t (RandomizedSearchCV, Refit by ROC-AUC) — VLO Nex

Generated: 2026-02-26 03:23:24
TRAIN period: ≤ 2022-12-31 | TEST period: ≥ 2023-01-01
TRAIN size: 2,672 | TEST size: 787
Features — scaled (float + VOL_*): 72 | passthrough: 5
RandomizedSearchCV: n_iter=80, refit='roc_auc'
mples_split': 25, 'clf__min_samples_leaf': 2, 'clf__max_features': 'sqrt', 'clf__max_depth': 30, 'clf__class_weight':
Best CV ROC-AUC: 0.5415
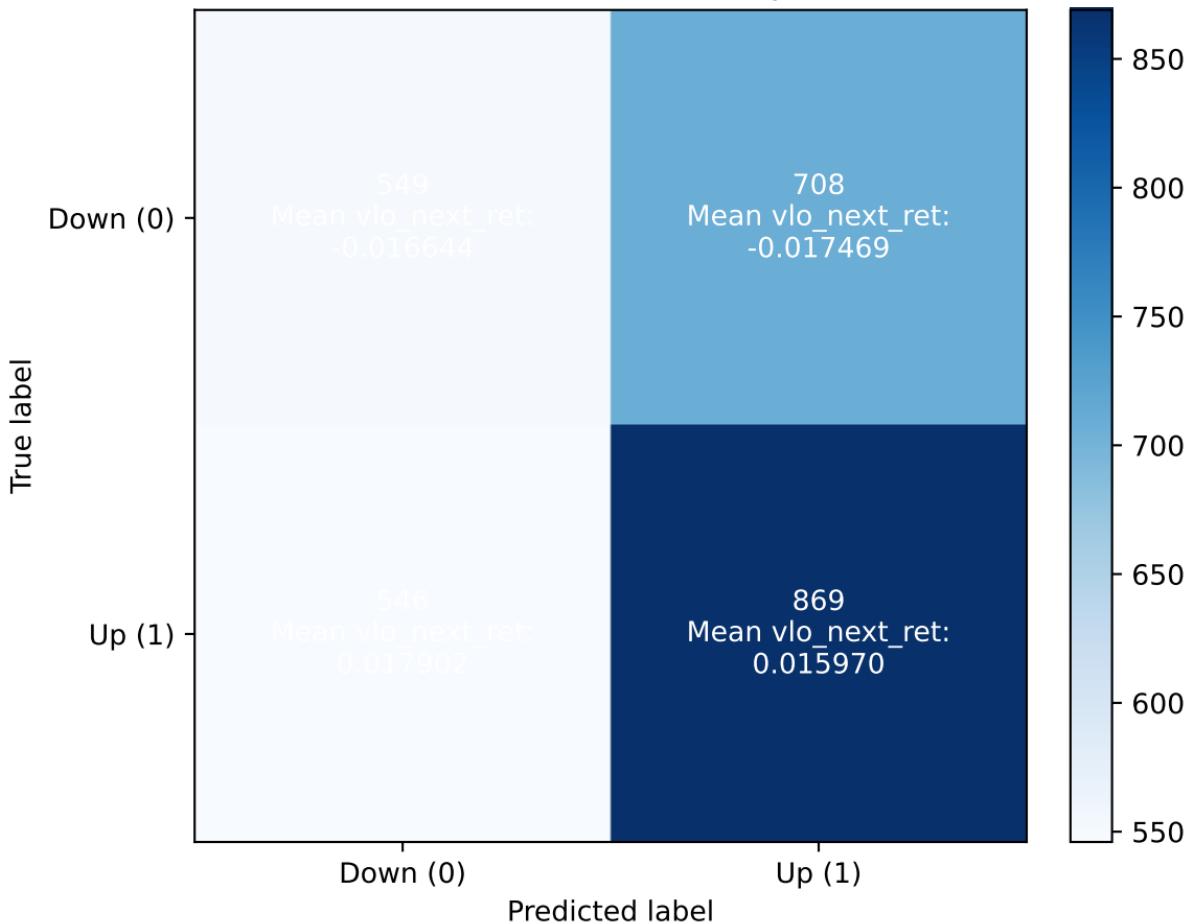
Randomized Search — max mean ROC-AUC over other Random Forest params
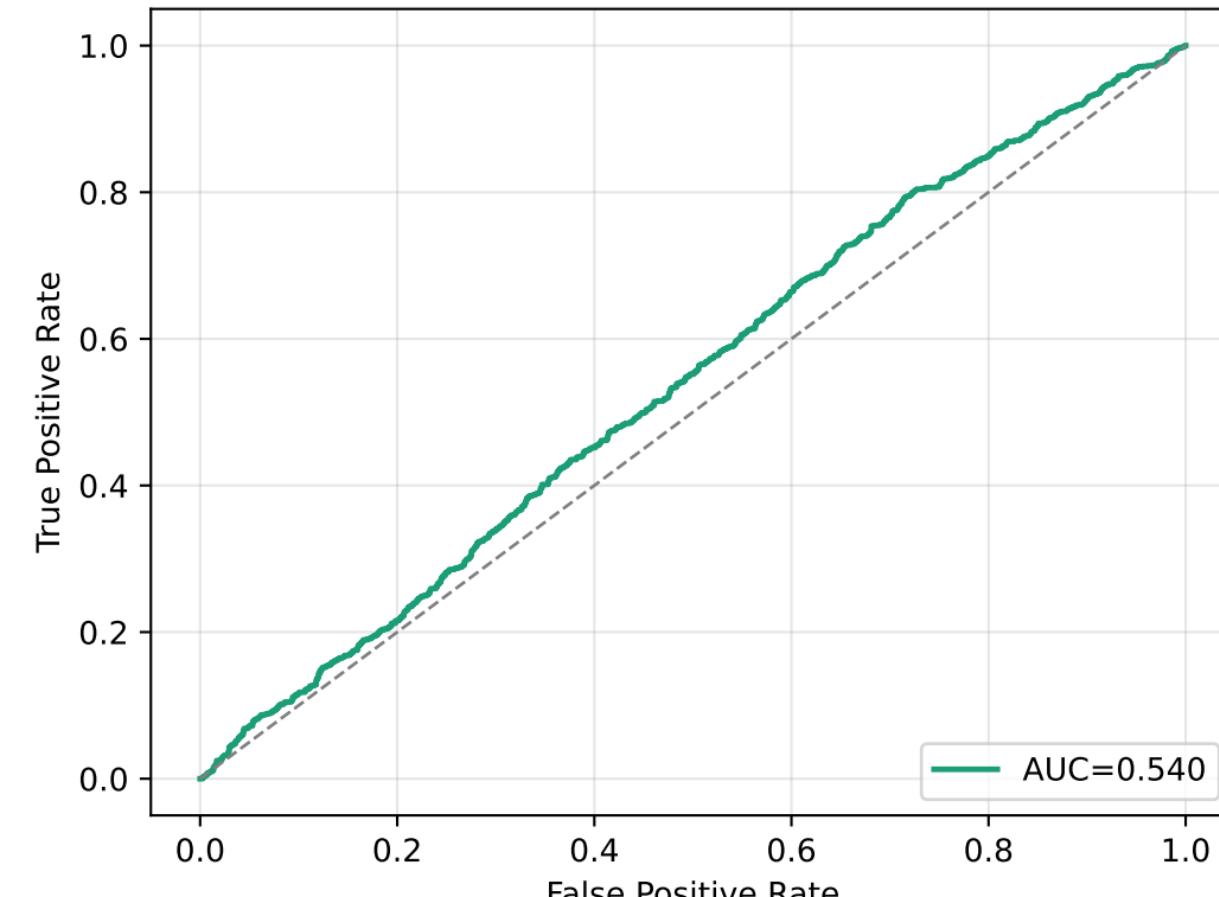
Cross-Validation ROC-AUC — Training Set (≤ 2022-12-31)

