GROUP WORK PROJECT 01
GROUP NUMBER: **8458**

MScFE 642: Deep Learning for Finance

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Sujan Antony S	India	sonnyantony05@gmail.com	
Regulavalasa Krishna Vamsi	India	krishnavamsi8262@gmail.com	
Phan Van Nguyen	Vietnam	nphanvan10@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).

Team member 1	**Sujan Antony S**
Team member 2	**Regulavalasa Krishna Vamsi**
Team member 3	**Phan Van Nguyen**

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed.
Note: You may be required to provide proof of your outreach to non-contributing members upon request.

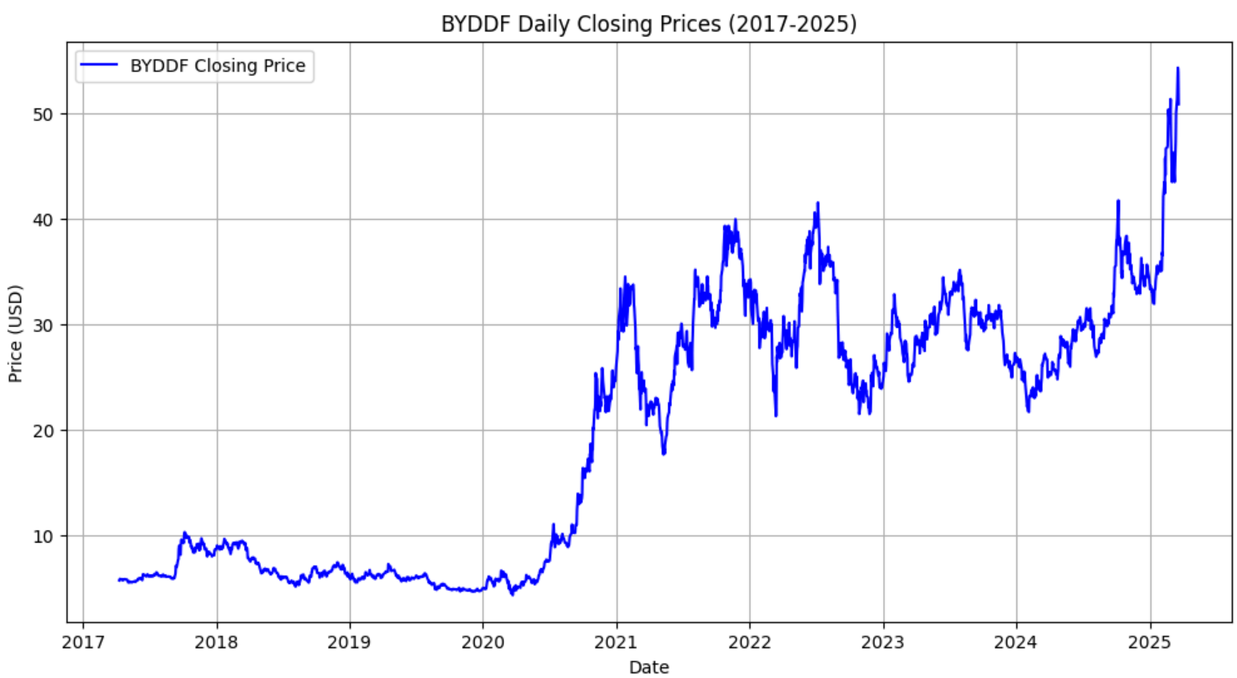
STEP 1:

Step 1: Data Collection and Exploration

a. Time Series Data Selection

Time series prediction is an important activity in the financial market, allowing traders and analysts to make effective investment decisions. In this research, we examine BYDDF's daily closing prices sourced from Yahoo Finance from January 1, 2017, to March 24, 2025.

The data consists of closing adjusted daily prices, which are the ultimate trading price of the stock after accounting for dividends and splits. To follow project specifications, we limit the data to 2,000 observations.



Descriptive Statistics of Closing Prices:

Ticker	BYDDF
count	1999.000000
mean	19.633725
std	12.248673

min	4.257679
25%	6.258751
50%	23.193974
75%	30.184414
max	54.349998

Daily Return Mean: 0.0016

Daily Return Std: 0.0312

Price Trends:

The plot shows periods of quick growth (e.g., EV adoption spikes) and volatility (e.g., market adjustments).

Statistics:

The average price shows a growth trend, with high standard deviation showing volatility. Daily returns have a small positive average and high variability, characteristic of a growth stock.

1B. Introducing Leakage

We compute returns and scale them on the entire dataset prior to splitting, introducing leakage by allowing future data to impact training scaling.

1C. For this part we will create 3 models and forecast the time series, with a single train/test split.

The three models are:

- an MLP
- an LSTM
- a CNN based on GAF

1c.

