

Final Project Proposal

Research Questions:

The project is to optimize a stock portfolio using deep learning to predict stock movements. The approach will focus on the top 7 and bottom 7 performing stocks within the S&P 500, using the overall S&P 500 index as a baseline for comparison. The goal is first to learn how these methods work and then further use deep learning to predict entrance price, etc.

I want to test this on hourly data; however, there are limitations since this data is not publicly available without payment. Therefore, I will begin testing on daily data first and later implement the approach on hourly data once access is secured.

The goal is twofold:

Buy/Sell Signals – Identify optimal entry and exit points by applying deep learning methods to stock data transformed into images.

Entrance Price Prediction – Use machine learning models to predict precise entry prices, thereby improving portfolio performance relative to the S&P 500 benchmark.

Literature Review:

B, Heaton J, et al. “Deep Learning in Finance.” *ArXiv.org*, 2016, arxiv.org/abs/1602.06561.

Cohen, Naftali, et al. “Trading via Image Classification.” *ArXiv.org*, 2019, arxiv.org/abs/1907.10046. Accessed 15 Sept. 2025.

Feng, Guanhao, et al. “Deep Learning for Predicting Asset Returns.” *ArXiv.org*, 26 Apr. 2018, arxiv.org/abs/1804.09314.

Zhang, Cheng, et al. *Deep Learning Techniques for Financial Time Series Forecasting: A Review of Recent Advancements: 2020-2022*. 20 Apr. 2023, <https://doi.org/10.48550/arxiv.2305.04811>.

Across these studies, there is a clear trajectory: early work established deep learning as a viable alternative to shallow econometric methods; subsequent empirical research showed measurable improvements in predictive accuracy; and most recent reviews emphasize diversification into more sophisticated architectures (e.g., Transformers, GANs, GNNs). Collectively, they illustrate the promise of deep learning in portfolio optimization and forecasting while cautioning about interpretability and computational costs.

Dataset Idea:

Buy/Sell Signals – I will use Yahoo Finance data (yfinance) and possible <https://api.polygon.io>. Yahoo Finance limitation would be daily and poly would be hourly but only for two years.

Federal Reserve data – In Stage 2, I plan to incorporate Federal Reserve data to strengthen the accuracy of price prediction models.

Methodology:

Many of these details may change once I begin the process. For example, I may decide to use an unsupervised method instead of a supervised one, depending on how the data and problem evolve.

Stage 1 – Buy/Sell Signals from Images

- Data: Daily OHLCV for top 7 and bottom 7 S&P 500 stocks (2017–now).
- Transformation: Convert rolling windows of prices into candlestick/line images.
- Labels: Generate rule-based buy signals (Bollinger Bands, RSI, MACD) for supervision.
- Model: Train a CNN to classify “buy” vs “no buy.”
- Output: Probabilities of buy signals → basic trading rules (enter on high-probability signals, exit after fixed horizon).
- Evaluation: Precision/recall of signals, plus strategy returns vs. S&P 500.

Stage 2 – Entry Price Prediction + Portfolio Optimization

- Data/Features: Extend Stage 1 by combining CNN image features with numerical indicators (momentum, volatility, cross-sectional info).
- Targets: Predict next-day/next-week returns or entry prices.
- Models: LightGBM as baseline; deep models like LSTMs/Transformers for sequence data.
- Integration: Convert predictions into expected returns → feed into a mean–variance portfolio optimizer with risk/turnover constraints.
- Output: Optimized weights across 14 stocks.
- Evaluation: Sharpe ratio, drawdown, and stability vs. S&P 500 benchmark.

Stage 3 – Profits (hopefully)**Expectations:**

I am expecting that there will be a difference in signals between the three groups—the top 7 stocks, VOO, and the bottom 7 stocks—that will allow me to rebalance the portfolio and optimize returns.

Challenges:

The first challenge would be the method of labeling buys and sells. As this would lay the foundation for my case study here. I expect to spend significant time reading and testing methods that yield the best outcomes. comes.

I am actually in the stage of testing out different methods.