Received 24 May 2024, accepted 14 June 2024, date of publication 24 June 2024, date of current version 8 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3418510

RESEARCH ARTICLE

A Multifaceted Approach to Stock Market Trading Using Reinforcement Learning

YASMEEN ANSARI[1], SAIRA GILLANI[2], MARYAM BUKHARI[3], BYEONGCHEON LEE[4], MUAZZAM MAQSOOD[3], AND SEUNGMIN RHO[5]

[1]Department of Finance, College of Administrative and Financial Sciences, Saudi Electronic University, Riyadh 13323, Saudi Arabia
[2]Faculty of Information Technology and Computer Science (FoIT & CS), University of Central Punjab, Lahore 54000, Pakistan
[3]Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock 43600, Pakistan
[4]Department of Security Convergence, Chung-Ang University, Seoul 06974, Republic of Korea
[5]Department of Industrial Security, Chung-Ang University, Seoul 06974, South Korea

Corresponding authors: Seungmin Rho (smrho@cau.ac.kr) and Muazzam Maqsood (muazzam.maqsood@cuiatk.edu.pk)

This work was supported by the BK21 FOUR Program (Education and Research Center for Securing Cyber-Physical Space) through the National Research Foundation (NRF) funded by the Ministry of Education of Korea under Grant 5199990314137.

ABSTRACT In the recent past, algorithmic stock market trading for financial markets has undergone significant growth and played a major role in investment decisions. Several methods have been proposed with the objective of designing optimum trading strategies to maximize profitability, economic utility, and risk-adjusted returns. Although traditional methods including mean reversion, momentum, and trend following approaches show good results, but have poor generalization and often perform well in specific time frames. Presently, Reinforcement Learning (RL) approaches are more adaptable and continually perceive the environment by making optimum trading decisions. However, it is still difficult to develop a lucrative trading approach in a complicated and dynamic stock market. The primary challenges in RL methods are effective state representation to reflect current market situations and a suitable trading reward to encourage agents to make more informed decisions. To address such challenges, this research presented a multifaceted strategy for multi-stock market trading using RL that incorporates enhanced state representation based on daily historical data, technical indicators, and fundamental indicators from balance sheets, income statements, and cash flow statements. To inform the agent about the impact of decisions taken on a day-to-day basis by considering risk, a novel reward function named PSR is also proposed. The proposed RL agent is trained in a multi-stock environment in which investors have multiple shares and trading signals are needed with the quantity of shares by using Advantage Actor-Critic (A2C), and Deep Deterministic Policy Gradient (DDPG) algorithms. Furthermore, the proposed multifaceted strategy is validated on 30 Dow Jones stocks and the proposed model outperforms the benchmark Dow Jones Industrial Average index during backtesting.

INDEX TERMS Artificial Intelligence, Algorithmic trading, stock market trading, reinforcement learning, technical indicators, fundamental analysis.

I. INTRODUCTION

Stock trading is the process of purchasing and selling a company's shares on the financial market. By means of recurring orders to purchase and sell, the major objective is to maximize capital returns by leveraging market volatility. The main theme is buying at a lower price and then selling for a higher price generates profit. However, in the age of rapid scientific breakthroughs, financial markets have experienced revolutionary shifts, affecting the fundamentals of stock trading techniques and market behavior [1]. This ever-changing environment of stock market conditions continues to draw the interest of investors, economists, as well as researchers. Designing an appropriate trading strategy is the most important aspect of stock trading in order to make more optimal decisions at the right time to attain more profits and

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

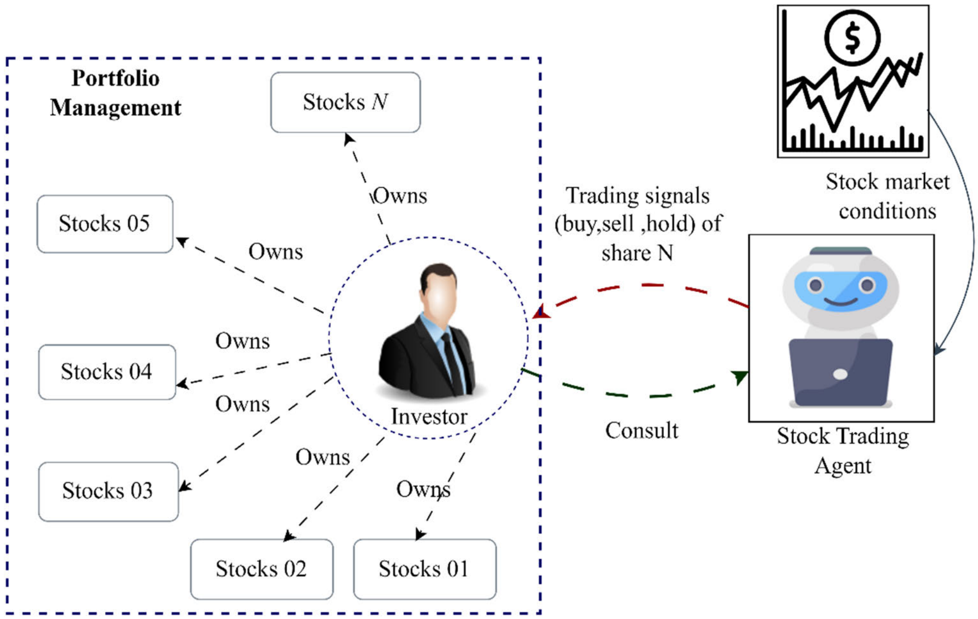


FIGURE 1. Stock trading system assisting in buying and selling of multiple stocks owned by investors.

prevent financial losses on time [2]. In this context, expert-designed trading techniques may not yield positive returns under all market situations [3]. With the emergence of computational methods in finance in the early 1990s, different researchers have introduced the use of Artificial intelligence (AI) to financial transactions in the stock market [4]. More precisely, to assist investors, different AI-assisted approaches have been devised including stock market prediction models [5], algorithmic trading agents [1], portfolio optimization and management [6], etc. Such strategies are quite useful in making better informed and profitable decision-making.

One of the primary benefits of employing these methods is to get rid of emotional decision-making, judging patterns that are overlooked by humans, and consuming information more quickly [4]. It is observed from existing studies that human trader's decision-making is often biased because of their sentiments and as a result, final profits are frequently lower than expected [7]. In addition, stock prices fluctuate often, making it difficult for human traders to respond effectively.

To address such issues, the concept of Algorithmic Trading (AT) is developed and going to evolve from time to time. A pictorial representation of stock trading systems assisting investors in buying, selling, and holding multiple stocks to increase the portfolio of an investor or entity is depicted in Figure 1. Algorithmic trading can be seen as a computer program that executes trades based on rules and logic specified by a programmer. In terms of analysis, a computer-based trading agent took less time in comparison with humans. In the existing literature, different algorithms for trading have been proposed e.g. mean reversion [8], momentum [9], and rules discovery methods [10]. Nevertheless, these rules-based methods show poor generalization and their performance is usually good in specific market situations [11]. Some forecasting-based methods have also been devised, in which future stock values are predicted and later on, rules for trading have been defined [12]. Various kinds of supervised learning techniques including linear models [13], tree-based models [14], and deep neural networks [15], are also employed. However, because of different factors of the stock market such as high volatility and the noisy nature of the stock market make it difficult to estimate future prices properly [16]. Furthermore, there is a significant disparity is observed between forecast signals and profitable trading actions [17], [18].

The usage of supervised machine learning for trading is quite challenging since the training of the model is done to anticipate prices with the objective of minimizing the prediction error [19]. Moreover, mapping between predictive signals to trading positions is non-trivial e.g. horizons of predictions are usually short (one to few days based on daily data) [20]. In practical trading, the handling of risk and portfolio are also important while trading. Hence, to improve stock trading strategies, more advanced approaches, such as Reinforcement Learning (RL), are also employed. One key advantage of using RL in Quantitative Trading (QT) is its ability to do market analysis and decision-making without explicitly anticipating future prices of stock. The usage of RL for Algorithmic trading can be thought of as a decision-making process, with the problem being to continually perceive the environment and make optimum trading decisions. These algorithms have also the capability to enhance their policy over time through self-learning which ultimately makes them suitable and adaptable to be employed for AT. Various methods have been used to build the trading algorithm utilizing reinforcement learning. Some research studies has just used historical stock data [21], while others have used news sentiment trends [22] and approached data fusion models i.e. technical indicators and candlestick charts [2], and fusion of macro-economic and sentiment data [23]. On the other side, several algorithmic advancements have also been devised, such as ensemble models [24]. The fundamental challenge of all these algorithms is the precise awareness of the stock environment, the more knowledgeable an agent is about the stock market, the more accurate decisions it can make. In this aspect, environment perception necessitates accurate and strong feature representation of the stock market. In accordance with that, designing a good reward function is highly critical to aid the RL agents in making more informed judgments. In comparison to the existing methods, although most adaptive and flexible trading techniques based on reinforcement learning have been developed, but still the research gap in improving the strong feature representation of the stock market exists to be filled.

In this paper, a multifaceted approach to stock market trading has been designed i.e. integrating multi-factors analysis in which states of RL-agent are comprised of daily historical data, technical indicators, and fundamental indicators computed from income statements, cash flow statements, and balance sheets. The rationale for taking many aspects into account is to make RL agents capable of capturing a larger variety of market characteristics and making better trading decisions. To be more specific, fundamental data involving companies' balance sheets, income statements, and cash flow statements also have a crucial role in demonstrating the financial performance of companies [25]. A very strong positive relationship between stock market returns and fundamental ratios has been observed in [26]. Hence, integrating fundamental data into stock trading methods helps investors make more informed decisions by providing a full view of a company's fundamentals.

Hence, here in this study, one major research question arises what if these factors including daily data, technical indicators, and fundamental data from companies' balance sheets, income statements, and cash flow statements are modeled as state space features in the RL-based agents? More precisely, in this research study, we have proposed an RL-based stock trading agent that assists investors in making trading decisions by not only considering the historical stock price records but also the technical indicators and fundamental data conditions of the stock market and companies. Secondly, this study also approaches the problem of stock market trading as a multi-stock environment in which investors have diverse portfolios made up of shares in various companies. The designed agent not only gives buying and selling signals but also specifies how many shares to purchase and sell for particular stocks per day. In addition, we have designed a PSR (Portfolio-Sharpe-Returns) based novel reward function to make agents make informed decisions considering risk management, daily returns, and increase in portfolio values. The pin-point contributions of this study are written as follows:

- A multifaceted DRL-based trading agent is designed to perform multi-stock trading with the objective of increasing the overall portfolio returns.
- Daily historical stock data, technical indicators, and fundamental data of companies from balance sheets, income statements, and cash flow statements are merged to create a rich feature representation of the stock market.
- A novel PSR reward function is proposed to provide agents with insights into the short-term profitability and impact of decisions taken on a day-to-day basis while also modeling risk associated with trading decisions.

The remainder of the article has been divided into different sections: Section II provides the related work, Section III discusses the proposed work, and Section IV shows the experimental results with a discussion, followed by Section V, which provides a comparative analysis, limitations, future directions, and conclusions.

II. RELATED WORK

This part includes a thorough review of the literature, as well as discussions of research gaps, limitations, and analysis. Currently, there exist different methods for stock market trading including traditional and rule-based methods, machine learning methods, and deep reinforcement learning techniques. Following is a detailed discussion of those methods:

Algorithmic trading, also referred to as quantitative trading is a subdomain of finance that can be thought of as a technique to make decisions regarding stock's buying and selling to attain maximum profits. The underlying algorithms are reliant on the collection of rules used to derive these decisions. In the initial stages, the majority of the approaches for algorithmic trading are designed by mathematicians, economists as well as traders in which there is no use of AI was found [27]. In this aspect, some common examples include trend following [28], momentum [9], and mean reversion strategies [8]. In these methods, expert knowledge from the domain of finance is of ultimate requirement to identify the financial market's fundamental trend. Following on, some research studies have employed the TTR (Technical Trading rule) based methods [29]. For instance, Metghalchi et al. [29] employs five TTR methods including moving average, relative strength index, momentum, etc. for trading decisions and validates their approach to Turkish indices. In line with that fuzzy-rules-based approaches are also exploited. For instance, Lauguico et al. [30] proposed the fuzzy-logic controller to generate the trading signals of buy, sell, and hold. In addition, features of candlesticks as well as Bollinger Bands (BB) were also employed as technical indicators. Likewise, Kim et al. [31] proposed a hybrid method for stock trading in which trading rules are constructed using rough sets and genetic algorithms. To validate the performance of their trading algorithm, data from the Korea Composite Stock Market Index 200 (KOSPI 200) was utilized and good performance is reported. Although these rule-based methods show good performance, they are limited to performing well in different market situations and have limited generalization ability.