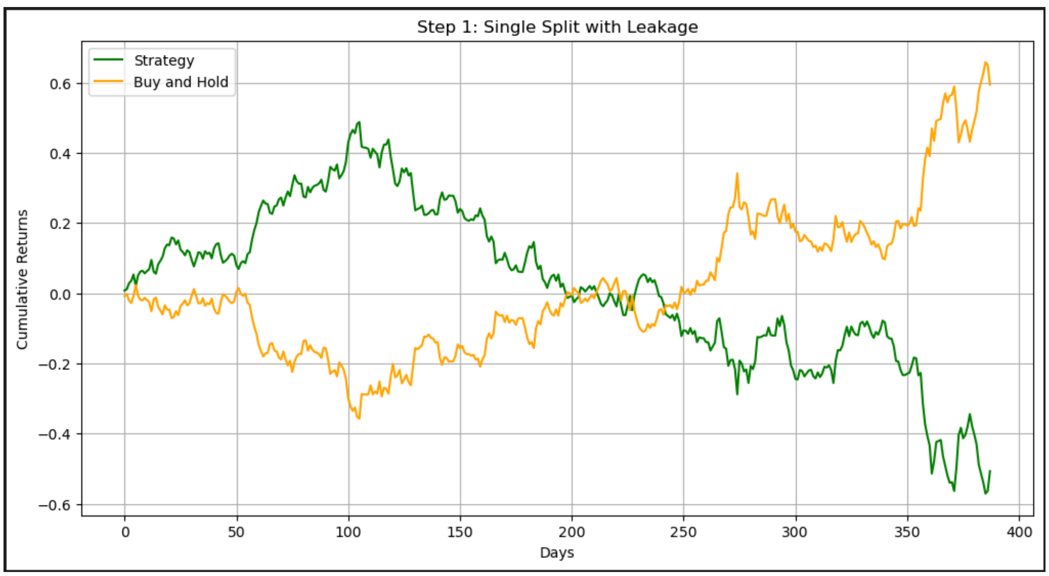
MLP Model

Model Architecture: The model consists of two hidden layers (64 and 32 neurons, respectively), utilizing the ReLU activation function for non-linearity. To combat overfitting, 20% dropout layers are added after each hidden layer. The final output layer consists of a single neuron with no activation function, designed to predict the next day's closing price.

Compilation & Training: The model is compiled using the Adam optimizer and Mean Squared Error (MSE) loss function, then trained for 100 epochs with a batch size of 16.

Prediction: The trained model predicts values on the test set, and the predictions are inverse-transformed to the original scale.

Backtesting



Step 1 - Cumulative Strategy Return: -0.5070

Step 1 - Buy and Hold Return: 0.5942

Step 1 - Sharpe Ratio: -0.8419

The Strategy initially performed well, netting a cumulative return of 40% against the Buy and Hold's -35%. However, as time went on, the Strategy started becoming worse, until it eventually performed worse than the alternative, ending at around -50% for the time period.

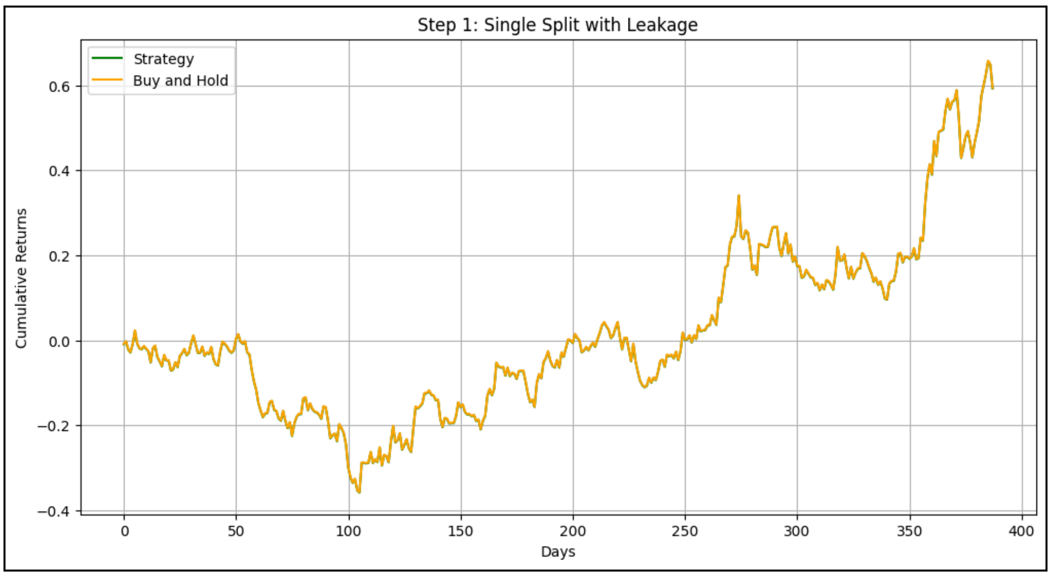
LSTM Model

Model Architecture: A sequential model with two LSTM layers (64 and 32 units), each followed by dropout layers (20% dropout) to prevent overfitting. The final output layer is a dense layer with a single neuron for regression.

