Can an LSTM Beat HAR? Deep Learning vs the Classic Volatility Model, S&P 500, 2008–2026

Gaurav Gupta · Research note · 28 October 2026 · Walk-forward, refit every January; 4,688 out-of-sample days; SPY OHLC and CBOE VIX; PyTorch on CPU

0.325 vs 0.392
Forecast loss (QLIKE): HAR + VIX vs LSTM; DM t = 2.6

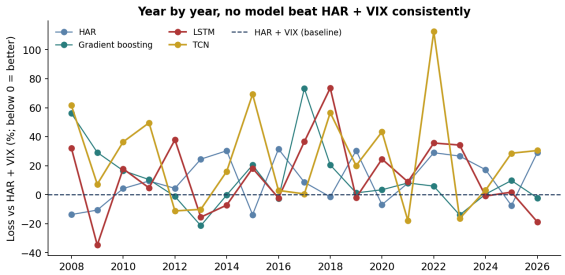
7 of 19
Years the LSTM won: all calm; lost 2008, 2015, 2018, 2020

−19.0 pts
LSTM under-forecast when realised vol > 30% (HAR + VIX −13.6)

$38M
Deeper worst drawdown per $1bn, LSTM vs HAR sizing

Forecaster	QLIKE	DM t vs HAR+VIX	Years better	Vol-target max DD	Miss vs 15%	VaR99 breaches
HAR + VIX	0.325	n/a	n/a	−32.6%	1.8 pts	1.5%
HAR	0.328	+0.2	6/19	−30.3%	1.7 pts	1.7%
VIX only	0.340	+2.2	8/19	−33.3%	2.0 pts	1.5%
Gradient boosting	0.371	+3.1	6/19	−31.5%	2.0 pts	1.8%
LSTM (PyTorch)	0.392	+2.6	7/19	−34.2%	2.1 pts	2.0%
Last month	0.392	+2.8	n/a	−28.0%	1.2 pts	2.0%
TCN (PyTorch)	0.450	+2.9	4/19	−36.3%	2.0 pts	2.5%

Buy-and-hold SPY: worst drawdown −53.0%, Sharpe 0.45; vol-targeted books 0.51–0.55. VaR: normal 99% one-day from each forecast (target 1%). Vol target: 15%, weekly rebalance, leverage cap 2, 2 bp per unit turnover.



Loss relative to HAR + VIX by year: networks win calm years, lose the shocks.

SOURCES

- SPY daily OHLC: Yahoo via defeatbeta/yahoo-finance-data (ODC-BY); CBOE VIX: github.com/datasets/finance-vix (ODC-PDDL); sha256-pinned
- Corsi (2009) JFEC 7(2); Garman & Klass (1980); Hochreiter & Schmidhuber (1997); Bai, Kolter & Koltun (2018); Patton (2011); Diebold & Mariano (1995); Bucci (2020); Christensen et al. (2023); Moreira & Muir (2017)

METHOD

- Daily variance = overnight² + Garman–Klass; target = log mean variance next 22 days
- Refit each January 2008–2026 on all prior data, 22-day embargo (asserted in code); LSTM/TCN 3 seeds, deterministic, fixed settings
- QLIKE, Diebold–Mariano (Newey–West 22 lags), vol-target and VaR decision tests

KEY CALCULATIONS

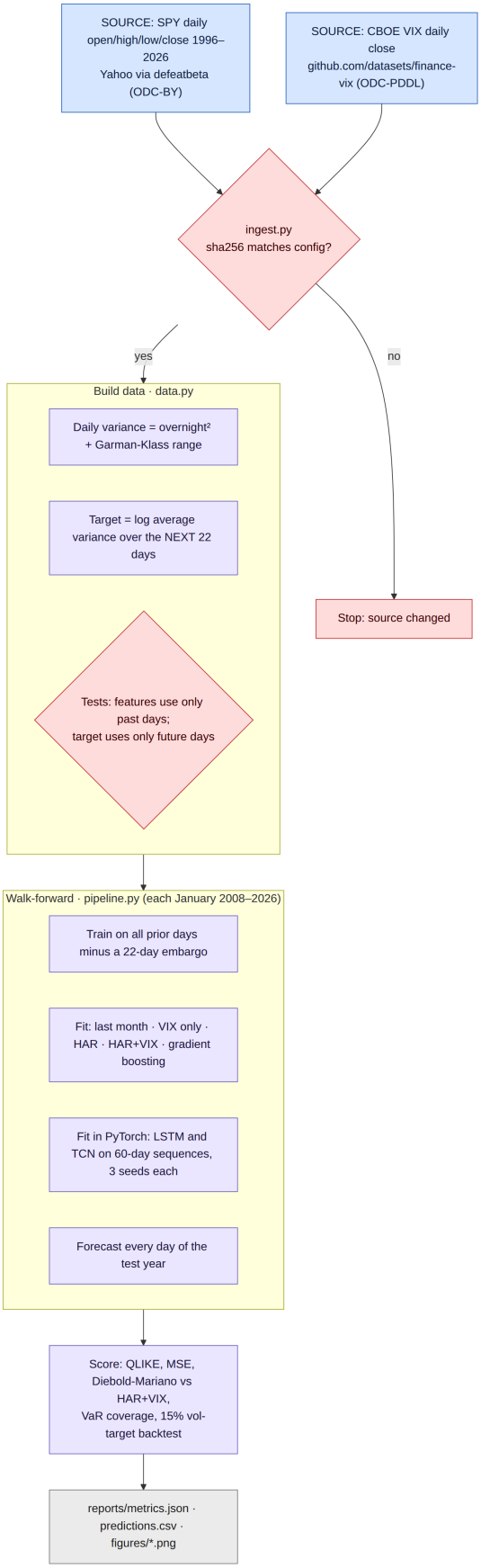
- Stress years, LSTM loss vs HAR+VIX: 2008 +32%, 2015 +18%, 2018 +73%, 2020 +25%
- VaR understatement on stress days ≈ 5.4 pts ÷ √252 × 2.326 ≈ $7.9M per $1bn
- Compute: all 19 refits ≈ 36 min on 2 CPU threads

RECOMMENDATIONS

- Keep HAR + VIX in production; run neural models in shadow
- Promote only on a significant, stable gain across regimes
- Choose the forecaster by the decision; add a stress overlay and fat-tail VaR adjustment

Independent research on public data; not investment advice; not deployed. Code: github.com/guptagaurav-info/lstm-vs-har-volatility · guptagaurav.info/articles/can-an-lstm-beat-har

Pipeline


SOURCE: SPY daily open/high/low/close 1996–2026 Yahoo via defeatbeta (ODC-BY)
SOURCE: CBOE VIX daily close github.com/datasets/finance-vix (ODC-PDDL)
ingest.py sha256 matches config?
yes
no
Stop: source changed
Build data · data.py
Daily variance = overnight² + Garman-Klass range
Target = log average variance over the NEXT 22 days
Tests: features use only past days; target uses only future days
Walk-forward · pipeline.py (each January 2008–2026)
Train on all prior days minus a 22-day embargo
Fit: last month · VIX only · HAR · HAR+VIX · gradient boosting
Fit in PyTorch: LSTM and TCN on 60-day sequences, 3 seeds each
Forecast every day of the test year
Score: QLIKE, MSE, Diebold-Mariano vs HAR+VIX, VaR coverage, 15% vol-target backtest
reports/metrics.json · predictions.csv · figures/*.png

How a risk committee would choose and govern the forecaster

