Flow-Matching Synthetic Data Augmentation for Imbalanced Tabular Classification

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: What is the impact of using flow-matching generated synthetic samples on classification recall for imbalanced tabular datasets compared to VAE and GAN baselines. 7 claims were extracted from source literature; 7 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.1/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: CTGAN-MOS: Conditional Generative Adversarial Network Based Minority-Class-Augmented Oversampling Scheme for Imbalanced Problems. Research question: What is the impact of using flow-matching generated synthetic samples on classification recall for imbalanced tabular datasets compared to VAE and GAN baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.1/10.

3 Results

4 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 6.1/10.

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
CTGAN-MOS is a novel data augmentation scheme for solving class imbalance problems.	✓	0.25
CTGAN-MOS encompasses six key steps: data engineering, identifying vulnerabilities, curating synthetic data using CTGAN,	✓	0.25
CTGAN-MOS maintains higher structural similarity between the original and the resampled data by adding high-quality samp	✓	0.33
CTGAN-MOS removes noisy samples from the data, which has remained unexplored in the CTGAN-based data augmentation.	✓	0.26
CTGAN-MOS augments data by adding fewer records compared to existing schemes, while offering comparable performance.	✓	0.26
Experiments are conducted on benchmark datasets to prove the feasibility of the proposed CTGAN-MOS in realistic scenario	✓	0.24
CTGAN-MOS has yielded accuracy values of 100% in experiments.	✓	0.18

References

- https://doi.org/10.1109/access.2023.3303509
- https://doi.org/10.3390/electronics13173509
- https://doi.org/10.1186/s13040-024-00366-0