In the recent past, machine learning algorithms have developed much interest among researchers in different kinds of applications such as computer vision, natural language processing, and voice-based systems [32], [33], [34], [35]. In accordance with that, machine learning algorithms have also been used in financial markets [36], [37], [38]. Different studies have carried out stock market predictions or future trends prediction to generate stock trading decisions. For instance, Banik et al. [36] proposed LSTM (Long short-term memory networks) to predict future stock prices along with technical indicators such as MFI, relative RSI, MACD, etc. An investment success score is computed based on the report of these predictions which is later on used by traders to make their trading decisions. Shah et al. [37] proposed a deep learning approach coupled with CNN and LSTM models to perform a time-series modeling having a look-up duration of 20 trading days. Following this forecasting, trading rules have been designed to assist investors regarding trading decisions. Thakur and Kumar [38] proposed a decision support system in which historical data of stock along with technical indicators were used to carry out algorithmic trading using a weighted multicategory generalized eigenvalue support vector machine (WMGEPSVM) and achieves the lowest MDD in comparison with existing methods. Most recent approaches have investigated directional changes-based algorithmic trading as well as machine learning algorithms [39], [40].

Following on, Deep Reinforcement Learning (DRL) is regarded as the third pillar of learning when it comes to stock market trading. For instance, Li et al. [41] proposed DRL strategy based deep-Q-Networks for stock market trading. Their models exhibit good profits on data of historical daily prices of different U.S stocks. Carta et al. [42] proposed multi-layer and multi-ensemble stock trader using DRL. They have employed historical stock price data along with the fusion of Gramian Angular Field images generated through time-series data. These two types of data are fed into the two Deep-learning-based agents followed by an assembling layer to generate intra-day trading signals. Kwak et al. [43] proposed recurrent reinforcement learning along with a self-attention mechanism for stock market trading. The experimentation is performed on historical data of stocks of S&P500, and their method shows good results in terms of Sharpe ratio. It is also indicated in their findings that the performance of their approach will eventually be determined by stock selection as well as portfolio allocation. In order to extract hidden features from stock data, Zou et al. [44] proposed Cascaded Long Short-Term Memory (CLSTM-PPO Model) as a feature extractor in the DRL algorithm. In addition to stock historical data some other technical indicators involving Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Commodity Channel Index (CCI) are also computed, and Average Directional Index (ADX) is also involved in the state-space of DRL algorithm in terms of different evaluation measures such as Sharpe ratio, and Cumulative Return (CR), etc. Other than involving only historical price data and technical indicators, some research studies have also involved the news sentiments into the states of DRL algorithms as feature representatives of the stock market [45], [46]. It is observed from these studies that the inclusion of news sentiments makes DRL agents more powerful. For instance, Chen and Huang [45] proposed a sentiment-aware DRL agent and the findings of their study show that the intended RL agents increased their revenues by integrating pricing information and news sentiment across many channels. However, in their model, only price as well as financial news is employed but other kinds of social media data and government policies should also need to be taken into account. In addition, a reward function based on only profits is not sufficient as there is a need to consider some other metrics e.g. Sharpe ratio. Similarly, Nan et al. [46] also proposed the sentiment and knowledge-based algorithmic trading technique based on Q-learning. In this method, time series data of stock prices is merged with news headline sentiments along with knowledge graphs for investigating news regarding implicit associations. It is observed that without sentiment the Sharpe ratio is about -1.357 while with sentiment, the Sharpe ratio is improved up to the value of 2.432 in the case of MSFT stock. Their proposed method, although shows good results but views trading as a single stock environment at one time.

In comparison to the existing studies, the proposed method captures a wider view of stock market conditions enriching the state representation of RL with rich features. The proposed makes trading decisions by taking into account daily returns, profits, and risk-aware trading decisions.

III. PROPOSED METHODOLOGY

In this section, the working of the proposed RL agent is discussed step by step. A pictorial representation of the proposed work is depicted in Figure 2.

Reinforcement learning, a branch of machine learning, involves powerful agents interacting with the world around them to increase their cumulative rewards. For instance, In 2015, Alpha Go outperformed the human professional players [47]. It also facilitates driven-objective learning and optimal decision-making processes. More explicitly, RL can be defined as a learning strategy in which interaction with environments is done iteratively to self-adjust the learned policies of trading agents. The environment is represented as Markov Decision Process (MDP) denoted as (S, A, T, R, γ) i.e. states S, action A, state transition T, reward R, and discounted factor γ. The return represents the total of future discounted rewards having a discount factor $\gamma \in (0, 1]$. The agent will be awarded for correct actions as well as punished for improper acts. Compared to humans, the agent develops its learning by increasing rewards while minimizing penalties. RL is primarily divided into different types including the value-based reinforcement learning methods

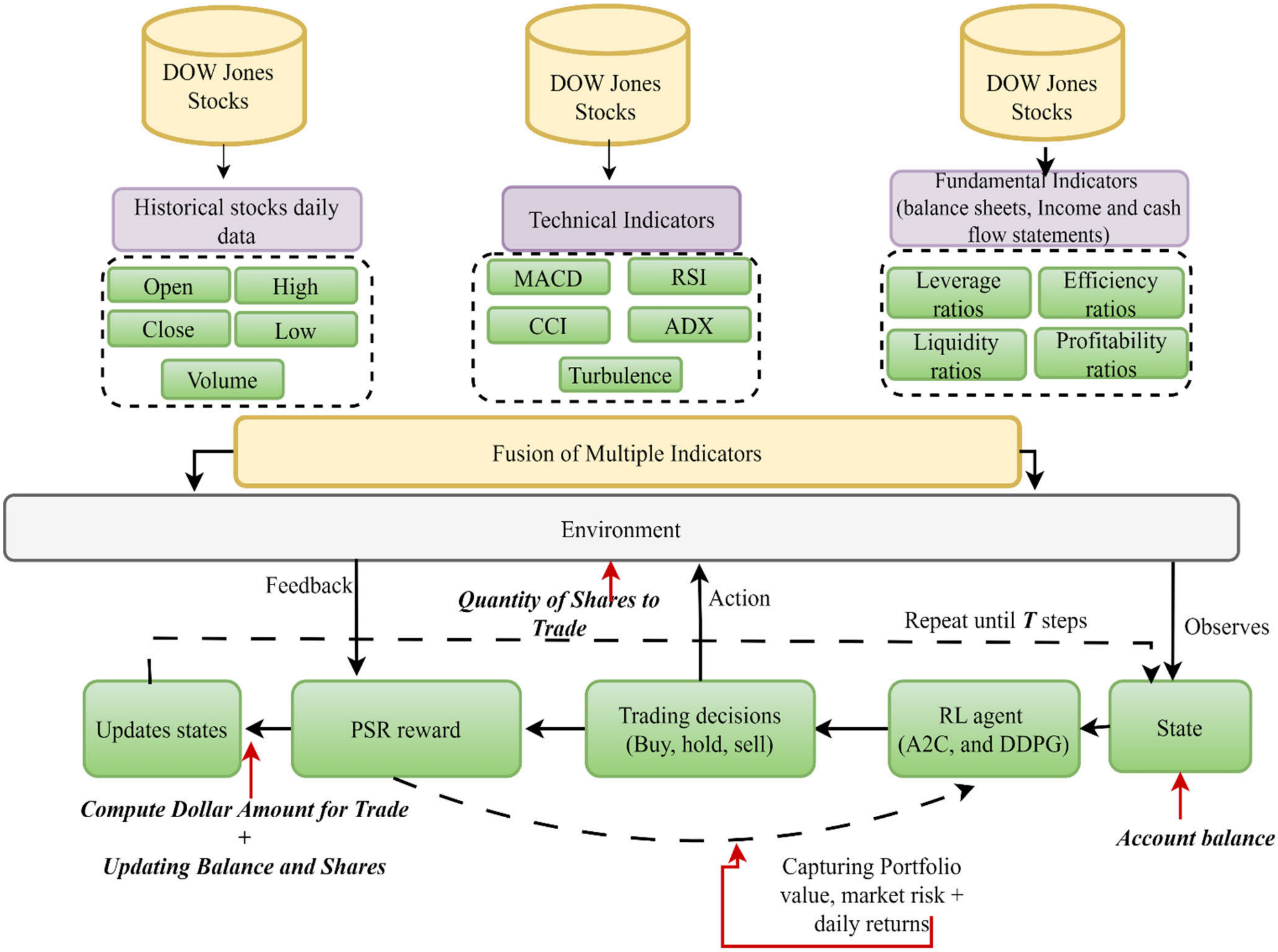


FIGURE 2. A pictorial overview of the proposed methodology for automated stock market trading.

as well as the policy-based reinforcement learning methods. Rather than these, there also exist actor-critic reinforcement learning methods that combine both value-based as well as policy-based methodologies.

To solve any problem using RL, it is necessary to formulate states, rewards, actions, and environment based on the design approach and desired outcome. In algorithmic trading, the state is formed through a series of observations as they are not directly given. To compensate for this, the MDP model has been amended with an observation probability $P(o|s, a)$. This enhanced model is known as the partially observable MDP (POMDP) paradigm [48].

A. MARKOV DECISION PROCESS (MDP) FORMULATION

The MDP consists of random variables that are employed to model stochastic procedures, switching from one state to another based on a few assumptions as well as probabilistic principles. In order to define RL, MDPs are the best mathematical models. In this aspect, the agent is referred to as the decision maker or learner, and the world around the agent in which the agent performs interactions is referred to as the environment. At each time step, i.e. $t \in \{1, 2, 3, \ldots, T\}$ the agent interacts with the environment. In this study, the MDP of multiple stock market trading is a trading environment with the objective to imitate real-world trading procedures.

1) STATES FORMULATION

Accurate representation of states of the environment is crucial for assisting agents in learning optimal policies. The agent's environment in the stock market is determined by the present market conditions. Selecting a set of data inputs is crucial for traders to understand the stock market and make trade rules. Similar to investors who take into different factors while making trading decisions, the proposed RL agent also considers different factors in its observation space. The proposed RL 's states include daily historical data of stocks including open, high, low, close, and volume i.e. $o_t, h_t, l_t, c_t, v_t \in \mathbb{R}^n_+$. More exactly, o_t shows the price of the stock's first transaction when the market opens, h_t and l_t is the highest and lowest price of the stock being traded at time t, c_t is the terminal price on which stocks are traded prior to market close, while v_t is the total number of shares being traded within a particular trading session. At any time in step t, the sum of the balance remaining in the account is denoted as $b_t \in \mathbb{R}^+$, and available shares for every stock in the portfolio are denoted as $h_t \in \mathbb{Z}_+^n$ is also encoded into states. In addition, the technical and fundamental data listed below is calculated and incorporated into the state representation.

a: TECHNICAL INDICATORS

Technical factors are heuristic as well as mathematical computations reliant on historical data of stocks, often utilized by investors to make trading decisions. The following are some technical indicators computed in this study:

- Moving Average Convergence Divergence (MACD)

It is one of the popular indicators used for technical analysis to exploit the changes in strength, direction, momentum as well as the duration of patterns in prices of stocks [49]. It is also referred to as a momentum indicator which recognizes moving averages. We have computed MACD for each stock present in a portfolio denoted as $M_t \in \mathbb{R}_+^n$ where n is the number of stocks.

- Relative Strength Index (RSI)

In financial markets, RSI is another technical indicator used by investors [49]. Depending upon the closing prices in the latest trading session, it is designed to show the present as well as historical strength or weakness of a stock or market. It also indicates the level of fluctuations happening in the recent prices. According to RSI, the stock has been oversold if within the support line, the price swings. This suggests that we can make buying decisions. Similarly, the stock is overbought, if the price swings around resistance and this implies the decision to sell. We have computed RSI for each stock present in a portfolio denoted as $R_t \in \mathbb{R}_+^n$ where n is the number of stocks.

- Commodity Channel Index (CCI)

This metric is a momentum-based oscillator and is computed by employing high, low, and close prices [50]. To show the buying and selling decisions, this metric contrasts the current price to a mean of prices within the specific window size. Moreover, when this metric is zero then it shows price is higher than the mean of historical prices. We have computed CCI for each stock present in a portfolio denoted as $C_t \in \mathbb{R}_+^n$ where n is the number of stocks.

