

Synthetic-Pretrained Tabular Foundation Models Under Distribution Shift in TabBench

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the performance degradation of synthetic-pretrained tabular foundation models under distribution shift scenarios in the TabBench robustness benchmarks. 15 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. 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: What is the performance degradation of synthetic-pretrained tabular foundation models under distribution shift scenarios in the TabBench robustness benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

14 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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 affected more than 10% of households (13.5 million) across the United States in 2021.	×	0.02
The relationship between other forms of shift showed a correlation of roughly $\rho \approx -0.2$.	×	0.03
TableShift does not host or distribute the data; each data source is publicly available.	×	0.04
The Food Stamps task uses person-level data from the American Community Survey (ACS).	×	0.03
The Food Stamps dataset filters for low-income adults aged 18-62 in households with at least one child.	×	0.01
The income threshold used for the Food Stamps task is \$30,000.	×	0.03
The ACS includes 10 regions used for domain shift in the Food Stamps task.	×	0.05
East South Central is used as the holdout domain for the Food Stamps task.	×	0.02
The ASSISTments Next Answer Correct task shows a performance drop of -34.49% with a shift based on School.	×	0.02
The College Scorecard Low Degree Completion Rate task shows a performance drop of -11.16% with a shift based on Institut	×	0.02
The ICU Hospital Mortality task shows a performance drop of -6.30% with a shift based on Insurance Type.	×	0.02
The Hospital Readmission task shows a performance drop of -5.94% with a shift based on Admission source.	×	0.03
The Diabetes diagnosis task shows a performance drop of -4.48% with a shift based on Race.	×	0.02
The ICU Length of Stay task shows a performance drop of -3.39% with a shift based on Insurance Type.	×	0.02
The Voting task shows a performance drop of -2.58% with a shift based on Geographic Region.	×	0.02

References

- <http://arxiv.org/abs/2504.20900v1>
- <http://arxiv.org/abs/2512.03307v1>
- <http://arxiv.org/abs/2312.07577v3>