Online Resource 1

Supplementary material for: Estimand-Relative Validity and Context-Driven Instability of Shapley Explanations for Tabular Foundation Models

Emin Talip Demirkiran

International Journal of Machine Learning and Cybernetics

Affiliation: Department of Computer Engineering, Eskişehir Technical University, Iki Eylul Campus, 26555 Eskisehir, Türkiye
Corresponding author email: etd@eskisehir.edu.tr
ORCID: https://orcid.org/0000-0001-6063-8178

S1. Prespecified protocol and analysis plan

- `Supporting_Source/Protocol/study_protocol.md`
- `Supporting_Source/Protocol/study_protocol.yaml`
- `Supporting_Source/Protocol/analysis_plan.md`
- `Supporting_Source/Protocol/sampling_and_data_protocol.md`
- `Supporting_Source/Protocol/robustness_and_sensitivity_plan.md`
- `Supporting_Source/Protocol/outcome_and_metric_registry.yaml`
- `Supporting_Source/Protocol/comparator_registry.yaml`
- `Supporting_Source/Protocol/study_matrix.csv`
- `Supporting_Source/Protocol/claim_study_traceability.csv`

S2. Study matrix

arm	units
F2_dose	24
F3_labelrand	24
armA	120
armA_d4	40
armA_real	40
armB	384

model	units
tabicl_v2	316
tabpfn_v2	316

Total planned units: 632. Full matrix: `study_matrix.csv` (shipped).

S3. Full result tables

asset	files	label
`B1_config`	Supporting_Source/Supplementary_Tables/B1_config.csv	
`B2_variance`	Supporting_Source/Supplementary_Tables/B2_variance.csv	
`completeness`	Supporting_Source/Supplementary_Tables/completeness.csv	
`error_contradictions`	Supporting_Source/Supplementary_Tables/error_contradictions.csv	
`error_cost_accuracy`	Supporting_Source/Supplementary_Tables/error_cost_accuracy.csv	
`error_exclusions`	Supporting_Source/Supplementary_Tables/error_exclusions.csv	

asset	files	label
error_hard_cases	Supporting_Source/Supplementary_Tables/error_hard_cases.csv	
error_misfit_strata	Supporting_Source/Supplementary_Tables/error_misfit_strata.csv	
error_numerical_health	Supporting_Source/Supplementary_Tables/error_numerical_health.csv	
error_shared_difficulty	Supporting_Source/Supplementary_Tables/error_shared_difficulty.csv	
error_status_counts	Supporting_Source/Supplementary_Tables/error_status_counts.csv	
error_unit_outcomes	Supporting_Source/Supplementary_Tables/error_unit_outcomes.csv	
L1_instance	Supporting_Source/Supplementary_Tables/L1_instance.csv	
L2_unit	Supporting_Source/Supplementary_Tables/L2_unit.csv	
multiplicity_ledger	Supporting_Source/Supplementary_Tables/multiplicity_ledger.csv	
robustness_leave_one_regime_out	Supporting_Source/Supplementary_Tables/robustness_leave_one_regime_out.csv	
robustness_r1_transform	Supporting_Source/Supplementary_Tables/robustness_r1_transform.csv	
robustness_r3_secondary_consistency	Supporting_Source/Supplementary_Tables/robustness_r3_secondary_consistency.csv	
robustness_r4_scale	Supporting_Source/Supplementary_Tables/robustness_r4_scale.csv	
robustness_r5_rank_distances	Supporting_Source/Supplementary_Tables/robustness_r5_rank_distances.csv	
secondary_C6	Supporting_Source/Supplementary_Tables/secondary_C6.csv	
secondary_Delta_real	Supporting_Source/Supplementary_Tables/secondary_Delta_real.csv	
secondary_F1	Supporting_Source/Supplementary_Tables/secondary_F1.csv	
secondary_F1_pairwise	Supporting_Source/Supplementary_Tables/secondary_F1_pairwise.csv	
secondary_F2	Supporting_Source/Supplementary_Tables/secondary_F2.csv	
secondary_F3	Supporting_Source/Supplementary_Tables/secondary_F3.csv	
secondary_F3_pairwise	Supporting_Source/Supplementary_Tables/secondary_F3_pairwise.csv	
secondary_metrics_rank_topk_sign	Supporting_Source/Supplementary_Tables/secondary_metrics_rank_topk_sign.csv	
secondary_runtime	Supporting_Source/Supplementary_Tables/secondary_runtime.csv	
secondary_scale_d4_vs_d8	Supporting_Source/Supplementary_Tables/secondary_scale_d4_vs_d8.csv	

