# From Headlines to Holdings:
# AI for Smarter Portfolio Decisions

Literature Review

Yun Lin, Jiawei Lou, Jinghe Zhang

June 2025

## 1 Introduction

The purpose of this literature review is to understand how Graph Neural Networks (GNNs), Graph Attention Networks (GATs), and Long Short-Term Memory (LSTM) networks are applied to stock portfolio optimization. This directly supports the objective of our project: to design an end-to-end deep learning framework that regularly adjusts portfolio allocation by modeling both temporal patterns and cross-asset relationships among stocks.

Traditional optimization methods rely on strong assumptions that often fail to hold in today's fast-paced and interconnected financial markets [2]. To overcome these limitations, recent research has explored deep learning approaches that learn portfolio weights directly from data, bypassing unstable intermediate steps like return forecasting [4, 7]. Moreover, incorporating sentiment data—such as news sentiment—has been shown to enhance stock price prediction accuracy, as demonstrated by Srinivas et al.[6].

This review focuses on three recent papers that apply deep learning techniques to portfolio optimization. In particular, we examine studies that directly predict portfolio weights and use models such as LSTM and GNN. We analyze their model architectures, datasets, and evaluation metrics. Gaining insight into the current research landscape in deep learning for portfolio optimization helps position our work within the broader academic and industry context.

# 2 Key Papers and Reviews

In this section, we discuss three key papers that are highly relevant to our research question and methodology.

## 2.1 Leveraging BiLSTM-GAT for enhanced stock market prediction

Lu et al. aim to enhance the prediction of stock prices by including technical and fundamental relationships between stocks through a dual graph neural network [4]. Their model captures short-term patterns within the stock and cross-stock dependencies. They predicted prices of each stock, then constructed optimized portfolios.

They use a BiLSTM-GAT-AM model that combines Bidirectional LSTM, Graph Attention Networks (GAT), and an attention mechanism to process both temporal and relational patterns in stock data. Their dual-graph approach incorporates technical similarity (with DTW distance) and fundamental relationships (based on industry sectors). Experimental results show that this integrated model significantly outperforms other models—including single-graph and non-attention architectures—on both prediction accuracy and portfolio returns. Their approach achieved the highest Sharpe ratio, the lowest drawdown, and an annual return of over 302%, well above the S&P 500 benchmark and competing models.

We use a similar approach: we use GAT and LSTM to model the stock relationships and temporal trends of markets. Our graph also uses multi-relational graphs to model the relationships between stocks, and we update them weekly to better capture evolving market conditions.

Their research has several strengths. First, they combine technical and fundamental relationships, capturing a comprehensive view of inter-industry connections. Next, their model outperforms single-graph baseline models and yields very high portfolio returns. On the other hand, their fundamental graph is fixed, which ignores changing relationships between industries over time. This model also ignores transaction costs and short-selling limits. Lastly, it is a two-step model: it uses the predicted price and then creates the portfolio, which introduces instability and uncertainty.

## 2.2 Deep Learning for Portfolio Optimization

Zhang et al. propose an end-to-end deep learning approach for portfolio optimization, aimed at directly maximizing the Sharpe ratio without forecasting individual asset returns [7]. Instead of individual stocks, they use ETFs (spanning equities, bonds, commodities, and volatility) to ensure diversification and reduce dimensionality. The authors experiment with several neural network architectures, including Fully Connected Networks (FCN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks, and finally select a single-layer LSTM as their main model. The input features consist of historical prices and returns over a fixed lookback window for each ETF. Historical prices and returns serve as inputs, and portfolio weights are produced via a softmax layer, optimized through gradient ascent on the Sharpe ratio.

Their method is benchmarked against traditional strategies, such as fixed-allocation, mean-variance optimization (MV), Maximum Diversification (MD), and Diversity Weighted Portfolio (DWP). Results over a long test period (2011–2020), including the COVID-19 crisis, show consistent outperformance in terms of Sharpe ratio and resilience to market shocks.

This paper is highly relevant to our project, as we also propose an end-to-end model for directly learning portfolio weights, benchmarked against standard baselines. Their use of LSTM supports our design choice for modeling temporal patterns.

The approach avoids unstable return forecasting, is rigorously benchmarked against traditional baselines, and demonstrates strong robustness during periods of market stress. However, the model uses only a single-layer LSTM, which limits its ability to capture cross-asset relationships, and it relies solely on basic price and return data without incorporating alternative data sources like sentiment or news. As a result, we will expand the model architecture to address these weaknesses.