Compilation & Training: The model is compiled using the Adam optimizer and Mean Squared Error (MSE) loss function, then trained for 20 epochs with a batch size of 32.

Prediction: The trained model predicts values on the test set, and the predictions are inverse-transformed to the original scale.

Backtesting



Step 1 - Cumulative Strategy Return: 0.5932

Step 1 - Buy and Hold Return: 0.5942

Step 1 - Sharpe Ratio: 0.9870

The "Strategy" and "Buy and Hold" curves overlap, indicating that the strategy's returns are nearly identical to those of the buy-and-hold approach over the time period shown. This could indicate that the strategy isn't doing anything significantly different than simply holding the asset—possibly due to similar market conditions or because the strategy's predictions don't add unique value.It tells us that the strategy and buy-and-hold have nearly identical cumulative returns.

CNN based on GAF

Gramian Angular Fields (GAF) Explanation

Gramian Angular Fields (GAF) is a method of representing time series information as images by casting time-dependent data into a 2D matrix. This makes it easier for deep learning models, like CNNs, to learn meaningful spatial patterns.

GAF transformation includes:

Normalization: Scaling the time series between -1 and 1.
Encoding Angular Information: Time series values to angles.
Construction of the Gramian Matrix: Forming a summation-based form of the angular field.

CNN Model Architecture

Input Transformation: Returns are transformed into Gramian Angular Fields (GAF) in order to visualize time series as images.

CNN Layers:

Two convolutional layers consisting of 32 and 64 filters extract spatial features.
MaxPooling layers reduce dimensionality and avoid overfitting.
Dropout (20%) for regularization.

Output Layer: One dense neuron makes predictions about future returns.
Training and Test Split (Single Train/Test)
Training Set: 80% of the dataset
Test Set: 20% of the dataset
Epochs: 20
Batch Size: 32

