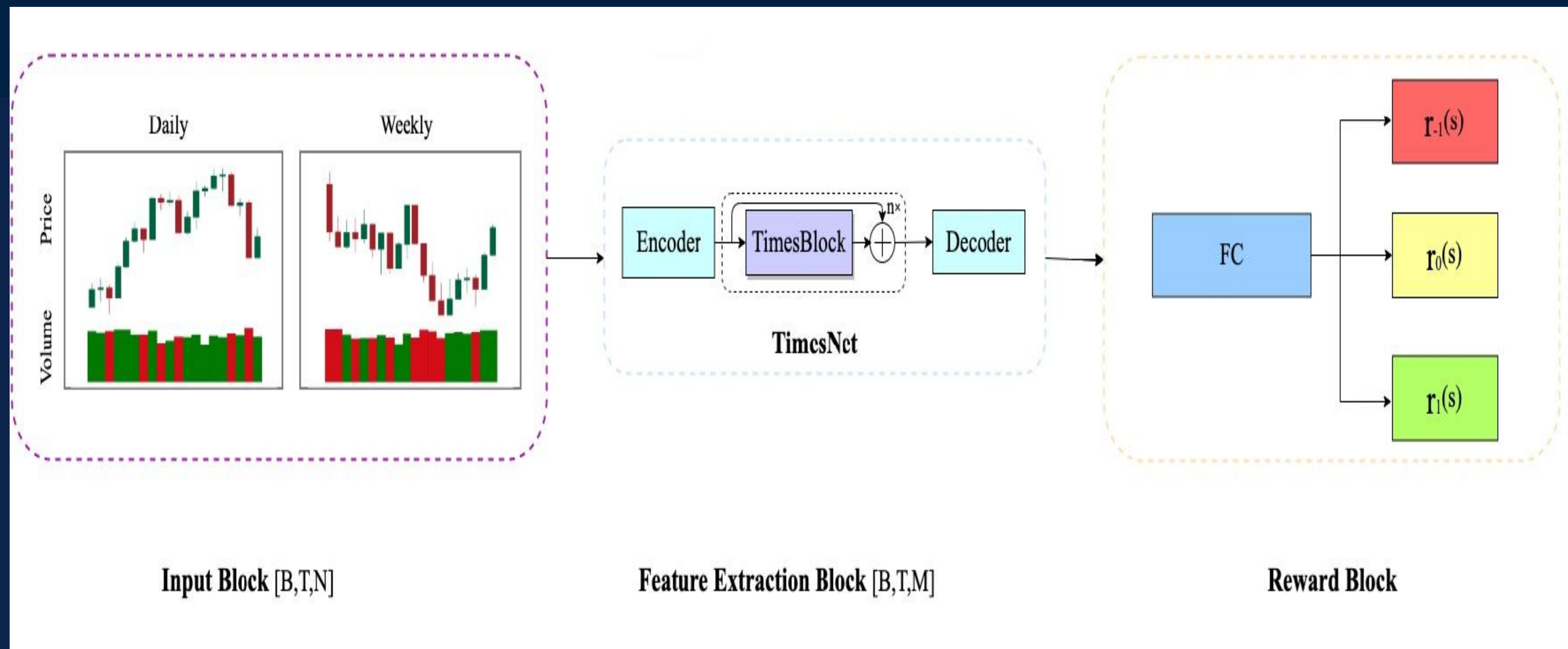
Optimizing Trading Strategy Using Self-Generated Rewards in Deep RL

Tanmay Naik – Sohan Shankar

Under the Mentorship of - Prof. Greg Chism

College of Information Science



Category	Hyperparameter	Value	Description
SRDDQN Agent	State Dim	200 (10 features × 20 steps)	Flattened sequence input dimension
	Action Dim	3	Number of actions: Sell, Hold, Buy
	Gamma	0.9	Discount factor for future reward
	Epsilon	0.9	Exploration rate (ε-greedy policy)
	Learning Rate	0.0001	For DQN optimizer
	Replay Memory Size	1000	Max size of replay buffer
	Batch Size	32	Samples drawn per update
	Episodes	100	Total training episodes