metric	value
mean ROC-AUC (CV folds)	0.5415
std ROC-AUC (CV folds)	0.0132
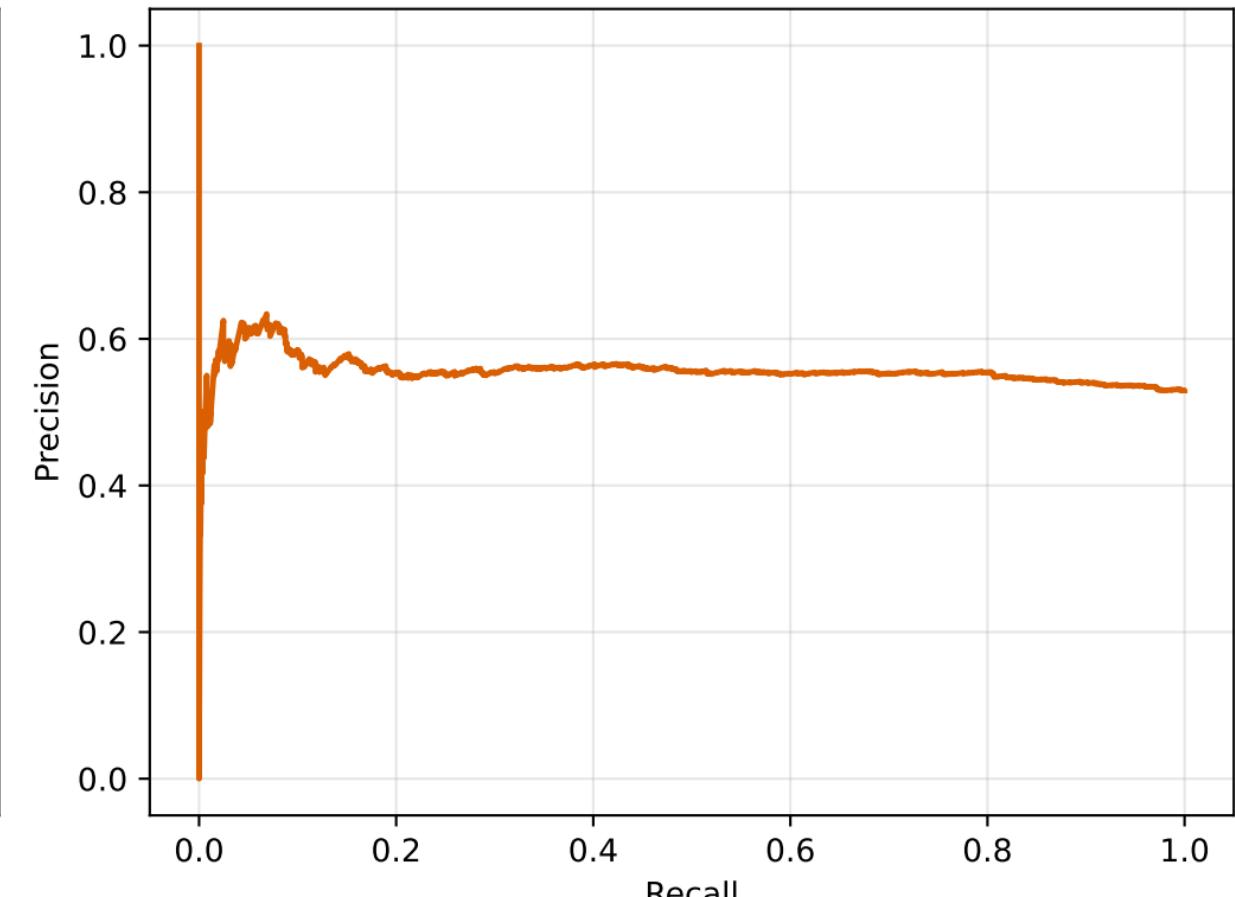
Confusion Matrix — TRAIN (CV predictions)