## 2.3 Large-scale Time-varying Portfolio Optimisation using Graph Attention Networks

Korangi et al. (2022) propose a novel portfolio optimization method utilizing Graph Attention Networks (GATs), tailored specifically for managing large portfolios comprising thousands of mid-cap companies, including risky, illiquid firms[3]. Unlike traditional mean-variance optimization—which assumes normally distributed returns, complete datasets, and static asset sets—their deep learning model directly learns portfolio weights without explicitly forecasting returns.

The methodology constructs graphs where each node represents a company, and edges indicate relationships based on return volatility. Distance correlation measures these relationships, effectively capturing nonlinear dependencies even with incomplete data. The Triangulated Maximally Filtered Graph (TMFG) technique filters weak and noisy connections, preserving only significant relationships. These refined graphs serve as inputs for the GAT model, which assigns importance scores to neighboring companies, converting them into portfolio weights through a custom layer. The model optimizes returns and risk using a Sharpe ratio-based loss function.

By applying this approach to daily stock price data from about 20,000 U.S. mid-cap firms (1990–2021), structured into 3-year rolling windows, Korangi et al. demonstrate their GAT model outperforms equal-weight portfolios, mean-variance optimization, and centrality-based graph methods. Specifically, the GAT model achieves higher Sharpe ratios and lower portfolio turnover, thus reducing transaction costs. Additionally, it maintains balanced investments, avoiding overly central or isolated companies.

Despite these strengths, Korangi et al.'s method has significant limitations. Notably, it lacks explicit temporal modeling components such as those provided by LSTM or BiLSTM architectures, limiting its ability to capture time-series patterns. Furthermore, relying solely on return volatility graphs without incorporating sentiment data might reduce its responsiveness to external market events and short-term volatility. Nevertheless, given the demonstrated superior performance of the GAT model compared to traditional methods, we choose GAT for portfolio optimization in our project.

# 3 Critical Analysis and Synthesis

| Feature | Zhang et al. | Lu et al. | Korangi et al. |
|---|---|---|---|
| Asset/Stock | ETFs (VTI, AGG, DBC, VIX) | S&P500 Future | 20,000 mid-cap stocks |
| Time-series Modeling | LSTM | BiLSTM | N/A |
| Graph Component | N/A | GAT + AM | GAT + TMFG |
| Graph Dynamics | N/A | Static | Rolling |
| Sentiment Data | N/A | N/A | N/A |
| Output | Portfolio Weight | Step 1:Predicted price; Step 2: Optimized weight | Portfolio Weight |
| Evaluation metrics | Sharpe Ratio, Maximum Drawdown, Turnover, Cumulative Return | MSE, MAE, Sharpe Ratio, Annual Return and Maximum Drawndown | Negative Logarithmic Sharpe Ratio |
| Findings | LSTM-based model outperforms traditional methods including MV, MD, DWP | LSTM combined with GAT will give the best result compared to other single layer model | GAT-based models out perform the traditional approach, and the model performance is robust across market conditions. |
| Strength | End-to-end optimization framework, more robust | Dual graph appraoch, multilayer model design | Graph construction reflects non-linear dependencies, end-to-end optimization framwork |
| Weakness | No modeling of cross-asset relationships | Static relationship between industries, two step optimization creates unstability | Lack of explicit temporal modeling, no public sentiment |

Figure 1: This table compares key features, modeling approaches, outputs, evaluation metrics, findings, strengths, and weaknesses of three portfolio optimization methods proposed by Zhang et al., Lu et al., and Korangi et al.

In summary, Lu et al.[4] demonstrated that a multilayer LSTM-GAT-AM model incorporating dual graph relationships delivers the best performance among the models they evaluated. Zhang et al.[7], on the other hand, proposed an end-to-end deep learning approach using LSTM, which not only generated higher returns than the traditional mean-variance method but also proved more stable than the two-step approach used by Lu et al.[4]. Similarly, Korangi et al.[3] showed that GAT-based models outperform traditional portfolio optimization techniques, reinforcing the potential of graph attention mechanisms in capturing complex asset relationships.

Additional literature also offers valuable insights into model design and data usage. For instance, Pacreau et al.[5] recommended using a dynamic graph updated weekly, rather than a static one, to better capture evolving market conditions. Srinivas et al.[6] highlighted the importance of incorporating sentiment data from news headlines, which significantly improved predictive accuracy.

