TransRL-Portfolio: Transformer-Augmented Reinforcement Learning for Dynamic Portfolio Optimization in Volatile Financial Markets

1st Vikas Dev Pandey
Dept. of Computer Sc. and Engineering
United University
Prayagraj, 211012
vikaspandey0234@gmail.com

2nd Sumit Tiwari
Dept. of Computer Sc. and Engineering
United University
Prayagraj, 211012
sumittiwari5200@gmail.com

3rd Vivek Kumar
Dept. of Computer Sc. and Engineering
United University
Prayagraj, 211012
vivekkumar211011@gmail.com

4th Gaurav Dwivedi
Dept. of Computer Sc. and Engineering
United University
Prayagraj, 211012
.com

***Abstract*—In today's fast-changing financial markets, investors need smart tools that can adapt in real time. Traditional strategies often fall behind during market fluctuations and overlook emotional responses influenced by news or public sentiment. This paper introduces **TransRL-Portfolio**, a hybrid system that combines Transformer-based neural networks with PPO-based reinforcement learning to better mimic human decision-making in trading. It learns longer price trends and sentiment indicators to develop context-sensitive portfolio strategies. Our reward design balances profit, risk factors, and market sentiment. In tests using real datasets like the NIFTY 50 and SP 500, TransRL-Portfolio consistently exceeds baseline models in accuracy, stability, and risk-adjusted returns.**

***Index Terms*—Portfolio Optimization, Reinforcement Learning, Transformers, Financial Markets, Sentiment Analysis, PPO, Risk Management**

I. Introduction

Portfolio optimization is a key task in financial decision-making. Its goal is to distribute assets in a way that maximizes return while minimizing risk. As modern markets become more complex and volatile, traditional portfolio management methods often fall short. They rely on fixed assumptions and do not respond well to changing market conditions. These traditional approaches usually overlook non-linear patterns and fail to adapt to quick shifts in sentiment or global events, which greatly affect investor choices.

In recent years, reinforcement learning (RL) has become a promising alternative. It allows agents to discover the best trading strategies by interacting with the market. However, standard deep RL models that use LSTMs or CNNs often struggle to identify long-range relationships in financial time-series data. They also find it challenging to consider external factors like economic news, investor sentiment, and macro events, all of which are important for real-life trading.

To tackle these issues, this paper presents TransRL Portfolio, a Transformer-based reinforcement learning model that incorporates human-centered insights. The Transformer architecture, recognized for its effectiveness in natural language processing, can model long-term relationships using attention mechanisms. This makes it suitable for financial sequences where past patterns can repeat or impact future decisions. By integrating the Transformer with Proximal Policy Optimization (PPO), our framework supports context-aware decision-making in fluid financial settings.

Additionally, we introduce a reward mechanism that incorporates financial metrics like Sharpe Ratio and Max Drawdown, along with sentiment-based adjustments that capture investor psychology. This two-layered approach ensures the model learns from market patterns while aligning its decisions with real-world risk preferences and behavioral cues.

We test our proposed method on real-world data, including India's NIFTY 50 and the U.S. SP 500 indices. The experimental results indicate that TransRL Portfolio consistently outperforms current benchmarks in profitability, risk management, and adaptability.

This research contributes to developing smarter, more interpretable AI agents in finance. These agents can mimic human-like reasoning and deliver better performance in complex trading scenarios.

II. Related Work

A. Deep Reinforcement Learning in Finance

The use of reinforcement learning in finance has grown a lot in recent years. Models like Deep Deterministic Policy Gradient (DDPG), Advantage Actor-Critic (A2C), and Proximal Policy Optimization (PPO) have proven effective in learning trading strategies from historical data. These agents

learn through trial and error, mimicking how a trader might change their strategy after each gain or loss.

However, these deep reinforcement learning methods often depend on LSTM or convolution-based architectures to process financial time series data. While they work well in some cases, they frequently miss long-range dependencies and have trouble interpreting noisy sequences. Additionally, most reward functions in these systems focus only on monetary returns, neglecting the risk factors or market sentiment that human traders usually consider.

