CausalMixFT Integration with Transformer Architectures for Tabular Benchmark Performance

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the integration of CausalMixFT with different Transformer architectures affect the fine-tuning performance on tabular benchmarks like TabLM-2B or TabBench, measured by validation accuracy. 10 claims were extracted from source literature; 0 were 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: How does the integration of CausalMixFT with different Transformer architectures affect the fine-tuning performance on tabular benchmarks like TabLM-2B or TabBench, measured by validation accuracy and training convergence time?.

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 4.5/10.

3 Results

13 papers retrieved. 10 claims extracted; 0 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	×	0.10
Default fine-tuning has a variability of $\pm$0.98, while CausalMixFT has a variability of $\pm$0.63, indicating greater instabi	×	0.07
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline, with pur	×	0.07
The performance normalization strategy used is scorenormalized = metricsign $\times$ (scoremethod / scorebaseline -1) $\times$ 100%, w	×	0.04
CausalMixFT extends the fine-tuning framework by mixing real and causally grounded synthetic samples into the fine-tunin	×	0.12
SCM-Based Synthetic Augmentation (CausalMixFT) uses SCMs fitted to the target dataset, enabling the model to learn joint	×	0.12
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	×	0.04
The structural relations between the features are estimated using the PC and FCI algorithms, producing a probabilistic a	×	0.05
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 in the SCM framework.	×	0.04

References

- http://arxiv.org/abs/2601.04110v2
- http://arxiv.org/abs/2602.09439v1
- http://arxiv.org/abs/2110.06500v2