Bayesian Ensemble Models vs. Variational Autoencoders for OOD Detection in High-Dimensional Tabular Data

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How do Bayesian ensemble models like BE-SNNs compare to variational autoencoders (VAEs) in terms of AUC-ROC for out-of-distribution (OOD) detection on high-dimensional tabular datasets when evaluated. 14 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.1/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Zero-Shot Image Anomaly Detection Using Generative Foundation Models. Research question: How do Bayesian ensemble models like BE-SNNs compare to variational autoencoders (VAEs) in terms of AUC-ROC for out-of-distribution (OOD) detection on high-dimensional tabular datasets when evaluated on established benchmarks like UCI or OpenML?.

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 3.1/10.

3 Results

16 papers retrieved. 14 claims extracted; 0 independently verified. Quality review score: 3.1/10.

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
DIFFPATHV2 achieves notably higher distinguishing CIFAR-10 inliers from CIFAR-100 anomalies compared to DiffPath.	×	0.02
The proposed improvement methodology performs best on average over all dataset combinations.	×	0.03
MSMA, a score-based diffusion model, is the runner-up model in the experiments.	×	0.07
The trajectory of the diffusion process contains sufficient information to near-perfectly distinguish datasets.	×	0.04
The proposed method shows that the paradigm translates to even unseen datasets.	×	0.06
The baseline without error computation or SSIM performs competitively on most benchmarks, except under near-semantic shi	×	0.05
Directly incorporating score errors improves performance, particularly on datasets such as CelebA and SVHN.	×	0.08
Modulating the original baseline with the inverse SSIM degrades performance.	×	0.03
Combining the Stein error with SSIM yields the strongest results overall, with an average AUROC of 94.9.	×	0.03
Distance-based methods assume that OOD samples lie further from the ID learned embedding manifold.	×	0.03
Density-based methods attempt to explicitly estimate the data log-likelihood using generative models.	×	0.08
Approaches based on Normalizing Flows (NFs) and Energy-based models (EBMs) often exhibit counterintuitive behavior, assi	×	0.05
Recent work has proposed using log-likelihood ratios (LLR) to mitigate the bias toward OOD regions.	×	0.02
Another line of research reframes the detection task from evaluating the likelihood to the information-theoretic notion	×	0.03

References

- http://arxiv.org/abs/2507.22692v1
- http://arxiv.org/abs/2504.20900v1
- http://arxiv.org/abs/2502.17119v2