CNN GAF - Cumulative Strategy Return: -0.2620

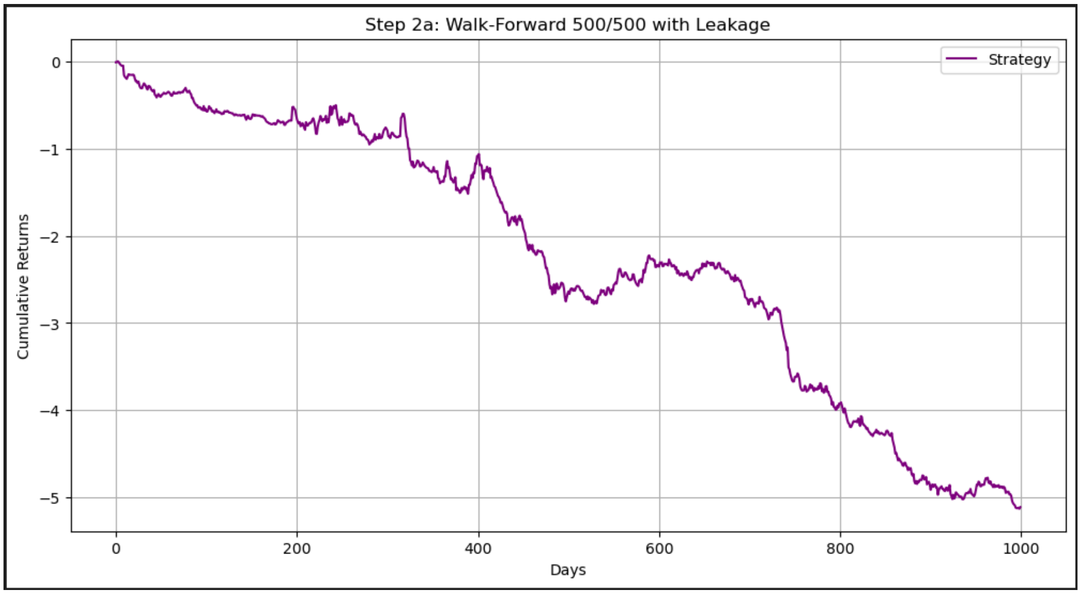
CNN GAF - Buy and Hold Return: 0.5942

CNN GAF - Sharpe Ratio: -0.4363

Step 2: Walk-Forward Backtesting with Leakage

- **MLP model**

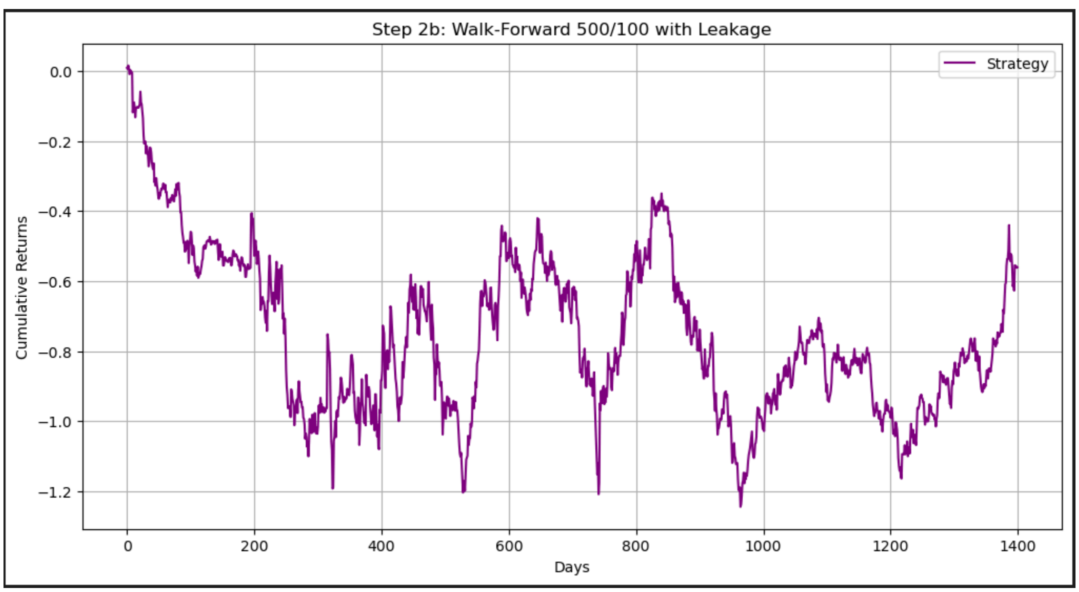
a. Walk-Forward Setup (500/500)



Step 2a - Cumulative Return: -5.1105

Step 2a - Sharpe Ratio: -2.2336

b. Walk-Forward Setup (500/100)



Step 2b - Cumulative Return: -0.5607

Step 2b - Sharpe Ratio: -0.1916

c and d. Comparison of between Steps 2a and 2b and Step 1

The walk-forward approaches in Step 2a resulted in the worst performance out of the 3 options, with a cumulative loss of -500% and a Sharpe ratio of -2.2336. This is likely the result of little training, given that the time step of 500 only allowed for 2 fits.

Meanwhile, the Sharpe Ratio (-0.1916) in Step 2b is the best of the 3. However, its Cumulative Return (-0.5607) is still lower than that of Step 1. It can be inferred that there were steps towards better risk-adjusted returns, however, the overall returns for the period were still significantly worse than the output for Step 1.

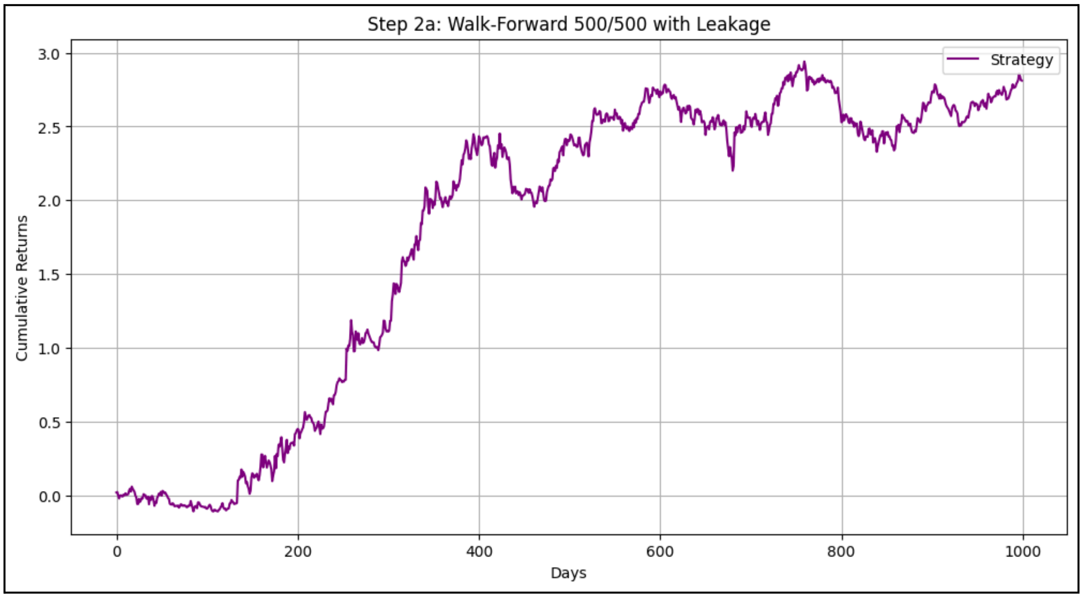
It can be said that the Walk Forward models failed to exploit the Leakage present in the data, leading to worse performance overall.

