# CausalMixFT and Fine-Tuning Robustness in Tabular Data Augmentation

Assignee Research

June 12, 2026

**Abstract**

Machine learning (ML) systems have introduced significant advances in various fields, due to the introduction of highly complex models. Despite their success, it has been shown multiple times that machine learning models are prone to imperceptible perturbations that can severely degrade their accuracy. So far, existing studies have primarily focused on models where supervision across all classes were available. In constrast, Zero-shot Learning (ZSL) and Generalized Zero-shot Learning (GZSL) tasks inherently lack supervision across all classes. In this paper, we present a study aimed on evaluat

## 1 Introduction

This paper examines: A Deep Dive into Adversarial Robustness in Zero-Shot Learning. Research question: How does CausalMixFT compare to other data augmentation techniques (e.g., SMOTE, MixUp) in terms of fine-tuning robustness on tabular datasets like TabMNAR or Monash, when evaluated using accuracy degradation under adversarial perturbations?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.8/10.

## 3 Results

12 papers retrieved. 9 claims extracted; 9 independently verified. Quality review score: 8.8/10.

## 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 CUB dataset has 312 attributes, 200 classes, and 11788 images. | ✓ | 0.18 |
| The SUN dataset has 102 attributes, 717 classes, and 14340 images. | ✓ | 0.18 |
| The AWA2 dataset has 85 attributes, 50 classes, and 37322 images. | ✓ | 0.17 |
| The standard per-class top-1 accuracy is used for ZSL evaluation. | ✓ | 0.16 |
| For GZSL, per-class top-1 accuracy values for seen and unseen classes are used to compute harmonic-scores. | ✓ | 0.23 |
| The reproduced values of ALE are denoted as original, although there are slight variations compared to the original resu | ✓ | 0.16 |
| The model selection focuses on models that aim to transfer the knowledge learned from seen classes to unseen classes. | ✓ | 0.24 |
| The Attribute-label embedding (ALE) model is formulated as F(x, y; W) = $\theta$(x)W T $\varphi$(y). | ✓ | 0.23 |
| ALE is one of the earlier studies that showed direct mapping by exploiting data and auxiliary information is more effect | ✓ | 0.27 |

## References


- http://arxiv.org/abs/2103.15670v3
- http://arxiv.org/abs/2507.05904v1
- http://arxiv.org/abs/2008.07651v1