

Comparative Impact of Causal Versus Correlational Synthetic Pretraining on Adversarial Robustness of Tabular Foundation Models

Assignee Research

June 11, 2026

Abstract

The development of tabular foundation models (TFMs) has accelerated in recent years, showing strong potential to outperform traditional ML methods for structured data. A key finding is that TFMs can be pretrained entirely on synthetic datasets, opening opportunities to design data generators that encourage desirable model properties. Prior work has mainly focused on crafting high-quality priors over generators to improve overall pretraining performance. Our insight is that parameterizing the generator distribution enables an adversarial robustness perspective: during training, we can adapt the

1 Introduction

This paper examines: Robust Tabular Foundation Models. Research question: What is the comparative impact of causal versus correlational synthetic data pretraining on the adversarial robustness of tabular foundation models as measured by accuracy degradation on perturbed benchmark subsets?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.2/10.

3 Results

14 papers retrieved. 13 claims extracted; 9 independently verified. Quality review score: 7.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Tabular foundation models (TFMs) rely on in-context learning (ICL) for classification and regression tasks with structure	✓	0.20
TFMs can produce high-quality predictions on new datasets in milliseconds when GPU-accelerated.	✓	0.18
Training TFMs relies on generating large amounts of diverse synthetic datasets constructed from structural causal models	✓	0.20
All current publicly available, competitive TFMs have been pretrained on datasets generated from a fixed prior distribution	✓	0.20
Fixed priors in TFM training underrepresent certain regions of the parameter space, potentially degrading performance on	✓	0.22
State-of-the-art TFMs still lag behind tree-based methods on some benchmarks.	×	0.13
The proposed method, RTFM, was applied to TabPFN V2.	×	0.10
Training TabPFN V2 with RTFM using only 90k additional training datasets significantly improved its ranking on several r	✓	0.18
The maximization stage of the RTFM algorithm freezes the model weights (gW) to maximize the optimality gap.	✓	0.18
The RTFM methodology uses a black-box optimization algorithm to search the SCM parameter space for parameters with large	✓	0.23
In the described implementation, the estimated optimality gap could be computed in a matter of seconds when parallelized	✓	0.18
The benchmark table includes synthetic dataset configurations with activation functions such as tanh, identity, elu, and	×	0.02
The benchmark table includes synthetic dataset configurations with noise distributions including uniform, exponential, a	×	0.03

References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2507.07829v1>
- <http://arxiv.org/abs/2512.03307v1>