What is the impact of varying levels of class skew in synthetic data augmentation on the F1-score stability of

Assignee Research

June 10, 2026

Abstract

Fine-tuning tabular foundation models (TFMs) under data scarcity is challenging, as early stopping on even scarcer validation data often fails to capture true generalization performance. We propose CausalMixFT, a method that enhances fine-tuning robustness and downstream performance by generating structurally consistent synthetic samples using Structural Causal Models (SCMs) fitted on the target dataset. This approach augments limited real data with causally informed synthetic examples, preserving feature dependencies while expanding training diversity. Evaluated across 33 classification datas

1 Introduction

This paper examines: Causal Data Augmentation for Robust Fine-Tuning of Tabular Foundation Models. Research question: What is the impact of varying levels of class skew in synthetic data augmentation on the F1-score stability of tabular foundation models (TFMs) when fine-tuned and evaluated on the UCI ML datasets (e.g., Breast Cancer, Heart Disease) compared to the original real data?.

2 Methodology

Systematic literature search across multiple databases yielded 10 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

10 papers retrieved. 14 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
Default fine-tuning has a median improvement of (+0.10 $\pm$0.98) over the pre-trained model on the same datasets.	×	0.08
Purely synthetic augmentation methods (CT-GAN, SCM, TabEBM, TableAugment, Mixed-Model) show negative median improvements o	×	0.07
CausalMixFT has a variability of $\pm$0.63, while default fine-tuning has a variability of $\pm$0.98, indicating greater stabili	×	0.07
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	✓	0.16
The normalization strategy used to compare performance across different data generators is based on Gorishniy et al. [12	×	0.07
CausalMixFT extends the fine-tuning framework of Bhler et al. [5] by mixing real and causally grounded synthetic sample	×	0.11
SCM-Based Synthetic Augmentation (CausalMixFT) uses Structural Causal Models (SCMs) fitted to the target dataset to gene	✓	0.18
SCMs in CausalMixFT encode causal dependencies among features through a directed acyclic graph (DAG) and a set of struct	×	0.05
The PC and FCI algorithms are used to estimate structural relations between features in CausalMixFT.	×	0.03
DoWhy's SCM framework with additive noise models is used to sample and fit DAGs in CausalMixFT.	×	0.02
Numerical features in CausalMixFT are modeled with regressors, and categorical features with classifiers.	×	0.04
Synthetic samples in CausalMixFT are generated by sampling exogenous noise and propagating it through the fitted SCM.	×	0.05

References

- http://arxiv.org/abs/1910.13757v1
- http://arxiv.org/abs/2512.03307v1
- http://arxiv.org/abs/2601.04110v2