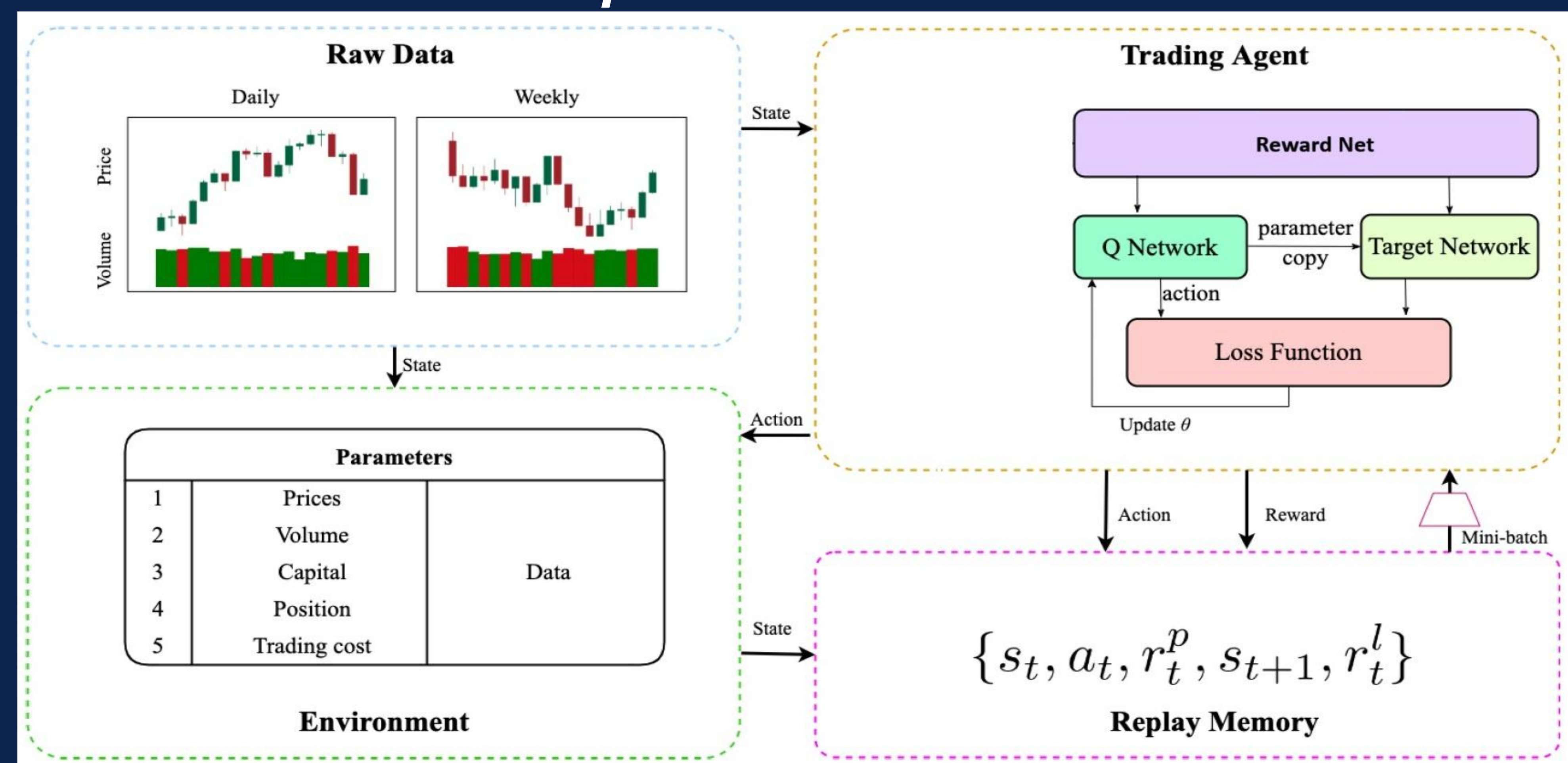
- The Reward Net comprises of three components: Input Block, Feature Extraction Block, and the Reward Block.
- It leverages supervised learning to mimic expert behavior through a self-generated reward function.

Portfolio results show that a $10,000 investment using the proposed algorithm (DJI, Jan 2021–Dec 2023) yields a 43.55% return (~$4,355 ROI).