- Directional Movement Index (DMI)

This metric is employed to determine the direction whereby the price of an asset moves [51]. It measures both the direction and magnitude of price movement and can be computed using high, low, and close prices. It has multiple lines i.e. positive and negative directional movement line (+DI) and (−DI), directional index (DX), average directional index (ADX), and EMA for ADX. In this study, we have computed DX for each stock present in a portfolio denoted as $D_t \in \mathbb{R}_+^n$ where n is the number of stocks.

- Turbulence

This metric is defined as periods of high volatility, unpredictability, and discord in financial markets. In this study, we have computed turbulence for each stock present in a portfolio denoted as $T_t \in \mathbb{R}_+^n$ where n is the number of stocks.

b: FUNDAMENTAL INDICATORS

Fundamental data consists of company data from their balance sheets, income statements, and cash flow statements to compute several ratios. This analysis is carried out to identify the security's intrinsic or true value, therefore it could be contrasted to the security's value on the market. Following are the fundamental ratios computed from fundamental data of different stocks collected from Alpha Vantage API in terms of balance sheets, income, and cash flow statements [52].

- ***Current Ratio*** The current ratio is one of the liquidity ratios which indicates the capacity of a company to repay immediate liabilities using present assets.
- ***Acid Test Ratio*** Acid-test ratio is also one of the liquidity ratios that indicates the capacity of a company to repay immediate liabilities with cash and cash equivalents.
- ***Operating Cash Flow Ratio*** The operating cash flow ratio is also one of the liquidity ratios that is an indicator of in a specific time frame, how many times an organization is capable of paying present liabilities utilizing the cash that is earned.
- ***Debt Ratio*** This ratio is one of the leverage financial ratios which indicates the total number of assets of an organization that are provided from debt.
- ***Debt to Equity Ratio*** In this ratio, the ratio of all debt's overall financial obligations to stockholders' equity is determined.
- ***Interest Coverage Ratio*** Interest Coverage ratio indicates how much a company is able to submit its interest expenses.
- ***Asset Turnover Ratio*** The asset turnover ratio is the ability of a company to generate sales with the help of its assets. It is one of the metrics of efficiency ratios.
- ***Inventory Turnover Ratio*** The inventory turnover ratio is the total number of times an inventory is sold as well as updated in a given time frame.
- ***Day Sales in Inventory Ratio*** The day Sales in Inventory ratio is the mean of total days for which a company holds their inventories prior to selling them to its clients.
- ***Return on Ratio*** The return on Asset ratio is a profitability ratio and it is computed to indicate how well a particular company is successful in employing their assets to produce profits.
- ***Return on Equity Ratio*** The return on Equity ratio is also the profitability ratio, and it is calculated to determine how well a particular company is successful in employing its equity to produce profits.

Above all, technical and fundamental indicators are computed and merged with historical data of stocks to generate a rich representation of the stock market.

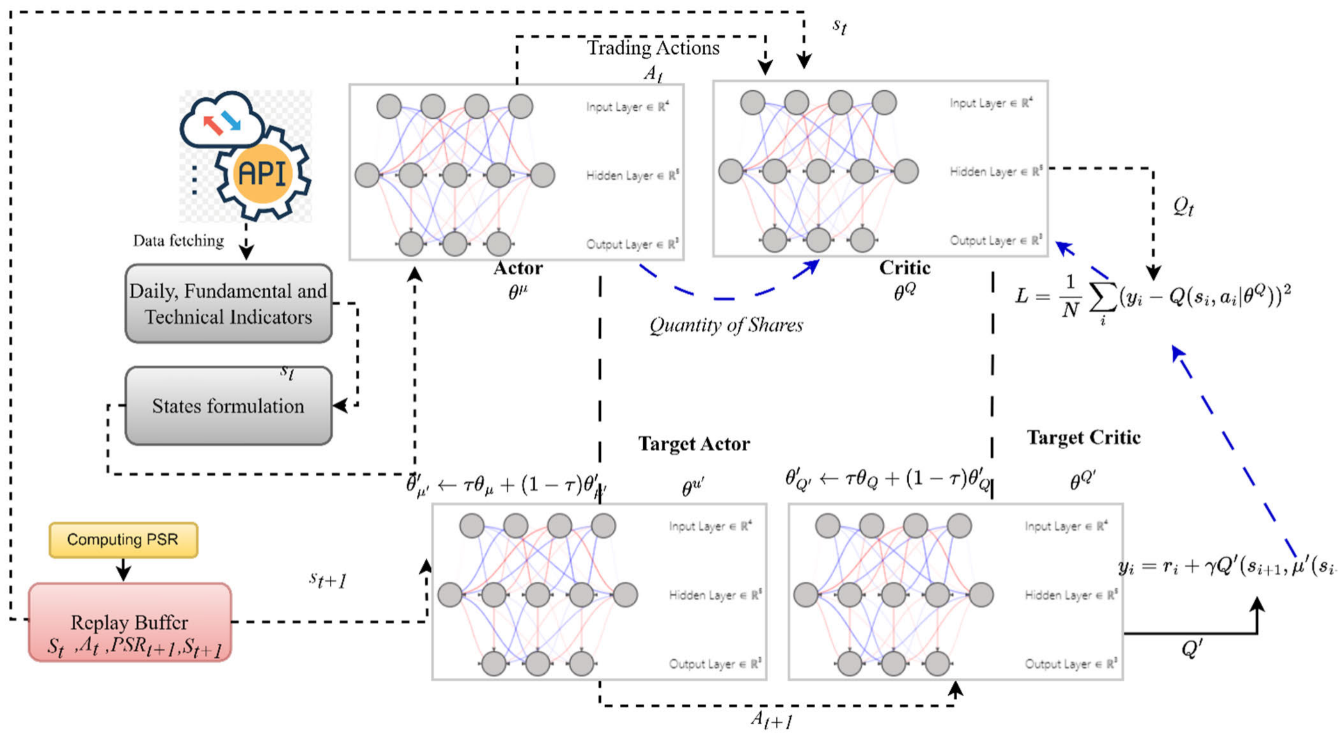


FIGURE 3. Detailed pictorial representation of proposed reinforcement model.

2) ACTIONS IN THE TRADING ENVIRONMENT

Actions are the agent's response after interacting with the trading environment. In this, the agent makes buy sell, and hold decisions after observing different aspects of the stock market including fundamental, technical, and daily data.

These actions $a \in A$ denoted as $a \in \{-1, 0, 1\}$ in which -1 shows to sell, 0 shows hold while 1 show buy. Here, the actions are taken over different shares, hence action space is $\{-k, \ldots, -1, 0, 1, \ldots, k\}$, in which symbol k shows the total shares. In this study, we set $k = 100$.

3) PSR REWARD

Agents are rewarded with numerical points based on the quality of their actions at each step. Rewards are feedback from the environment to the agent to guide its learning. In this study, we have proposed the PSR (Portfolio Sharpe Returns) based reward function. The proposed reward is capable of informing the agent about the short-term daily return it made after making buying, selling, and holding decisions over multiple shares. Mathematically, the proposed reward function is defined below in equation (1):

$$Reward = Change\ in\ Portfolio + Sharpe\ ratio + 0.9 * daily\_returns \quad (1)$$

where Sharpe ratio is given below:

$$Sharpe\ ratio = \frac{E[R_a - R_b]}{\sigma_a} \quad (2)$$

In the above equation, E denotes expected value, R_a denotes asset return, R_b denotes risk-free return, and σ indicates the standard deviation of asset excess return. Here, value 0.9 is tuned by carrying out different simulations.

4) STATE TRANSITIONS

After deciding on an action by the RL agent according to its present state and policy (for example, buying or selling a specific number of shares of an asset), the RL agent executes that action in the market. As an example, if the agent chooses to purchase 100 shares of a company, a market transaction is executed, altering the agent's portfolios as well as the financial situation. Hence, after taking action, the dollar amount of shares, balance, and trading shares are updated followed by updating the states of RL for the next day. For instance, if the agent purchases stock, the dollar value of shares owned in that asset increases, whereas the current balance falls by the sum of money paid on the transaction. After updating the portfolio as well as trading shares, the RL agent moves to the subsequent state in the state space. The state provides an instantaneous view of the outside world and the agent's position following performing the action.

B. PROPOSED REINFORCEMENT AGENTS

In this study, we have employed two different reinforcement learning models namely: Advantage Actor Critic (A2C), and Deep Deterministic Policy Gradient (DDPG). The detailed

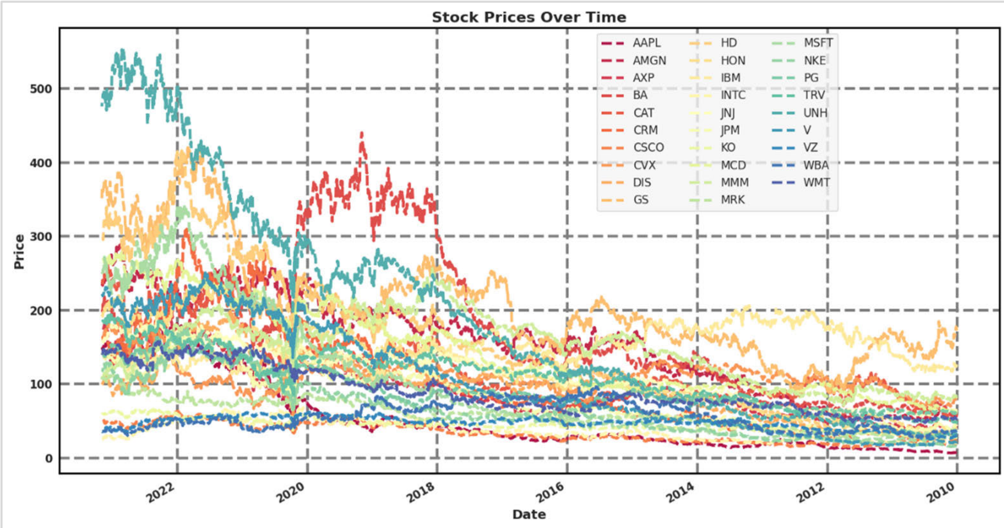


FIGURE 4. Stock prices over time for DOW Jones 30 companies.

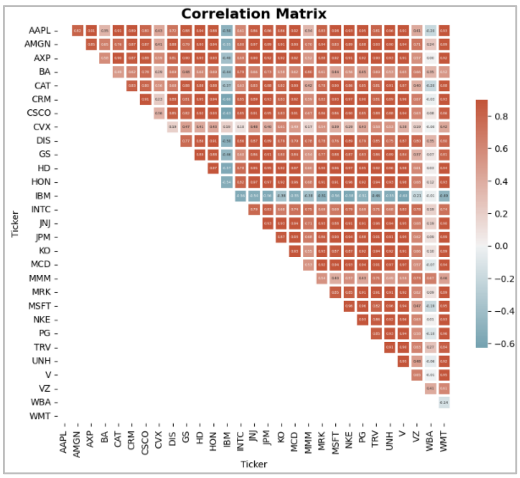


FIGURE 5. Correlation matrix of DOW jones stocks.

pictorial representation of the proposed model is depicted in Figure 3. Following is a detailed description of those models:

1) ADVANTAGE ACTOR-CRITIC (A2C)

Actor-critic algorithms of RL integrate the features of both value-based and policy-reliant techniques for instance actor-critic algorithms address the problem of high variance which occurs in the back propagation of policy-based techniques [53]. Hence, integrating actor-critic approaches with generalized advantage estimation considerably minimizes the variability of gradient updates. In this context, A2C [54] is one of the popular algorithms designed to enhance policy gradient updates. In A2C, the advantage function is added to overcome the variance of the policy gradient. More precisely, the critic module computes the advantage function, rather than solely estimating the value function. The assessment of an action taken by a trading agent is not only reliant on the benefit of the action i.e. by selling, buying, and holding stocks, but also its potential for further improvement. This decreases policy network variance as well as improves model robustness. By employing different data instances, the A2C model employs the same network for agents to update the gradients of the model. For interaction, every agent observes the environment on their own and takes action. In order to pass over the mean gradients from all agents to a global model, a coordinator is employed by A2C in every iteration when the agents are finished with the computation of their gradients. This global model later on can be used to update the actor and critic model. Synchronized gradient updates are efficient, quicker, and suitable for larger batch sizes of stock data. Hence, the A2C model is ideal for stock trading due to its reliability. The mathematical form of the A2C objective function is given below in equation (3):

$$\nabla J_{\theta}\left(\theta\right) = \mathbb{E}\left[\sum_{t=1}^{T} \nabla_{\theta} log\pi_{\theta}\left(a_t \mid s_t\right) A\left(a_t \mid s_t\right)\right], \tag{3}$$

In the above equation, $\pi_{\theta}\left(a_t \mid s_t\right)$ indicates the policy network, while the advantage function is shown by $A\left(a_t \mid s_t\right)$. This advantage function can be re-written as:

$$A\left(a_t \mid s_t\right) = Q\left(a_t \mid s_t\right) - V(s_t) \tag{4}$$

Or

$$A\left(a_t \mid s_t\right) = r(s_t, a_t, s_{t+1}) + \gamma V(s_{t+1}) - V\left(s_t\right) \tag{5}$$

The term $r(s_t, a_t, s_{t+1}) + \gamma V(s_{t+1})$ is the temporal difference error of a state-value function. In this study, the actor and critic network are based on a multi-layer perceptron.