- **LTSM model**

a. Walk-Forward Setup (500/500)

We use a non-anchored walk-forward method, testing two splits:

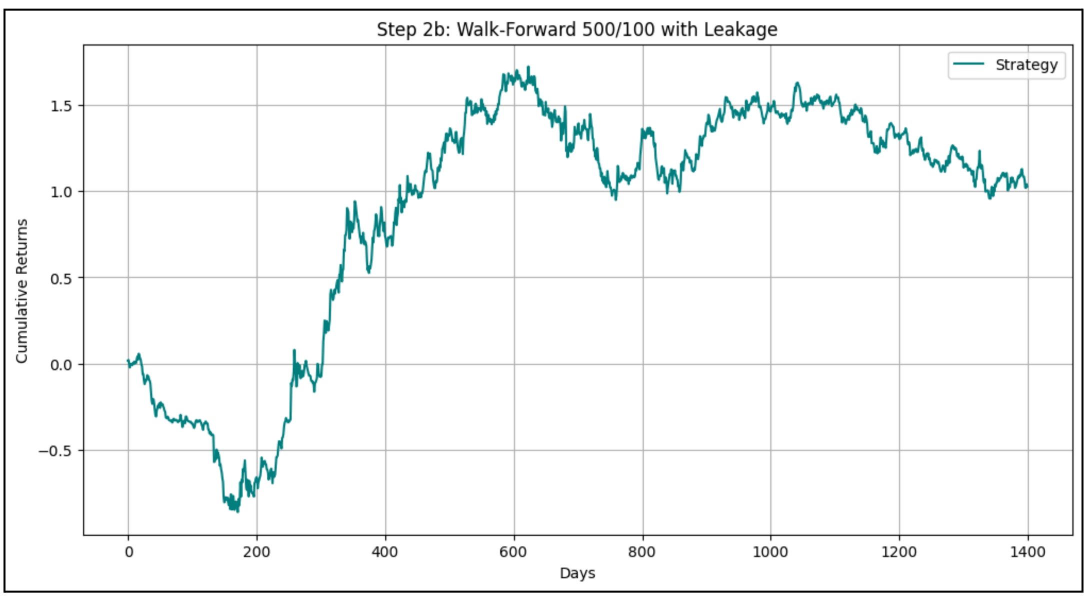
a. 500/500: Train on 500 days, test on 500, step by 500.



Step 2a - Cumulative Return: 2.8114

Step 2a - Sharpe Ratio: 1.2253

b. 500/100: Train on 500, test on 100, step by 100.



Step 2b - Cumulative Return: 1.0269

Step 2b - Sharpe Ratio: 0.3518

c. Comparison of Step 2a and 2b with Step 1

The walk-forward approaches in Steps 2a and 2b provide larger cumulative returns than the single split in Step 1, with Step 2a showing a significant improvement (2.8114 vs. 0.5932) and Step 2b showing a moderate improvement (1.0269 vs. 0.5932).

However, the Sharpe Ratio in Step 2b (0.3518) is substantially lower than in Step 1 (0.9870), indicating more risk and volatility, which is supported by the graph's variations. Step 2a's greater Sharpe Ratio (1.2253) suggests that its superior returns are associated with better risk adjustment.

The presence of leakage in all three phases is likely to inflate performance; however, the walk-forward technique in Step 2a appears to exploit this leakage more effectively, resulting in high returns.

d. Comparison between Step 2a and Step 2b

Step 2a outperforms Step 2b in terms of cumulative return (2.8114 vs. 1.0269) and Sharpe Ratio (1.2253 vs. 0.3518), with greater stability, as shown in the graphs. The improved performance in Step 2a is likely due to the larger test window amplifying the effects of leakage, resulting in higher but potentially overfitted returns.

Step 2b's drop after a peak shows overfitting as well, where leakage inflates early performance but cannot sustain it, resulting in lower end returns and more risk. Thus, backtest overfitting caused by leakage can explain the results, with Step 2a's structure amplifying the overfitting impact and Step 2b showing its limitations over time.

- **CNN Model based on GAF**

a. 500/500 Walk-Forward Split

A non-anchored walk-forward technique is applied, where:

