Counterfactual Explanation Integration Enhances Causal Model Generalization on Tabular Data

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: Can counterfactual explanation integration improve the generalization of causal models on out-of-distribution tabular datasets like CinC. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: CLEAR: Generative Counterfactual Explanations on Graphs. Research question: Can counterfactual explanation integration improve the generalization of causal models on out-of-distribution tabular datasets like CinC?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.5/10.

3 Results

16 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 2.5/10.

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 validity of CLEAR is 1.	×	0.04
CLEAR achieves a validity of 0.94 $\pm$ 0.02 on the Community dataset.	×	0.03
CLEAR achieves a proximityX of 0.91 $\pm$ 0.01 on the Community dataset.	×	0.02
CLEAR achieves a proximityA of 0.77 $\pm$ 0.00 on the Community dataset.	×	0.02
CLEAR achieves a causality of 0.65 $\pm$ 0.03 on the Community dataset.	×	0.04
CLEAR achieves a time of 0.01 $\pm$ 0.01 on the Community dataset.	×	0.05
The validity of CLEAR-NC degrades dramatically due to the lack of counterfactual prediction loss.	×	0.04
The performance w.r.t. proximity is worse in CLEAR-NPA, CLEAR-NPX, and CLEAR-NP as the similarity loss is removed.	×	0.02
Removing the similarity loss can also hurt the performance of causality when the variables in the causal relation of int	×	0.03
The performance w.r.t. causality is impeded in CLEAR-VAE.	×	0.04
The auxiliary variable promotes causality.	×	0.15

References

- http://arxiv.org/abs/1912.03277v3
- http://arxiv.org/abs/2604.15297v2
- http://arxiv.org/abs/2210.08443v2