

Noise Scaling Effects on Calibration Error in Tabular Foundation Models Under Distribution Shift

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: What is the impact of varying noise scaling factors in synthetic data generators on the calibration error of tabular foundation models under distribution shift. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Tackling covariate shift with node-based Bayesian neural networks. Research question: What is the impact of varying noise scaling factors in synthetic data generators on the calibration error of tabular foundation models under distribution shift?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.3/10.

3 Results

8 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 3.3/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
The datasets used are CIFAR and TINYIMAGENET, which have corrupted versions of the test set provided by Hendrycks & Diet	×	0.02
The architectures used are VGG16, RESNET18, and PRACTRESNET18.	×	0.02
Three structures of latent variables are tested: in, out, and both.	×	0.08
$K \in \{1, 2, 4\}$ Gaussian component(s) are used in the variational posterior.	×	0.02
Fig. 7 shows the optimal performance on ID data is quite similar between different latent architectures.	×	0.04
On OOD, the optimal performance of using both input and output latent variables is similar to using only output latent v	×	0.08
The optimal γ is lower when the model uses both types of latent variables (z, s).	×	0.06
Fig. 8 shows that as γ increases, the NLL of noisy labels increases much faster than that of clean labels even when the	×	0.03
Fig. 9 shows that as high γ prevents learning from noisy labels, it leads to improved performance on clean test sets.	×	0.05
Fig. 3 shows that the model with higher entropy M32 performs better than the one with lower entropy M16 across all corru	×	0.03
Fig. 4 shows that λ controls the severity of the generated corruptions.	×	0.04

References

- <http://arxiv.org/abs/2206.02435v2>

- <http://arxiv.org/abs/2512.03307v1>
- <http://arxiv.org/abs/2312.07577v3>