Comparing LSTMs and GRUs With and Without Reinforcement Learning for Stock Price Prediction and Trading Decisions

Alex Yepiz, Industrial Engineering
Mentor: Dr. Michael Clough, Associate Teaching Professor
School of Computing and Augmented Intelligence

QR CODE

Research Question

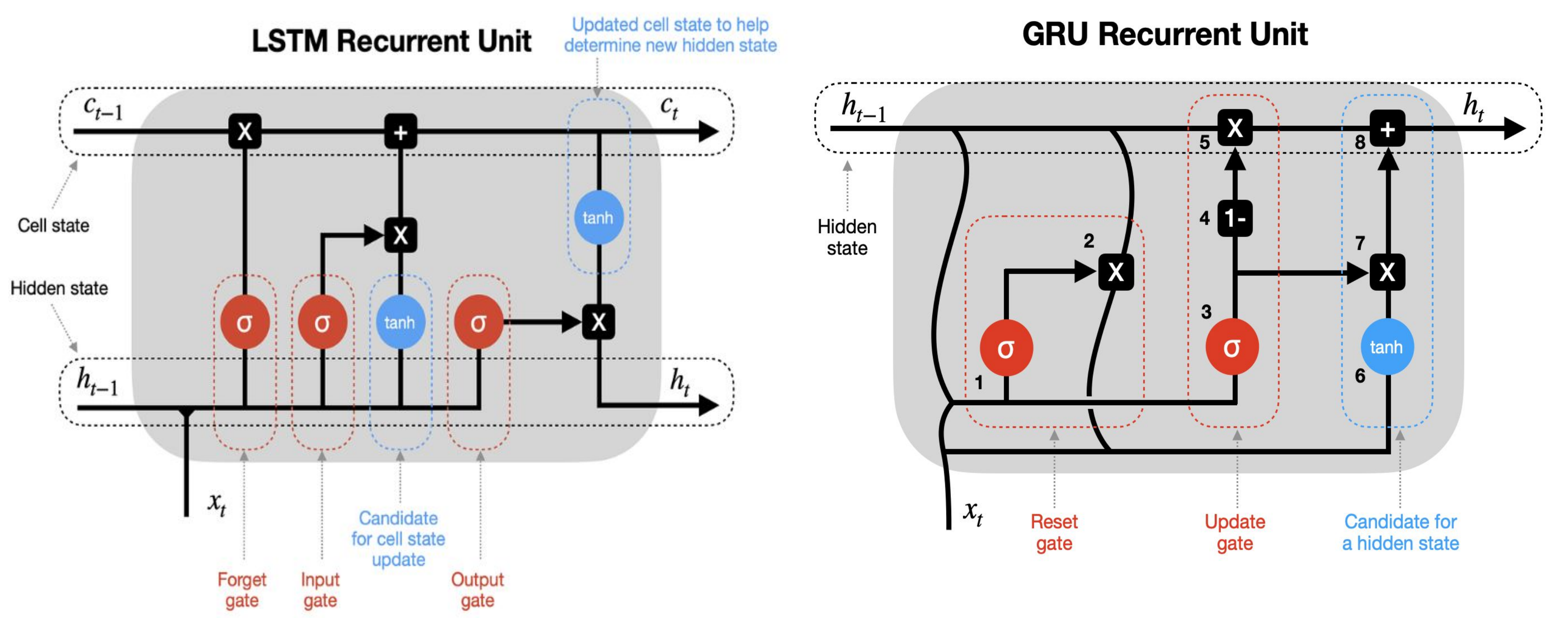
Does adding reinforcement learning (RL) to Long-Short Term Memory (LSTM) and Gated Recurrent Units (GRU) models improve trading performance compared to their standalone versions?

Introduction

Financial markets are notoriously noisy and difficult to predict. Networks such as LSTMs and GRUs are commonly used for time-series forecasting, with RL often assumed to improve decision-making. This study quantifies the effect RL has on these models.

