Performance Comparison of CausalMixFT with SMOTE and MixUp in Tabular Foundation Model Fine-Tuning

Assignee Research

June 11, 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 datas

1 Introduction

This paper examines: Causal Data Augmentation for Robust Fine-Tuning of Tabular Foundation Models. Research question: How does the performance of CausalMixFT compare to other data augmentation techniques like SMOTE or MixUp in fine-tuning tabular foundation models, as measured by validation accuracy and downstream task generalization on TabularGLUE benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

3 Results

7 papers retrieved. 13 claims extracted; 9 independently verified. Quality review score: 7.5/10.

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.	✓	0.23
Default fine-tuning has a variability of $\pm$0.98, while CausalMixFT has a variability of $\pm$0.63.	×	0.15
CausalMixFT ranks first overall in average ranks across datasets.	✓	0.15
The experiments were conducted on the Mitra model across 33 classification datasets with 10 folds each from the TabArena	✓	0.26
SCM-based augmentation stabilizes fine-tuning under small-data conditions by introducing causally structured synthetic d	✓	0.28
The normalization strategy suggested by Gorishniy et al. [12] is used to compare the performance across different data g	×	0.06
The base model's (Mitra's) zero-shot performance is used as the performance baseline.	×	0.06
The normalized performance is computed as scorenormalized = metricsign $\times$ (scoremethod / scorebaseline -1) $\times$ 100%.	×	0.02
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	✓	0.27
The PC and FCI algorithms are used to estimate the structural relations between the features.	✓	0.17
DoWhy's SCM framework with additive noise models is used to sample and fit DAGs.	✓	0.16
Numerical features are modeled with regressors, and categorical features with classifiers.	✓	0.18
Synthetic samples are generated by sampling exogenous noise and propagating it through the fitted SCM.	✓	0.22

References

- http://arxiv.org/abs/2508.11281v3
- http://arxiv.org/abs/2601.04110v2

- http://arxiv.org/abs/2504.20900v1