

# Impact of Synthetic Data Ratios from SCMs on Robustness in CausalMixFT Fine-Tuning

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: What is the impact of varying the synthetic data ratio generated by SCMs in CausalMixFT on the robustness of fine-tuned TFMs across domain-shifted tabular datasets, evaluated using metrics like test. 17 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. 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: What is the impact of varying the synthetic data ratio generated by SCMs in CausalMixFT on the robustness of fine-tuned TFMs across domain-shifted tabular datasets, evaluated using metrics like test AUC and calibration error?.

## 2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

## 3 Results

9 papers retrieved. 17 claims extracted; 1 independently verified. Quality review score: 4.2/10.

## 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
Experiments were conducted on the Mitra model across 33 classification datasets with 10 folds each from the TabArena ben	×	0.07
The study totaled 2,310 fine-tuning runs.	×	0.11
Model performance is reported as normalized ROC-AUC relative to the pre-trained model.	×	0.07
CausalMixFT achieves a median improvement of $+0.12 \pm 0.63$ over the pre-trained model.	×	0.05
The default fine-tuning baseline achieves a median improvement of $+0.10 \pm 0.98$ over the pre-trained model.	×	0.09
Purely synthetic augmentation methods (CTGAN, SCM, TabEBM, TableAugment, and MixedModel) show negative median improvement	×	0.08
CausalMixFT has a performance variability of $\pm 0.63$ , while default fine-tuning has a variability of $\pm 0.98$ .	×	0.10
In average rank analysis across datasets, CausalMixFT ranks first overall.	×	0.03
The normalization strategy uses the base model's (Mitra's) zero-shot performance as the baseline.	×	0.03
The normalization formula is defined as: $\text{score\_normalized} = \text{metric\_sign} \times (\text{score\_method} / \text{score\_baseline} - 1) \times 100\%$ .	×	0.00
The metric_sign is set to 1 for metrics where higher is better (e.g., ROC-AUC) and -1 for metrics where lower is better	×	0.05
The method generates synthetic data using Structural Causal Models (SCMs) fitted to the target dataset.	✓	0.22
Structural relations between features are estimated using the PC and FCI algorithms.	×	0.04
The estimation of structural relations produces a probabilistic adjacency matrix encoding edge strengths between variabl	×	0.02
DAGs are sampled and fitted using DoWhy's SCM framework with additive noise models.	×	0.03
In the SCM framework, numerical features are modeled with regressors and categorical features with classifiers.	×	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/2406.04295v2>
- <http://arxiv.org/abs/2603.10254v1>