Results and Discussion

Yun Lin, Jiawei Lou, Jinghe Zhang

July 2025

1 Model Results and Comparison

We implemented five versions of our LSTM-GAT model to investigate the effects of different feature sets and graph structures on portfolio performance.

Feature / Graph	**Model v1**	**Model v2**	**Model v3**	**Model v4**	**Model v5**
Graph Type	Static	Static	Static	Dynamic	Dynamic
1. Close/Volume	✓	✓	✓	✓	✓
2. Log Return	✓	✓	✓	✓	✓
3. Annualized Returns (1W/2W/1M)		✓	✓	✓	✓
4. 5D Rolling Volatility		✓	✓	✓	✓
5. MACD (1W–1M)		✓	✓	✓	✓
6. News Count					✓
7. Average Sentiment					✓
8. Sentiment Variance			✓	✓	✓
9. News Frequency					
10. Weighted Sentiment			✓	✓	✓

Table 1: Feature usage and graph type across model versions. **Note:** Model v5 applies PCA to features 1–6, 7–8, and 10, retaining the top 6 principal components.

Table 2: Performance Comparison (Test Period)

Metric	**Model V1**	**Model V2**	**Model V3**	**Model V4**	**Model V5**	**Equal-Weight**	**CAPM-MVO**
Total Return	28.11%	29.86%	**37.32%**	33.50%	33.04%	24.73%	21.99%
Annualized Return	23.65%	25.10%	**31.23%**	28.10%	27.72%	20.85%	18.58%
Volatility	**26.02%**	26.29%	27.19%	26.60%	28.45%	24.89%	22.03%
Sharpe Ratio	0.91	0.95	**1.15**	1.06	0.98	0.83	0.84
VaR (95%)	-2.62%	-2.57%	**-2.50%**	-2.68%	-2.64%	-2.53%	-2.04%
Max Drawdown	-23.38%	-23.35%	-22.05%	-21.70%	**-20.99%**	-22.60%	-21.59%

