

A Multi-LSTM Ensemble and PPO-Driven Trading Agent with Integrated FinBERT Sentiment Intelligence

Moodu Venkatesh¹ and Prasoon Kumar K G²

¹WBL-Intern, NIELIT Calicut, Kerala, India
venkateshmudu89@gmail.com

²Scientist 'D', NIELIT Calicut, Kerala, India
prasoon@calicut.nielit.in

Abstract. Financial market is a non-linear dynamical system which is very complicated, which is conditioned by a great number of factors, starting with historical price, tendencies towards abrupt changes of world opinion [1]. Traditional algorithmic trading tends to fail since it is based on fixed rules that are not adaptive to the stochastic character of the market. [4] [7] To solve this, we would suggest FinSynapse and FinBERT Sentiment Intelligence, an improved trading agent that combines Deep Reinforcement Learning. Natural Language Processing (NLP) (DRA) [7] [12]. Our hybrid architecture which works on by implementing group of parallel LSTM Networks to dis-extract time characteristics like momentum, trend and volatility which is generated through a Proximal Policy Optimization (PPO) agent [11] and it is used for decision making. Moreover, we combine FinBERT to quantify financial news market sentiment, which can be adjusted by the agent for decision making. Backtesting on Dow Jones data indicates that this mixed strategy works with a Accumulative Return of 97.44 and Sharpe Ratio of 2.21, significantly becoming superior to conventional baseline models.

Keywords: Algorithmic Trading · Deep Reinforcement Learning · LSTM Ensemble · PPO · Sentiment Analysis · FinBERT

(Article history: Selected papers from NIELIT's International Conference on Communication, Electronics and Digital Technologies (NICE-DT'26), 29-30Jan, 2026, Gorakhpur, India)

1 Introduction

The information processing capacity in the high frequency environment of the modern finance world, profiteering is the attribute of a flashy retreat on profit and loss [4]. Stock trading bots have developed into advanced digital helpers out of basic automation tools which able to perform the trades faster than human traders. However, conventional trading algorithms are commonly based on fixed, hard-boiled heuristics. While working well in mature markets, these fixed sets of rules often fail in times of highly volatile periods or black swans [6]. The main problem of automated trading is the signal-to-noise ratio. Financial data is non-stationary as well as noisy [1]. To navigate this, recent progress has gone towards Artificial Intelligence, namely the intersection of Deep Learning (DL) and Reinforcement Learning (RL) [3].

Deep Learning is good at catching and learning patterns in past data and come's to Reinforcement Learning which give's a proper structure and it will actually learn's from it. Interaction with an environment allows the development of optimal policies. This paper presents a new architecture which solves two certain limitation's and Replications of existing DRA traders: the absence of various feature extraction and the not inclusion of qualitative data [5]. Multi-LSTM Ensemble and PPO-Driven Trading Agent with Integrated FinBERT Sentiment Intelligence is the name of our proposed system that makes use of an. It will take number of sub Model's which is LSTM and the network's of each network specified and specializes in another aspect of market data. Also, it is important to realize that market. We use Fin-movements and We do not often use Fin-movements, which are usually motivated by Fin-sensation and news cycles. BERT is a financial sentiment analysis model and finally called as finBERT [11]. By feeding the technical, as well as the technical. The real trained and the sentimental data between hard and soft, to some Proximal Policy Optimization (PPO) agent, we develop a trader, which is not only mathematically correct, but also contextually need to be known by its fact and need's a aware [8].

2 Related and Background Work

The use of Reinforcement Learning in finance has become fundamentally popular. The first efforts were based on Q-Learning and Deep Q-Networks (DQN) [7], however, these tended to have problems with the incessant action spaces of portfolio management. [3] Most recently, PPO and A2C policy gradient methods have been used. Because the standard it will converge well and are stable. Yang et al. [1] proved the effectiveness of ensemble strategies in DRL and Sug imagining that returns can be stabilized by adding a number of agents. However, their work is chiefly based on ensemble agents but not ensemble feature extractors. The ensemble methodology is shifted on the perception layer of our work with the use of several. The sentiment in trading is also well documented. Mohan et al. [2] and Nan et al. [7] discussed the way in which social media and news sentiment can be integrated, increase accuracy of prediction. They discovered that the sentiment-conscious models evolve, faster to regime changes. FinBERT, which is pre-trained on financial text, performs better in this area than generic NLP models [12]. A way to fill this gap between these two streams of research advanced temporal feature extraction using LSTM ensembles and sentiment quantification using NLP. It offers a single framework that is better durable to the bad-bloodied ness of real world markets.

