Multimodal Dataset Distillation Effects on Model Alignment and Adversarial Robustness

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How do multimodal dataset distillation approaches affect model alignment and robustness against adversarial attacks on VQA and image-captioning tasks. 15 claims were extracted from source literature; 0 were 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: Geometry-Aware Uncertainty Coresets for Robust Visual In-Context Learning in Histopathology. Research question: How do multimodal dataset distillation approaches affect model alignment and robustness against adversarial attacks on VQA and image-captioning tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 15 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

15 papers retrieved. 15 claims extracted; 0 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 CRC-100K benchmark contains 100,000 H&E-stained colorectal tissue patches spanning 9 classes.	×	0.02
The study omits the background class in CRC-100K and reports 8-class performance.	×	0.03
The MHIST benchmark consists of 3,152 colorectal polyp patches annotated as hyperplastic polyp (HP) or sessile serrated	×	0.02
Experiments use Qwen and LLaVA VLM families loaded with default ImageNet pre-trained weights from HuggingFace.	×	0.04
Coresets are optimized via greedy selection for 1,000 iterations with α = 0.1 and β = 0.1.	×	0.05
Baselines include random sampling, kNN retrieval, mutual-information-informed retrieval (DR), and the shift-vector metho	×	0.05
Dataset-distillation baselines include Trajectory Matching (TM), Distribution Matching (DM), and diffusion-based distill	×	0.11
Statistical significance is assessed using the two-sided Wilcoxon signed-rank test over paired per-run metric values.	×	0.02
On CRC-100K with Qwen at 3-shot, GAUC achieves an accuracy of 0.610$\pm$0.030.	×	0.04
On CRC-100K with Qwen at 3-shot, GAUC achieves an F1 score of 0.588$\pm$0.015.	×	0.03
GAUC improves over the MIMIC baseline by 1.16 percentage points in accuracy on CRC-100K with Qwen at 3-shot.	×	0.04
The improvement of GAUC over MIMIC on CRC-100K with Qwen at 3-shot is statistically significant with $p < 0.05$.	×	0.03
GAUC reduces Expected Calibration Error (ECE) from 0.153$\pm$0.015 to 0.145$\pm$0.012 compared to the DR baseline on CRC-100K wi	×	0.03
The GAUC method optimizes the coreset by minimizing a composite objective over three terms without any gradient-based pa	×	0.09
GAUC uses Maximum Mean Discrepancy (MMD) with an RBF kernel to measure distributional discrepancy between the ful datas	×	0.09

References

- http://arxiv.org/abs/2605.23482v1
- http://arxiv.org/abs/2307.10350v2
- http://arxiv.org/abs/2605.18419v1