# How to leverage Reinforcement Learning and its algorithm to trade in a stock market environment effectively.

Nahid Abdolrahmanpour
110107635
Fall 2022
University of Windsor
Canada
Email: abdolran@uwindsor.ca

Namarta Vij
110120312
Fall 2022
University of Windsor
Canada
Email:vijn@uwindsor.ca

Rajath Bharadwaj
110099435
Fall 2022
University of Windsor
Canada
Email: bharadw3@uwindsor.ca

## I. ABSTRACT

**The goal of stock market investment is to obtain more profits. Many researchers have attempted implementing machine learning based stock trading in recent years. Implementing dynamic trading strategies and getting valuable information from multiple sources is challenging when dealing with the complex stock market. In this paper, we compared three reinforcement learning algorithms such as Advantage Actor Critic (A2C) Proximal Policy Optimization (PPO), and Deep Q-Network (DQN) for stock trading and analyzed the performance using stock data and technical indicators. The agent in reinforcement learning bases its trading decisions on a fusion of the deep neural network's from various features like Open, High, Low, Close, Volume and the indicators like Parabolic Stop Reverse (PSAR), Relative Strength Index (RSI), SuperTrend and Exponential Moving Average (EMA), and the current state of the stock market—studies using data from the Indian stock (Reliance) and data fetched from yahoo finance. Results show that the PPO is able to yield higher returns than the other two algorithms.**

## II. KEYWORDS

Artificial Intelligence(AI), Reinforcement Learning(RL), Stock Trading.

## III. INTRODUCTION

Stock trading is buying and selling stocks to obtain investment profit. The key to stock trading is to make the right trading decisions at the correct times and develop a suitable trading strategy [1].In recent years, many studies have been based on machine learning methods to predict stock trends or prices to implement stock trading.Due to the increasing complexity of trading in stock markets, automating trading operations becomes a significant aim for individual traders, portfolio managers, and financial organizations. Further, the growing availability of stock market digital records (i.e., prices and trading volumes) has motivated and directed more research efforts toward automated trading. Although forecasting future expenses is the fundamental problem in stock markets, other issues are also of the essence since they combine future costs with associated actions to be taken[2]. Among the state-of-the-art techniques, machine learning techniques are the most widely chosen techniques in recent years, given the rapid development of the machine learning community. The other reason is that traditional statistical learning algorithms can not cope with the non-stationary and non-linearity of the stock markets [3]. Generally, two main approaches exist to analyze and predict stock price: technical analysis [4] and fundamental analysis [5]. The technical analysis looks into the market's past data only to predict the future. On the other hand, the fundamental analysis considers additional information such as the economic status, news, financial reports, and meeting notes of the CEO discussion. The technical analysis relies on the efficient market hypothesis (EMH) [6]. The EMH states that it will reflect all the fluctuation in the market very quickly in the price of stocks. In practice, the cost can be updated in milliseconds [7], leading to the very high volatility of the stocks. In recent years, technical analysis has attracted much attention because we have enough information just by looking at the historical stock market, which is public and well-organized, compared to the fundamental analysis, where we need to analyze the unstructured dataset.Compared to supervised learning techniques and at a certain level, unsupervised learning algorithms are widely used in stock price prediction. To the best of our knowledge, the reinforcement learning for stock price prediction has yet to receive enough support as it should. The main issue with supervised learning algorithms is that they are inadequate to deal with time-delayed rewards [8,9]. In other words, supervised learning algorithms focus only on the accuracy of the prediction at the moment without considering the delayed penalty or reward. Furthermore, most supervised machine learning algorithms can only provide