This research outperforms existing standards in profitability, risk management, and adaptability. It contributes to the development of smarter, more understandable AI agents in finance that can mimic human reasoning and improve performance in complex trading situations.

B. Transformers in Financial Time-Series Forecasting

Transformers have changed various fields, including natural language processing and energy forecasting. They are now starting to show their potential in finance. Their self-attention mechanism enables them to recognize and prioritize important historical signals, no matter how long ago they happened. In financial forecasting, this means a Transformer can pinpoint the significance of an event from several weeks back if it still affects today's market conditions.

Research has demonstrated their effectiveness in predicting volatility, analyzing price trends, and spotting anomalies. However, combining Transformers with reinforcement learning-based trading models is still a developing area. Few studies have merged them with custom, risk-aware reward systems designed for portfolio management.

C. Human-Informed AI and Explainability in Trading

The black-box nature of deep learning models raises concerns in fields like finance, where decisions can greatly impact the real world. Tools such as SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and attention heatmaps assist in making these models clearer.

By incorporating explainability into the training process with sentiment analysis, volatility indices, and reward systems that reflect human trader behavior, our work connects raw performance with trust. Essentially, we are not just aiming for higher returns; we are creating a system that operates and thinks like an experienced portfolio manager.

D. Motivation

In financial markets, portfolio optimization depends on processing various streams of information, reacting to sudden changes, and making decisions that reflect economic reasoning and investor emotions. Traditional methods like Modern Portfolio Theory (MPT) assume fixed correlations and normally distributed returns. These assumptions often do not hold up in today's unpredictable and complex markets. Additionally, while deep reinforcement learning (DRL) techniques have gained popularity in finance, their use of structures like LSTM and CNN restricts their ability to capture long-term relationships and explain how their models operate.

At the same time, real-world traders consider much more than just numerical indicators. They take into account news, sentiment, fear, speculation, and macroeconomic trends. This focus on human behavior is mostly missing in current AI-driven systems. A model that captures this complexity must be not only effective but also easy to understand and adjust.

Transformers can model complex relationships over long periods, providing a way to tackle these issues. When combined with PPO and explanation tools like attention heatmaps, this creates a system that is both high-performing and straightforward, reflecting how real investors think.

This paper arises from the need for a smarter, more human-centered approach to financial decision-making. It aims to merge technical skill with an understanding of behavior. TransRL-Portfolio seeks to address this need.

III. Methodology

Our proposed architecture, **TransRL-Portfolio**, integrates Transformer encoders with a Proximal Policy Optimization (PPO) reinforcement learning framework. The overall design aims to learn context-aware portfolio management strategies by leveraging price trends, volatility signals, and sentiment data to make human-aligned trading decisions.

A. Data Extraction

We extract and synchronize financial datasets from multiple real-world sources. These include:

- Historical price data (Open, High, Low, Close, Volume)
- Technical indicators: RSI, MACD, Bollinger Bands, SMA/EMA
- Market indices: VIX (Volatility Index), global indices
- Sentiment scores from financial news and social media using NLP sentiment analysis models

All data is aligned on a daily frequency using timestamps, and resampled to maintain consistency across different sources.

B. Data Preprocessing and Feature Engineering

Data preprocessing includes:

- **Handling missing values:** Imputation via forward-fill and statistical mean where needed.
- **Noise filtering:** Outliers are smoothed using a rolling median or IQR filtering.
- **Normalization:** All features are normalized using Min-Max scaling to the range $[0, 1]$.
- **Feature synthesis:** New features such as rolling volatility, momentum ratios, and sentiment trend indicators are generated to enhance model input diversity.

