Structural Causal Model Complexity and F1 Stability in Tabular Foundation Models under Distribution Shift

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 3 peer-reviewed papers addressing the following research question: How does varying the complexity of Structural Causal Models (SCMs) used in CausalMixFT affect the downstream F1 score stability of tabular foundation models on out-of-distribution (OOD) datasets like. 10 claims were extracted from source literature; 6 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Cross-Cultural Crash Pattern Analysis in the United States and Jordan Using BERT and SHAP. Research question: How does varying the complexity of Structural Causal Models (SCMs) used in CausalMixFT affect the downstream F1 score stability of tabular foundation models on out-of-distribution (OOD) datasets like TabFact or TabMWP?.

2 Methodology

Systematic literature search across multiple databases yielded 3 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.8/10.

3 Results

3 papers retrieved. 10 claims extracted; 6 independently verified. Quality review score: 6.8/10.

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 study applies BERT and SHAP techniques to examine traffic crash patterns in the United States and Jordan.	✓	0.31
In the USA, vehicle overturns are a major factor in fatal crashes.	×	0.11
In the USA, roadway conditions such as guardrails are a major factor in fatal crashes.	×	0.15
In Jordan, technical defects play a more critical role in fatal crashes than in the USA.	✓	0.17
In Jordan, driver behavior plays a more critical role in fatal crashes than in the USA.	×	0.13
SHAP analysis identifies 'driver' and 'damage' as pivotal terms across both the USA and Jordan.	✓	0.22
SHAP analysis identifies 'overturn' as a country-specific term for the USA.	×	0.14
SHAP analysis identifies 'technical' as a country-specific term for Jordan.	✓	0.15
The study uses BERT/Bi-LSTM models for crash severity prediction.	✓	0.17
The BERT/Bi-LSTM models achieved up to 99.5% accuracy in crash severity prediction.	✓	0.19

References

- https://doi.org/10.3390/electronics14020272
- https://doi.org/10.3390/electronics13173509
- https://doi.org/10.1613/jair.1.13200