Scaling Causal Synthetic Data for Tabular Foundation Models on Tab-OOD

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does increasing the size of causally-generated synthetic datasets affect the out-of-distribution accuracy of tabular foundation models on the Tab-OOD benchmark. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. 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: How does increasing the size of causally-generated synthetic datasets affect the out-of-distribution accuracy of tabular foundation models on the Tab-OOD benchmark?.

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 5.2/10.

3 Results

14 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 5.2/10.

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
CausalMixFT achieves the highest median improvement of (+0.12$\pm$0.63) over the pre-trained model, outperforming both the d	×	0.12
Default fine-tuning occasionally achieves higher peak performance on individual datasets, but its variability is substan	×	0.07
CausalMixFT ranks first overall in average ranks and corresponding critical difference (CD) intervals across datasets.	×	0.03
Early stopping based on limited validation data leads to significant validation set overfitting depending on the fine-tu	✓	0.17
The normalization strategy suggested by Gorishniy et al. [12] is used to compare the performance across different data g	×	0.05
CausalMixFT extends the fine-tuning framework of Bhler et al. [5] by mixing real and causally grounded synthetic sample	×	0.12
SCM-Based Synthetic Augmentation (CausalMixFT) uses SCMs fitted to the target dataset, enabling the model to learn joint	×	0.12
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	×	0.15
The structural relations between the features are estimated using the PC and FCI algorithms [34, 35], producing a probab	×	0.02
DAGs are sampled and fitted using DoWhy's SCM framework with additive noise models [32].	×	0.02
Numerical features are modeled with regressors, and categorical features with classifiers in the SCM framework.	×	0.04
The complexity of the internally used model types can be controlled through a quality hyperparameter.	×	0.05
Synthetic samples are generated by sampling exogenous noise and propagating it through the fitted SCM.	×	0.04

References

- http://arxiv.org/abs/2601.04110v2
- http://arxiv.org/abs/2512.03307v1
- http://arxiv.org/abs/2312.01792v1