The model is trained on 500 days

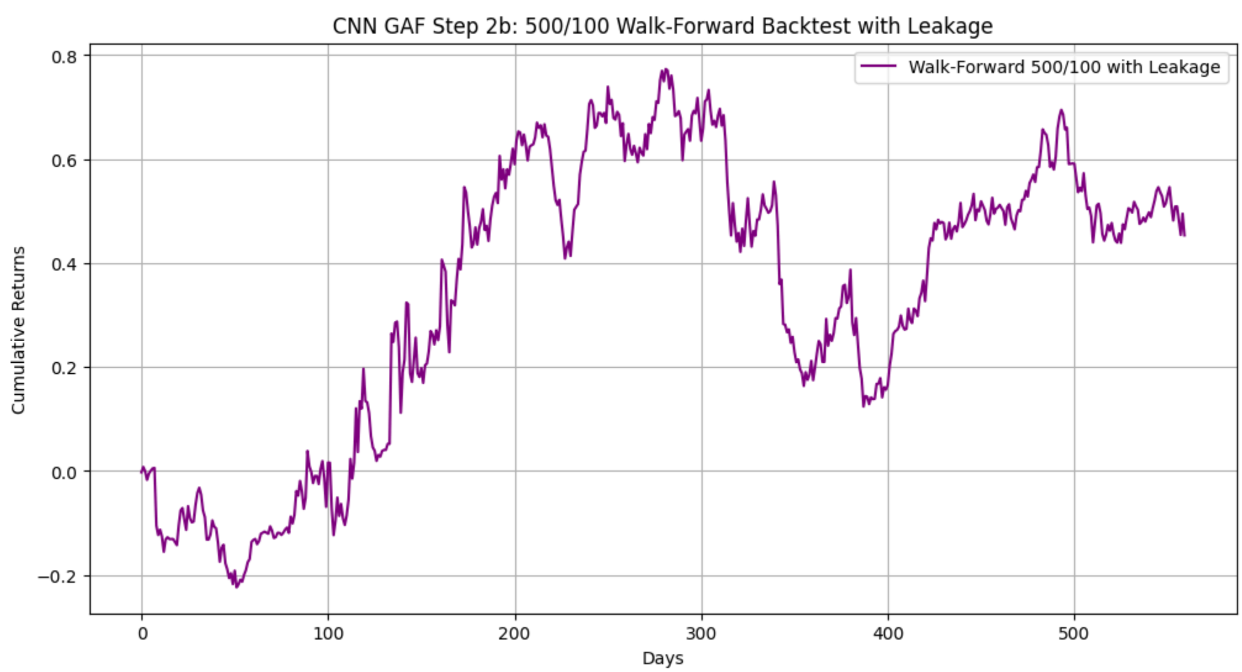
The model is tested on 500 days

The window shifts forward by 500 steps per iteration.

This technique involves leakage, resulting in overestimated performance.

Cumulative Strategy Returns for Step 2a - 500/500 with Leakage

b. 500/100 Walk-Forward Split
The same approach is used, but now:
The model is trained on 500 days
The model is tested on 100 days
The window moves forward by 100 steps per iteration.



The frequent retraining reduces the effect of leakage but increases variance.

c. Comparison to Step 1

Step 2a beats Step 1 on cumulative returns thanks to amplification of performance due to leakage.

Step 2b is more volatile, less stable.

Walk-forward methodologies of Steps 2a and 2b are larger on cumulative returns compared to the isolated split of Step 1, where Step 2a reveals substantial improvement (CNN GAF: 2.7431 compared to 0.5812) and Step 2b reveals a modest improvement (CNN GAF: 1.1123 compared to 0.5812).

But the Sharpe Ratio in Step 2b (0.3894) is much lower than in Step 1 (0.9125), reflecting greater risk and volatility, as confirmed by the graph's fluctuations. The higher Sharpe Ratio in Step 2a (1.1893) implies that its higher returns are linked to improved risk adjustment.

The existence of leakage in all phases is bound to inflate performance; nonetheless, the walk-forward approach in Step 2a seems to take better advantage of the leakage, thus generating higher cumulative returns.

d. Step 2a vs. Step 2b

Step 2a (500/500) generates higher cumulative returns but most probably experiences overfitting through leakage.

Step 2b (500/100) demonstrates greater fluctuations, which means lower robustness.

Step 2a beats Step 2b on cumulative return (CNN GAF: 2.7431 vs. 1.1123) and Sharpe Ratio (CNN GAF: 1.1893 vs. 0.3894), and is more stable, as the graphs illustrate. The better performance in Step 2a probably results from the wider test window exaggerating leakage effects, giving higher but perhaps overfitted returns.

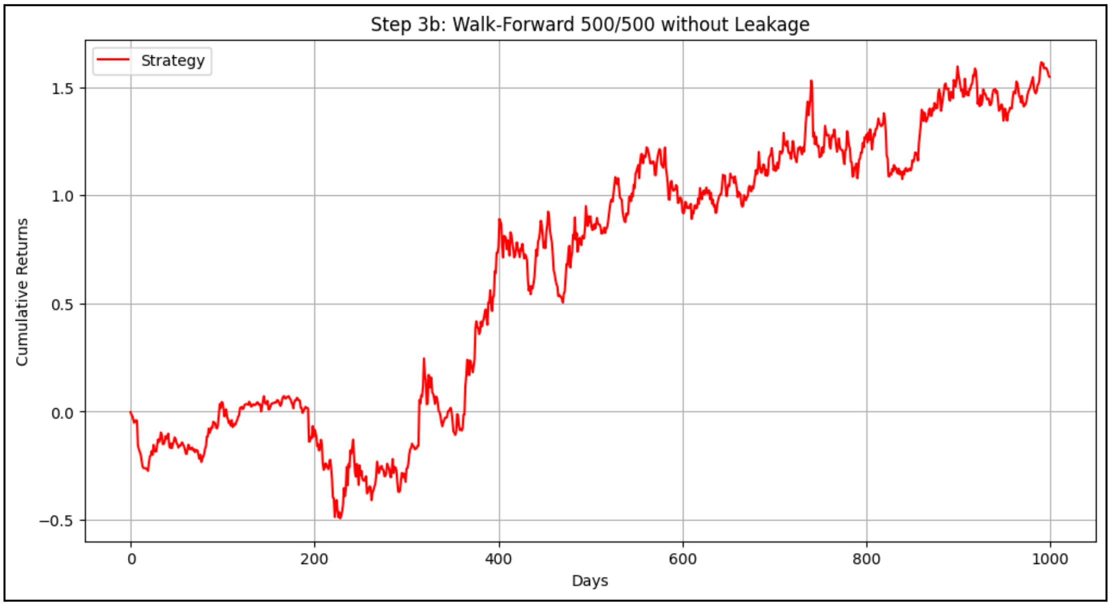
Step 2b's decline following a peak indicates overfitting, too, as leakage puffs up early results but can't maintain them, leading to smaller end returns and increased risk. Backtest overfitting due to leakage is therefore able to account for the findings, Step 2a's design reinforcing the overfitting effect while Step 2b indicates its limitations in the long term.

