

# Synthetic Tabular Data Diversity and Cross-Domain Generalization Performance

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does increasing the diversity of synthetic tabular pretraining data affect cross-domain generalization accuracy compared to scaling dataset size alone. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: A Comprehensive Benchmark of Machine and Deep Learning Across Diverse Tabular Datasets. Research question: How does increasing the diversity of synthetic tabular pretraining data affect cross-domain generalization accuracy compared to scaling dataset size alone?.

## 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 5.2/10.

## 3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.2/10.

## 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.

## References

- <http://arxiv.org/abs/2306.11636v1>
- <http://arxiv.org/abs/2507.05904v1>
- <http://arxiv.org/abs/2408.14817v1>