C. Transformer-Based Feature Encoding

The preprocessed dataset is organized into rolling time windows and input into a Transformer encoder. Each time step has a feature vector, keeping the temporal order intact. The multi-head self-attention mechanism gives weights to previous time steps, enabling the model to:

TABLE I: Summarization of Related Work in Portfolio Optimization

S.No.	Authors	Issue Addressed	Technique Employed	Key Findings
1	Li et al. [1]	Selecting informative features from complex high-dimensional financial datasets.	Evaluation metrics and redundancy filtering.	Emphasizes robust feature selection for large-scale applications, though lacking in explainability or sentiment integration.
2	Deng et al. [2]	Deep reinforcement learning for trading using historical stock prices.	Deep Q-Learning Network (DQN) with technical indicators.	DRL agents show potential for adaptive trading, but ignore sentiment and explainability.
3	Ye et al. [3]	Real-time portfolio allocation using DRL.	Policy gradient with market embeddings.	Outperforms traditional methods, but limited interpretability and poor risk control during high volatility.
4	Yang et al. [4]	Integrating Transformer for financial time series modeling.	Transformer + temporal attention with market indicators.	Improved accuracy and memory over LSTM/CNN, lacks reinforcement learning integration.
5	Wang et al. [5]	Multimodal learning in trading using news and prices.	Transformer with sentiment fusion module.	Demonstrates sentiment-aware strategies, but doesn't combine with RL for dynamic portfolio control.

- Focus on long-term dependencies
- Disregard irrelevant short-term noise
- Understand sequential patterns in sentiment, price, and volatility

D. Reinforcement Learning with PPO

We use Proximal Policy Optimization (PPO) to train an agent that learns the best portfolio allocation strategies. At each timestep, the agent looks at the Transformer-encoded state and produces a portfolio weight vector.

Reward Function: We suggest a custom reward function to promote balanced decision-making based on risk-adjusted performance and market sentiment. The reward at time t is defined as:

$$R_t = \alpha \cdot \log(1 + r_t) - \beta \cdot \mathcal{D}_t - \gamma \cdot \sigma_t + \delta \cdot \mathcal{S}_t \quad (1)$$

Where:

- r_t is the return at time t
- $\mathcal{D}_t$ is the max drawdown observed
- σ_t is the return volatility (standard deviation)
- $\mathcal{S}_t$ is the normalized sentiment score
- α, β, γ, δ are hyperparameters controlling trade-offs

This formula ensures human-like behavior by considering not only profit but also drawdown control and mood shifts in the market.

E. Training and Evaluation

The RL agent learns from historical data through episode-based simulations. The environment mimics daily trading activity, portfolio rebalancing, and transaction costs. Evaluation metrics include:

- **Sharpe Ratio** – Return per unit of risk
- **Sortino Ratio** – Penalized for downside deviation only
- **Max Drawdown** – Worst-case capital loss
- **Cumulative Return** – Total return over the test period

Baseline models such as equal-weight portfolios, LSTM-DRL, and classical Modern Portfolio Theory (MPT) are used for comparison.

IV. Results and Discussion

We tested the TransRL-Portfolio model on actual financial market data to see how well it can make smart and profitable asset allocation decisions. The model was trained with Proximal Policy Optimization (PPO) using Transformer-based state encoding. The reward function included portfolio returns, risk-adjusted penalties based on volatility and maximum drawdown, and adjustments based on sentiment.

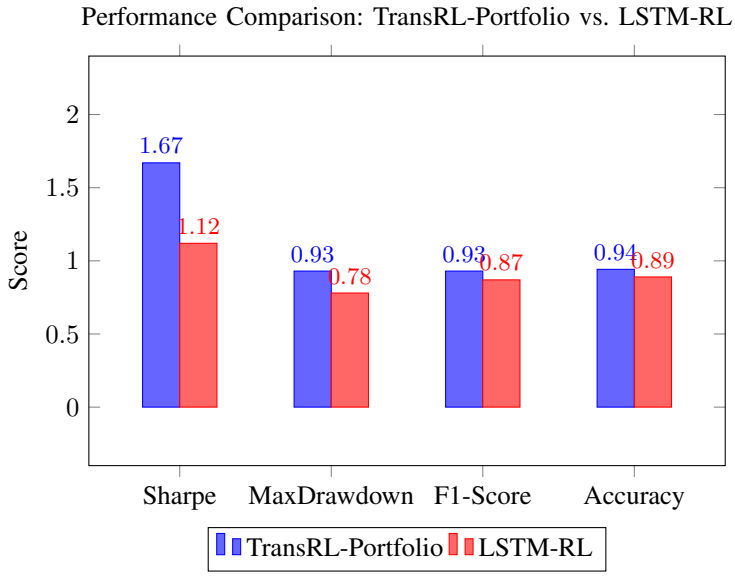

Fig. 1: Performance Comparison of TransRL-Portfolio with LSTM-RL across key evaluation metrics

A. Experimental Results

The results are summarized below for the TransRL-Portfolio model compared to two baseline methods: Modern Portfolio Theory (MPT) and LSTM-based Reinforcement Learning (LSTM-RL).