action recommendations on particular stocks; reinforcement learning can lead us directly to the decision-making step, i.e., to decide how to buy, hold or sell any stock.Reinforcement learning is a machine learning subfield and artificial intelligence process which establishes learning techniques to train agents in a trial-and-error environment.Reinforcement learning means learning "what can be done to maximize the numerical benefit signal." The reinforcement model is not telling what actions should be taken but must determine which produces the richest benefits. [10].AI and machine learning applications can rely heavily on learning with reinforcement. Software and computer engineers frequently use reinforcement machine learning to create operational standards and parameters for soft AI to follow when fetching and displaying information. Reinforcement Learning is a robust mathematical framework where the agents interact directly with the environment. It is experience-driven autonomous learning where the agent enhances its efficiency by trial and error to optimize the cumulative reward. It does not require labeled data to do so. For Autonomous learning policy, search and value function approximation are vital tools. Reinforcement Learning applies a gradient-based or gradient-free approach to detect an optimal stochastic policy for continuous and discrete state action settings(Sutton and Barto,2014). While considering an economic problem, despite traditional approaches, reinforcement learning methods prevent suboptimal performance by imposing significant market constraints that lead to finding an optimal market analysis and forecast strategy. Despite Reinforcement Learning successes in recent years, these results need more scalability and cannot manage high-dimensional problems. By combining Reinforcement Learning and Deep Learning methods, the Deep Reinforcement Learning technique, where Deep Learning is equipped with vigorous function approximation, representation learning properties of deep neural networks (DNN), and handling complex and nonlinear patterns of economic data, can efficiently overcome these problems. This paper suggests a deep reinforcement learning model. It integrates multi feature data to implement stock trading to conduct a deeper analysis of the performance of different RL algorithms and find the best dynamic trading strategy. We can learn the best trading strategy by analyzing stock data and technical indicators. Additionally, it is essential to consider how the reward function is set up in reinforcement learning. When trading stocks, it's necessary to evaluate investment risk and returns and strike a reasonable balance between them. In this study, we optimize the explained variance and minimize the value loss. We did the comparison between three different RL algorithms to get the best result. The main contributions of this paper are as follows: We built the environment to mimic the stock market to be applicable on the all different RL algorithms We used Reliance stock data to measure the performance of the agent on different RL algorithms We also used Weights and Biases to track the algorithmic specific metrics along with the GPU utilization during the running of episodes. We track the best performant agent based on the 'value$_loss$'$and$'$explained_variance$'$and for DQN we use exploration rate

## IV. AI History

Artificial intelligence (AI) aims to endow machines with human intelligence. Machine learning (ML) is a method for implementing AI by using algorithms to parse data, learn from data, and make decisions and predictions regarding real-world events. Deep learning (DL) is a technology for realizing ML, which enables ML to recognize many applications and expands the scope of AI. Reinforcement learning (RL), also known as evaluation learning, is a technique of ML. Deep reinforcement learning (DRL) is the combination of DL and RL. It aims at realizing the optimization objective of RL with the operation mechanism of DL to advance toward general AI. In contrast to traditional ML, RL does not have an immediate end result; only a temporary reward (set primarily according to human experience) is observed. Therefore, RL can also be regarded as delayed supervised learning [11]. In the case of small state space and behavior space, RL technology can be used to enable network entities to identify the optimal strategy for decision-making or behavior. However, in a complex, large-scale network, for improving learning efficiency, a learning method that combines RL with DL, namely DRL, is regarded as a potential solution [12]. The three key elements of RL are the system status, the system actions, and the rewards. In RL, the environment is typically represented as a Markov decision process (MDP). Agents interact with the unknown environment through repeated observation, action, and reward to construct the optimal strategy [13]. Due to the limited data that are obtained from outside, DRL systems often rely on their own experience to learn by themselves. Via this approach, knowledge is acquired, and solutions are adapted to the environment.

## V. Literature Survey

Many machine-learning techniques have been used in stock trading recently. Investors base their trading choices on their evaluation of the stock market.However, they cannot make timely trading decisions based on stock market changes due to the influence of numerous factors. Machine learning techniques have more advantages over conventional trading strategies because they can learn trading strategies by examining stock market data and finding profit patterns that non-financial experts are unaware of. Studies that use deep learning techniques to implement stock trading exist. Deep learning techniques typically use stock trading to forecast future stock prices or trends. With the advancement,many researchers have started using a reinforcement learning approach to predict the stock value . Chakole et al. used reinforcing learning to interact with the real stock market as its environment to find the best dynamic trading strategy. A novel LSTM-based framework was put forth by Rundo et al. to carry out accurate stock price prediction. The suggested method consists of two pipelines, the first designed to predict trends and the

second created to forecast stock price values. The proposed system is based on LSTM, and a mathematical price correction method carried out by Markov statistical model concerning stock close price prediction. The proposed framework can accurately forecast stock prices according to the results. In terms of accuracy, the suggested method performs better than statistical models like SMA. Therefore, In most studies, it is evident that deep-Q-Networks are used for algorithmic trading decisions; however, some researchers have employed more complex and extended versions, such as Dueling DQN and Double DQN. Additionally, variations in state construction are observed. Also, Li et al. conducted an intriguing study in which candlestick charts were used as an image feature. Mehtab et al. suggest a stock price forecasting model that uses investor sentiment from social media to supplement the results of a deep learning framework and achieve a very high level of prediction accuracy. Dang et al. for trading applications, researchers mainly examined low frequency (typically daily) active trading on single assets with simulated or real data. Deng et al. proposed a model Deep Direct Reinforcement Learning and added fuzzy learning, which is the first attempt to combine deep learning and reinforcement learning in the field of financial transactions.