2) DEEP DETERMINISTIC POLICY GRADIENT (DDPG)

Deep Deterministic Policy Gradient (DDPG) algorithms integrate the ideas of both deep Q-learning as well as deterministic policy gradients and for function approximation,

TABLE 1. Data collected from income statements of each stock of DOW jones.

Reported Currency	Research And Development	Other Non-Operating Income
Gross Profit	Operating Expenses	Depreciation
Total Revenue	Investment Income Net	Depreciation And Amortization
Cost Of Revenue	Net Interest Income	Income Before Tax
Cost of Goods and Services Sold	Interest Income	Income Tax Expense
Operating Income	Interest Expense	Interest And Debt Expense
Selling General and Administrative	Non-Interest Income	Net Income From Continuing Operations
Comprehensive Income Net of Tax	ebit	ebitda
Net Income	Date	

TABLE 2. Data collected from balance sheets of each stock of DOW jones.

reported Currency	Total Assets	Total Current Assets
Long Term Investments	Short Term Investments	Other Current Assets
Deferred Revenue	Current Debt	Short Term Debt
Long Term Debt Non-current	Short Long Term Debt Total	Other Current Liabilities
Common Stock	Common Stock Shares Outstanding	Current Net Receivables
Cash And Cash Equivalents at Carrying Value	Cash And Short-Term Investments	inventory
Intangible Assets	Intangible Assets Excluding Goodwill	goodwill
Other Non-Current Assets	Total Liabilities	Total Current Liabilities
Total Non-Current Liabilities	Capital Lease Obligations	Long Term Debt
investments	Retained Earnings	Current Long-Term Debt
Total Non-Current Assets	Property Plant Equipment	
Current Accounts Payable	Accumulated Depreciation Amortization PPE	

TABLE 3. Data collected from cash flow statements of each stock of DOW jones.

Reported Currency	Operating Cash flow	Payments For Operating Activities
Proceeds From Operating Activities	Change In Operating Liabilities	Change In Operating Assets
Depreciation Depletion And Amortization	Change In Receivables	Profit Loss
Capital Expenditures	Change In Inventory	Cash flow from Investment
Cash flow From Financing	Proceeds From Repayments of Short-term Debt	Payments For Repurchase of Common Stock
Payments For Repurchase of Equity	Payments For Repurchase of Preferred Stock	Dividend Payout
Dividend Payout Common Stock	Dividend Payout Preferred Stock	Proceeds From Issuance of Common Stock
Proceeds From Issuance of Preferred Stock	Proceeds From Repurchase of Equity	Proceeds From the Sale of Treasury Stock
Change In Cash and Cash Equivalents	Change In Exchange Rate	Net Income
Proceeds From Issuance of Long-Term Debt And Capital Securities Net		

neural networks are employed [55]. In comparison with deep-Q-learning, in which learning of the model takes place via Q-values, it faces the challenges of the curse of dimensionality. However, opposing that, DDPG carries out learning in a direct manner from observations via policy gradients. In DPG, a parameterized actor function denoted as $\mu(s|\theta^{\mu})$ is retained to indicate present policies by performing the mapping of states into particular actions in a deterministic manner. The critic $Q(s, a)$ performs learning by employing the equation of Bellman. By utilizing the chain rule on the anticipated returns from the initial distribution J depending upon the weights of the actor are indicated in the below equations (6)-(7):

$$\nabla_{\theta_{\mu}} \approx Es_t \sim \rho^{\beta} \left[\nabla_{\theta_{\mu}} Q\left(s, a \,|\, \theta^{Q}\right) |_{s=s_t, a=\mu(s_t|\theta^{\mu})} \right], \tag{6}$$

$$= Es_t \sim \rho^{\beta} \left[\nabla_{a} Q\left(s, a \,|\, \theta^{Q}\right) |_{s=s_t, a=\mu(s_t) \nabla_{\theta^{\mu}} \mu(s \,|\, \theta^{\mu})|s=s_t} \right] \tag{7}$$

However, because of non-linear function approximations, the performance is not as good. Hence, in DDPG neural networks are employed for function approximates. For this, consider a stock trading agent based on DDPG taking a decision of either

TABLE 4. List of DOW jones stocks.

Symbol	Description	Symbol	Description	Symbol	Description
AXP	American Express Co	GS	Goldman Sachs Group Inc	JPM	JPMorgan Chase & Co
AMGN	AMGN Inc	HD	Home Depot Inc	MCD	McDonald's Corp
Apple	Apple Inc	HON	Honeywell International Inc	MMM	3M Co
BA	Boeing Co	IBM	International Business Machines Corp	MRK	Merck & Co Inc
CAT	Caterpillar Inc	INTC	Intel Corp	MSFT	Microsoft Corp
CISCO	Cisco Systems Inc	JNJ	Johnson and Johnson	NKE	Nike Inc
CVX	Chevron Corp	KO	Coca-cola Co	PG	Procter & Gamble Co
TRV	Travelers Companies Inc	UNH	UnitedHealth Group Inc	CRM	Salesforce Inc
VZ	Verizon Communications Inc	V	Visa Inc	WBA	Walgreens Boots Alliance Inc
WMT	Walmart Inc	DIS	Walt Disney Co	DOW	Dow Inc

buy, sell, and hold i.e. action a_t at s_t and obtains PSR reward r_t and moved to state s_{t+1} at time step t. Such collection of transitions (s_t, a_t, r_t, s_{t+1}) are kept in in replay buffer R. Following on, N collection of transitions is pulled from the buffer and change the Q-value y_i as follows:

$$y_i = r_i + \gamma Q'\left(s_{i+1}, \mu'\left(s_{i+1} \mid \theta^{\mu'}\right) \mid \theta^{Q'}\right) \tag{8}$$

The critic model is updated by minimizing the loss function as follows:

$$L = \frac{1}{N}\sum_i \left(y_i - Q(s_i, a_i|\theta^Q\right)^2 \tag{9}$$

While the actor is updated as follows:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q\left(s, a \mid \theta^Q\right) |_{s=s_i, a=\mu(s_i)\nabla_{\cdot\theta^\mu}\mu(s\mid\theta^\mu)|s_i} \tag{10}$$

In the last, the target networks of DDPG are updated as follows:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \tag{11}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \tag{12}$$

The DDPG's ability for handling continuous action space makes it ideal for stock trading.

IV. EXPERIMENTS AND RESULTS

In this section, we describe the data, evaluation criteria utilized to study the model, and the findings of our proposed technique, as well as analysis of findings.

A. DATA PREPROCESSING

In this research, the proposed model is validated on the Dow Jones 30 constituent stocks as a multi-stock trading strategy. The daily data is collected from Yahoo Finance API within the time frame of 2010-01-01 to 2023-03-01 from which training data is from 2010-01-01 to 2021-10-01 while the backtesting is done from 2021-10-01 to 2023-03-01. Similarly, the fundamentals of companies including balance sheets, income statements, and cash flow statements are collected from Alpha Vantage [52]. The data available in balance sheets, income, and cash flow statements are provided in Tables 1, Table 2, and Table 3. Given that fundamental data is given quarterly, we used rolling techniques to change their frequency to daily, allowing smooth integration with daily stock data and technical indicators. The stocks for which income statements, cash flow statements, and balance sheet data are scrapped from Alpha Vantage are given in Table 4. Every value in Table 4 corresponds to a certain stock symbol, which acts as an abbreviated identification for the company's shares, and the full name of the company it represents. Moreover, the historical stock prices of DOW Jones stocks and the correlation matrix are depicted in Figure 4 and Figure 5.

B. EVALUATION METRICS

We used several kinds of evaluation metrics to determine how effective the suggested agent is at generating stock trading decisions. Following are our evaluation metrics:

- ***Cumulative Return*** The number of returns at the ending of trading session is referred to as a cumulative return.
- ***Sharpe Ratio*** The metric measures the performance of an agent in terms of returns on an investment by considering risk.
- ***Max Drawdown*** This metric measures the agent robustness by determining the greatest loss from a peak to a trough throughout a certain time period, showing the worst fall in the value of the portfolio that the proposed agent might have suffered.
- ***Annual Return*** This metric measures the total returns in terms of profit and loss for an investment strategy within a 1-year time frame.
- ***Annual Volatility*** This metric also measures model robustness by computing the standard deviation of portfolio returns.
- ***Calmar Ratio*** This measures the efficiency of investment companies like hedge funds or commodities trading advisors (CTAs).

TABLE 5. Results of proposed agents and baseline ^DJI index using DTF.

Model	Cumulative return	Stability	Annual return	Annual volatility	Max Drawdown
A2C	-0.478±4.28	0.164742	-0.358±3.04	18.574±1.122	-20.478±2.479
DDPG	1.648%±5.72	0.147945	1.146%± 4.03	18.098%±1.44	-19.492%±3.03
Baseline ^DJI Index	-0.048644	0.280983	-0.034876	18.16%	-22%

TABLE 6. Results of proposed agents and baseline ^DJI index in terms of ratios using DTF.

Model	Sharpe ratio	Calmar ratio	Omega ratio	Tail ratio	Sortino ratio
A2C	0.0789690±0.15	-0.003322±0.15	1.013462±0.027	0.9956±0.05	0.110748±0.22
DDPG	0.17±0.23	0.087573±0.24	1.027647±0.04	0.949873±0.33	0.229468±0.08
Baseline ^DJI Index	-0.105351	-0.158953	0.982546	0.970602	-0.146974

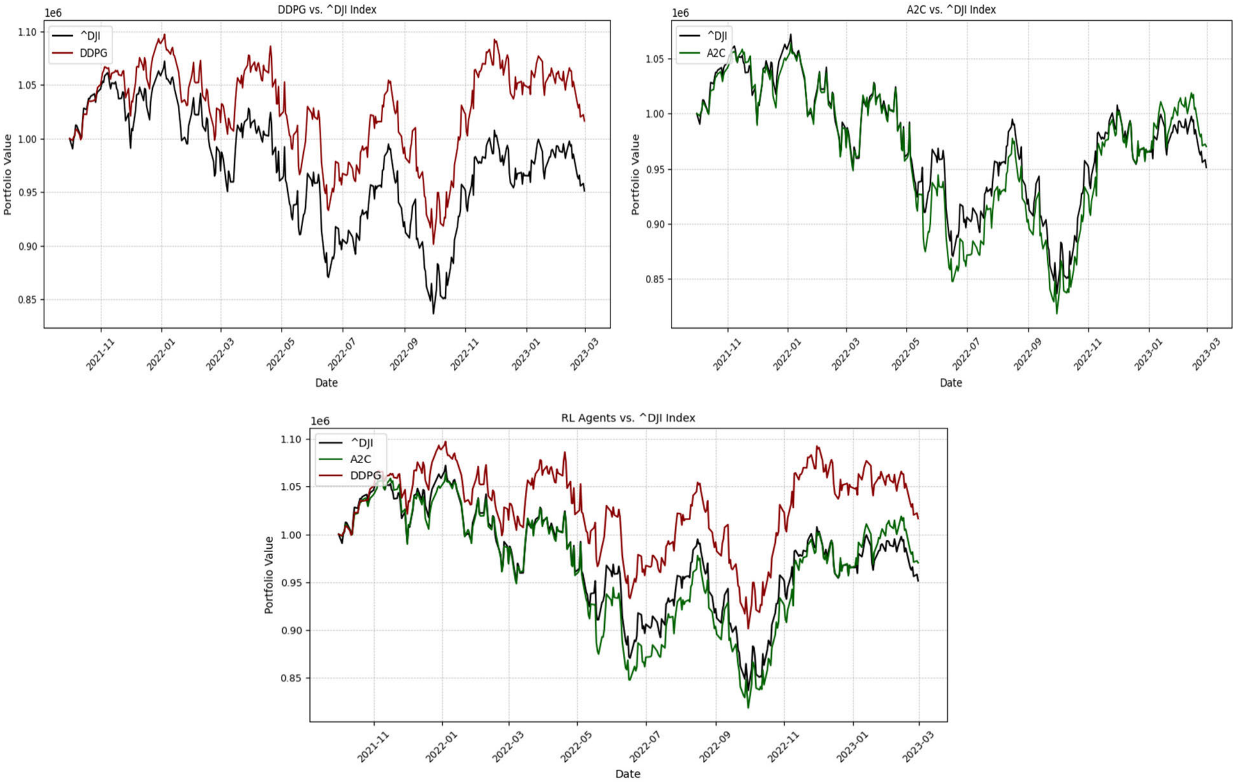


FIGURE 6. Performance of RL agent's vs baseline ^DJI index during backtesting with DFT experimental setup.