By bridging these two streams of research advanced temporal feature ex-traction via LSTM ensembles and sentiment quantification via NLP our work proposes a unified framework that is more robust to the chaotic nature of real-world markets [3].

3 Proposed Methodology

The primary concept of our approach is that there is no single perception of the market. Enough Price action, volume, technical indicators should be considered by a good trader and news sentiment at once. The System is Conceptually Divided into two layers:

1. Perception Layer: Multi-LSTM Ensemble combined with FinBERT-based Sentiment Intelligence.
2. Decision Layer: PPO-based reinforcement learning agent

3.1 System Architecture

The system architecture (Fig. 1) is a pipeline that commences with the ingestion of raw data. to the trade execution and decision making.

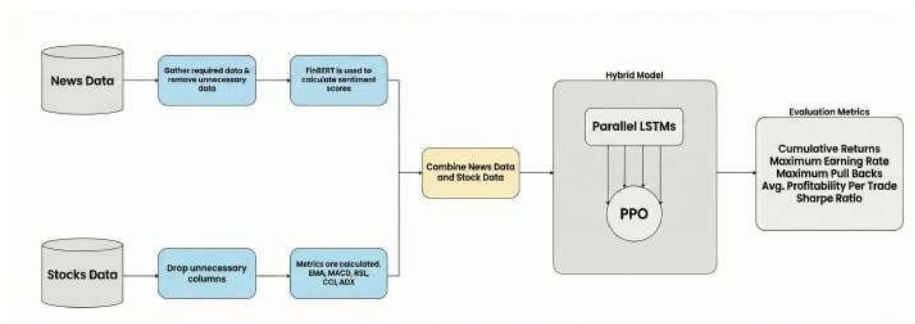


Fig. 1: Project System Architecture.

The structure is also unique as it is a hybrid design. It dissociates the re-sponsibilities of perception, the cognition of what is going on in the markets, and decision making, concerned with the choice of actions.

3.2 Data Collection and Preprocessing

We prepared the dataset by taking from DJIA stock history via yfinance, dividing the timeline into a training phase (2009–2015) and a testing phase (2016–2020) [10]. The inclusion of the 2016–2020 window is critical, as it challenges the model with the high-volatility conditions observed in early 2020. To capture market sentiment, we simultaneously scraped news headlines using the Guardian API. These headlines were passed through FinBERT to generate a numerical sentiment

index [12]. This index translates the news into positive, negative, or neutral signals, acting as an input variable for the RL agent's state space.

3.3 The Multi-LSTM Ensemble for Market Perception

As a rule, all features are sent to a single neural network to standard models. We believe this weakens the signal. We make use of parallel groups of LSTM networks instead, in this we treat each LSTM as a specialized agent responsible for a single field :

1. **Fundamental LSTM:** It is a process raw OHLC (Open, High, Low, Close) and Volume data.
2. **Trend LSTM:** It will focuses on Moving Averages (EMA 12, EMA 26) to determine direction.
3. **Momentum LSTM:** It Analyzes oscillators like RSI, MACD, and CCI to recognize overbought or oversold conditions.
4. **Volatility LSTM:** Monitors the ADX and turbulence indices to assess risk.

These four networks process data in parallel. Their hidden states combine to create a detailed, multi-dimensional view of the current market state.