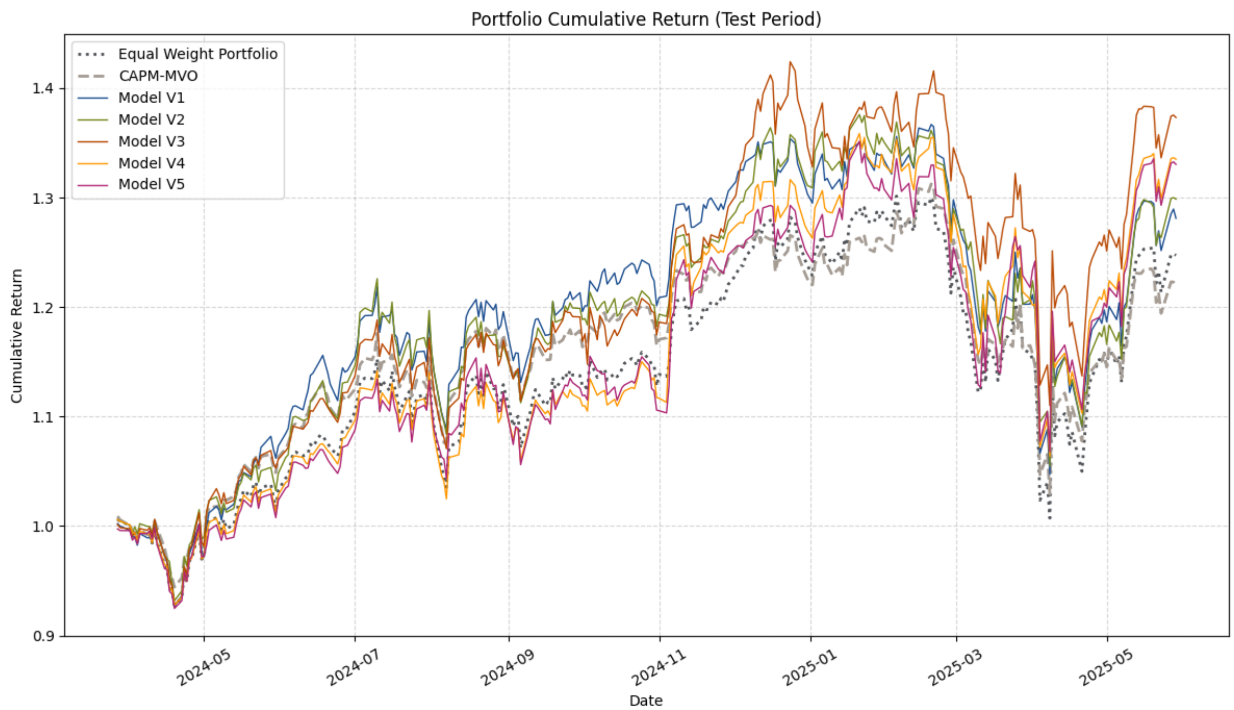


Figure 1: Comparison of Cumulative Return

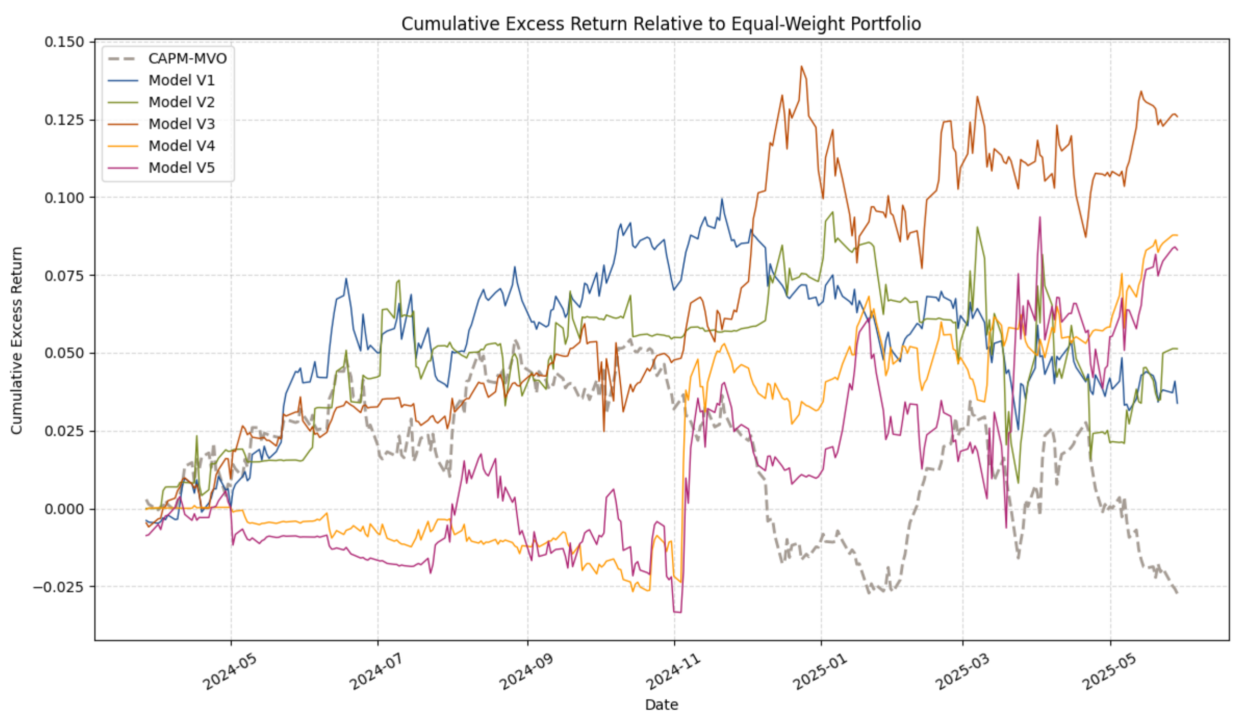


Figure 2: Comparison of Cumulative Excess Return

Model v1 uses only three basic features: close price, volume, and log return. It performs slightly better than the equal-weight benchmark, achieving a 13.43% higher annualized return and a 9.64% higher Sharpe ratio. It also exhibits the **lowest volatility** among all LSTM-GAT models, suggesting stable performance. However, it lags behind the more advanced models in both return and risk-adjusted return, reflecting the limitations of using only basic features.

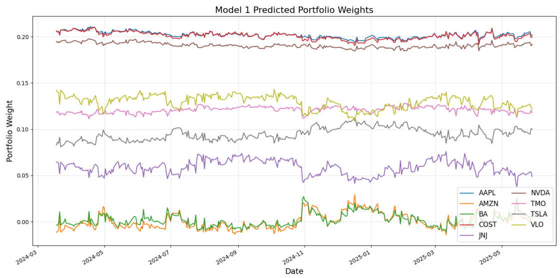
Model v2 extends Model v1 by including additional price-based features such as annualized returns, 1-week rolling volatility, and MACD. This results in improved performance: a 20.38% increase in annualized return and a 14.46% higher Sharpe ratio compared to the equal-weight benchmark. However, the volatility of Model v2 is higher than that of Model v1 and the equal-weight portfolio. This suggests that while more price-based features enrich the model's understanding of market trends, they may also introduce greater variability and noise.

Model v3 adds sentiment-based features—sentiment variance and weighted sentiment—on top of Model v2. This model achieves the **best overall performance**, with a 31.23% annualized return and a Sharpe ratio of 1.15. It also has the lowest Value at Risk (VaR) at -2.50%, indicating strong downside protection. These results highlight the value of integrating **financial news sentiment**, especially during periods of market turbulence, as it provides complementary signals beyond price-based indicators.