- ***Omega Ratio*** This is defined as the risk-return performance of an investment strategy executed by a trading agent.
- ***Tail Ratio*** The tail ratio also indicates risk-adjusted performance which is employed to determine the investment's downside risk.
- ***Sortino Ratio*** The Sortino ratio also indicates the risk-adjusted performance of a trading strategy by doing penalization of returns that fall under the needed rate of return.

C. RESULTS OF PROPOSED TRADING AGENTS WITH DTF PROTOCOL

To validate the proposed method, we have designed different experimental setups. For instance, in the DTF protocol or setup, we have taken daily historical data, technical indicators, and fundamental indicators into state representation. Following on, we trained A2C and DDPG models, on DOW Jones comprising 30 companies. The stock trading agents are trained from 2010-01-01 to 2021-10-01 by taking into different information including historical stock data, fundamental indicators, and technical indicators. The feedback provided to the model is in terms of PSR which considers portfolio value, Sharpe ratio, and daily returns. By combining these components, PSR provides a comprehensive evaluation of portfolio outcomes, managing risks, and investment plan efficacy. The results in terms of evaluation metrics are computed after training. More precisely, both models are trained for 50000-time steps, with an initial amount of 1000000, with an allowable quantity of shares selling and buying set to

TABLE 7. Results of proposed agents and baseline ^DJI index using DT.

Model	Cumulative return	Stability	Annual return	Annual volatility	Max Drawdown
A2C	-2.752%±2.73	0.166±0.11	-1.974%±1.96	19.910%± 0.032	-22.3%±6.90
DDPG	-0.478±4.28	0.164742±0.15	-0.36±3.04	18.574%± 1.12	-20.408±2.47
Baseline ^DJI Index	-0.048644	0.280983	-0.034876	18.16%	-22%

TABLE 8. Results of proposed agents and baseline ^DJI index in terms of ratios using DT.

Model	Sharpe ratio	Calmar ratio	Omega ratio	Tail ratio	Sortino ratio
A2C	-0.0043±0.12	-0.089±0.12	0.999358±0.020	0.938245±0.05	-0.00543±0.17
DDPG	0.078969±0.15	-0.003322±0.15	1.013462±0.020	0.995675±0.05	0.11074±0.22
Baseline ^DJI Index	-0.105351	-0.158953	0.982546	0.970602	-0.146974

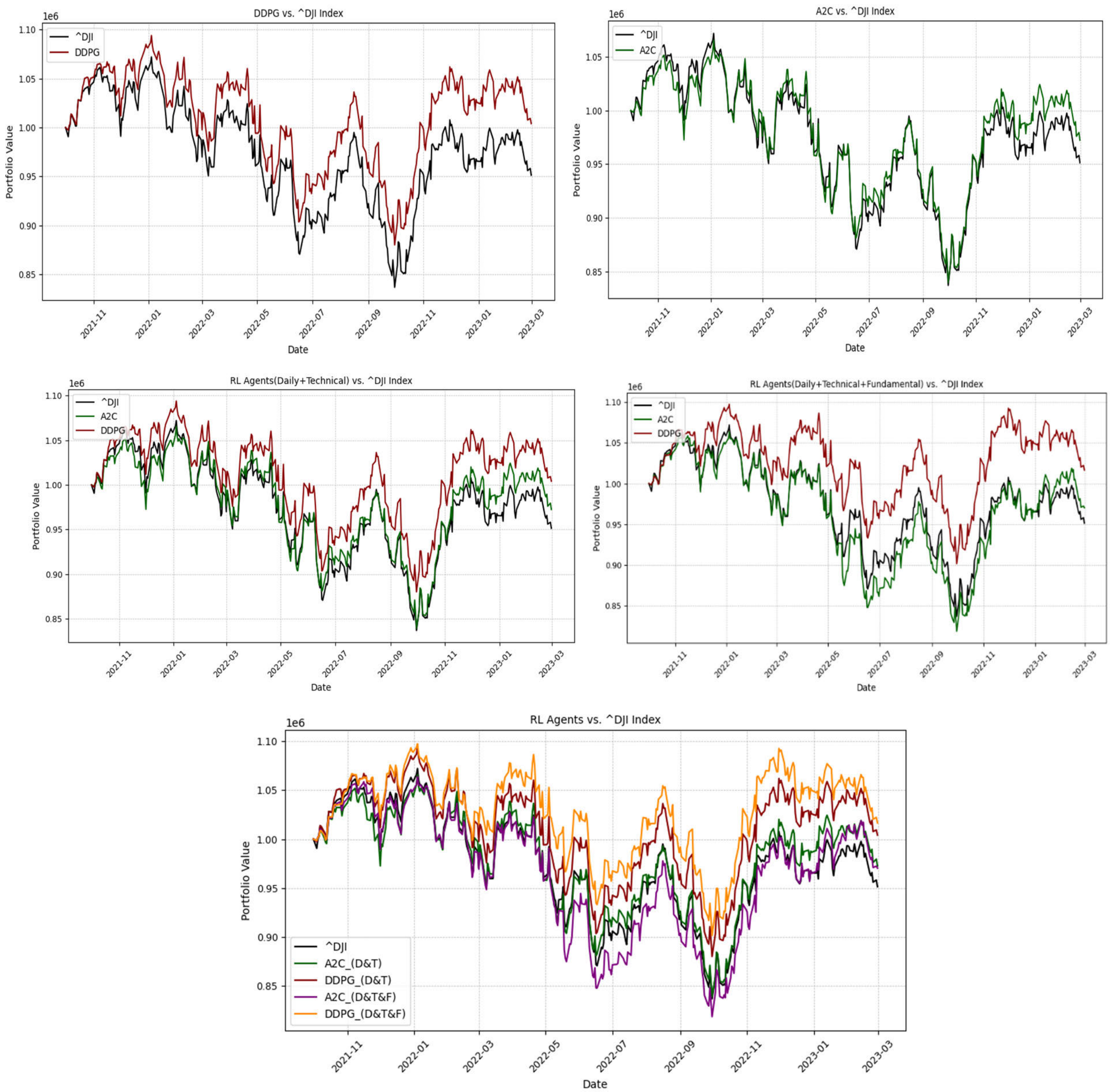


FIGURE 7. Performance of RL agent's vs baseline ^DJI index during backtesting with DT experimental setup.

TABLE 9. Results of proposed agents and baseline ^DJI index using TF.

Model	Cumulative return	Stability	Annual return	Annual volatility	Max Drawdown
A2C	1.152±6.56	0.147±0.14	0.784±4.6	17.750±0.87	-17.760±3.50
DDPG	-0.004±5.88	0.14±0.150	-0.030±4.17	17.700±1.43	-19.434±3.38
Baseline ^DJI Index	-0.048644	0.280983	-0.034876	18.16%	-22%

TABLE 10. Results of proposed agents and baseline ^DJI index in terms of ratios using TF.

Model	Sharpe ratio	Calmar ratio	Omega ratio	Tail ratio	Sortino ratio
A2C	0.131421±0.27	0.082909±0.25	1.023034±0.04	0.997267±0.07	0.188528±0.38
DDPG	0.095946±0.23	0.021578±0.22	1.016437±0.03	1.018637±0.08	0.134832±0.32
Baseline ^DJI Index	-0.105351	-0.158953	0.982546	0.970602	-0.146974

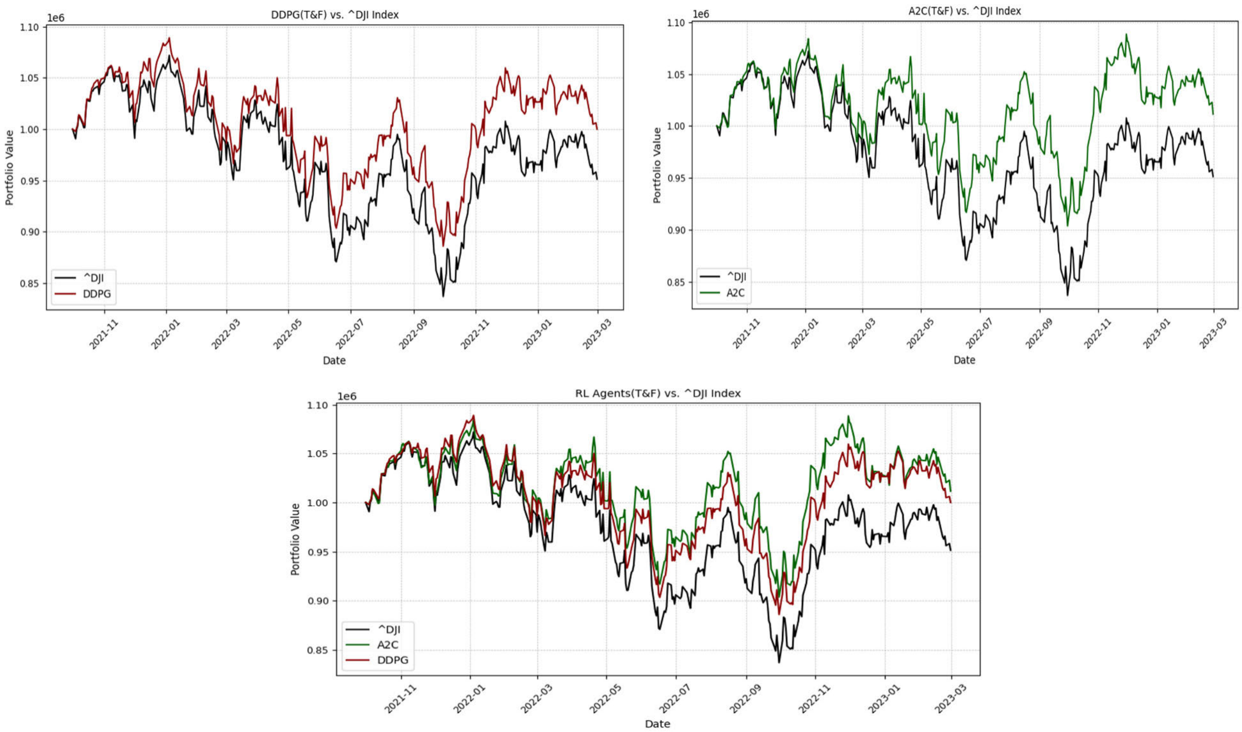


FIGURE 8. Performance of RL agent's vs baseline ^DJI index during backtesting with TF experimental setup.

100. The results of the A2C and DDPG models are shown in Tables 5 and 6. In Table 5, the performance in terms of cumulative returns, stability, annual returns, max drawdown, and annual volatility is given while the performance in terms of ratios including Sharpe, Calmar, Sortino, tail, and Omega is given in Table 6. In addition to that, the comparison with the baseline ^ DJI index has also been carried out. The results in Table 5 show that both models perform better than the baseline ^ DJI Index particularly, in terms of the Sharpe ratio, and Sortino ratio. The annual return generated by DDPG is also higher in comparison with the baseline ^DJI index. This means that during the trading period, investors who followed the DDPG approach outperformed and were able to generate more profits than those who invested in the baseline index. However, if performance has been contrasted then results of DDPG are better than A2C. Moreover, the Sharpe ratio is a fundamental indicator for assessing the risk-adjusted return of an investment technique. A Sharpe ratio of 0.17 as shown in Table 6, for DDPG, indicates that the strategy's returns, after considering the amount of risk, are good and encouraging. The baseline index ^DJI has a Sharpe ratio of around -0.10, indicating that its returns could not effectively account for the degree of risk involved, leading to inferior risk-adjusted performance.

