the financial field.

**Xiangpei Hu** received his B.S. (1983), M.S. (1987) and Ph.D. (1996) degrees from Harbin Institute of Technology, China, respectively. He is a Professor of Management Science at Dalian University of Technology, China, "Distinguished Young Scholars" of National Natural Science Foundation of China (NNSFC), and "Chang-jiang Scholars Distinguished Professor" of Ministry of Education (MOE) of China.

His research and teaching interests are electronic commerce, supply chain and logistics management, intelligent operations research and the real-time optimization control for dynamic systems. He has published over 200 scholarly papers in refereed journals.