IOWA STATE UNIVERSITY
Department of Computer Science
Research Poster

When Do Engineered Features Beat Raw Prices?

A Walk-Forward Benchmark of Classical, Neural, and World-Model Predictors on a Global Equity Dataset

Nilutpaul Sarker Yash[1]†, Tousif Bashar[1], Md Mahmudul Islam Maruf[2], Masrur Mouno[1], Sayantan Chakraborty[1], Koushik Howlader[3], Ushashi Bhattacharjee[1], Tirtho Roy[1]

[1] Iowa State University, USA • [2] University of Global Village (UGV), Barishal, Bangladesh • [3] University of Dhaka, Bangladesh

† Presenting author

FIGURE 1. CONTROLLED BENCHMARK OVERVIEW — FROM INPUT REPRESENTATION TO EXPERIMENTAL ANSWER

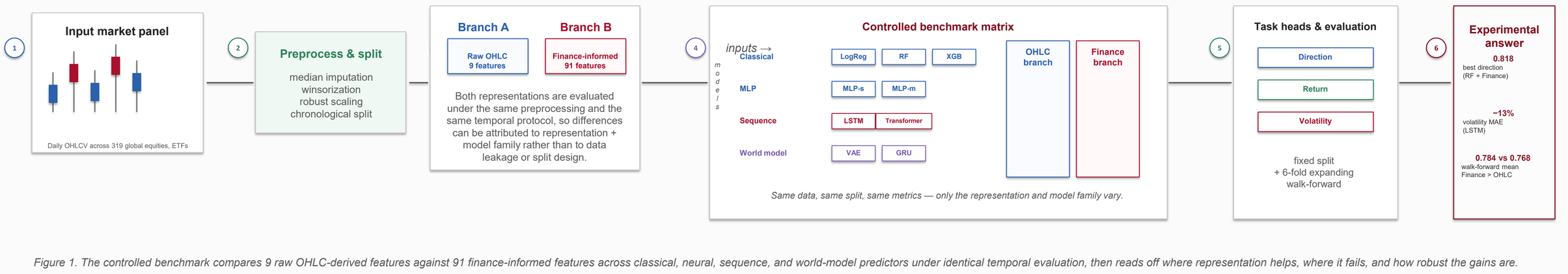


Figure 1. The controlled benchmark compares 9 raw OHLC-derived features against 91 finance-informed features across classical, neural, sequence, and world-model predictors under identical temporal evaluation, then reads off where representation helps, where it fails, and how robust the gains are.

DATASET, TASKS & SCOPE

319	12	2018–2026	9 vs 91
equities & ETFs	sectors	daily OHLCV	features

Market universe

Technology · Financials · Healthcare · Consumer · Industrials · Energy · Materials · Real Estate · Comm. Serv. · Utilities · ETFs / Indices

Prediction tasks

Direction: 3-class {up, flat, down} with neutral band τ = 0.005
Return: next-day log return (secondary target)
Volatility: forward 5-day realized volatility

Preprocessing and controls

Median imputation and winsorization
Robust scaling for volatile distributions
Cross-sectional normalization in the controlled benchmark
Strict chronology to prevent look-ahead leakage

Why this benchmark matters

It isolates representation from architecture by keeping the dataset, preprocessing, and evaluation protocol fixed.
This lets us ask a cleaner question: are gains coming from better features or just from more complex models?

Key numbers at a glance

0.818	best directional accuracy
-13%	LSTM volatility MAE reduction
0.784 vs 0.768	walk-forward mean accuracy
−0.039	transformer delta with added features

PROTOCOL & WORLD-MODEL INSET

Temporal evaluation protocol

Train → 2024-06-30 | Val | Test → 2026-03-31

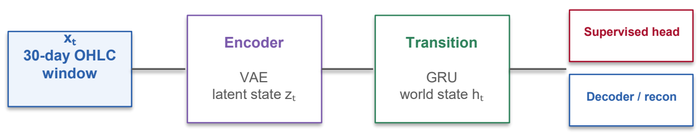
Fixed split + 6-fold expanding-window walk-forward analysis

Experimental comparison grid

models \ representations →	OHLC (9)	Finance (91)
Classical	LR·RF·XGB	LR·RF·XGB
MLP	S · M	S · M
Sequence	LSTM·TF	LSTM·TF
World model	VAE+GRU	VAE+GRU

This grid encodes the core study design: same data and protocol, two competing representations, and multiple **model families**.