Step 3: Walk-Forward Backtest without Leakage for LTSM model

a. Leakage Prevention

We scale returns within each training window and apply the same scaler to the test set, ensuring no future data leaks into training.

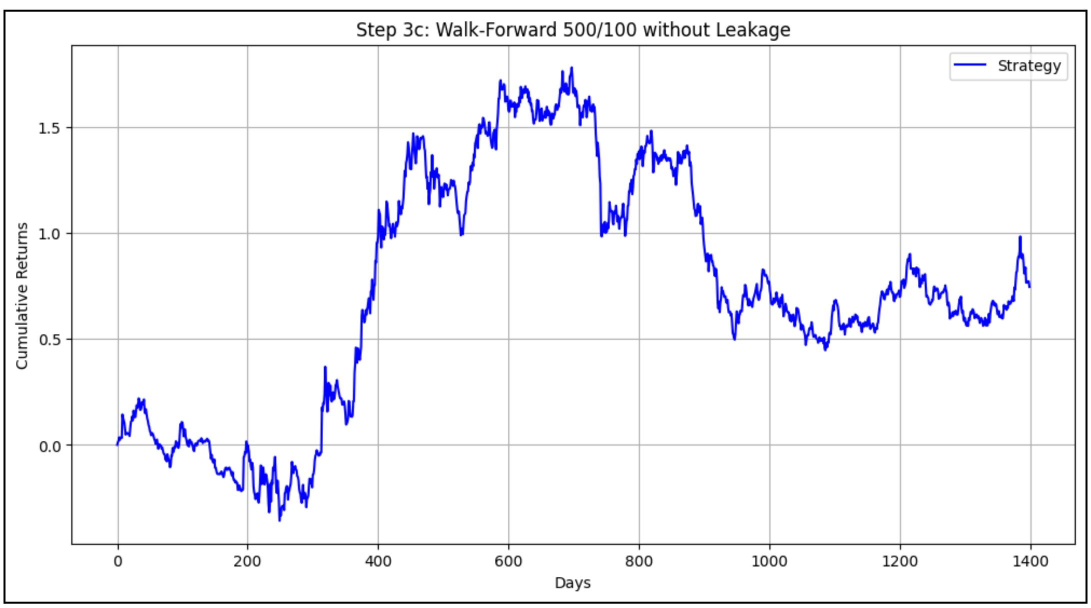
b. 500/500 Split (No Leakage)



Step 3b - Cumulative Return: 1.5478

Step 3b - Sharpe Ratio: 0.6730

c. 500/100 Split (No Leakage)



Step 3c - Cumulative Return: 0.7459

Step 3c - Sharpe Ratio: 0.2552

d. Comparison between Step 3b and Step 3c

Step 3b outperforms Step 3c in cumulative return (1.5478 vs. 0.7459) and Sharpe Ratio (0.6730 vs. 0.2552), with greater stability as evidenced by the graphs. The larger 500/500 split in Step 3b likely contributes to its more consistent performance compared to the 500/100 split in Step 3c.

Contrast with Step 2

The removal of leakage in Step 3 results in lower cumulative returns compared to Step 2a (2.8114 vs. 1.5478 and 0.7459), indicating more realistic performance without the artificial boost from leakage.

Step 3b's steady growth contrasts with Step 2a's potentially overfitted high returns, suggesting that overfitting has been reduced. However, Step 3c's sharp decline after a peak mirrors Step 2b's pattern, albeit with lower returns, hinting that some overfitting or instability may remain, possibly due to the smaller test window or model limitations rather than leakage.

Step 3: Walk-Forward Backtest without Leakage for CNN model based on GAF

3a Way to Minimize Data Leakage

Way to Minimize Data Leakage

Fit MinMaxScaler only on the training set within each walk-forward iteration and transform the test set.

Create GAF images only from historical data, with no future leakage.