TABLE I
THE LITERATURE SURVEY TABLE

| Authors | Technique | Performance | Dataset |
|---|---|---|---|
| Chakole et al | Q-learning algorithm of Reinforcement Learning | Average annual return(23.57%) | Indian and the American experimental index stocks (NIFTY,SENSEX) |
| Xu et al | GRU-RL with attention mechanism | Average annual return(33.57%) | Chinese stocks |
| Koratamaddi et al | Deep RL | Annualized Investment return(22.5%) | Dow Jones companies |
| Tan et al | RL | Average annual return (36.91%) | US- stocks |
| Rundo et al | LSTM+RL | Average annual Return (36.23%) | Tickstory Web-site |
| Our Test Results | Deep Reinforcement Learning | DQN=45%> PPO, DQN=57% > A2C | Reliance |

## VI. METHODOLOGY

In this section, we provide the details of the datasets that we used. We did a comparative analysis between different reinforcement learning algorithms and implemented stock trading by analyzing the stock market with multisource data. In this section, first, we introduce the overall deep reinforcement learning model, then the feature extraction process of different data sources is described in detail. Finally, the specific application of reinforcement learning in stock trading is introduced.

```
In [126]: env.df[5012:6700]
Out[126]:
```

| Date | Open | High | Low | Close | Volume | Dividends | Stock Splits | SUPERT_7_1.0 | SUPERTd_7_1.0 | SUPERTl_7_1.0 | ... | SMA_5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2015-10-20 00:00:00+05:30 | 459.562354 | 460.371938 | 453.847585 | 455.609650 | 8349406 | 0.0 | 0.0 | 447.276525 | 1 | 447.276525 | ... | 442.098999 |
| 2015-10-21 00:00:00+05:30 | 455.038162 | 464.229410 | 453.847597 | 454.800049 | 6647651 | 0.0 | 0.0 | 449.126899 | 1 | 449.126899 | ... | 447.056555 |
| 2015-10-23 00:00:00+05:30 | 459.092324 | 463.396038 | 448.204285 | 455.109619 | 5608802 | 0.0 | 0.0 | 449.126899 | 1 | 449.126899 | ... | 452.004590 |
| 2015-10-26 00:00:00+05:30 | 458.086032 | 458.609609 | 448.275890 | 449.013855 | 5048242 | 0.0 | 0.0 | 464.061327 | -1 | NaN | ... | 454.676245 |
| 2015-10-27 00:00:00+05:30 | 447.085138 | 451.585517 | 446.418421 | 447.442322 | 5354228 | 0.0 | 0.0 | 458.841720 | -1 | NaN | ... | 452.395099 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2022-08-16 00:00:00+05:30 | 2638.256764 | 2649.323542 | 2619.413994 | 2643.341553 | 3049304 | 0.0 | 0.0 | 2586.633866 | 1 | 2586.633866 | ... | 2597.190527 |
| 2022-08-17 00:00:00+05:30 | 2657.000450 | 2668.864632 | 2647.578865 | 2657.149902 | 3044037 | 0.0 | 0.0 | 2613.660135 | 1 | 2613.660135 | ... | 2616.731689 |
| 2022-08-18 00:00:00+05:30 | 2660.000000 | 2669.000000 | 2640.199951 | 2651.300049 | 3419730 | 8.0 | 0.0 | 2613.660135 | 1 | 2613.660135 | ... | 2634.042090 |
| 2022-08-19 00:00:00+05:30 | 2650.000000 | 2661.699951 | 2604.949951 | 2613.850098 | 4366123 | 0.0 | 0.0 | 2613.660135 | 1 | 2613.660135 | ... | 2640.147607 |
| 2022-08-22 00:00:00+05:30 | 2617.000000 | 2631.300049 | 2586.000000 | 2606.600098 | 4371179 | 0.0 | 0.0 | 2653.155309 | -1 | NaN | ... | 2636.448340 |

1688 rows × 21 columns

Fig. 1. Reliance stock data structure example

### A. The overall structure of the model

The best dynamic trading strategy is more easily learned by implementing stock trading based on deep reinforcement learning and accurately analyzing the state of the stock market. We combine the features of stock data and technical indicators to obtain a more profound feature representation of the stock market state and learn the best dynamic trading strategy. The general layout of the model is depicted in the flowchart. The deep learning module for extracting features from the stock data and the reinforcement learning module for making the trading decision, are the two components of our project.