Building on these findings, our project aims to integrate the strengths of these studies into a unified end-to-end framework. Specifically, we propose a multilayer LSTM-GAT model based on a tri-dynamic graph that captures relationships derived from return correlation, sentiment similarity, and sector membership. Additionally, we include sentiment data from stock-related news headlines as input features to enhance the model's ability to directly predict portfolio allocations.

We also observed that most of the reviewed studies use the Sharpe ratio as the primary evaluation metric for portfolio performance. In contrast, we propose using a variance-adjusted return maximization objective. Ekmekçioğlu et al.[1] have argued that the Sharpe ratio and mean-variance objectives are similar in nature as they both account for the balance between returns and volatility. However, we believe that explicitly using a mean-variance objective allows for more direct comparison with the traditional mean-variance portfolio optimization baseline.

# 4 Data Description

## 4.1 Data Overview & Sources

All of our data was collected from publicly accessible APIs. We obtained stock price data—including open, close, volume, and other metrics at 60-minute intervals—from the AlphaVantage API. Timestamped news data, including the title, snippet, relevance score, and sentiment score, was collected from the Marketaux API. Both datasets cover the period from January 2021 to May 2025.

## 4.2 Data Characteristics & Relevance

The news dataset includes a relevance score, sentiment score, title, and snippet for each article linked to our selected stocks. We convert the textual content into dense vectors using the sentence-transformers library, allowing us to extract semantic features for modeling. These sentiment scores and embeddings are used both to update the dynamic graph—via sentiment similarity—and as inputs to the LSTM to capture temporal sentiment trends.

The stock price dataset is used to compute key features such as returns, volatility, and rolling statistics. Rolling return correlations are also calculated to dynamically update the asset relationship graph. These price-based features, along with sentiment data, are fed into the LSTM to model temporal patterns.

All datasets undergo initial quality checks and summary analysis to assess completeness, distribution, and stability. This ensures that both price and news features are clean, informative, and suitable for training the portfolio optimization model.

# 5 Conclusion

This literature review explores recent studies that apply deep learning techniques—specifically LSTM, GAT, and GNN models—to stock portfolio optimization. We closely examine three representative works: Zhang et al.[7], Lu et al.[4], Korangi et al.[3], all of which highlight the effectiveness of deep learning in financial markets. Additional literature emphasizes the use of graph structures to capture inter-asset dependencies, recurrent networks for modeling time-series behavior, the integration of sentiment data, and the shift toward end-to-end optimization frameworks.

Building on these insights, our project integrates key strengths from prior work into a unified framework. We incorporate both price and sentiment data and propose a multilayer model that captures dynamic relationships among assets to directly predict portfolio allocations.

# References


[1] Ömer Ekmekçioğlu and Mustafa Ç. Pınar. "Graph neural networks for deep portfolio optimization". In: *Neural Computing and Applications* 35 (2023), pp. 20663–20674. DOI: 10.1007/s00521-023-08862-w. URL: https://doi.org/10.1007/s00521-023-08862-w.

[2] J. B. Heaton, N. G. Polson, and J. H. Witte. "Deep learning in finance". In: *arXiv preprint arXiv:1602.06561* (2016). URL: https://arxiv.org/abs/1602.06561.

[3] Kamesh Korangi, Christophe Mues, and Cristián Bravo. *Large-scale time-varying portfolio optimisation using graph attention networks*. arXiv preprint arXiv:2407.15532. 2024. URL: https://arxiv.org/abs/2407.15532.

[4] Xiaobin Lu, Josiah Poon, and Matloob Khushi. "Leveraging BiLSTM-GAT for enhanced stock market prediction: a dual-graph approach to portfolio optimization". In: *Applied Intelligence* 55 (2025), p. 601. DOI: 10.1007/s10489-025-06462-w. URL: https://doi.org/10.1007/s10489-025-06462-w.

[5] Grégoire Pacreau, Edmond Lezmi, and Jiali Xu. *Graph neural networks for asset management*. SSRN preprint. 2021. URL: https://ssrn.com/abstract=3976168.

[6] S. Srinivas et al. "Effects of Daily News Sentiment on Stock Price Forecasting". In: *arXiv preprint arXiv:2308.08549* (2023). URL: https://arxiv.org/abs/2308.08549.

[7] Zihao Zhang, Stefan Zohren, and Stephen Roberts. *Deep learning for portfolio optimization*. arXiv preprint arXiv:2005.13665. 2020. URL: https://arxiv.org/abs/2005.13665.