S4. Additional figure

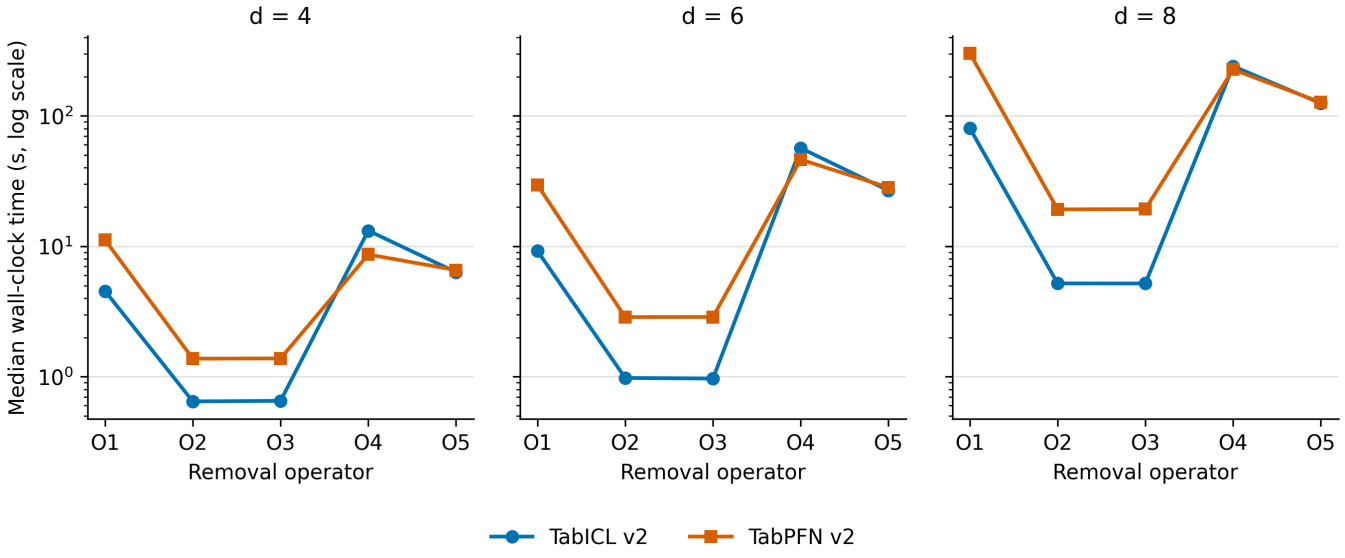


Figure 1: Median wall-clock time by removal operator, model and feature dimension.

Supplementary Figure S1. Median wall-clock time by removal operator, model and feature dimension. The vertical axis uses a logarithmic scale. Values are descriptive medians without uncertainty intervals and reflect the specified CPU environment.

S5. Per-unit index

The complete machine-readable index is supplied as `Supporting_Source/unit_index.csv`. It contains all 632 pre-specified study units and records the unit identifier, study arm, model, condition, seed, terminal status, and runtime. All 632 units reached `COMPLETED` status and there were no execution failures. The CSV is provided instead of a typeset 632-row table to preserve exact identifiers and machine readability.

S6. Environment and resource identity

Pinned environment: `configs/requirements-research.lock.txt` (exact versions).

Resource identity:

resource	identity
R-MOD-TABPFN-V2-CLF	tabpfn-v2-classifier-finetuned-zk73skhh.ckpt
R-MOD-TABPFN-V2-REG	tabpfn-v2-regressor.ckpt
R-MOD-TABICL-V2-CLF	tabicl-classifier-v2-20260212.ckpt
R-MOD-TABICL-V2-REG	tabicl-regressor-v2-20260212.ckpt
R-SW-TABPFN	.venv (project virtual environment, –system-site-packages)
R-SW-TABICL	.venv (project virtual environment, –system-site-packages)
R-SW-TORCH	.venv (project virtual environment, –system-site-packages)
R-SW-NUMPY	.venv (project virtual environment, –system-site-packages)
R-SW-SHAPIQ	.venv (project virtual environment, –system-site-packages)
R-SW-XGBOOST	.venv (project virtual environment, –system-site-packages)
R-DATA-OPENML-1462	banknote-authentication.arff
R-DATA-OPENML-40983	wilt.arff
R-DATA-OPENML-1489	phoneme.arff
R-DATA-OPENML-1046	mozilla4.arff
R-DATA-OPENML-151	electricity.arff

Model weights are identified by sha256 and are **not** redistributed.

Protocol bundle: `fea6f25013345eb79eb89da84649d8acf9c15a0c5ab3648032967a0cea875f4a`

S7. Reproduction instructions

```
1. python -m venv .venv && .venv/Scripts/pip install -r configs/requirements-research.lock.txt
2. place the pinned checkpoints and OpenML files (digests in S6); they are verified on load
3. python -c "from tav.config import load_protocol; print(load_protocol(verify=True).bundle_sha256)"
   expected: fea6f25013345eb79eb89da84649d8acf9c15a0c5ab3648032967a0cea875f4a
4. python -m tav.campaign.cli plan --write        # 632 units, plan hash recorded
5. python -m tav.campaign.cli make-jobs           # then launch shards via scripts/job_supervisor.py
6. python -m tav.analysis.cli all                 # frozen analysis, after embargo release
7. python -m tav.publication.cli all              # documents and assets
```

S8. Licence notices

- TabPFN v2 is used under the Prior Labs License; this work is “Built with PriorLabs-TabPFN”.
- TabICL v2 is used under the BSD-3-Clause license.
- OpenML datasets are referenced by id, version and MD5; they are not redistributed here.
- Model weights are not redistributed; they are identified by sha256 and upstream location.