**End to End RL pipeline of our comparative analysis of different algorithms**

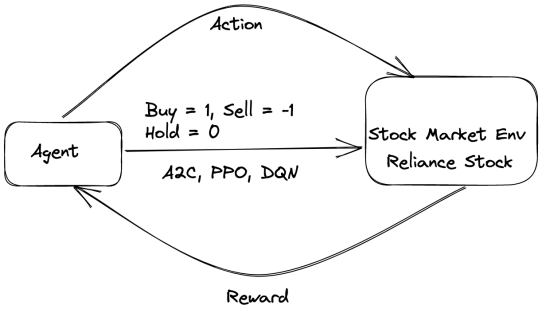


Fig. 2. Flowchart of our study

- **Build and load the environment for stock market**
- **Agent has three possible actions such as Sell, Buy and Hold**
- **The policy is multilayer Perceptron(MLP)**
- **The objective for our agent is to maximize the profit and minimize the value loss during taking the action**

### B. Deep Learning module

This study aims to learn the best dynamic trading strategy by fusing data from multiple sources to create a more profound feature representation of the stock market's environmental state. Raw stock data contain a lot of noise, even though they can reflect changes in the stock market. Relevant technical indicators are used as data sources for

analyzing the stock market in this study to lessen the impact of noise and perceive changes in the stock market more objectively and accurately.

### C. Stock Data and Technical Indicator Feature Extraction

Due to the noise in stock data, we use appropriate technical indicators to decrease the impact of noise. Technical indicators reflect the changes in the stock market from different views. Raw stock data we use include opening price, closing price, high price, low price, and trading volume. Technical indicators used in this paper are the PSAR, EMA, RSI and Supertrend. Indicators are calculated by mathematical formulas based on stock prices and trading volumes.

## VII. Experiments and Results

This section mainly introduces the dataset, evaluation metrics, comparison methods, implementation details, and experimental result analysis.

### A. Dataset

In this study, we did comparative analysis between three different RL algorithms for the stock market and validated it using a dataset of Reliance stock. The dataset's time frame spans the months of January 1996 to January 2023. The testing period is from January 2019 to January 2021, and the training period is from January 2012 to December 2018. As shown in Table 1, the daily open price, high, low, close price, and trading volume of the stock are all included in stock data.

## VIII. Metrics

Three evaluation indicators used in this paper are:

- **PSAR (Parabolic stop and reverse)**
  Traders use the parabolic SAR indicator to establish trend direction and potential reversals in price.
  A rising PSAR has a slightly different formula than a falling PSAR.
  PSAR = Prior PSAR+[Prior AF (Prior EP – Prior PSAR)]
  FPSAR = Prior PSAR–[Prior AF (Prior PSAR – Prior EP)]

  Where :

  RPSAR=Rising PSAR
  AF= Acceleration Factor, it starts at $0.02$ and increases by $0.02$ , up to a maximum of $0.2$ , each time the extreme point makes a new low (falling ASR) or high (rising SAR)
  FPSAR = Falling PSAR
  EP = Extreme Point, the lowest low in the current downtrend (falling SAR) or the highest high in the current uptrend (rising SAR)

- **RSI (Relative Strength Index)**
  The relative strength index (RSI) is a momentum indicator used in technical analysis and measures the speed and magnitude of a security's recent price changes to estimate overvalued or undervalued conditions in the price of that security. The RSI uses a two-part calculation :

  $RSI_{step.one}$ =100−$[\frac{100}{1+\frac{AverageGain}{AverageLoss}}]$

  $\text{RSI}_{Step.Two}$ =100−$[\frac{100}{1+\frac{(PreviousAverageGain*13)+CurrentGain}{(PreviousAverageLoss*13)+CurrentLoss}}]$

- **EMA ( Exponential Moving Average )**
  EMA is a class of moving average (MA) that sets a greater weight on the most recent data points.

  $\text{EMA}_{Today}$ =[$\text{Value}_{Today} * \frac{Smoothing}{1+Days}$] + $[EMA_{yesterday} * (1 - \frac{Smoothing}{1+Days})]$

- **Supertrend indicator**
  It is a trend-following indicator. This indicator works on only two parameters: Periods: Traders usually use 10 periods – Average True Range number of days (ATR – yet another indicator that gives you market volatility value by decompressing the range of prices of a security for a particular time). Multiplier: A multiplier is a value by which ATR would be multiplied. Three multipliers are used.

