Scaling Synthetic Tabular Data for Cross-Domain Foundation Model Generalization

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: Does increasing the scale of synthetic tabular training data improve the cross-domain generalization of foundation models on structured data benchmarks. 15 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Causal Data Augmentation for Robust Fine-Tuning of Tabular Foundation Models. Research question: Does increasing the scale of synthetic tabular training data improve the cross-domain generalization of foundation models on structured data benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

11 papers retrieved. 15 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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
CausalMixFT achieves the highest median improvement of (+0.12$\pm$0.63) over the pre-trained model on 33 classification data	×	0.08
Default fine-tuning has a variability of $\pm$0.98, while CausalMixFT has a variability of $\pm$0.63.	×	0.07
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline.	×	0.07
Purely synthetic generators, including CTGAN, SCM, TabEBM, TableAugment, and Mixed-Model, show negative median improvemen	×	0.08
The experiments were conducted on the Mitra model across 33 classification datasets with 10 folds each, totaling 2,310 f	×	0.12
SCM-based augmentation stabilizes fine-tuning under small-data conditions by introducing causally structured synthetic d	×	0.10
Early stopping based on limited validation data leads to significant validation set overfitting depending on the fine-tu	✓	0.16
The normalization strategy suggested by Gorishniy et al. [12] is used to compare the performance across different data g	×	0.05
The base model's (Mitra's) zero-shot performance is used as the performance baseline.	×	0.05
The normalized performance is computed as scorenormalized = metricsign $\times$ (scoremethod / scorebaseline -1) $\times$ 100%.	×	0.03
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	×	0.05
The PC and FCI algorithms are used to estimate the structural relations between the features.	×	0.03
DoWhy's SCM framework with additive noise models is used to sample and fit DAGs.	×	0.04
Numerical features are modeled with regressors, and categorical features with classifiers.	×	0.04
Synthetic samples are generated by sampling exogenous noise and propagating it through the fitted SCM.	×	0.05

References

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