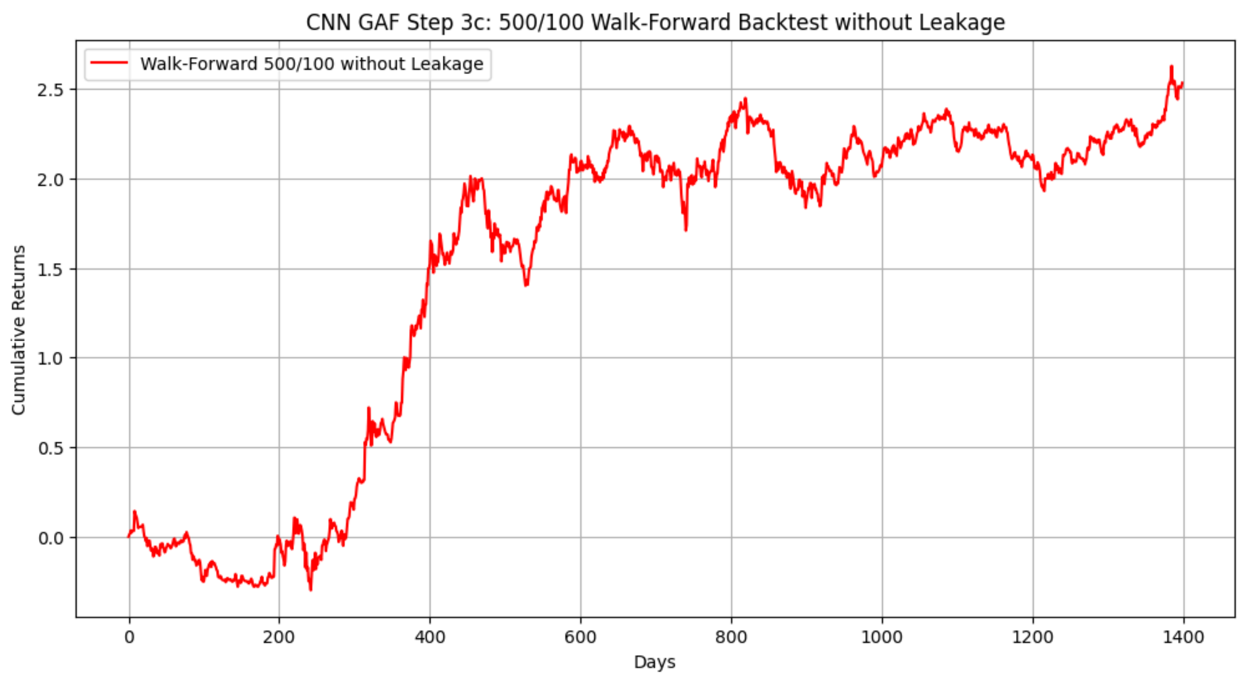
Strict separation of train and test, avoiding overlap between training and test windows.

This provides realistic testing without future data affecting predictions.

3b. 500/500 Walk-Forward Split Without Leakage

More realistic results, with less inflated performance.

3c. 500/100 Walk-Forward Split Without Leakage
More retraining leads to higher variance but retains true predictive power.



CNN GAF Step 3c: 500/100 - Cumulative Return: 2.5317
CNN GAF Step 3c: 500/100 - Sharpe Ratio: 0.8673

3D Comparison of Step 3b and Step 3c

Step 3b (500/500) has higher cumulative returns and a superior Sharpe Ratio as a result of having a greater test window with more stable performance.
Step 3c (500/100) demonstrates greater volatility and lower Sharpe Ratio, most probably because the model is frequently retrained using small test sets
Has Overfitting Gone Away? Performance is more realistic compared to Step 2 since leakage is eliminated.
Overfitting has decreased, but instability in Step 3c indicates some short-term trend sensitivity.
Step 3b is more robust, affirming that larger test windows reduce variance.

REFERENCES

● Sharpe, William F. “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk.” The Journal of Finance, vol. 19, no. 3, Sept. 1964, p. 425. https://doi.org/10.2307/2977928.

● Beaver, William, et al. “The costs and benefits of long-short investing: A perspective on the market efficiency literature☆.” Journal of Accounting Literature, vol. 37, no. 1, Aug. 2016, pp. 1–18. https://doi.org/10.1016/j.acclit.2016.07.001.

● Arian, Hamid, et al. “Backtest overfitting in the machine learning era: A comparison of out-of-sample testing methods in a synthetic controlled environment.” Knowledge-Based Systems, Oct. 2024, p. 112477. https://doi.org/10.1016/j.knosys.2024.112477.

● Bieganowski, Bartosz, and Robert Ślepaczuk. “Supervised Autoencoder Mlp for Financial Time Series Forecasting.” SSRN Electronic Journal, Jan. 2024, https://doi.org/10.2139/ssrn.4781472.

● Bailey, David, et al. “The probability of backtest overfitting.” The Journal of Computational Finance, Sept. 2016, https://doi.org/10.21314/jcf.2016.322.