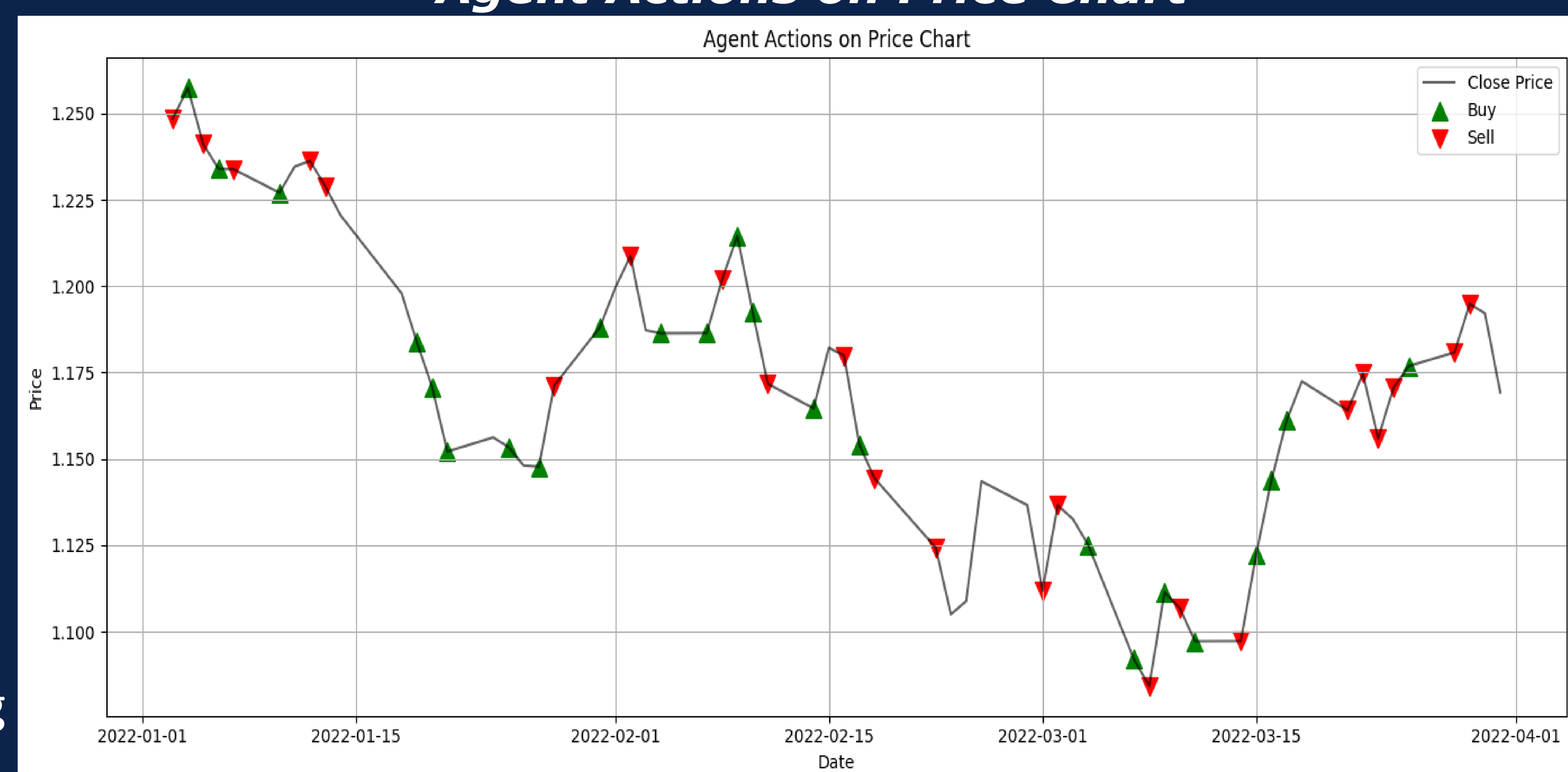
- The proposed algorithm would yield approximately 43.55% return (~$4,355 ROI).
- In contrast, a conventional strategy yields up to 23% (~$2,300 ROI).
- The proposed approach significantly outperforms traditional trading methods.