Similarly, the Sharpe ratio of A2C is also positive in comparison with the baseline ^DJI Index. If the analysis has been performed using Max drawdown, then baseline ^DJX generates a somewhat bigger Max drawdown of around

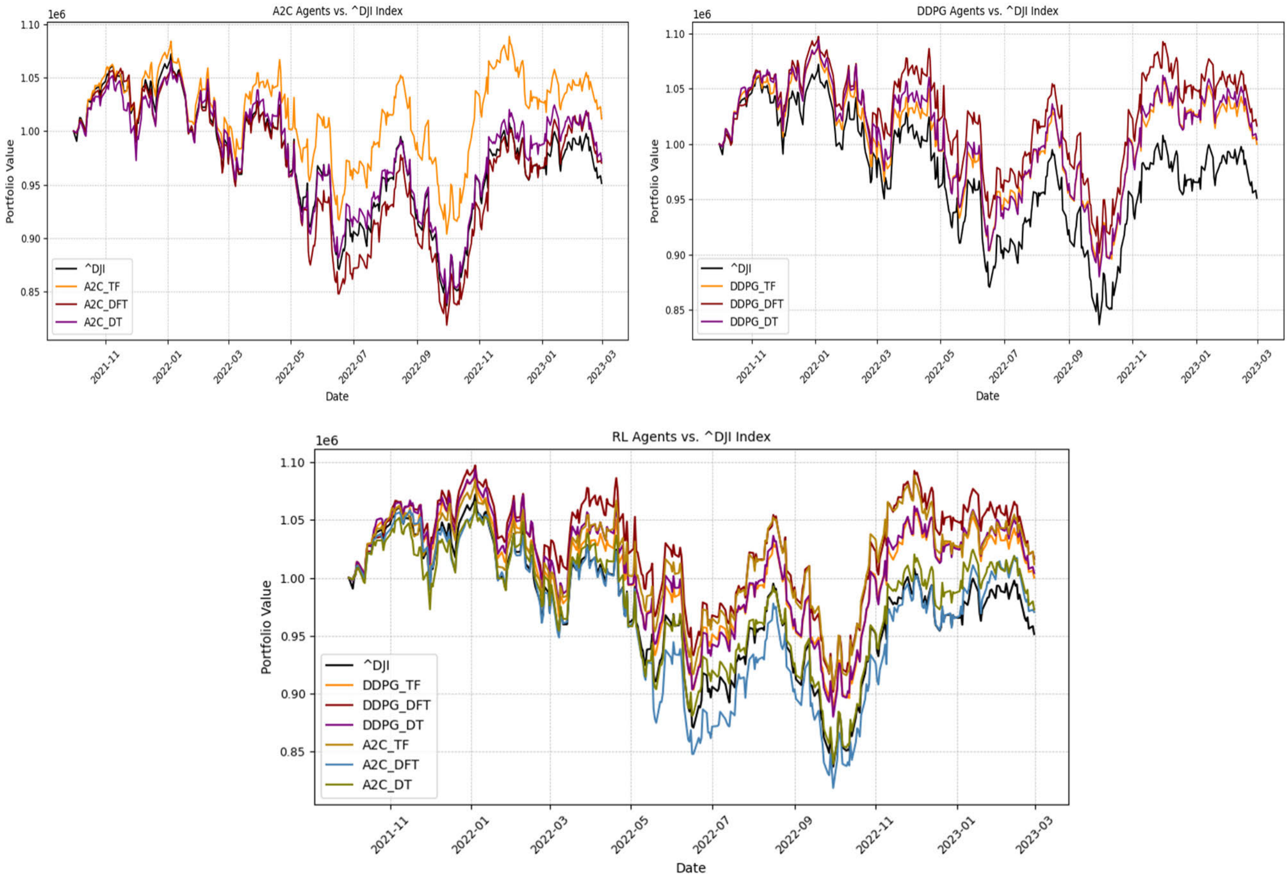
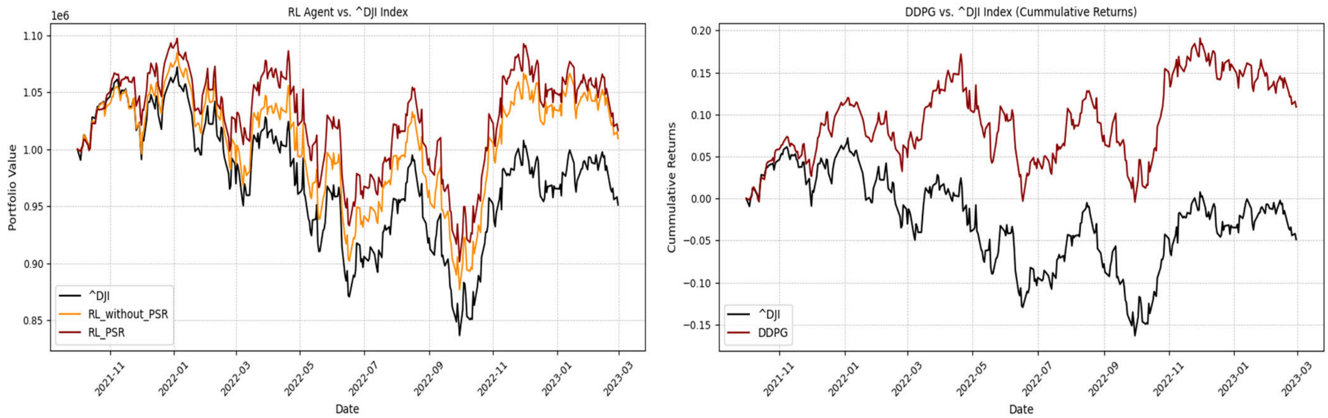


FIGURE 9. **Comparative analysis of RL agent's vs baseline ^DJI index during backtesting.**

−22%, indicating a greater degree of loss for investors tracking the index. But the RL agents, especially, DDPG have less drawdown of about −19.492%, showing that the DDPG approach demonstrates stronger abilities to manage risks or more resistance to market downturns, making it possibly more appealing to investors looking to reduce downside risk. Moreover, the proposed RL agents also perform well in terms of Omega, Sortino, and Calmar ratios. For instance, the Omega ratio of baseline ^DJI Index is about 0.98 while the RL agents have 1.027647 and 1.013462 respectively. This shows the good performance of RL agents in comparison with the baseline ^DJI index over the testing trading period during back testing. Likewise, the portfolio values of DDPG, A2C, and baseline index ^DJI have also been plotted as shown in Figure 6. It is clear from Figure 6 that the performance of the DDPG agent is excellent over the complete trading period, however, the A2C performance is not as much better, particularly at the time frame 2022-07, but performs goods at the end of the trading period i.e. 2023-01 to 2023-03. These results show that RL agents have a strong ability to not only generate trading buying, selling, and holding signals but also have a good ability to generate how many shares should be sold, bought, and held on a particular day.

D. RESULTS OF PROPOSED TRADING AGENTS WITH DT PROTOCOL

In the next experimental setup, we have evaluated the proposed RL agent on DT protocol. In DT protocol or setup, we have taken only daily historical data, and technical indicators into state representation. Similar to the previous setup, we trained the A2C and DDPG models on the Dow Jones Index, which includes 30 companies. The stock trading agents have been trained from January 1, 2010, to October 1, 2021, using historical market data and technical indicators. The major rationale behind this experimental setup is to determine whether fundamental indicators have any impact on improving stock trading strategy or not. Through rigorous testing and evaluation, the study strives to determine the extent how fundamental factors impact portfolio returns, volatility, and risk-adjusted performance. Similar to the DTF protocol, the feedback provided to the model is in terms of PSR which considers different components including portfolio value, Sharpe ratio, and daily returns. By merging such components, PSR provides a comprehensive evaluation of portfolio outcomes, managing risks, and investment plan efficacy. The results in terms of evaluation metrics are computed after training. More precisely, all parameters are kept same such as both models are trained for 50000-time steps, with an initial amount of



(a) (b)

FIGURE 10. Comparative analysis of RL agent's vs baseline ^DJI Index during backtesting in terms of (a) Comparative analysis "with" and "without" PSR (b) Comparative analysis in terms of Cumulative returns.

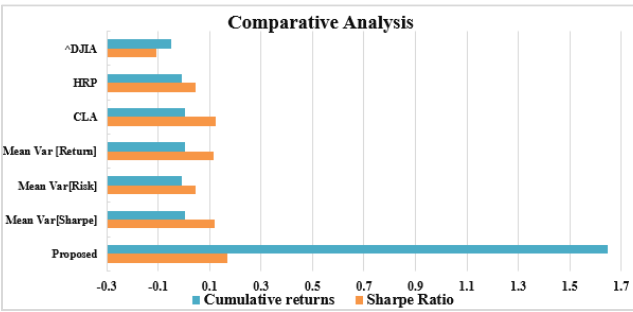


FIGURE 11. Comparative analysis of RL agent's vs baseline methods.

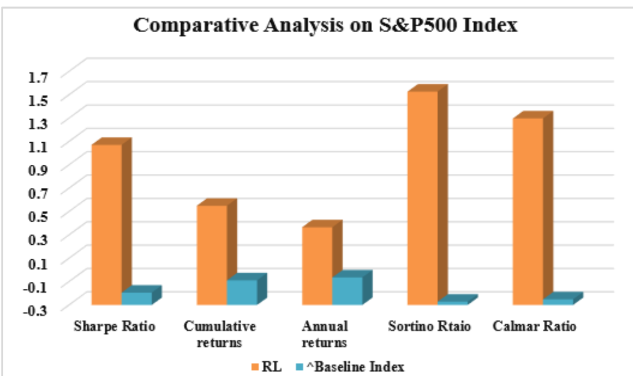


FIGURE 12. Comparative analysis of RL agents on S&P500 index.

10000000, with permissible quantity of shares selling and buying is set to 100. This indicates that at any time **t**, the agent is allowed to buy and sell of shares between 1 to 100. The results of the A2C and DDPG models are shown in Tables 7 and 8. Table 7 shows results with respect to cumulative returns, stability, annual returns, max drawdown, and annual volatility, whereas Table 8 shows performance in terms of Sharpe, Calmar, Sortino, tail, as well as Omega ratios. In addition to that, the comparison with the baseline ^DJI index has also been carried out. The results in Table 8 show that both models perform better than the baseline ^DJI Index particularly, in terms of Sharpe ratio, and Sortino ratio. However, if performance has been contrasted then the results of DDPG are still better than A2C in this experimental setting. The Sharpe ratio of DDPG is 0.07, whereas the baseline index ^DJI has a dismal Sharpe ratio of around −0.10. Figure 7 illustrates the portfolio values of DDPG, A2C, and baseline index ^DJI. More precisely, the first graph in Figure 7 shows that the outcome of the DDPG agent is excellent across the whole trading period; nevertheless, the A2C performance is not as good, especially during the time frame 2022-07, but performs well at the end, from 2023-01 to 2023-03. If the analysis is done between two different experimental protocols, then it is observed that the highest results are obtained with DDPG in the DTF protocol. This is due to the reason that in this case all factors are taken into account including daily stock data, technical indicators, and fundamental indicators. Hence, in the DTF experimental setting, the RL agent is also aware of the fundamental ratios of each stock or company to make trading decisions instead of only considering daily data and technical indicators.

E. RESULTS OF PROPOSED TRADING AGENTS WITH TF PROTOCOL

In the third experimental setting, we tested the suggested RL agents using the TF technique.

In the TF protocol or setup, we solely used technical indicators, close values, and fundamental indicators for state representation. Similar to the previous setups, we trained the A2C and DDPG models on the Dow Jones Index, which included 30 companies. The stock trading agents have been trained from January 1, 2010, to October 1, 2021, using technical indicators, close values, and fundamental indicators. The primary goal of this experimental setting is to examine whether daily stock data has an influence on stock trading techniques. Similar to the DTF and DT protocol, the feedback provided to the model is in terms of PSR which considers different components including portfolio value, Sharpe ratio, and daily returns. By merging such components, PSR provides a comprehensive evaluation of portfolio outcomes, managing risks, and investment plan efficacy. The results in terms of evaluation metrics are computed after training. More precisely, all parameters are also kept the same in this third experimental setup, such as both models are trained for 50000-time steps, with an initial amount of 10000000, with the permissible quantity of shares selling and buying set to 100. The results of the A2C and DDPG models are shown in Tables 9 and 10. Table 9 shows results with respect to cumulative returns, stability, annual returns, max drawdown, and annual volatility, whereas Table 10 shows performance in terms of Sharpe, Calmar, Sortino, tail, as well as Omega ratios. In addition to that, the comparison with the baseline ^DJI index has also been carried out. The results in Table 10 show that both models perform better than the baseline ^DJI Index particularly, in terms of the Sharpe ratio, and Sortino ratio. However, if performance has been contrasted then in this third experimental setup the results of A2C are better than DDPG. The Sharpe ratio of A2C is 0.13, whereas the baseline index DJI has a poor Sharpe ratio of approximately -0.10. This assessment demonstrates the efficacy of the A2C approach to obtaining higher risk-adjusted compared to the baseline index. Moreover, Figure 8 shows that the outcome of the A2C agent is good, especially within the time frame of 2022-03 to 2023-03. It is observed that although A2C performance in this setup is good but still it is lower than DDPG with the DTF protocol. Hence, it is concluded that DDPG agents with DTF data i.e. daily stock data, technical indicators, and fundamental indicators will make good trading decisions. More precisely, if the analysis is done between all experimental protocols, then it is observed that the highest results are obtained with DDPG in the DTF protocol. This is due to the reason that when an RL agent incorporates a variety of data sources, such as fundamental data, technical indicators, and daily stock information, its performance improves dramatically. This enhancement can be given to the complete structure of the data input, which gives the agent a full picture of the market environment and underlying information useful for managing asset values. In boosting the capacities and efficacy of RL-based investing methods, this DTF information emphasizes the significance of data diversity and thorough analysis of the stock market. Following on, agent-by-agent analysis has also been done with all different experimental setups. Figure 9 shows the comparative analysis of the baseline ^DJI Index with all stock trading RL agents. More precisely, first graph in Figure 9 shows the performance of A2C with different protocols, second graph in Figure 9 shows the performance of DDPG agents across different protocols, and third graph in Figure 9 shows the combined results indicating that DDPG performs best with DTF protocol.

