

Novel Structural Fidelity Metrics and Their Correlation with Downstream Classifier Performance in Synthetic Tabular Data

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How do novel evaluation metrics for structural fidelity correlate with downstream classifier performance when assessing synthetic tabular data from CausalMixFT and diffusion baselines. 9 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Evaluating Generative Models for Tabular Data: Novel Metrics and Benchmarking. Research question: How do novel evaluation metrics for structural fidelity correlate with downstream classifier performance when assessing synthetic tabular data from CausalMixFT and diffusion baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.8/10.

3 Results

9 papers retrieved. 9 claims extracted; 3 independently verified. Quality review score: 5.8/10.

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
FAED effectively captures generative modeling issues overlooked by existing metrics.	✓	0.29
FPCAD exhibits promising performance but requires further refinements to enhance its reliability.	✓	0.18
FAED successfully detects all synthesized problems in the experimental results on three standard network intrusion detection datasets.	×	0.14
FPCAD shows promising but improvable performance in detecting synthesized problems.	×	0.04
Existing metrics fail to identify key issues in generative modeling for tabular data.	×	0.14
FAED and FPCAD are novel and robust metrics for evaluating generative models in the tabular data domain.	✓	0.29
The study simulates real-world generative modeling challenges by embedding Quality Decrease, Mode Drop, and Mode Collaps.	×	0.05
TSTR is particularly useful for detecting cases where synthetic data only partially represents real data.	×	0.05
TRTS assesses whether synthetic samples introduce patterns absent in real data.	×	0.06

References

- <http://arxiv.org/abs/2601.04110v2>
- <http://arxiv.org/abs/2504.20900v1>
- <http://arxiv.org/abs/2512.03307v1>