Scalability and Performance Trade-offs of BE-SNNs vs. Deterministic Models for Large-Scale OOD Detection

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the scalability of BE-SNNs for OOD detection compare to single deterministic models when applied to large-scale tabular datasets (e.g., >1M samples) in terms of training time and memory. 14 claims were extracted from source literature; 6 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.4/10. 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 does the scalability of BE-SNNs for OOD detection compare to single deterministic models when applied to large-scale tabular datasets (e.g., >1M samples) in terms of training time and memory efficiency, while maintaining competitive AUC-ROC performance?.

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 6.4/10.

3 Results

12 papers retrieved. 14 claims extracted; 6 independently verified. Quality review score: 6.4/10.

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
The experimental analysis was conducted on three standard network intrusion detection datasets.	✓	0.24
The study compares proposed metrics against established evaluation methods including Fidelity, Utility, TSTR, and TRTS.	✓	0.21
FAED effectively captures generative modeling issues that are overlooked by existing metrics.	✓	0.30
FPCAD exhibits promising performance but requires further refinements to enhance reliability.	✓	0.20
The study introduces three novel evaluation metrics named FAED, FPCAD, and RFIS for tabular data.	✓	0.23
The study simulates three specific challenges in real datasets: Quality Decrease, Mode Drop, and Mode Collapse.	×	0.03
Experimental results show that FAED successfully detects all synthesized problems (Quality Decrease, Mode Drop, and Mode	×	0.04
Existing metrics (SDV Fidelity, Utility, TSTR, TRTS) fail to identify key generative modeling issues in the conducted ex	✓	0.18
Inception Score (IS) and Frchet Inception Distance (FID) are standard quantitative metrics for evaluating generative mo	×	0.12
TSTR involves training a classifier on synthetic data and testing it on real data.	×	0.04
TRTS involves training a classifier on real data and testing it on synthetic data.	×	0.04
A high TSTR accuracy suggests that synthetic data effectively approximates real-world distributions.	×	0.04
A high TRTS score indicates that the synthetic data retains key characteristics of the real data.	×	0.03
The utility quantification in the study uses a Random Forest model for classifier-based evaluations.	×	0.02

References

- http://arxiv.org/abs/2506.16791v4
- http://arxiv.org/abs/2504.20900v1
- http://arxiv.org/abs/2509.09030v2