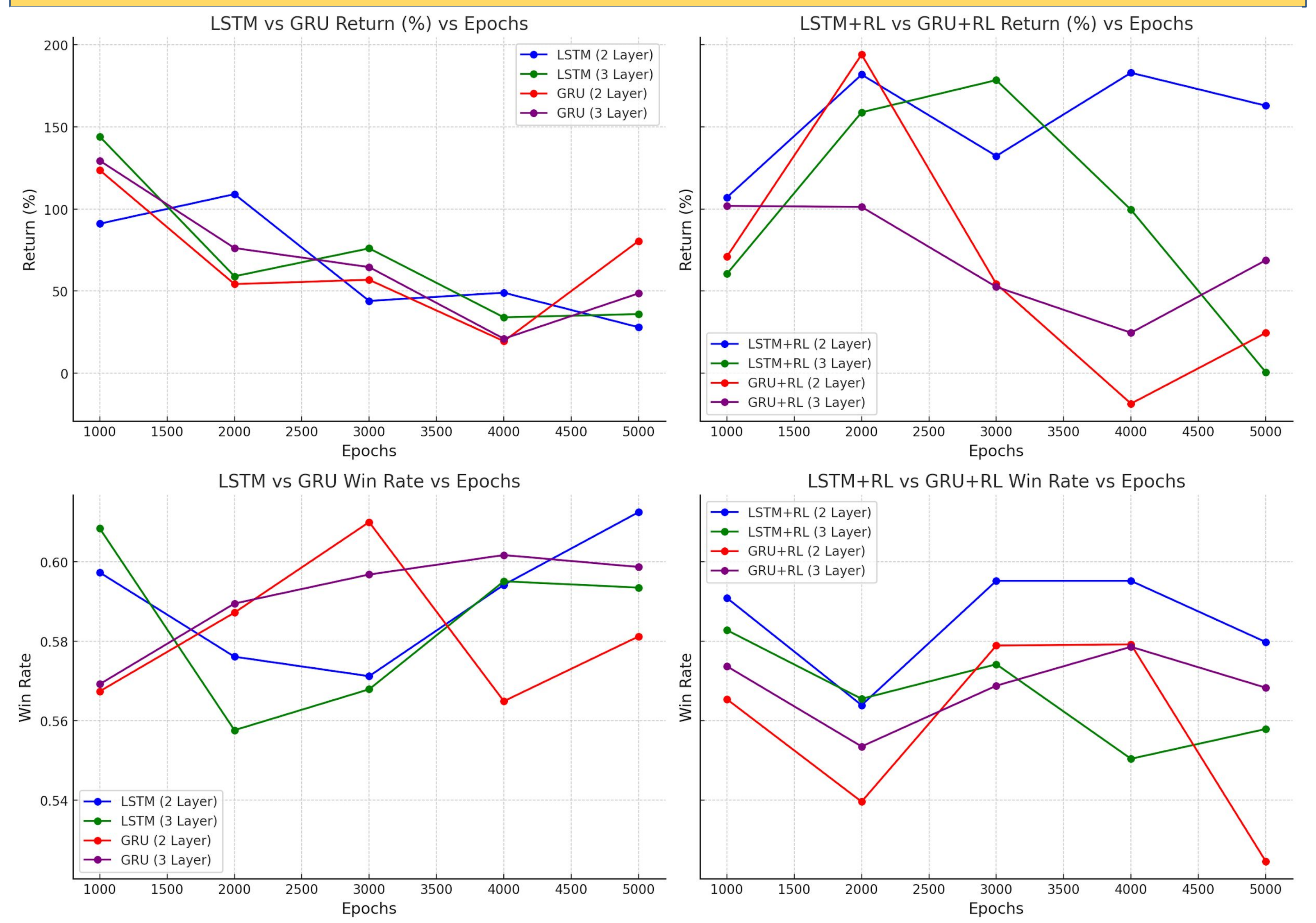
Methodology

This project compares four models: LSTM, LSTM+RL, GRU, and GRU+RL. All models were trained on SPY data from 2000–2015 using a 5-day rolling window of past returns to predict the next day's return. The key architectural difference is that LjanSTMs use both long-term and short-term memory gates, while GRUs simplify this with a combined gate mechanism. A RL policy was introduced only during the trading period, not during training, to make real-time decisions (Buy, Sell, Hold) based on predicted returns. Each model was backtested from 2015–2025 with a simulated starting capital of $100,000 to assess trading performance.



