Diffusion-Based Tabular Generative Models vs. CTGAN and SMOTE in Cross-Domain Adaptation

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How does the performance of diffusion-based tabular generative models compare to CTGAN and SMOTE in cross-domain adaptation tasks, as measured by downstream classification F1-scores on imbalanced. 10 claims were extracted from source literature; 10 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.9/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. Research question: How does the performance of diffusion-based tabular generative models compare to CTGAN and SMOTE in cross-domain adaptation tasks, as measured by downstream classification F1-scores on imbalanced datasets?.

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 7.9/10.

3 Results

4 papers retrieved. 10 claims extracted; 10 independently verified. Quality review score: 7.9/10.

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
Class imbalance (CI) in classification problems arises when the number of observations belonging to one class is lower t	✓	0.34
Ensemble learning combines multiple models to obtain a robust model and has been prominently used with data augmentation	✓	0.42
In the last decade, a number of strategies have been added to enhance ensemble learning and data augmentation methods, a	✓	0.37
A combination of ensemble learning and data augmentation methods has been applied in many studies.	✓	0.25
The evaluation of different combinations of ensemble learning and data augmentation methods would enable a better unders	✓	0.35
The paper presents a computational study to evaluate data augmentation and ensemble learning methods used to address pro	✓	0.36
The paper presents a general framework that evaluates 9 data augmentation and 9 ensemble learning methods for CI problem	✓	0.31
The objective of the paper is to identify the most effective combination for improving classification performance on imb	✓	0.29
The results indicate that combinations of data augmentation methods with ensemble learning can significantly improve cla	✓	0.40
Traditional data augmentation methods such as the synthetic minority oversampling technique (SMOTE) and random oversampl	✓	0.43

References

- https://doi.org/10.48550/arxiv.2304.02858
- https://doi.org/10.3390/make5010019
- https://doi.org/10.1186/s40537-021-00414-0