  The formula for the Supertrend indicator is :

  Up=$[\frac{(high+low)}{2}] + (multiplier * ATR)$

  Down=$[\frac{(high+low)}{2}] - (multiplier * ATR)$

### A. Comparative Experiment on Given Dataset

The following part includes the training graphs from Weights and Biases

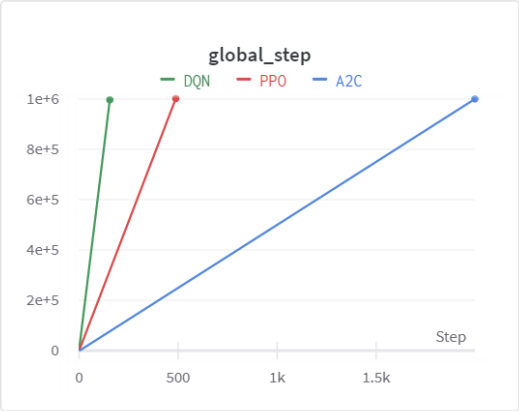


Fig. 3. No. of iteration/ epochs

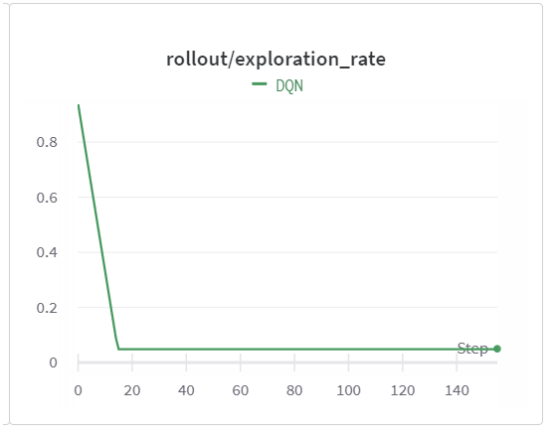


Fig. 6. Mean episode length (averaged over 100 episodes)

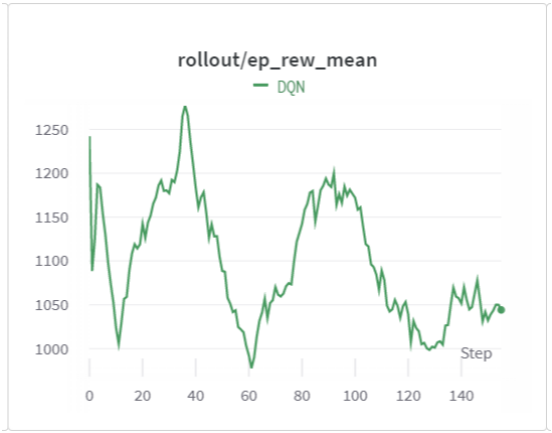


Fig. 4. Mean episodic training reward (averaged over 100 episodes))

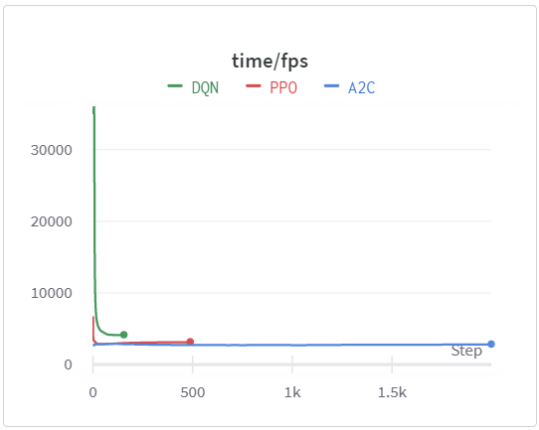


Fig. 7. Number of frames per seconds (includes time taken by gradient update)

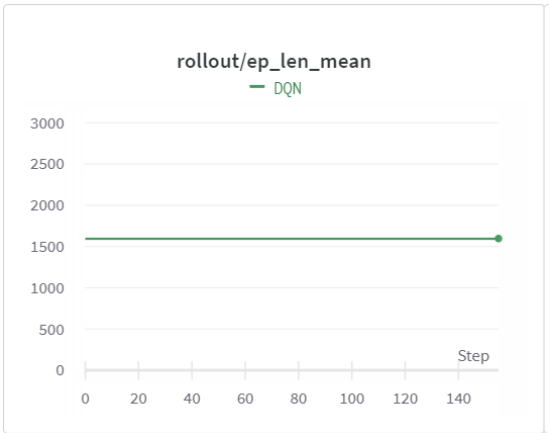


Fig. 5. Mean episode length (averaged over 100 episodes)