Side-by-Side Comparison of LSTM and GRU Internal Structures

Results



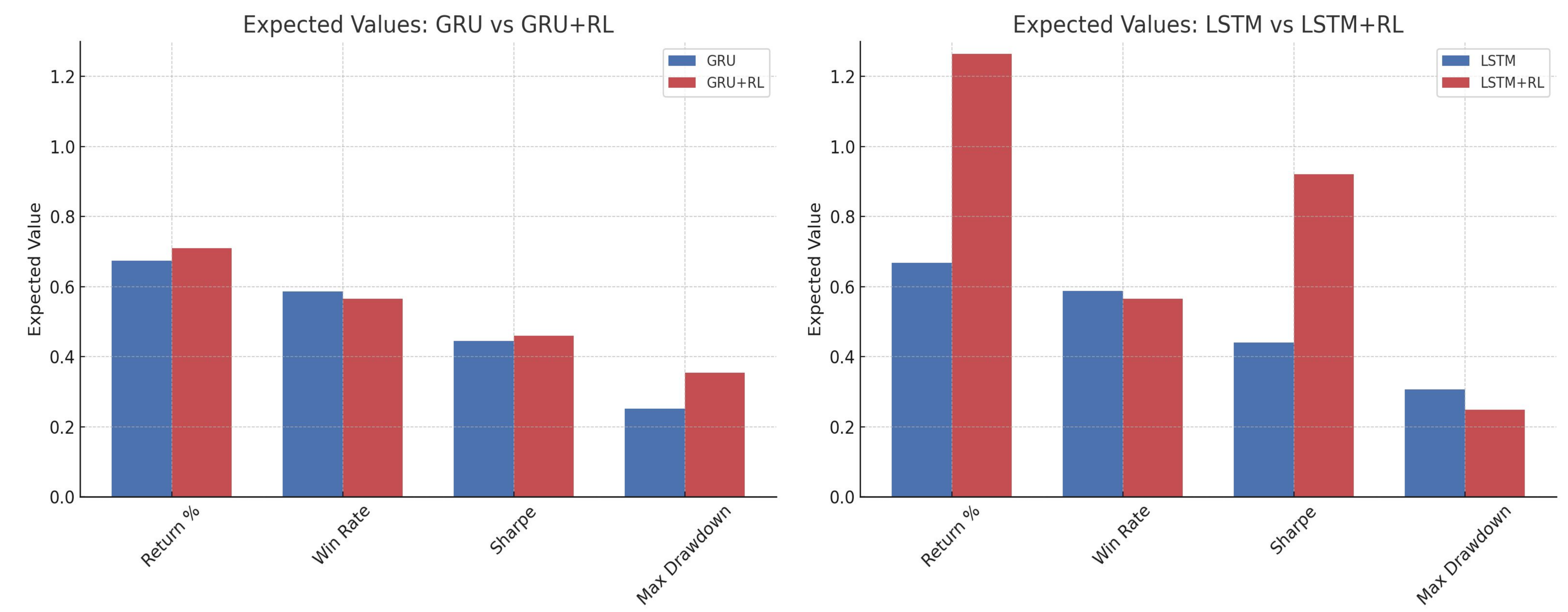
Return (%) and Win Rate vs Epochs for LSTM and GRU Models With and Without RL

Two-sample t-tests compared GRU and LSTM means before and after RL.

Test	P-Value	Significant Difference (α=0.01)?
Return %, GRU	0.999	N
Return %, LSTM	0.016	N
Win Rate, GRU	0.006	Y
Win Rate, LSTM	0.139	N
Sharpe, GRU	0.160	N
Sharpe, LSTM	0.020	N
Max Drawdown, GRU	0.307	N
Max Drawdown, LSTM	0.040	N

P-values from two-sample t-tests Comparing Model Performance Before and After RL

The two-sample t-tests demonstrate that GRU win rate is the only performance metric that has a significant difference before and after the addition of RL. This suggests that RL benefits the GRU unit greater than the LSTM unit, likely due to the GRU's lack of a designated cell state.



Expected Values for GRU and LSTM Models With and Without RL

Although the two-sample t-tests showed no statistically significant difference in LSTM Return % or Sharpe Ratio at the 0.01 level, the expected value bar chart reveals that both metrics nearly doubled after RL was added. This indicates that while RL may improve LSTM performance in practice, the observed increases are not strong enough to reach statistical significance. RL might be more effective when paired with the dual-memory architecture of LSTM, but further study is needed to validate this claim. LSTM+RL achieved the lowest expected drawdown of all models, while GRU+RL experienced a higher drawdown than GRU alone. Although these differences were not statistically significant, they reinforce the idea that LSTM may respond more stably to RL, whereas RL could have introduced volatility into the GRU due to its simpler architecture.

Conclusion

While RL is often assumed to improve trading performance, the results of this study suggest that its impact is architecture-dependent. LSTM+RL achieved the strongest performance overall, yet the lack of statistical significance in its performance highlights the need for further investigation. To gain a deeper understanding of RL's effect on LSTM and GRU models, future research should incorporate additional metrics such as Mean Squared Error (MSE), R^2, and directional accuracy during both training and trading periods.