	Down (0)	Up (1)
Down (0)	549 Mean vlo_next_ret: -0.016644	708 Mean vlo_next_ret: -0.017469
Up (1)	546 Mean vlo_next_ret: 0.017902	869 Mean vlo_next_ret: 0.015970

True label

Predicted label
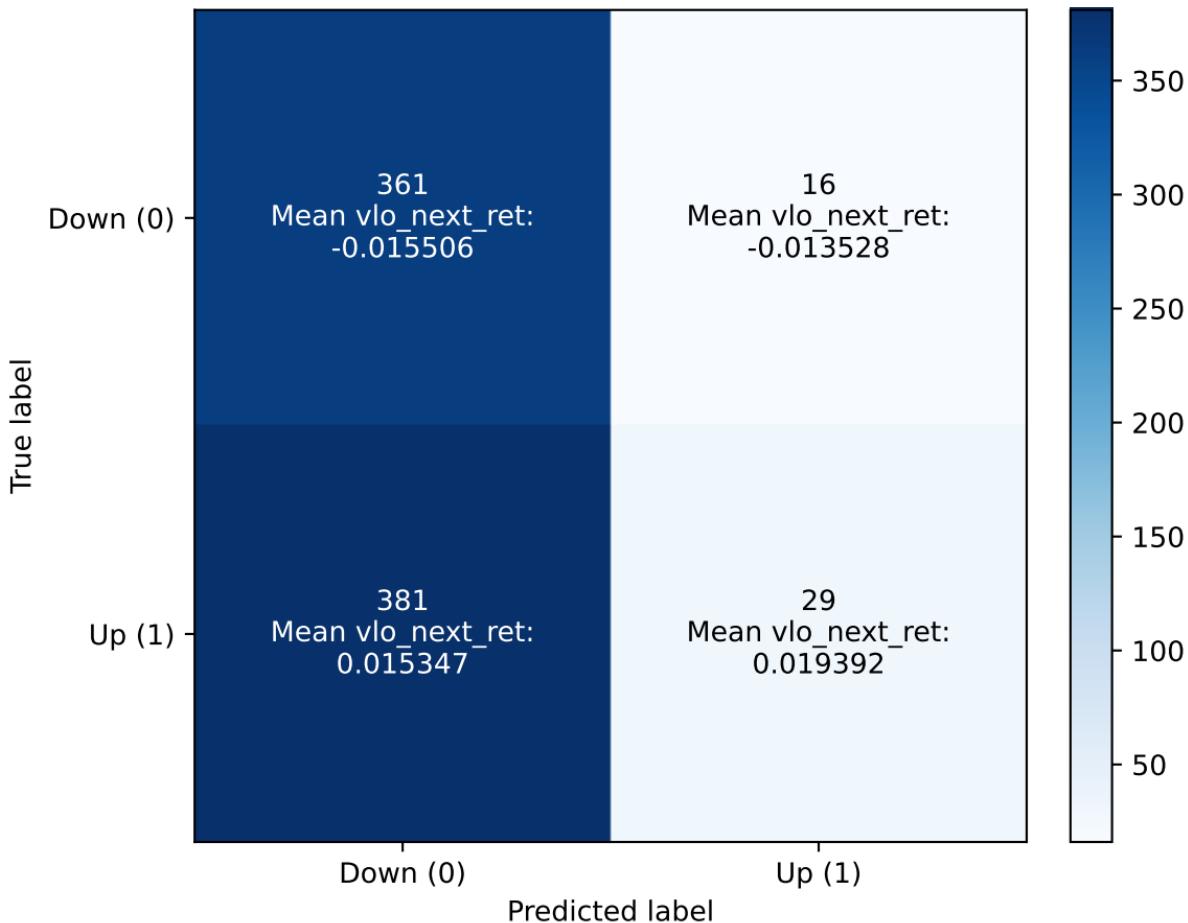
ROC Curve — TRAIN (CV pooled)

