

CausalMixFT and Traditional Augmentation Effects on TabLLM Calibration and Robustness

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the impact of CausalMixFT versus traditional data augmentation on the calibration error and robustness of TabLLM under varying degrees of label noise. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.6/10. 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 CausalMixFT versus traditional data augmentation on the calibration error and robustness of TabLLM under varying degrees of label noise?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.6/10.

3 Results

12 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 4.6/10.

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 ± 0.98 , while CausalMixFT has a variability of ± 0.63 , indicating greater stability	×	0.07
CausalMixFT ranks first overall in average ranks across datasets, followed by the default fine-tuning baseline.	×	0.06
Purely synthetic augmentation methods, including CTGAN, SCM, TabEBM, TableAugment, and MixedModel, show negative median	×	0.08
Early stopping based on limited validation data leads to significant validation set overfitting depending on the fine-tuning	✓	0.16
The normalization strategy used to compare performance across different data generators is based on Gorishniy et al. [12]	×	0.05
CausalMixFT extends the fine-tuning framework of Böhler et al. [5] by mixing real and causally grounded synthetic sample	×	0.11
SCM-Based Synthetic Augmentation (CausalMixFT) uses SCMs fitted to the target dataset, enabling the model to learn joint	×	0.11
SCMs explicitly encode causal dependencies among features through a directed acyclic graph (DAG) and a set of structural	×	0.05
The structural relations between the features are estimated using the PC and FCI algorithms, producing a probabilistic a	×	0.02
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.	×	0.01
Synthetic samples are generated by sampling exogenous noise and propagating it through the fitted SCM.	×	0.06

References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2406.16966v1>
- <http://arxiv.org/abs/2110.06500v2>