How does the anomaly detection F1-score of Deepseek R1 compare to Mistral 7B on time-series datasets with distribution shifts versus in-domain benchmarks

Assignee Research

May 29, 2026

Abstract

Anomaly detection presents a unique challenge in machine learning, due to the scarcity of labeled anomaly data. Recent work attempts to mitigate such problems by augmenting training of deep anomaly detection models with additional labeled anomaly samples. However, the labeled data often does not align with the target distribution and introduces harmful bias to the trained model. In this paper, we aim to understand the effect of a biased anomaly set on anomaly detection. Concretely, we view anomaly detection as a supervised learning task where the objective is to optimize the recall at a given

1 Introduction

This paper examines: Understanding the Effect of Bias in Deep Anomaly Detection. Research question: How does the anomaly detection F1-score of Deepseek R1 compare to Mistral 7B on time-series datasets with distribution shifts versus in-domain benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.3/10.

3 Results

11 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 2.3/10.

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/2112.14436v1
- http://arxiv.org/abs/2111.11108v1
- http://arxiv.org/abs/2105.07346v1