

What is the impact of scaling the number of pre-training causal models on the fairness-accuracy trade-off in T

Assignee Research

June 10, 2026

Abstract

Training machine learning models for fair decisions faces two key challenges: The *fairness-accuracy trade-off* results from enforcing fairness which weakens its predictive performance in contrast to an unconstrained model. The incompatibility of different fairness metrics poses another trade-off – also known as the *impossibility theorem*. Recent work identifies the bias within the observed data as a possible root cause and shows that fairness and predictive performance are in fact in accord when predictive performance is measured on unbiased data. We offer a causal explanation for

1 Introduction

This paper examines: Overcoming Fairness Trade-offs via Pre-processing: A Causal Perspective. Research question: What is the impact of scaling the number of pre-training causal models on the fairness-accuracy trade-off in TabPFN, evaluated using disparity metrics and accuracy scores on biased tabular benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 6 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.3/10.

3 Results

6 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.3/10.

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.

References

- <http://arxiv.org/abs/2506.01594v2>
- <http://arxiv.org/abs/2501.14710v1>
- <http://arxiv.org/abs/2601.17912v2>