

CausalMixFT Robustness Across Domains via Synthetic Data Fine-Tuning

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: Can the robustness improvements from CausalMixFT be replicated in cross-domain settings by evaluating synthetic data-aided fine-tuning on both OpenML-CC18 and ImageNet-C benchmarks. 14 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Benchmarking Distribution Shift in Tabular Data with TableShift. Research question: Can the robustness improvements from CausalMixFT be replicated in cross-domain settings by evaluating synthetic data-aided fine-tuning on both OpenML-CC18 and ImageNet-C benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

3 Results

16 papers retrieved. 14 claims extracted; 0 independently verified. Quality review score: 3.2/10.

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
Food insecurity affects more than 10% of households (13.5 million) across the United States in 2021.	×	0.01
The American Community Survey (ACS) is used as the data source for the Food Stamps task.	×	0.03
The Food Stamps task filters data for low-income adults aged 18-62 in households with at least one child.	×	0.02
An income threshold of \$30000 is used based on the U.S. poverty threshold for a family with one child.	×	0.04
The ACS includes 10 regions: Puerto Rico; New England (Northeast region); Middle Atlantic (Northeast region); East North	×	0.01
The East South Central (South region) is used as the holdout domain for the Food Stamps task.	×	0.03
The ASSISTments Next Answer Correct task shows a Δ of -34.49% with a p-value of 0.011.	×	0.01
The College Scorecard Low Degree Completion Rate task shows a Δ of -11.16% with a p-value of 0.010.	×	0.02
The ICU Hospital Mortality task shows a Δ of -6.30% with a p-value of 0.008.	×	0.01
The Hospital Readmission task shows a Δ of -5.94% with a p-value of 0.002.	×	0.01
The Diabetes task shows a Δ of -4.48% with a p-value of 0.001.	×	0.01
The ICU Length of Stay task shows a Δ of -3.39% with a p-value of 0.015.	×	0.01
The Voting task shows a Δ of -2.58% with a p-value of 0.016.	×	0.01
The Food Stamps task shows a Δ of -4.48% with a p-value of 0.001.	×	0.01

References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2403.10075v2>

- <http://arxiv.org/abs/2312.07577v3>