Fig. 8. Current learning rate value

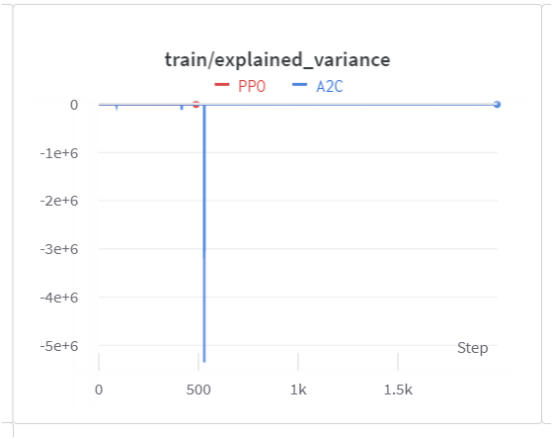


Fig. 9. Fraction of the return variance explained by the value function

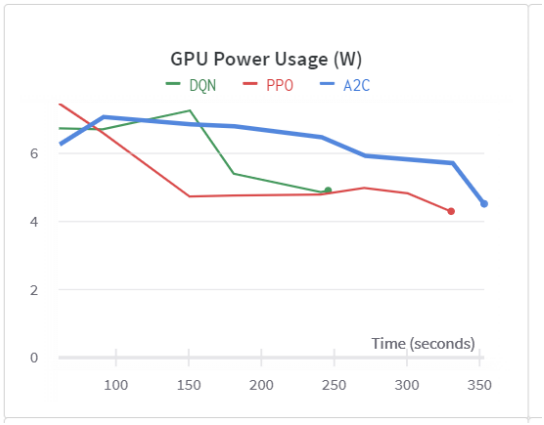


Fig. 12.

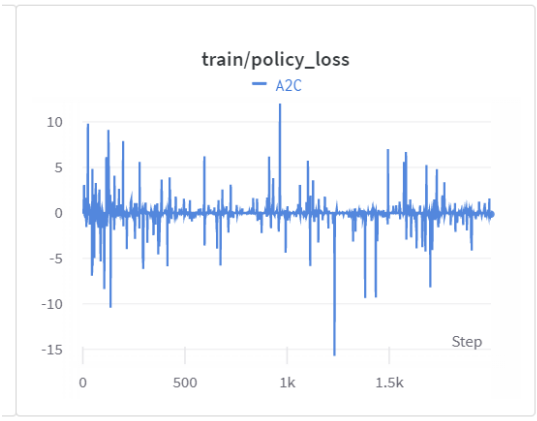


Fig. 10. Current value of the policy gradient loss (its value does not have much meaning

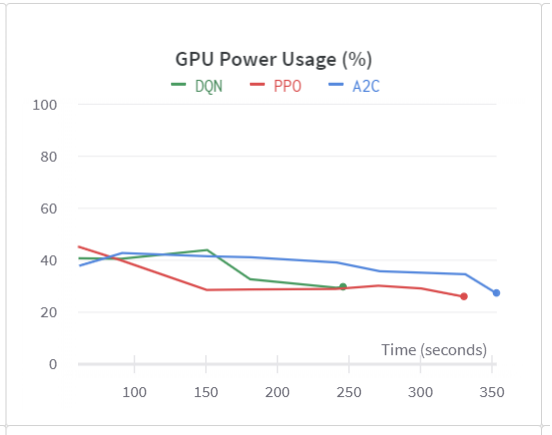


Fig. 13.


Fig. 11. Mean value of the entropy loss (negative of the average policy entropy

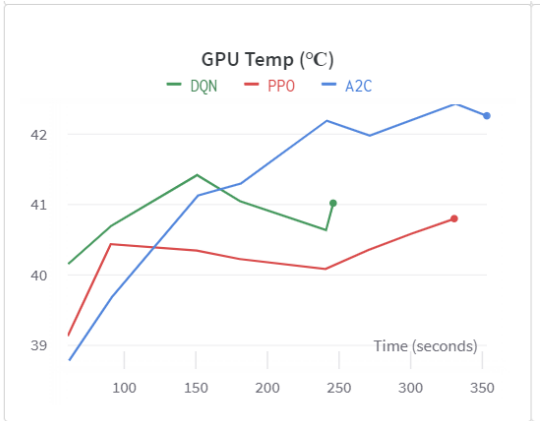


Fig. 14.

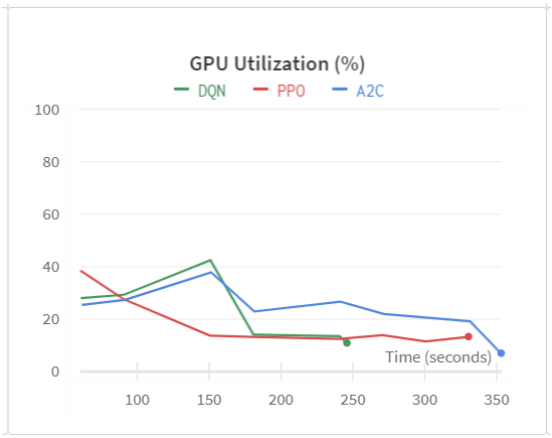


Fig. 15.

