

How does the F1-score performance of tabular foundation models (TFMs) fine-tuned on diffusion-generated synthetic

Assignee Research

June 10, 2026

Abstract

Fine-tuning tabular foundation models (TFMs) under data scarcity is challenging, as early stopping on even scarcer validation data often fails to capture true generalization performance. We propose CausalMixFT, a method that enhances fine-tuning robustness and downstream performance by generating structurally consistent synthetic samples using Structural Causal Models (SCMs) fitted on the target dataset. This approach augments limited real data with causally informed synthetic examples, preserving feature dependencies while expanding training diversity. Evaluated across 33 classification datasets

1 Introduction

This paper examines: Causal Data Augmentation for Robust Fine-Tuning of Tabular Foundation Models. Research question: How does the F1-score performance of tabular foundation models (TFMs) fine-tuned on diffusion-generated synthetic data compare to those fine-tuned on SCM-generated synthetic data in imbalanced classification tasks, as evaluated on the Adult and Credit datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.0/10.

3 Results

12 papers retrieved. 10 claims extracted; 1 independently verified. Quality review score: 5.0/10.

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	×	0.09
Default fine-tuning has a variability of ± 0.98 , while CausalMixFT has a variability of ± 0.63 , indicating greater stability	×	0.08
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline.	×	0.08
Purely synthetic generators, including CTGAN, SCM, TabEBM, TableAugment, and Mixed-Model, show negative median improvement	×	0.08
The normalization strategy used to compare performance across different data generators is based on Gorishniy et al. [12]	×	0.06
CausalMixFT extends the fine-tuning framework by mixing real and causally grounded synthetic samples, using SCMs fitted	✓	0.19
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	×	0.15
The structural relations between features are estimated using the PC and FCI algorithms, producing a probabilistic adjac	×	0.03
DAGs are sampled and fitted using DoWhy’s SCM framework with additive noise models.	×	0.02
Numerical features are modeled with regressors, and categorical features with classifiers in the SCM framework.	×	0.04

References

- <http://arxiv.org/abs/2502.17119v2>
- <http://arxiv.org/abs/2401.00965v1>
- <http://arxiv.org/abs/2601.04110v2>