Impact of Structural Causal Model Complexity on CausalMixFT Generalization to Out-of-Distribution Tabular Data

Assignee Research

June 12, 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: To what extent does the causal structure complexity (e.g., number of confounders or mediators) in the SCM used for CausalMixFT affect the generalization performance on OOD tabular datasets like CovidShift or WMTD?.

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 8.5/10.

3 Results

14 papers retrieved. 10 claims extracted; 9 independently verified. Quality review score: 8.5/10.

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 pretrained model on 33 classification data	✓	0.27
Default fine-tuning has a variability of $\pm$0.98, while CausalMixFT has a variability of $\pm$0.63, indicating greater instabi	✓	0.19
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline, with pur	✓	0.26
Early stopping based on limited validation data leads to significant validation set overfitting depending on the fine-tu	✓	0.26
The normalization strategy for performance comparison is defined as scorenormalized = metricsign $\times$ (scoremethod / scoreb	×	0.04
CausalMixFT extends the fine-tuning framework by mixing real and causally grounded synthetic samples, using SCMs fitted	✓	0.25
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	✓	0.26
The PC and FCI algorithms are used to estimate the structural relations between features, producing a probabilistic adja	✓	0.21
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 in the SCM framework.	✓	0.20

References

- http://arxiv.org/abs/2001.04197v4
- http://arxiv.org/abs/1905.11374v5
- http://arxiv.org/abs/2601.04110v2