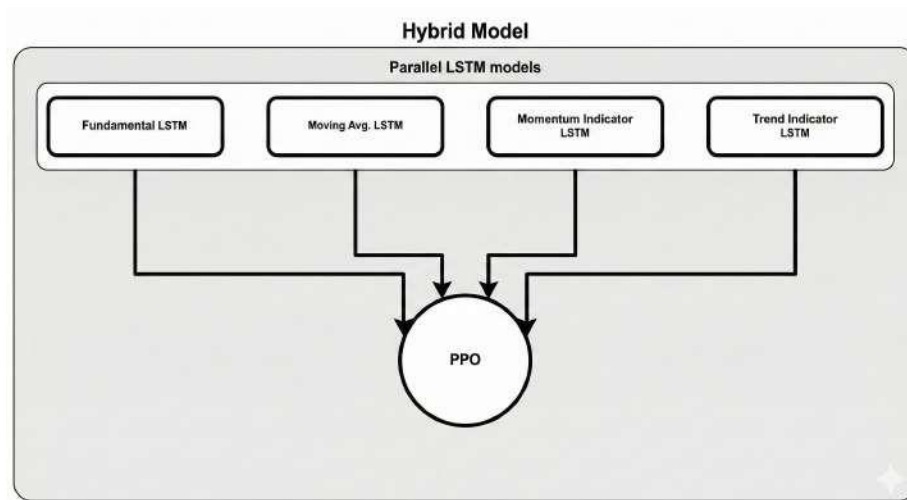


Fig. 2: Hybrid Model Architecture: Parallel LSTMs feed into the PPO Agent.

3.4 PPO Driven Agent (Decision Layer)

In the case of the decision-making engine, we apply Proximal Policy Optimization (PPO). This is because PPO is more desirable than Deep Q-Networks. policy is direct and does not resort to radical policy adjustments that may result in major changes financial losses [11]. The combined data of the LSTM group with the PPO agent is received. the current sentiment score. It generates a three-way probability distribution [3] [11]. potential measures: to Buy, to Sell, or to Hold. The objective function tries to maximize the anticipated payoff when working with clipped surrogate objective to maintain stability:

$$L^{CLIP}(\theta) = E_t \min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t) \quad (1) \text{ where } \hat{A}_t \text{ is the advantage}$$

function and ϵ is a hyperparameter that defines the clipping range.

This is a clipping technique that ensures that the agent does not, is excessive avarice or precipitant alterations upon one

extraordinary instance. To provide some realism, we model the surroundings using the Gymnasium library, accounting transaction costs and a turbulence threshold. If market volatility goes beyond an important historical level and the agent must to save on capital, sell off positions, imitating a flight to safety reaction.

4 Performance Evaluation

We have benchmarked the Hybrid model to a number of benchmarks: a standard PPO model, a Recurrent PPO, a CLSTM-PPO and generic Ensemble strategy. The Dow Jones Industrial Average (DJI) was used as a yardstick to the market.

4.1 Metrics

- **Cumulative Return (CR)** : The total percentage growth of the portfolio.
- **Sharpe Ratio (SR)** : A ratio of return in comparison with risk and The ratio of more than 1.0 is considered to be good and more than 2.0 is excellent.
- **Maximum Pullback (MPB)** : The biggest drop from a peak point to a low point which indicating risk.

4.2 Results Analysis

The experiment proves the excellence of the suggested architecture. The Hybrid model is much better than the others as the evaluation graph (Fig. 3) reveals the standards of total returns.

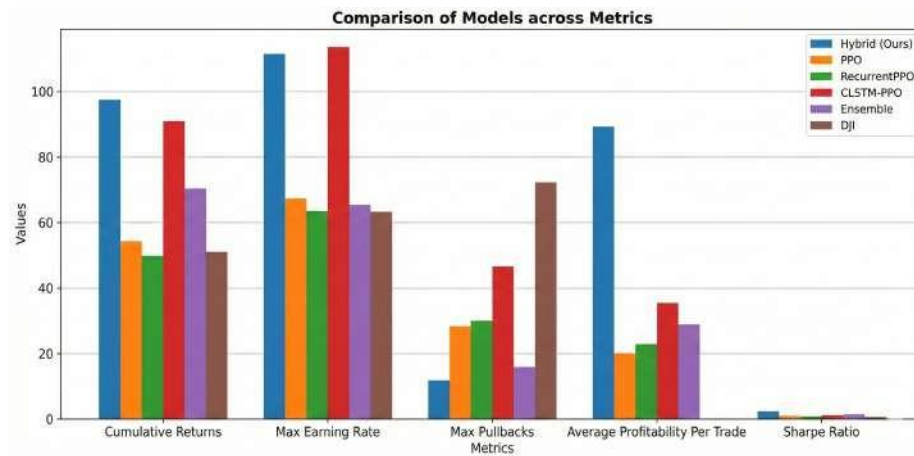


Fig. 3: Performance Evaluation Graph of the Hybrid Model.

