

# Generative AI-Augmented Tabular Data for Robust Multimodal Classification Under Distribution Shifts

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the impact of generative AI-based tabular data augmentation on the robustness of multimodal classifiers against distribution shifts in OpenML-CC18 datasets. 7 claims were extracted from source literature; 7 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Thinking Beyond Distributions in Testing Machine Learned Models. Research question: What is the impact of generative AI-based tabular data augmentation on the robustness of multimodal classifiers against distribution shifts in OpenML-CC18 datasets?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.7/10.

## 3 Results

13 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 8.7/10.

## 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
Testing practices within the machine learning (ML) community have centered around assessing a learned model's predictive	✓	0.48
Recent work on robustness and fairness testing within the ML community has pointed to the importance of testing against	✓	0.39
These efforts also focus on estimating the likelihood of the model making an error against a reference dataset/distribut	✓	0.30
The view of testing that focuses solely on estimating the likelihood of the model making an error against a reference da	✓	0.39
There are other sources of robustness failures, for instance corner cases which may have severe undesirable impacts.	✓	0.32
Decades of work within software engineering testing has focused on assessing a software system against various stress co	✓	0.48
The authors put forth a set of recommendations to broaden the view of machine learning testing to a rigorous practice.	✓	0.32

## References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2112.03057v1>
- <http://arxiv.org/abs/2502.17119v2>