AUC=0.540
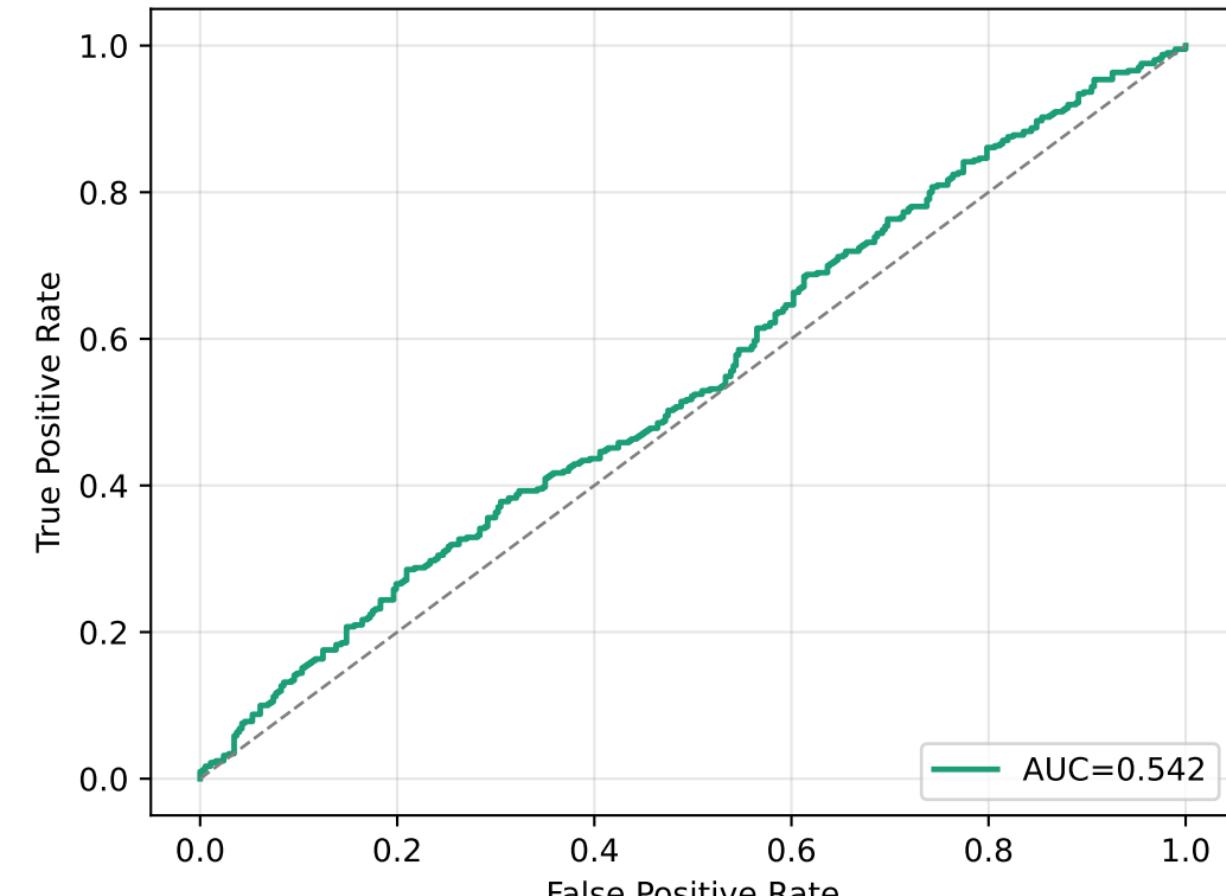
Precision-Recall — TRAIN (CV pooled)
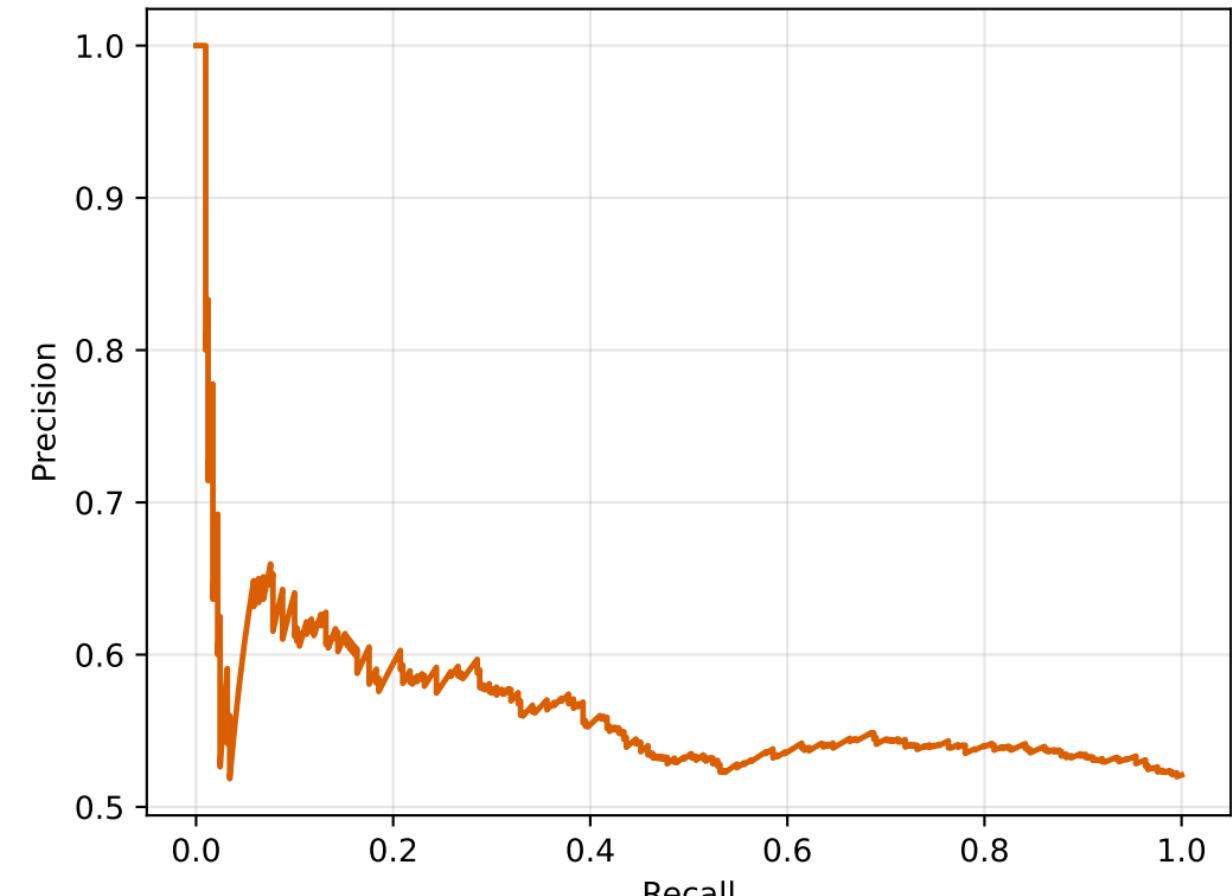
Confusion Matrix — TEST (2023+)

