

Can the conditional input methodology from ciDATGAN be extended to multimodal datasets, and how does it affect

Assignee Research

June 10, 2026

Abstract

Conditionality has become a core component for Generative Adversarial Networks (GANs) for generating synthetic images. GANs are usually using latent conditionality to control the generation process. However, tabular data only contains manifest variables. Thus, latent conditionality either restricts the generated data or does not produce sufficiently good results. Therefore, we propose a new methodology to include conditionality in tabular GANs inspired by image completion methods. This article presents ciDATGAN, an evolution of the Directed Acyclic Tabular GAN (DATGAN) that has already been sh

1 Introduction

This paper examines: ciDATGAN: Conditional Inputs for Tabular GANs. Research question: Can the conditional input methodology from ciDATGAN be extended to multimodal datasets, and how does it affect downstream task performance (e.g., accuracy in classification tasks) compared to baseline GANs?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.9/10.

3 Results

13 papers retrieved. 15 claims extracted; 0 independently verified. Quality review score: 3.9/10.

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
ciDATGAN was trained on the LPMC dataset five times.	×	0.03
DATGAN was trained on the LPMC dataset five times.	×	0.03
ciDATGAN generated five different synthetic datasets.	×	0.10
DATGAN generated five different synthetic datasets.	×	0.10
The LPMC dataset consists of 18 variables and 16,904 trips.	×	0.02
The conditional inputs for ciDATGAN are the variables age, female, and hh_region.	×	0.10
The SRMSE metric was used to assess the results.	×	0.07
ciDATGAN can be used to generate large synthetic populations.	×	0.10
ciDATGAN can correct bias using unbiased conditional inputs.	×	0.10
ciDATGAN can generate data representative of the original data.	×	0.05
ciDATGAN uses a DAG for the LPMC dataset as shown in Figure 15.	×	0.04
ciDATGAN uses frequency lists for each variable in the synthetic dataset and compares it to similar lists computed on th	×	0.03
ciDATGAN uses frequency lists for each combination of two variables for the synthetic and original datasets.	×	0.06
ciDATGAN uses a LightGBM model to predict columns in the original data.	×	0.04
ciDATGAN can learn from a feeder dataset and efficiently generate a population using a large sample for the conditional	×	0.15

References

- <http://arxiv.org/abs/2510.21391v1>
- <http://arxiv.org/abs/2104.09630v2>
- <http://arxiv.org/abs/2210.02404v1>