- ***Developed a Self-Rewarding Deep RL (SRDRL) mechanism with a self-generated reward function to enhance traditional RL.***
- ***Addressed reward design challenges with a scalable, adaptable method for dynamic environments.***
- ***Applied SRDQN to optimize trading strategies using expert metrics like Min-Max, Sharpe Ratio, and Returns.***

Complete Architecture

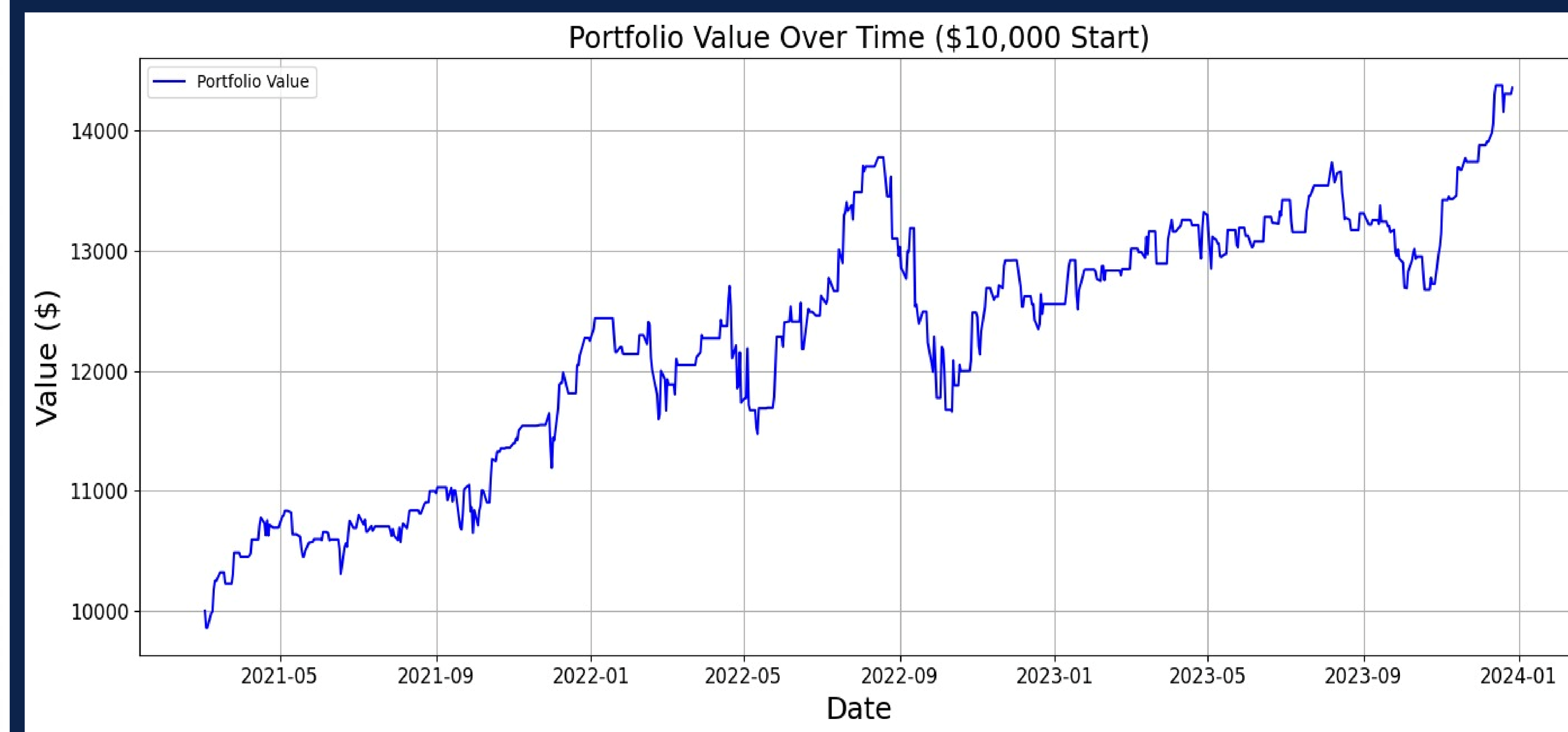


Agent Actions on Price Chart



Strategy	Cumulative Return (%)	Annualized Return (%)	Sharpe Ratio	Max Drawdown (%)
Self-Rewarding DQN (SRDQN)	43.55	13.67	0.98	-15.36
Buy & Hold	4.12	10.55	0.54	-20.23
Sell & Hold	-4.12	-10.38	-0.40	-25.37
Mean Reversion (Moving Averages)	1.54	4.84	0.27	17.95
Trading DQN	-43.04	-22.51	-1.26	45.57

Portfolio Performance



- Future work will refine the integration of self-rewarding mechanisms in more complex and dynamic trading environments.
- Multi-agent learning systems will be explored within the SRDRL framework.
- Large language models (e.g., GPT, FinBERT) will be incorporated for sentiment analysis and market trend prediction
- Sentiment-driven insights from LLMs will support more informed and context-aware trading decisions.