	Down (0)	Up (1)
Down (0)	361 Mean vlo_next_ret: -0.015506	16 Mean vlo_next_ret: -0.013528
Up (1)	381 Mean vlo_next_ret: 0.015347	29 Mean vlo_next_ret: 0.019392

ROC Curve — TEST (2023+)

AUC=0.542
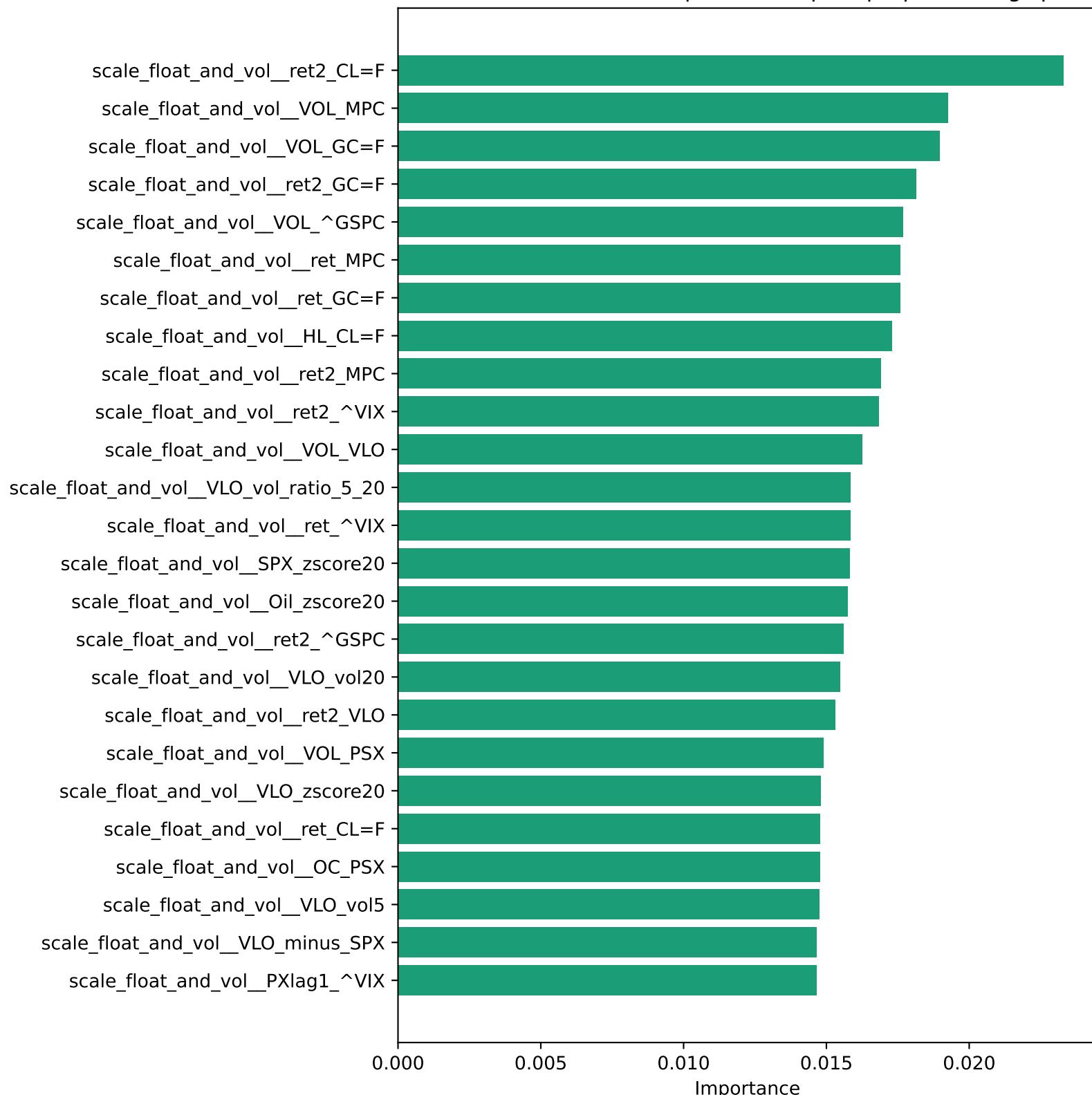
Precision-Recall — TEST (2023+)
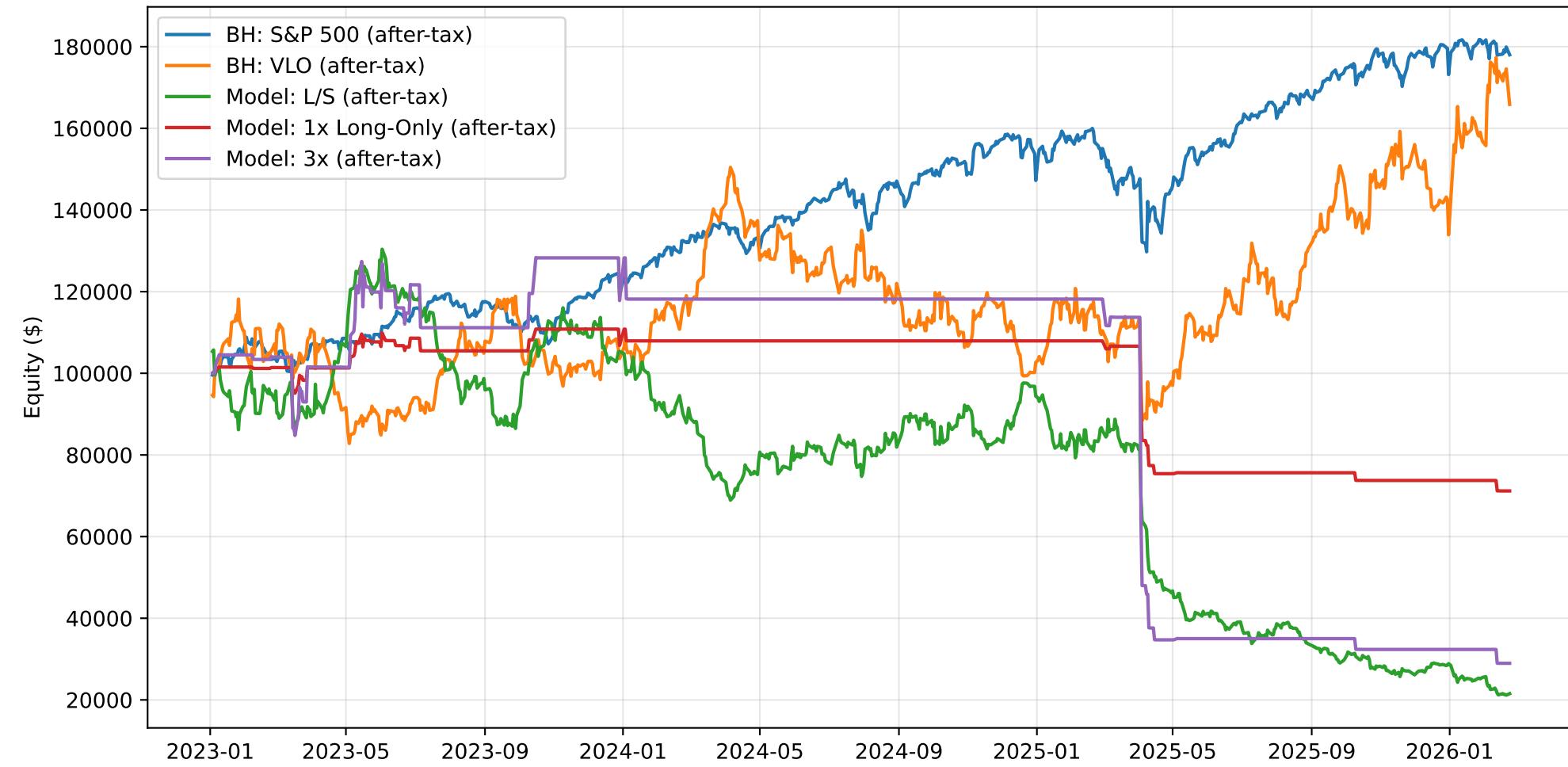
Classification Report — TEST (2023+)