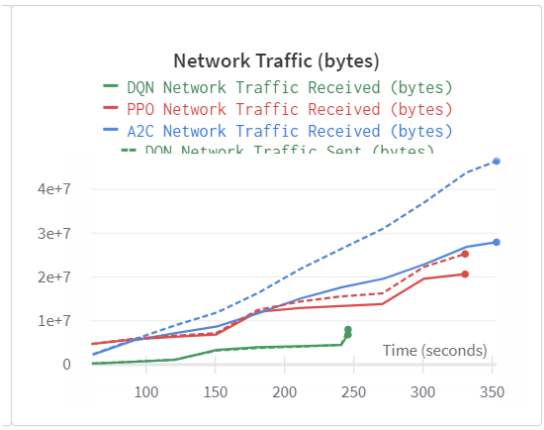


Fig. 16.

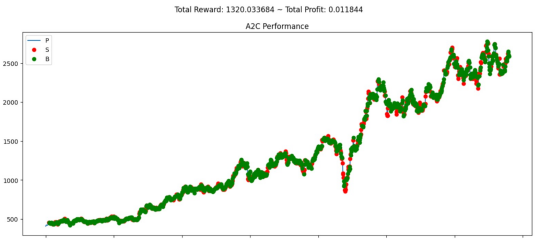


Fig. 17. As we can see from the green highlighted graphs, PPO outperforms both the algorithms i.e, DQN and A2C. However, when it comes to annual average return we can see that DQN performs 45% ¿ PPO and 57% ¿ A2C

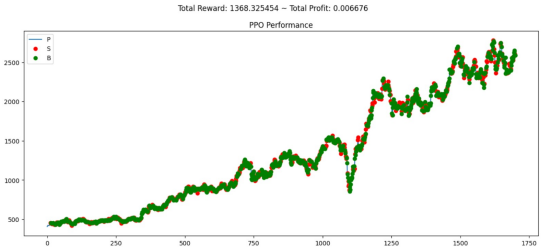


Fig. 18.


Fig. 19.

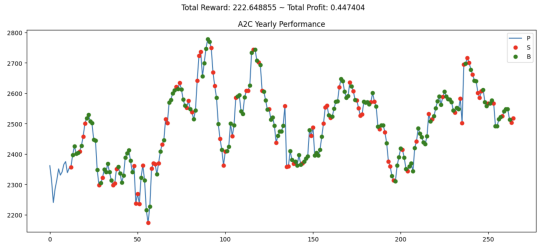


Fig. 20. yearly performance, 2016-03-02 - 2022-08-22 Testing period.

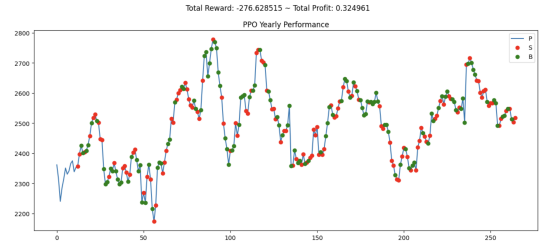


Fig. 21. yearly performance

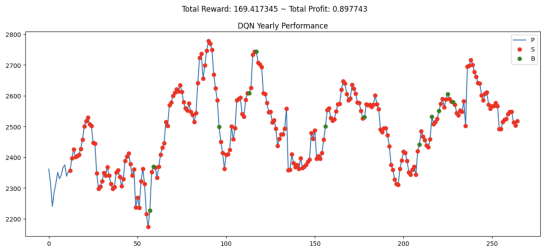


Fig. 22. yearly performance

## IX. Conclusion

Correct analysis of the stock market state is one of the challenges when implementing stock trading based on deep reinforcement learning. In this paper, we analyze the Reliance stock data based on deep reinforcement learning to implement stock trading. Stock data and technical indicators can reflect the changes in the stock market from different perspectives, we use different deep neural networks to extract the features of the data source and fuse those features, and the fused features are more helpful to learn the optimal dynamic trading strategy. Further, we also built an environment that mimics the stock market with different actions such as Sell and Buy and did a comparative study between different RL algorithms to get the best one. It can be concluded from the experimental results that the trading strategies learned based on the deep reinforcement learning method can be dynamically adjusted according to the stock market changes and have more advantages over just DL. It is important to obtain informative data from relevant sentiments and or the news for stock trading. In future research, we will consider using this information and train a more stable trading strategy.