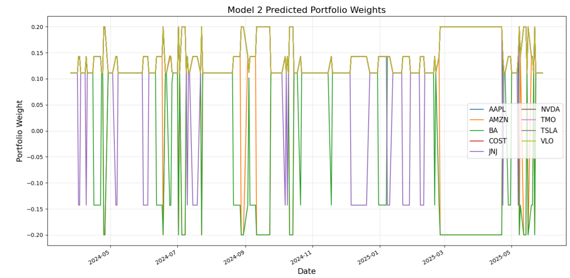
Model v4 introduces a dynamic graph structure that updates weekly based on return correlation, sentiment correlation, and sector membership. Compared to Model v3, it achieves **lower volatility and reduced drawdown**, indicating better risk control. However, its return and Sharpe ratio are slightly lower, possibly due to transient, short-term relationships. Still, it outperforms the equal-weight model with a 34.77% increase in annualized return and a 27.71% higher Sharpe ratio.

Model v5 applies Principal Component Analysis (PCA) to reduce feature dimensionality, retaining the top six components. While it records slightly lower returns than Model v4, it achieves the **smallest maximum drawdown** at -20.99%, especially during the tariff-induced market shock in April 2025. This suggests that PCA helps by filtering out noise and preserving the most informative signals, improving stability under adverse conditions.

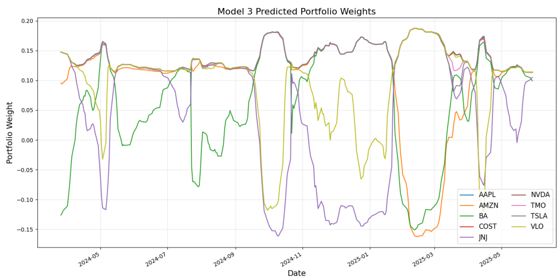
In summary, all LSTM-GAT models incorporating sentiment data and graph-based relationships outperformed the equal-weight and CAPM-MVO baselines in both return and risk metrics. These results underscore the **effectiveness of combining price and sentiment features**, along with graph neural network architectures, for enhanced portfolio optimization.



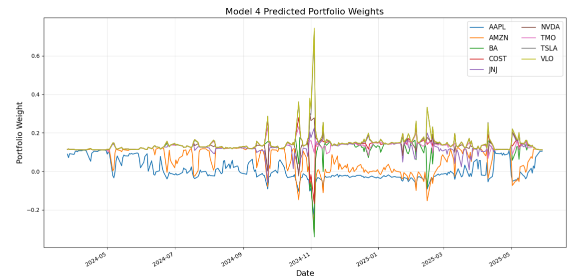
(a) Model v1



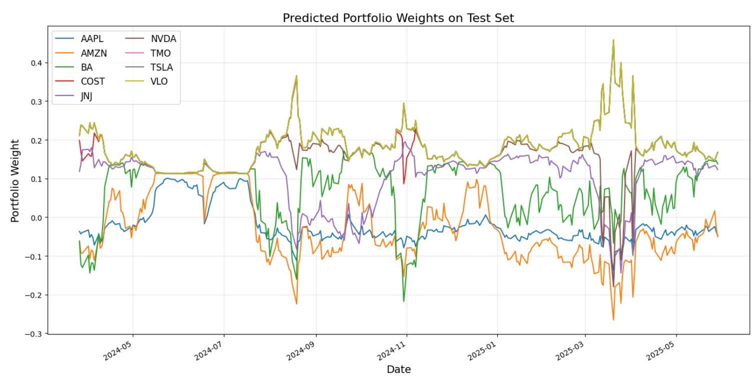
(b) Model v2



(c) Model v3



(d) Model v4



(e) Model v5

Figure 3: Predicted Portfolio Weights for Each Model Version

2 Limitations and Future Improvements

While our model demonstrates strong performance, several **limitations remain** that suggest areas for further development.

First, we limit our universe to **nine fixed stocks** from the S&P 500 and enforce a fully invested portfolio (sum of weights = 1) without leverage. During broad market downturns—such as the April 2025 tariff event—the model lacks the flexibility to allocate to cash or safe-haven assets. It can only redistribute weights among the selected stocks, which may all experience simultaneous declines. A more **diversified asset universe** and flexible allocation constraints could enhance the model's resilience.

Second, our sentiment features are derived directly from third-party API scores, which may oversimplify the underlying news content. This approach limits our control over how sentiment is extracted and interpreted. In future work, we aim to apply **NLP models** to raw news text to develop more robust and context-aware sentiment signals.

Third, the model currently does not incorporate **transaction costs**, slippage, or liquidity constraints, which limits its practical application. Integrating these **real-world trading frictions** would make the model more suitable for deployment in live trading environments.

Lastly, due to computational limitations, we were unable to conduct comprehensive **hyperparameter tuning**. The training process involves hundreds of thousands of parameter combinations, and current computing capacity limits the exploration space. Access to stronger hardware or distributed training infrastructure would enable more thorough optimization and experimentation.