CausalMixFT Synthetic Data Enhances Few-Shot Learning in Tabular Foundation Models

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: Can the synthetic samples generated by CausalMixFT improve the few-shot learning capabilities of tabular foundation models on the TabularGLUE benchmarks, and how does this compare to zero-shot. 17 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: Can the synthetic samples generated by CausalMixFT improve the few-shot learning capabilities of tabular foundation models on the TabularGLUE benchmarks, and how does this compare to zero-shot performance measured by accuracy and AUC-ROC?.

2 Methodology

Systematic literature search across multiple databases yielded 13 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

13 papers retrieved. 17 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
Experiments were conducted on the Mitra model across 33 classification datasets from the TabArena benchmark suite.	×	0.08
The experimental setup involved 10 folds for each of the 33 datasets, totaling 2,310 fine-tuning runs.	×	0.10
Model performance is reported as normalized ROC-AUC relative to the pre-trained model.	×	0.07
CausalMixFT achieved a median improvement of +0.12$\pm$0.63 over the pre-trained model.	×	0.05
The default fine-tuning baseline achieved a median improvement of +0.10$\pm$0.98 over the pre-trained model.	×	0.09
Purely synthetic augmentation methods (CT-GAN, SCM, TabEBM, TableAugment, and MixedModel) showed negative median improvem	×	0.08
CausalMixFT demonstrated lower performance variability ($\pm$0.63) compared to the default fine-tuning baseline ($\pm$0.98).	×	0.08
In average rank analysis across datasets, CausalMixFT ranked first overall.	×	0.04
The normalization strategy uses the base model's (Mitra's) zero-shot performance as the baseline.	×	0.10
The normalization formula is defined as: score_normalized = metricsign $\times$ (score_method / score_baseline - 1) $\times$ 100%.	×	0.01
In the normalization formula, metricsign is set to 1 for metrics where higher is better (e.g., ROC-AUC) and -1 for metri	×	0.03
The method generates synthetic data using Structural Causal Models (SCMs) fitted to the target dataset.	✓	0.22
Structural relations between features are estimated using the PC and FCI algorithms.	×	0.04
The estimation of structural relations produces a probabilistic adjacency matrix encoding edge strengths between variabl	×	0.01
DAGs are sampled and fitted using DoWhy's SCM framework with additive noise models.	×	0.03
In the SCM framework, 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/2112.10006v6
- http://arxiv.org/abs/2311.14544v1