Furthermore, the impact of PSR reward is shown in Figure 10 (a), and it becomes apparent that with PSR, the RL agent makes better trading decisions. More specifically, in the absence of PSR, we solely utilized change in portfolio value as a reward function, but we observe that agents make more accurate decisions when risk management and daily returns are encompassed. Likewise, Figure 10 (b) shows the cumulative return of the proposed RL agent vs the ^DJI index which is also high in comparison with the traditional buy-and-hold strategy.

To further validate the performance of the proposed RL, we validate it against non-reinforcement learning methods such as baseline index i.e. Dow Jones Industrial Average (DJIA), Mean-variance optimization with different objectives [56], machine learning-inspired Hierarchical Risk Parity algorithm [57] and Critical Line Algorithm [58]. In Figure 11, Mean Var (Sharpe) indicates that the objective is based on Sharpe, Mean Var (risk) indicates that the objective is set to maximize return for a provided target risk, and Mean Var (return) indicates that the objective is set to minimize risk for a provided target return, CLA stands for Critical Line Algorithm, and HRPA stands for Hierarchical Risk Parity algorithm. Figure 11 shows the Sharpe ratio and cumulative returns of several standard trading techniques as well as the proposed method. More explicitly, Figure 11 shows that the reinforcement learning technique presented in this study produces promising results, with higher performance in terms of both the Sharpe ratio and cumulative returns. In this context, the proposed trading strategy's elevated Sharpe ratio indicates that it not only produced significant returns but also efficiently managed risk-adjusted returns.

F. RESULTS OF PROPOSED TRADING AGENT ON S&P500 INDEX

To further evaluate RL-agent's performance and illustrate its usefulness across other markets, we have performed experimentation on the S&P 500 index, which included 500 stocks. Here, we focused on a selection of 23 stocks from the Energy sector. This experimentation allowed us to evaluate the performance of the proposed RL agent in a specific industry environment, revealing important information about its performance. To eliminate bias, the historical data is obtained within the same time range as that of DOW Jones. More specifically, the daily data is collected from Yahoo Finance API within the time frame of 2010-01-01 to 2023-03-01 where the training data range is from 2010-01-01 to 2021-10-01 while the back testing is done from 2021-10-01 to 2023-03-01. Following on, from Alpha Vantage API, we have collected data in terms of balance sheets, income statements, and cash flow statements for preparing fundamental data for this S&P500 Index. In the experimental phase, the reward function is modified to improve model performance. Particularly, the hyperparameter coefficient in PSR is reduced, which was originally set at 0.9 for the DOW Jones stocks, to 0.1 after extensive tuning of its value. The main reason for changing its value is to balance the importance of daily returns against risk-adjusted returns. Figure 12 shows the outcomes of the proposed RL agent in terms of yearly and cumulative returns, Sharpe ratio, Sortino ratio, and Calmar ratio. The results depicted in Figure 12 indicate that the proposed RL agent performs very well in different stock markets or on stocks belonging to a particular sector i.e. Energy sector, and also outperforming the baseline index.

G. DISCUSSIONS AND IMPLICATIONS FOR PRACTICE

In the finance domain, algorithmic trading is a very hot research topic, since it is of ultimate requirement for investors to assist them with the best trading decisions. These algorithmic trading techniques have several benefits over individual traders, including increased reliability, faster and more accurate execution, and not influenced by emotional bias. Traditional human trading approaches have natural constraints. For example, information cannot be properly priced, and rigid operational frameworks cannot adjust to changing environments. To design optimal trading strategies, although traditional methods have been designed, including mean reversion, and trend-following methods, their performance is limited and often performs well in specific timeframes. To cope with this, RL-based methods are designed to do stock trading with the objective of improving overall annual returns. However, the major challenge in stock trading is the effective representations of states which make agents more knowledgeable about the stock market. Hence, in this research, we have designed a multifaceted approach to stock trading by incorporating daily stock data, technical indicators, and fundamental indicators which are collected from Alpha Vantage API comprising balance sheets, income statements, and cash flow statements for each stock in DOW Jones. The RL agents comprising A2C and DDPG performance has been validated by designing different experimental protocols and it is observed that DDPG agents outperform the ^DJI Index. In addition, to make agents capable of handling increases in portfolio values, promoting daily returns, and handling risk management, PSR reward is designed. Furthermore, daily returns represent a percentage increase or decrease in the portfolio's value across one single trading day, providing RL agents with insights into the portfolio's immediate performance patterns. The inclusion Sharpe ratio in reward works as a significant measure of the portfolio's ability to generate returns in proportion to the risk absorbed.

The implications of proposed RL agents are significant for the finance industry involving practitioners and investors. One of the key benefits of the proposed RL is the absence of bias and emotions, and its decisions are not influenced by external factors that are often present in human traders. Moreover, due to RL paradigm, the proposed model has continuous learning properties and has the capability to evolve depending on the market dynamics. By taking into account multiple dimensions of the stock market, the proposed model is a holistic decision-making framework and allows practitioners to evaluate market dynamics from a variety of viewpoints, resulting in more thorough and well-informed trading decisions. The proposed model's multifaceted strategy for market analysis makes it more resilient to market swings and unpredictability. In today's evolving world, considering both technical and fundamental analysis are modern trading strategies, and modeling such themes into the world of RL agents brings a new line of strategies, consequently getting a competitive advantage in the marketplace. Fundamental data, like balance sheets cash flow statements, and income statements provide knowledge of a company's intrinsic worth and financial health, allowing for the recognition of undervalued or overpriced equities. On the other hand, the use of technical indicators in the proposed RL leads to the capturing of trends and patterns in the historical daily data and makes agents capable of trading signals for possible price changes. Since the underlying agents are designed for effective daily trading, the daily returns are also modeled into the reward to make it informed about the daily returns it gets from trading decisions. Moreover, the stock market is volatile and for this, adaptable methods to the dynamic conditions of the stock market are quite important, and the proposed RL paradigm perfectly suits this requirement. It will do continuous improvement from continuous feedback in different aspects, including portfolio values, risk management through Sharpe, and daily returns. This adaptive risk management strategy assists investors in protecting capital and preserving wealth during unstable market conditions. The proposed method is also quite helpful for investors having multiple shares in different companies instead of only single-stock trading strategies. This will lead to more generic and practical scenarios which are captured in the proposed RL-agents.

Alongside the strength of the proposed method, it is crucial to give some future work initiatives. Currently, the state representation of the RL agents is based on daily stock data, technical indicators, and fundamental ratios, however, one possible future endeavor is to exploit the performance of the proposed method under other factors such as macroeconomic variables, news, and Twitter sentiments and hybrid of different variables. Moreover, the design of different reward functions including e.g. Calmar ratios is also a potential future research direction. Furthermore, we acknowledge that reinforcement learning approaches take longer to execute than standard trading strategies. For example, RL-trading methods might include the time required to train the reinforcement learning model, optimize trading strategies, backtest, and evaluate performance. To address such issues, parallel processing, distributed computing, algorithmic improvements, and model compression are some of the options available for reducing execution time. In some studies, human feedbacks are combined with RL to speed up its training [59], [60]. For example, in stock trading, human traders can evaluate the RL agent's performance on a regular basis and offer feedback on its trading decisions. The agent can then utilize the input to enhance its policy in the following training rounds. Furthermore, in some studies, model level modifications were made, such as applying synchronous deep reinforcement learning to overcome the execution time problem [61]. Likewise, learning process of RL methods has also been adapted to improve learning time e.g. reduce exploration space of RL by supervisory control theory (SCT) [62]. Moreover, although speed or execution time matters a lot, especially in high frequency trading, accuracy is also not a negligible metric. Since, accurate trading signals in stock trading assist investors for risk management, capital preservation, as well as long-term performance. Hence, establishing a balance among execution speed and accuracy is a key research challenge for designing strong and efficient trading strategies.

V. CONCLUSION

Stock market trading serves as one of the study fields that has profited from the current popularity of reinforcement learning (RL) in handling complicated decision-making issues. Considering the huge number of studies conducted, modeling the stock trading issue in an RL context remained a potential research topic for a variety of reasons, such as effective state representation, optimum rewards, optimal RL agents, and MDP formulation. Hence, in this research study, a new multifaceted approach is designed which considers daily stock data, technical indicators, and fundamental indicators computed for each company's income statements, cash-flow statements, and balance sheets. Later on, all data is fused to create rich stock market representations that are used as input to the RL. In addition, to make the model aware of risk management, profitability on daily returns, and portfolio values, a PSR reward function is proposed. The performance of RL agents has been validated on DOW Jones stocks having a sufficient degree of volatility with different experimental setups. It is concluded from the results that proposed RL agents outperform the baseline ^DJI Index during backtesting in terms of Sharpe ratio, cumulative returns and annualized returns. In a nutshell, the adoption of reinforcement learning methods in stock trading has major implications for investors, including improved processes for making choices to the automation of trading strategies.

REFERENCES

[1] M. E. Hajj and J. Hammoud, "Unveiling the influence of artificial intelligence and machine learning on financial markets: A comprehensive analysis of AI applications in trading, risk management, and financial operations," *J. Risk Financial Manage.*, vol. 16, no. 10, p. 434, Oct. 2023.

[2] P. Liu, Y. Zhang, F. Bao, X. Yao, and C. Zhang, "Multi-type data fusion framework based on deep reinforcement learning for algorithmic trading," *Appl. Intell.*, vol. 53, no. 2, pp. 1683–1706, Jan. 2023.

[3] X. Wu, H. Chen, J. Wang, L. Troiano, V. Loia, and H. Fujita, "Adaptive stock trading strategies with deep reinforcement learning methods," *Inf. Sci.*, vol. 538, pp. 142–158, Oct. 2020.

[4] F. G. D. C. Ferreira, A. H. Gandomi, and R. T. N. Cardoso, "Artificial intelligence applied to stock market trading: A review," *IEEE Access*, vol. 9, pp. 30898–30917, 2021.

[5] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: A systematic literature review," *Proc. Comput. Sci.*, vol. 216, pp. 96–102, Jan. 2023.

[6] J. Sen and A. Dutta, "Portfolio optimization for the Indian stock market," in *Encyclopedia of Data Science and Machine Learning*. Hershey, PA, USA: IGI Global, 2022, pp. 1904–1951.

[7] G. Chen, K. A. Kim, J. R. Nofsinger, and O. M. Rui, "Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors," *J. Behav. Decis. Making*, vol. 20, no. 4, pp. 425–451, Oct. 2007.

[8] J. M. Poterba and L. H. Summers, "Mean reversion in stock prices: Evidence and implications," *J. Financial Econ.*, vol. 22, no. 1, pp. 27–59, 1988.

[9] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *J. Finance*, vol. 48, no. 1, p. 65, Mar. 1993.

[10] J.-L. Wang and S.-H. Chan, "Stock market trading rule discovery using pattern recognition and technical analysis," *Expert Syst. Appl.*, vol. 33, no. 2, pp. 304–315, Aug. 2007.

[11] Y. Deng, F. Bao, Y. Kong, Z. Ren, and Q. Dai, "Deep direct reinforcement learning for financial signal representation and trading," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 653–664, Mar. 2017.

[12] D. Jothimani and S. S. Yadav, "Stock trading decisions using ensemble-based forecasting models: A study of the Indian stock market," *J. Banking Financial Technol.*, vol. 3, no. 2, pp. 113–129, Oct. 2019.

