

Multimodal vs. Text-Only Models in Robustness to Distribution Shift on TableShift

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How do multimodal models incorporating tabular and textual data compare to text-only models in robustness to distribution shift, as measured by accuracy on cross-modal benchmarks like TableShift. 6 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: How do multimodal models incorporating tabular and textual data compare to text-only models in robustness to distribution shift, as measured by accuracy on cross-modal benchmarks like TableShift?.

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 3.7/10.

3 Results

12 papers retrieved. 6 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 affects more than 10% of households (13.5 million) across the United States in 2021.	×	0.00
The data source for the Food Stamps task is the American Community Survey (ACS).	×	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.	×	0.02
The TableShift benchmark includes tasks such as ASSISTments Next Answer Correct, College Scorecard Low Degree Completion	×	0.04
The models evaluated in TableShift include XGBoost, LightGBM, MLP, FT-Transformer, ResNet, SAINT, NODE, TabTransformer,	×	0.03

References

- <http://arxiv.org/abs/2309.10783v1>
- <http://arxiv.org/abs/2502.17119v2>
- <http://arxiv.org/abs/2312.07577v3>