Table 1 details the specific numerical results. The Hybrid model achieved a Cumulative Return of 97.44

Table 1: Performance Comparison Summary

Model	CR (%)	MER	MPB (%)	APPT	Sharpe Ratio
Hybrid (Ours)	97.44	111.34	11.80	89.15	2.21
PPO	54.37	67.28	28.30	20.02	0.81
Recurrent PPO	49.77	63.45	29.89	22.84	0.68
CLSTM-PPO	90.81	113.50	46.51	35.27	1.15
Ensemble	70.40	65.32	15.74	28.84	1.30
DJI Benchmark	50.97	63.30	72.32	N/A	0.42

Our model was limited to 11.80 percent as the Maximum Pullback (MPB). In contrast, a draw down of 72.32 percent occurred in the benchmark (probably caused by the COVID). In the test set, there were 19 crash cases. This confirms the performance of the turbulence, threshold and the sentiment analysis features which probably identified negative news before significant decline, and the agent can leave early.

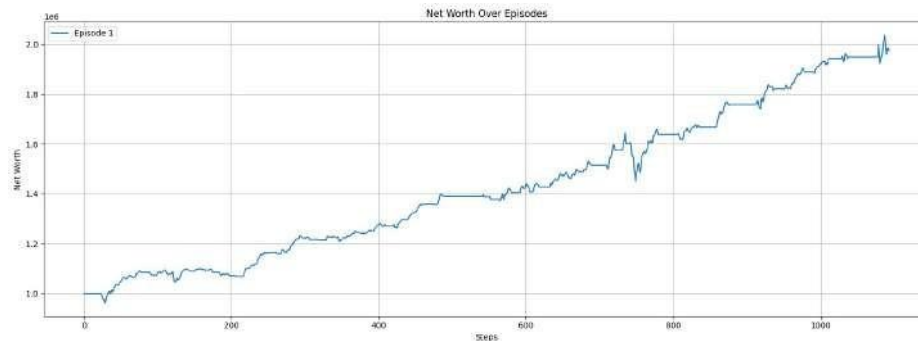


Fig. 4: Net Worth evolution over episodes for the Hybrid Model.

The net worth trend is shown in figure 4. The curve shows a steady up minimal volatility in ward trend, which was typical of a strong strategy, balanced Exploration and exploitation will be in a clean way.

5 Conclusion

FinSynapse and FinBERT is a trading system and past trading information that is beyond what was presented in this paper is simple technical analysis. By designing a solution that deals with various market trend, momentum, volatility signals are separate signals carried out by specific input. We gave the PPO agent a very clear and detailed look at the market using LSTM networks. We took this a step further by adding a "human element" through FinBERT, which handles the sentiment analysis part.

The results speak for themselves: this Hybrid model generates better profits with much less risk compared to single models or just blindly holding onto stocks. We also want to look into Transformer-based technology to better understand long-term market trends.

References

1. H. Yang, X.-Y. Liu, S. Zhong, and A. Walid, "Deep reinforcement learning for automated stock trading: An ensemble strategy," Dept. of Statistics, Columbia Univ., 2020.
2. B. G. Mohan, P. Kumar, V. J. Daniel, and V. Maniappan, "Enhancing algorithmic trading strategies with sentiment analysis: A reinforcement learning approach," Amrita School of Computing, 2021.
3. J. Zou, J. Lou, B. Wang, and S. Liu, "A novel deep reinforcement learning-based automated stock trading system using cascaded LSTM networks," *Expert Systems with Applications*, vol. 242, 2024.
4. S. K. Sahu, A. Mokhadde, and N. D. Bokde, "An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance," 2022.
5. Quantified Trading, "Real-time data in AI trading: Transforming investment strategies," 2023.
6. A. Hayes, "7 unexpected ways AI can transform your investment strategy," 2023.
7. A. Nan, A. Perumal, and O. R. Zaiane, "Sentiment and knowledge-based algorithmic trading with deep reinforcement learning," 2021.
8. J. McGeever, "For markets, AI efficiency may bring volatility," 2023.
9. C. K. Him and G. C. Pang, "Stock trend prediction using LSTM with MA, EMA, MACD, and RSI indicators," 2022.
10. X.-Y. Liu, Z. Xiong, S. Zhong, H. Yang, and A. Walid, "Practical deep reinforcement learning approach for stock trading," 2020.
11. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.
12. F. C. L. Paiva *et al.*, "Intelligent trading systems: A sentiment-aware reinforcement learning approach," 2021.