

Causal Structure Complexity and Cross-Domain Robustness in Tabular Foundation Models

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: How do varying levels of causal structure complexity in synthetic data generated by SCMs affect the cross-domain robustness of tabular foundation models when evaluated on F1 scores across multiple. 11 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.8/10. 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: How do varying levels of causal structure complexity in synthetic data generated by SCMs affect the cross-domain robustness of tabular foundation models when evaluated on F1 scores across multiple low-resource datasets?.

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 4.8/10.

3 Results

7 papers retrieved. 11 claims extracted; 2 independently verified. Quality review score: 4.8/10.

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
CausalMixFT outperforms the default fine-tuning baseline $(+0.10 \pm 0.98)$ and all purely synthetic augmentation methods, in	×	0.11
Default fine-tuning has a variability of ± 0.98 , while CausalMixFT has a variability of ± 0.63 , indicating greater instability	×	0.08
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline, with pur	×	0.07
Early stopping based on limited validation data leads to significant validation set overfitting depending on the fine-tu	✓	0.15
The normalization strategy for performance comparison is based on the zero-shot performance of the Mitra model, with nor	×	0.04
CausalMixFT extends the fine-tuning framework by mixing real and causally grounded synthetic samples, using SCMs fitted	✓	0.17
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	×	0.14
The PC and FCI algorithms are used to estimate structural relations between features, producing a probabilistic adjacenc	×	0.02
DAGs are sampled and fitted using DoWhy’s SCM framework with additive noise models, with numerical features modeled with	×	0.04
Synthetic samples are generated by sampling exogenous noise and propagating it through the fitted SCM.	×	0.04

References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2603.10254v1>
- <http://arxiv.org/abs/2507.03528v1>