## X. Data Availability

- The experimental data in this paper can be downloaded from Yahoo Finance (https://finance.yahoo.com/).
- https://wandb.ai/full-metal/sb3-Latest/reports/All-Runs-A2C-PPO-an-DQNVmlldzozMjgxMjA4access Token=xk5er4wn8mn1ckx625egt4ukypsf0p2yq1pat6kxe69743114vvg8j18wurj3fsi

## XI. Appendix

Namarta conducted a review of the literature to determine the topic, and Nahid added additional relevant work to the literature to select the method for our model. Rajath developed the model, and then Namarta and Nahid used different performance metrics to assess the model's accuracy. We three initially attempted to collect the dataset from diverse sources, and in the end we concluded with the proposed work where we did a comparative analysis. We collaborated synchronously and continuously throughout the project while researching, developing, and writing this report.

## References


[1] Y. Li, W. Zheng, and Z. Zheng, "Deep, robust reinforcement learning for practical algorithmic trading," IEEE Access, vol. 7, pp. 108014–108022, 2019.

[2] D. Bertsimas and A. Lo, "Optimal control of execution costs," Journal of Financial markets, 1998.

[3] Hiransha, M., Gopalakrishnan, E.A., Menon, V.K., Soman, K.: Nse stock market prediction using deep-learning models. Procedia computer science 132, 1351–1362 (2018).

[4] Lo, A.W., Mamaysky, H., Wang, J.: Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. The journal of finance 55(4), 1705–1765 (2000).

[5] Thomsett, M.C.: Getting started in fundamental analysis. John Wiley Sons (2006).

[6] Malkiel, B.G., Fama, E.F.: Efficient capital markets: A review of theory and empirical work. The journal of Finance 25(2), 383–417 (1970).

[7] lorescu, I., Mariani, M.C., Stanley, H.E., Viens, F.G.: Handbook of Highfrequency Trading and Modeling in Finance, vol. 9. John Wiley Sons (2016).

[8] Lee, J.W.: Stock price prediction using reinforcement learning. In: ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 01TH8570). vol. 1, pp. 690–695. IEEE (2001).

[9] Jangmin, O., Lee, J., Lee, J.W., Zhang, B.T.: Adaptive stock trading with dynamic asset allocation using reinforcement learning. Information Sciences 176(15), 2121– 2147 (2006).

[10] BELLMAN R. Dynamic programming and Lagrange multipliers [J]. Proceedings of the National Academy of Sciences, 1956, 42(10): 767-769.

[11] Z. Ning, P. Dong, X. Wang, L. Guo, J. J. Rodrigues, X. Kong, J. Huang, and R. Y. Kwok, "Deep reinforcement learning for intelligent internet of vehicles: An energy-efficient computational offloading scheme," IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 4, pp. 1060–1072, 2019.

[12] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," IEEE Communications Surveys Tutorials, vol. 21, no. 4, pp. 3133–3174, 2019.

[13] R. Q. Hu et al., "Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning," IEEE Transactions on Vehicular Technology, vol. 67, no. 11, pp. 10 190–10 203, 2018.

[14] Jagdish Bhagwan Chakole, Mugdha S. Kolhe, Grishma D. Mahapurush, Anushka Yadav, Manish P. Kurhekar,A Q-learning agent for automated trading in equity stock markets,Expert Systems with Applications,Volume 163,2021,113761, ISSN 0957-4174.

[15] Hongfeng Xu, Lei Chai, Zhiming Luo, Shaozi Li, Stock movement prediction via gated recurrent unit network based on reinforcement learning with incorporated attention mechanisms, Neurocomputing,Volume 467,2022,Pages 214-228,ISSN 0925-2312.

[16] Prahlad Koratamaddi, Karan Wadhwani, Mridul Gupta, Sriram G. Sanjeevi, Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation, Engineering Science and Technology, an International Journal, Volume 24, Issue 4, 2021, Pages 848-859, ISSN 2215-0986.

[17] Tan, Zhiyong, Chai Quek, and Philip YK Cheng. "Stock trading with cycles: A financial application of ANFIS and reinforcement learning." Expert Systems with Applications 38.5 (2011): 4741-4755.

[18] ]Rundo, Francesco. "Deep LSTM with reinforcement learning layer for financial trend prediction in FX high frequency trading systems." Applied Sciences 9.20 (2019): 4460.

[19] S. Mehtab and J. Sen, "A robust predictive model for stock price prediction using deep learning and natural language processing", In Proceedings of the 7th International Conference on Business Analytics and Intelligence, Bangalore, India, 2019.

[20] Dang, Quang-Vinh. "Reinforcement learning in stock trading." International conference on computer science, applied mathematics and applications. Springer, Cham, 2019.

[21] Y. Deng, F. Bao, and Y. Kong, "Deep direct reinforcement learning for financial signal representation and trading," IEEE Transactions on Neural Networks and Learning Systems , vol. 28, no. 3, pp. 653–664, 2016.