Mask Ratios and Robustness in Tabular Foundation Models via Self-Supervised Reconstruction

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How do different mask ratios in reconstruction-based self-supervised learning affect the robustness of tabular foundation models to high-dimensional noise. 11 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Distributionally robust self-supervised learning for tabular data. Research question: How do different mask ratios in reconstruction-based self-supervised learning affect the robustness of tabular foundation models to high-dimensional noise?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

12 papers retrieved. 11 claims extracted; 1 independently verified. Quality review score: 3.8/10.

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 utilizes the bank dataset from Moro et al. [2014] and the census dataset from Kohavi [1996].	×	0.01
DFR consistently outperforms JTT and ERM on the Bank and Census datasets across the reported metrics.	×	0.05
DFR achieves a 14% AUROC gain over ERM on the Bank dataset.	×	0.02
DFR achieves a 25% AUROC gain over ERM on the Census dataset.	×	0.02
The pretraining strategy consists of two stages: Stage 1 ERM pre-training and Stage 2 robust representation learning.	×	0.13
Stage 1 optimizes Masked Language Modeling (MLM) loss for feature reconstruction using ERM.	✓	0.16
Strategy 1 (JTT) identifies samples that are not reconstructed correctly to form an error set for each categorical featu	×	0.06
In Strategy 1, samples in the error set are upweighted for each category to learn specific models per category.	×	0.07
Strategy 2 (DFR) constructs a balanced validation dataset for each category during phase 2.	×	0.04
The downstream classification employs an ensemble approach to construct the representation.	×	0.10
The dataset features include k categorical features and c continuous features.	×	0.08

References

- http://arxiv.org/abs/2601.04110v2

- http://arxiv.org/abs/2410.08511v6
- http://arxiv.org/abs/2402.01204v4