[13] D. Bhuriya, G. Kaushal, A. Sharma, and U. Singh, "Stock market predication using a linear regression," in *Proc. Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, vol. 2, Apr. 2017, pp. 510–513.

[14] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3149–3157.

[15] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 1643–1647.

[16] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, p. 383, May 1970.

[17] R. Sawhney, S. Agarwal, A. Wadhwa, T. Derr, and R. R. Shah, "Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach," in *Proc. Conf. Artif. Intell.*, 2021, pp. 497–504.

[18] S. Sun, R. Wang, and B. An, "Reinforcement learning for quantitative trading," *ACM Trans. Intell. Syst. Technol.*, vol. 14, pp. 1–29, Jun. 2023.

[19] A. Alameer, H. Saleh, and K. Alshehri, "Reinforcement learning in quantitative trading: A survey," *Authorea Preprints*, Nov. 1, 2023, doi: 10.36227/techrxiv.19303853.v1.

[20] Z. Zhang, S. Zohren, and R. Stephen, "Deep reinforcement learning for trading," *J. Financial Data Sci.*, Mar. 16, 2020, doi: 10.3905/jfds.2020.1.030.

[21] L. Chen and Q. Gao, "Application of deep reinforcement learning on automated stock trading," in *Proc. IEEE 10th Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Oct. 2019, pp. 29–33.

[22] A. R. Azhikodan, A. G. Bhat, and M. V. Jadhav, "Stock trading bot using deep reinforcement learning," in *Innovations in Computer Science and Engineering: Proceedings of the Fifth ICICSE 2017*. Singapore: Springer, 2019, pp. 41–49.

[23] Y. Liu, D. Mikriukov, O. C. Tjahyadi, G. Li, T. R. Payne, Y. Yue, K. Siddique, and K. L. Man, "Revolutionising financial portfolio management: The non-stationary transformer's fusion of macroeconomic indicators and sentiment analysis in a deep reinforcement learning framework," *Appl. Sci.*, vol. 14, no. 1, p. 274, Dec. 2023.

[24] X. Yu, W. Wu, X. Liao, and Y. Han, "Dynamic stock-decision ensemble strategy based on deep reinforcement learning," *Appl. Intell.*, vol. 53, pp. 2452–2470, 2023.

[25] I. Contreras, J. I. Hidalgo, and L. Núñez-Letamendia, "A GA combining technical and fundamental analysis for trading the stock market," in *Applications of Evolutionary Computation: EvoApplications 2012: EvoCOMNET, EvoCOMPLEX, EvoFIN, EvoGAMES, EvoHOT, EvoIASP, EvoNUM, EvoPAR, EvoRISK, EvoSTIM, and EvoSTOC*. Berlin, Heidelberg: Springer, 2012, pp. 174–183.

[26] S. Rm Musallam, "Exploring the relationship between financial ratios and market stock returns," *Eurasian J. Bus. Econ.*, vol. 11, no. 21, pp. 101–116, May 2018.

[27] T. Théate and D. Ernst, "An application of deep reinforcement learning to algorithmic trading," *Expert Syst. Appl.*, vol. 173, Jul. 2021, Art. no. 114632.

[28] S. Fong and J. Tai, "The application of trend following strategies in stock market trading," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Aug. 2009, pp. 1971–1976.

[29] M. Metghalchi, N. Durmaz, P. Cloninger, and K. Farahbod, "Trading rules and excess returns: Evidence from Turkey," *Int. J. Islamic Middle Eastern Finance Manage.*, vol. 14, no. 4, pp. 713–731, Jul. 2021.

[30] S. Lauguico, J. Alejandrino, D. Macasaet, R. R. Tobias, A. Bandala, and E. Dadios, "A fuzzy logic-based stock market trading algorithm using bollinger bands," in *Proc. IEEE 11th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ., Manage.*, Nov. 2019, pp. 1–6.

[31] Y. Kim, W. Ahn, K. J. Oh, and D. Enke, "An intelligent hybrid trading system for discovering trading rules for the futures market using rough sets and genetic algorithms," *Appl. Soft Comput.*, vol. 55, pp. 127–140, Jun. 2017.

[32] M. Bukhari, S. Yasmin, S. Naz, M. Maqsood, J. Rew, and S. Rho, "Language and vision based person re-identification for surveillance systems using deep learning with LIP layers," *Image Vis. Comput.*, vol. 132, Apr. 2023, Art. no. 104658.

[33] Y. Qiu, Z. Song, and Z. Chen, "Short-term stock trends prediction based on sentiment analysis and machine learning," *Soft Comput.*, vol. 26, no. 5, pp. 2209–2224, Mar. 2022.

[34] M. Alhussein and G. Muhammad, "Voice pathology detection using deep learning on mobile healthcare framework," *IEEE Access*, vol. 6, pp. 41034–41041, 2018.

[35] M. Maqsood, S. Yasmin, I. Mehmood, M. Bukhari, and M. Kim, "An efficient DA-Net architecture for lung nodule segmentation," *Mathematics*, vol. 9, no. 13, p. 1457, Jun. 2021.

[36] S. Banik, N. Sharma, M. Mangla, and S. N. Mohanty, "LSTM based decision support system for swing trading in stock market," *Knowl.-Based Syst.*, vol. 239, Mar. 2022, Art. no. 107994.

[37] A. Shah, M. Gor, M. Sagar, and M. Shah, "A stock market trading framework based on deep learning architectures," *Multimedia Tools Appl.*, vol. 81, no. 10, pp. 14153–14171, Apr. 2022.

[38] M. Thakur and D. Kumar, "A hybrid financial trading support system using multi-category classifiers and random forest," *Appl. Soft Comput.*, vol. 67, pp. 337–349, Jun. 2018.

[39] A. Adegboye, M. Kampouridis, and F. Otero, "Algorithmic trading with directional changes," *Artif. Intell. Rev.*, vol. 56, no. 6, pp. 5619–5644, Jun. 2023.

[40] H. Ao and M. Li, "Exploiting the potential of a directional changes-based trading algorithm in the stock market," *Finance Res. Lett.*, vol. 60, Feb. 2024, Art. no. 104936.

[41] Y. Li, P. Ni, and V. Chang, "Application of deep reinforcement learning in stock trading strategies and stock forecasting," *Computing*, vol. 102, no. 6, pp. 1305–1322, Jun. 2020.

[42] S. Carta, A. Corriga, A. Ferreira, A. S. Podda, and D. R. Recupero, "A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning," *Appl. Intell.*, vol. 51, no. 2, pp. 889–905, Feb. 2021.

[43] D. Kwak, S. Choi, and W. Chang, "Self-attention based deep direct recurrent reinforcement learning with hybrid loss for trading signal generation," *Inf. Sci.*, vol. 623, pp. 592–606, Apr. 2023.

[44] J. Zou, J. Lou, B. Wang, and S. Liu, "A novel deep reinforcement learning based automated stock trading system using cascaded LSTM networks," *Expert Syst. Appl.*, vol. 242, May 2024, Art. no. 122801.

[45] Y.-F. Chen and S.-H. Huang, "Sentiment-influenced trading system based on multimodal deep reinforcement learning," *Appl. Soft Comput.*, vol. 112, Nov. 2021, Art. no. 107788.

[46] A. Nan, A. Perumal, and O. R. Zaiane, "Sentiment and knowledge based algorithmic trading with deep reinforcement learning," in *Proc. Int. Conf. Database Expert Syst. Appl.*, 2022, pp. 167–180.

[47] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[48] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artif. Intell.*, vol. 101, nos. 1–2, pp. 99–134, May 1998.

[49] T. Chong, W.-K. Ng, and V. Liew, "Revisiting the performance of MACD and RSI oscillators," *J. Risk Financial Manage.*, vol. 7, no. 1, pp. 1–12, Feb. 2014.

[50] M. Maitah, P. Prochazka, and M. Cermak, "Commodity channel index: Evaluation of trading rule of agricultural commodities," *Int. J. Econ. Financial Issues*, vol. 6, pp. 176–178, Apr. 2016.

[51] I. Gurrib, "Performance of the average directional index as a market timing tool for the most actively traded USD based currency pairs," *Banks Bank Syst.*, vol. 13, no. 3, pp. 58–70, Aug. 2018.

[52] *Alpha Vantage*. Accessed: Feb. 6, 2024. [Online]. Available: https://www.alphavantage.co/

[53] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," 2015, *arXiv:1506.02438*.

[54] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1928–1937.

[55] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.

[56] A. Ang, "Mean-variance investing," Columbia Bus. School Res. Paper Ser., New York, NY, USA, Tech. Rep. Paper No. 12/49, Aug. 2012, doi: 10.2139/ssrn.2131932.

[57] M. Lopez de Prado, "Building diversified portfolios that outperform out-of-sample," *J. Portfolio Manag.*, May 16, 2023, doi: 10.3905/jpm.2016.42.4.059.

[58] D. Bailey and M. López de Prado, "An open-source implementation of the critical-line algorithm for portfolio optimization," *Algorithms*, vol. 6, no. 1, pp. 169–196, Mar. 2013.

[59] D. Xu, M. Agarwal, E. Gupta, F. Fekri, and R. Sivakumar, "Accelerating reinforcement learning using EEG-based implicit human feedback," *Neurocomputing*, vol. 460, pp. 139–153, Oct. 2021.

[60] Z. Lin, B. Harrison, A. Keech, and M. O. Riedl, "Explore, exploit or listen: Combining human feedback and policy model to speed up deep reinforcement learning in 3D worlds," 2017, *arXiv:1709.03969*.

[61] R. AbdelKawy, W. M. Abdelmoez, and A. Shoukry, "A synchronous deep reinforcement learning model for automated multi-stock trading," *Prog. Artif. Intell.*, vol. 10, no. 1, pp. 83–97, Mar. 2021.

[62] J. Yang, K. Tan, L. Feng, A. M. El-Sherbeeny, and Z. Li, "Reducing the learning time of reinforcement learning for the supervisory control of discrete event systems," *IEEE Access*, vol. 11, pp. 59840–59853, 2023.

YASMEEN ANSARI is currently an Assistant Professor with the Department of Finance, College of Administrative and Financial Sciences, Riyadh, Saudi Arabia. Her teaching experience spans over 19 years. Her research interests include financial literacy, behavioral finance, e-learning, research tools and techniques, and individual well-being.

SAIRA GILLANI received the Ph.D. degree in information sciences from the Corvinus University of Budapest, Hungary. She joined the COMSATS Institute of Information Technology, Islamabad, Pakistan, in 2016. She was an Assistant Professor with Saudi Electronic University, Jeddah, Saudi Arabia. She was a Research Scholar with the Corvinno Technology Transfer Center of Information Technology and Services, Budapest, Hungary. She was a Research Associate with the Center of Research in Networks and Telecom (CoReNet), CUST, Pakistan. She is currently a Professor with the University of Central Punjab, Lahore, Pakistan. Her research interests include data sciences, text mining, data mining, machine learning, vehicular networks, mobile edge computing, and the Internet of Things.

MARYAM BUKHARI is currently pursuing the Ph.D. degree in computer science with COMSATS University Islamabad, Attock, Pakistan. Her research interests include computer vision, deep learning, and machine learning.

BYEONGCHEON LEE received the bachelor's degree from the Department of AI and Big Data, Soonchunhyang University, in 2024. He is currently pursuing the master's degree with the Department of Security Convergence, Chung-Ang University, Seoul, Republic of Korea. His research interests include data analysis, machine learning, intrusion detection, and NLP.

MUAZZAM MAQSOOD received the Ph.D. degree in software engineering with a keen interest in artificial intelligence and deep learning-based systems. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad, Attock Campus, Pakistan. He has published various top-ranked impact factor articles in the areas of image processing, medical imaging, recommender systems, stock exchange prediction, and big data analytics. His main research focus is to use the latest machine learning and deep learning algorithms to develop automated solutions, especially in the field of pattern recognition and data analytics. He is also a reviewer of many impact factor journals and a program committee member of various international conferences.

SEUNGMIN RHO is currently an Associate Professor with the Department of Industrial Security, Chung-Ang University. His current research interests include databases, big data analysis, music retrieval, multimedia systems, machine learning, knowledge management, as well as computational intelligence. He has published 300 papers in refereed journals and conference proceedings in these areas. He has been involved in more than 20 conferences and workshops as various chairs and more than 30 conferences/workshops as a program committee member. He has edited a number of international journal special issues as the Guest Editor, such as *Multimedia Systems*, *Information Fusion*, and *Engineering Applications of Artificial Intelligence*.