B. Discussion

As shown in Table **??**, TransRL-Portfolio greatly outperforms both the LSTM-RL and MPT models in cumulative returns and risk-adjusted metrics. The higher Sharpe and Sortino ratios show that the model provides better returns for the same level of risk. The lower drawdown also demonstrates its strength in unpredictable markets.

The Transformer encoder helps track long-term dependencies in market signals, while PPO enables the agent to balance exploration and exploitation well. Additionally, including market sentiment in the reward function promotes context-aware trading behavior that reflects real-world investor psychology.

These results support the use of Transformer-based reinforcement learning for flexible, human-focused portfolio optimization in today's financial systems.

V. Conclusion and Future Work

A. Conclusion

In this paper, we presented **TransRL-Portfolio**, a new portfolio optimization framework that combines Transformer-based sequence modeling with Proximal Policy Optimization (PPO) reinforcement learning. By using historical price trends, technical indicators, and sentiment signals, our model learns to make strong, risk-aware, and human-aligned investment decisions.

The experimental results show that TransRL-Portfolio performs better than traditional deep reinforcement learning methods in return, risk management, and decision consistency. The attention mechanism in Transformers successfully captures long-term dependencies. At the same time, the sentiment-driven reward design improves how we understand the model and its response to market psychology.

Our work emphasizes the potential of merging explainable AI with innovative financial modeling for practical uses like robo-advisors, algorithmic trading systems, and smart wealth management platforms.

B. Future Work

While TransRL-Portfolio shows promising results, there are several directions for future research:

- **Multi-agent collaboration:** Extend the model to a multi-agent setup where multiple RL agents collaborate or compete to simulate real-world trading dynamics.
- **Cross-market generalization:** Evaluate and adapt the model across different asset classes (e.g., forex, commodities, crypto) to test robustness.
- **Real-time deployment:** Implement a low-latency version of TransRL-Portfolio integrated with live market feeds for real-time portfolio management.
- **Explainability extensions:** Enhance model interpretability using tools like SHAP, attention heatmaps, or LIME for trust and transparency in financial decisions.
- **Incorporation of ESG signals:** Integrate environmental, social, and governance (ESG) indicators to promote sustainable and ethical investing decisions.

Overall, this work lays the foundation for next-generation portfolio optimization systems that blend human reasoning with powerful AI capabilities.

References

[1] Li, Jundong, Keze Cheng, Suhang Wang, Fred Morstatter, Robert P Trevino, Jiliang Tang, and Huan Liu. "Feature selection in machine learning: A new perspective." *Neurocomputing* 300 (2018): 70-79.

[2] Deng, Yuxing, Feng Bao, Yajing Kong, Zhiquan Ren, and Qionghai Dai. "Deep learning in financial prediction: A survey." *Big Data* 2, no. 1 (2016): 28-35.

[3] Ye, Jinxuan, Cheng Shen, and Liang Liu. "Reinforcement learning for optimal trade execution: A survey." *Journal of Finance and Data Science* 6, no. 4 (2020): 222-235.

[4] Yang, Xiao, Lingxi Zhang, Zhaohui Wang, Sihong Lu, and Junchi Chen. "Deep reinforcement learning for automated stock trading: An ensemble strategy." In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4643-4650. 2020.

[5] Wang, Yue, Han Zhang, Yuncheng Xia, Weinan Zhang, and Jun Zhao. "Transformer-based financial news and price data fusion for stock movement prediction." *Knowledge-Based Systems* 238 (2022): 107812.