World-model inset



Composite objective: $L = \lambda_{sup} \cdot L_{sup} + L_{recon} + \beta \cdot L_{KL} + \gamma \cdot L_{trans}$

with $\lambda_{sup} = 5.0$, $\beta = 0.5$, $\gamma = 1.0$

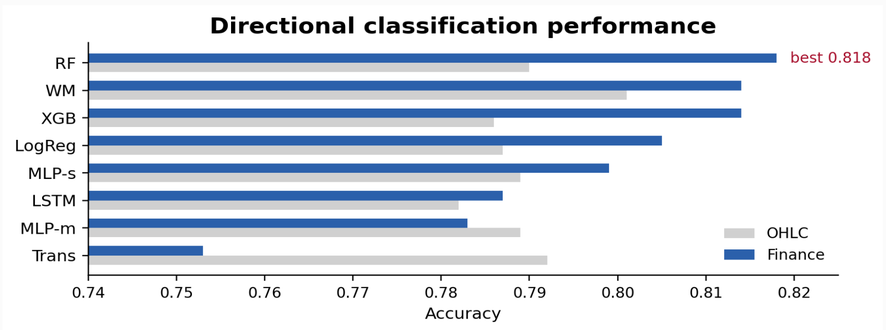
Why the world model matters here

It adds a latent-dynamics comparator rather than only standard tabular and sequence baselines.
In the benchmark, it is the strongest OHLC-only classifier.
This makes it a natural fit for WM@Booth, where modeling latent state and transition structure is central.

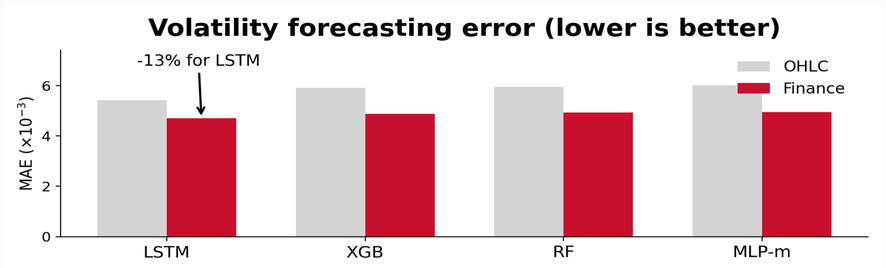
Protocol takeaway

The study is deliberately controlled: the same market panel, the same chronology, and the same evaluation protocol are used throughout. That makes the comparison interpretable. When gains appear, we can attribute them more credibly to representation quality and model family rather than to confounds in splitting or preprocessing.

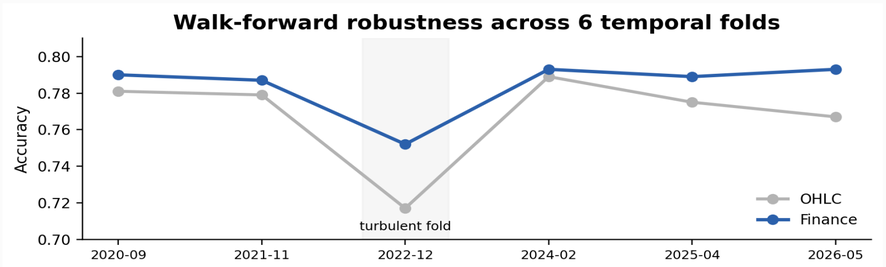
RESULTS & INTERPRETATION



0.818 — Best directional accuracy: Random Forest with finance-informed features. The two biggest positive deltas are RF and XGBoost (+0.028 each).



−13% — For volatility, engineered features reduce LSTM MAE by roughly 13%.



0.784 vs 0.768 — Finance-informed features produce a higher and steadier walk-forward mean accuracy.

Interpretation

Feature engineering helps direction and volatility prediction.
The gains concentrate in tree-based / simple models rather than in deeper architectures.
The compact world model is the strongest OHLC-only classifier.
Next-period return regression shows little exploitable signal.

TAKEAWAYS, LIMITATIONS & NEXT STEPS

1. Representation wins
Engineered finance features improve direction and volatility prediction, especially for tabular models.

2. Complexity is not enough
The transformer does not win, while the compact world model is the strongest OHLC-only classifier.

3. Returns remain hard
Raw return regression remains weak; the useful structure lies more in direction and volatility.

Limitations — Single data source, fixed hyperparameters, and no transaction-cost / portfolio-construction evaluation.

Next steps — Add cost-aware backtesting, extend to more asset classes, and tune models individually after the controlled benchmark stage.

Core contribution: a controlled financial ML benchmark showing that representation quality improves both peak performance and temporal robustness more reliably than architectural escalation alone.

Disclosure of interests: The authors have no competing interests to declare that are relevant to the content of this article.

Presenting Author: Nilutpaul Sarker Yash

IOWA STATE UNIVERSITY