

Robustness of Tabular Foundation Models Pretrained on Adversarial Synthetic Datasets versus Traditional Generative Priors

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: How does the robustness of tabular foundation models pretrained on adversarial synthetic datasets compare to those pretrained on traditional generative priors when evaluated on out-of-distribution tabular datasets, as measured by accuracy and calibration metrics?.

2 Methodology

Systematic literature search across multiple databases yielded 10 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

10 papers retrieved. 11 claims extracted; 8 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 structured	✓	0.20
TFMs can produce high-quality predictions on new datasets in milliseconds when GPU-accelerated.	✓	0.17
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 lag behind tree-based methods on some benchmarks.	×	0.14
The proposed method formalizes adversarial training over the SCM parameter space to adapt to challenging regions such as	✓	0.22
The proposed algorithm is named ROBUST TABULAR FOUNDATION MODELS (RTFM).	×	0.11
Applying RTFM to TabPFN V2 with only 90k additional training datasets significantly improves its ranking on several real	✓	0.19
The optimality gap is estimated by sampling a fixed number of generators and datasets and computing the difference between	×	0.10
In the described implementation, the estimated optimality gap could be computed in a matter of seconds when parallelized	✓	0.17

References

- <http://arxiv.org/abs/2311.11096v1>
- <http://arxiv.org/abs/2504.20900v1>
- <http://arxiv.org/abs/2512.03307v1>