

# Adversarial Perturbation Severity and Robustness in CodeT5 and JaCoText for QuixBugs Repair

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the effect of varying levels of adversarial perturbation severity on the robustness of CodeT5 and JaCoText in repairing code from the QuixBugs benchmark, measured by correct fix rates and. 13 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: A Hierarchical Assessment of Adversarial Severity. Research question: What is the effect of varying levels of adversarial perturbation severity on the robustness of CodeT5 and JaCoText in repairing code from the QuixBugs benchmark, measured by correct fix rates and grammatical coherence?.

## 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 3.8/10.

## 3 Results

14 papers retrieved. 13 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 authors introduce a new set of Hierarchy-aware Attacks that optimize adversarial examples by aiming at both diminish	×	0.07
The paper provides the first assessment on adversarial severity on the iNaturalist-H benchmark.	✓	0.16
The iNaturalist-H benchmark is characterized as highly challenging, large-scale, long-tailed, and hierarchically-structu	×	0.06
Employing CHAT to gradually learn classes from a tree results in enhanced defenses against hierarchical adversarial atta	×	0.10
Existing adversarial robustness benchmarks are only concerned with evaluating whether attacks can induce error in networ	×	0.11
Hacohen and Weinshall [18] demonstrated that progressive learning from easy to hard instances enhances performance compa	×	0.02
Graves et al. [17] created an automatic curriculum learning method to improve convergence time.	×	0.03
Bertinetto et al. [6] conducted the first large-scale evaluation of Accuracy and Average Mistakes of modern methods usin	×	0.05
Hierarchical-based approaches are categorized into three groups: Label-embedding methods, hierarchical losses, and hiera	×	0.04
In the defined hierarchical tree notation, $Y_0$ represents the set of leaf nodes and $Y_H-1$ represents the root.	×	0.05
The transition function $Ch,h'$ maps any label in $Y_h$ into a subset of nodes in $Y_{h'}$ , representing the label's offspring whe	×	0.04
The hierarchical distance $d_H$ is defined by the height of the least common ancestor between nodes.	×	0.03
The hierarchical distance $d_H$ is used to measure the semantic difference between an instance's prediction and the corresp	×	0.03

## References

- <http://arxiv.org/abs/2509.05372v2>
- <http://arxiv.org/abs/2108.11785v1>
- <http://arxiv.org/abs/2008.07651v1>