Defense-Free Federated Learning Accuracy Under Poisoning Attacks Across Benchmarks

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: Can defense-free federated learning frameworks maintain competitive accuracy on standard benchmarks compared to Byzantine-robust aggregators under varying poisoning rates. Data poisoning is a type of adversarial attack on training data where an attacker manipulates a fraction of data to degrade the performance of machine learning model. Therefore, applications that rely on external data-sources for training data are at a significantly higher risk. 11 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Influence Based Defense Against Data Poisoning Attacks in Online Learning. Research question: Can defense-free federated learning frameworks maintain competitive accuracy on standard benchmarks compared to Byzantine-robust aggregators under varying poisoning rates?.

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 4.5/10.

3 Results

9 papers retrieved. 11 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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
Experiments were conducted on a machine with Intel Core i7-8550U CPU, 4 cores, 8 logical processors, base speed of 2.0 G	×	0.00
The proposed defense was implemented using Python 3.6 with scikit-learn implementation of the classifiers.	×	0.03
The system was tested on 64-bit Windows 10 enterprise edition and Ubuntu 16.04 LTS version.	×	0.01
Six datasets were used for the experiments.	×	0.04
The poisoning budget computation is relative to the training data.	×	0.11
Three of the six datasets are common in reference [28].	×	0.01
Two existing poisoning attacks on online learning were considered: simplistic attack [27] and online attack [5].	✓	0.16
Constant learning rates of 0.01, 0.05, and 0.09 were used, along with the optimal learning rate provided by scikit-learn	×	0.02
The influence window size (winf) in the defense algorithm was empirically found using grid search.	×	0.03
The poisoned points are generated by the adversary using the baseline attacks.	×	0.07
The defender first uses slab defense for data sanitization.	×	0.12

References

- http://arxiv.org/abs/2308.03331v1
- http://arxiv.org/abs/2104.13230v1

- http://arxiv.org/abs/2506.01989v2