Adversarial Training with PGD vs. FGSM for CodeT5 Robustness on MBXP Python

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How does adversarial training with PGD attacks compare to FGSM attacks in improving the robustness of CodeT5 on the MBXP Python benchmark. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Robust Image Classification: Defensive Strategies against FGSM and PGD Adversarial Attacks. Research question: How does adversarial training with PGD attacks compare to FGSM attacks in improving the robustness of CodeT5 on the MBXP Python benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.3/10.

3 Results

9 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.3/10.

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/2408.13274v1
- http://arxiv.org/abs/2011.05157v2
- http://arxiv.org/abs/2008.03709v4