	precision	recall	f1-score	support
0	0.487	0.958	0.645	377
1	0.644	0.071	0.127	410
accuracy			0.496	787
macro avg	0.565	0.514	0.386	787
weighted avg	0.569	0.496	0.375	787

Random Forest Feature Importances (post-preprocessing space)

After-Tax Equity Curves — TEST: 2023 to most recent

Annual After-Tax Performance (2023+), incl. Average

	BH: S&P 500 (after-tax)	BH: VLO (after-tax)	Model: L/S (after-tax)	Model: 1x Long-Only (after-tax)	Model: 3x (after-tax)
2023	19.78%	9.50%	-0.02%	6.82%	17.81%
2024	19.21%	-3.45%	-9.94%	-2.60%	-7.87%
2025	13.32%	29.79%	-69.25%	-31.67%	-72.62%
2026	-0.32%	16.09%	-24.33%	-3.48%	-10.49%
Average	13.00%	12.98%	-25.88%	-7.73%	-18.29%

Run Metadata & Time-Split Summary

```
Report saved: /mnt/batch/tasks/shared/LS_root/mounts/clusters/dvandyk1/code/Users/dva
Scaler saved: /mnt/batch/tasks/shared/LS_root/mounts/clusters/dvandyk1/code/Users/dva
Model  saved: /mnt/batch/tasks/shared/LS_root/mounts/clusters/dvandyk1/code/Users/dva

TRAIN period: ≤ 2022-12-31 | TEST period: ≥ 2023-01-01
TRAIN size: 2,672 | TEST size: 787
CV folds: 5
RandomizedSearchCV n_iter: 80
Scaled features (float + VOL_*): 72
Passthrough numeric features: 5
Best params: {'clf__n_estimators': 200, 'clf__min_samples_split': 25, 'clf__min_sampl
Best CV ROC-AUC (refit metric): 0.5415
TEST ROC-AUC: 0.5423 | TEST Accuracy: 0.4956

Backtest assumptions (TEST period only, AFTER-TAX curves shown):
  Initial capital: $100,000
  Tx cost per side: 2.00 bps
  Taxes (annual on positive P&L, loss carry-forward):
    - Long-term (buy & hold): 20.0%
    - Short-term (active): 37.0%
  Short borrow APR (L/S shorts): 0.30%
  3x long-only daily cost (financing+expense): 0.0595%
  Stop-loss: NOT APPLIED

Feature exclusions (pre-model): vlo_next_ret, vlo_next_dir
```