Learning Rate Annealing Effects on Adversarial Robustness in Open-Source Code Models

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the impact of learning rate annealing on the robustness of open-source code models when evaluated on adversarial examples from the LiveCodeBench dataset, measured by pass@k scores and. 17 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Crafting Adversarial Examples for Deep Learning Based Prognostics (Extended Version). Research question: What is the impact of learning rate annealing on the robustness of open-source code models when evaluated on adversarial examples from the LiveCodeBench dataset, measured by pass@k scores and accuracy degradation?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.3/10.

3 Results

14 papers retrieved. 17 claims extracted; 3 independently verified. Quality review score: 5.3/10.

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
Adversarial attacks can cause a serious defect in the remaining useful life estimation.	✓	0.23
Crafted adversarial examples are highly transferable and may cause significant damages to PHM systems.	✓	0.30
Modern PHM techniques can help reduce downtime by 35%-45%, maintenance cost by 20%-25%, and can increase production by 2	×	0.08
IoT sensors are known for their vulnerability to cyber-attacks.	×	0.11
DL algorithms can be easily fooled by adversarial examples.	×	0.11
Cyber-threats against businesses/factories have increased by more than 200% over the past year.	×	0.02
Adversarial examples can often transfer from one model to another model.	×	0.09
Adversarial attacks have been extensively studied in the computer vision domain.	×	0.12
The impact of adversarial attacks on the PHM domain has not been studied yet.	×	0.14
Adversarial attacks can lead to a wrong prognostic decision, e.g., a wrong estimation of RUL can delay the maintenance o	×	0.06
Unexpected failures are considered a primary operational risk, as they can hinder productivity and can incur a huge loss	✓	0.25
In the modern automotive industry, an assembly line has several robots working on a car, and if even one robot fails, it	×	0.02
The paper presents an empirical study of adversarial attacks in real-life scenarios using NASA's C-MAPSS dataset.	×	0.08
The paper performs a comprehensive study of the transferability property of adversarial examples in DL-based PHM models.	×	0.12
This is the first work that shows the impact of adversarial attacks on DL enabled PHM systems.	×	0.14
The paper presents algorithms for crafting FGSM and BIM adversarial examples.	×	0.11
The visualization layer uses the data collected from the field, along with the results from the PHM model, and provides	×	0.03

References

- http://arxiv.org/abs/1910.08108v2
- http://arxiv.org/abs/2009.10149v2
- http://arxiv.org/abs/2303.15127v1