CausalMixFT Robustness and Generalization in Low-Data OpenML-CC18 Domains

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: Does the robustness of CausalMixFT in low-data regimes translate to improved generalization across diverse domains in OpenML-CC18 when compared to traditional data augmentation techniques like SMOTE. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. 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: Does the robustness of CausalMixFT in low-data regimes translate to improved generalization across diverse domains in OpenML-CC18 when compared to traditional data augmentation techniques like SMOTE and GAN-based methods?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

15 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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 on 33 classification data	×	0.09
Default fine-tuning has a variability of $\pm$0.98, while CausalMixFT has a variability of $\pm$0.63.	×	0.10
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline.	×	0.09
Purely synthetic generators, including CTGAN, SCM, TabEBM, TableAugment, and Mixed-Model, show negative median improvemen	×	0.08
Early stopping based on limited validation data leads to significant validation set overfitting depending on the fine-tu	✓	0.17
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.	×	0.03
The normalized performance is computed as scorenormalized = metricsign $\times$ (scoremethod / scorebaseline -1) $\times$ 100%.	×	0.03
SCM-Based Synthetic Augmentation (CausalMixFT) explicitly encodes causal dependencies among features through a directed	×	0.14
The structural relations between the features are estimated using the PC and FCI algorithms.	×	0.03
DAGs are sampled and fitted using DoWhy's SCM framework with additive noise models.	×	0.03
Numerical features are modeled with regressors, and categorical features with classifiers.	×	0.03
Synthetic samples are generated by sampling exogenous noise and propagating it through the fitted SCM.	×	0.05

References

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

- http://arxiv.org/abs/2412.00381v1
- http://arxiv.org/abs/2603.10254v1