

# Causal Structure Errors in Fine-Tuned Tabular Foundation Models and OOD Classification Performance

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 datasets

## 1 Introduction

This paper examines: Causal Data Augmentation for Robust Fine-Tuning of Tabular Foundation Models. Research question: What is the impact of causal structure discovery errors on the downstream classification performance of fine-tuned tabular foundation models when evaluated on out-of-distribution splits?

## 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 7.7/10.

## 3 Results

13 papers retrieved. 10 claims extracted; 8 independently verified. Quality review score: 7.7/10.

## 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	✓	0.27
Default fine-tuning has a variability of $\pm 0.98$ , while CausalMixFT has a variability of $\pm 0.63$ , indicating greater stability	×	0.13
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline.	✓	0.23
Early stopping based on limited validation data leads to significant validation set overfitting depending on the fine-tuning	✓	0.26
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 of Böhler et al. [5] by mixing real and causally grounded synthetic samples	✓	0.26
SCM-Based Synthetic Augmentation (CausalMixFT) uses SCMs fitted to the target dataset to generate synthetic data.	✓	0.18
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 in CausalMixFT.	✓	0.16
DoWhy’s SCM framework with additive noise models is used to sample and fit DAGs in CausalMixFT.	✓	0.16

## References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2410.19412v3>
- <http://arxiv.org/abs/2512.03307v1>