

# Scaling Synthetic Video Pretraining Data for Zero-Shot Action Recognition in Multimodal Models

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 5 peer-reviewed papers addressing the following research question: How does increasing the volume of synthetic video pretraining data affect zero-shot action recognition accuracy on Kinetics-700 for multimodal foundation models. 10 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Robust Tabular Foundation Models. Research question: How does increasing the volume of synthetic video pretraining data affect zero-shot action recognition accuracy on Kinetics-700 for multimodal foundation models?.

## 2 Methodology

Systematic literature search across multiple databases yielded 5 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

## 3 Results

5 papers retrieved. 10 claims extracted; 1 independently verified. Quality review score: 4.2/10.

## 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
Tabular foundation models (TFMs) have emerged as a promising direction for classification and regression tasks with structured data.	✓	0.16
TFMs rely on in-context learning (ICL).	×	0.02
TFMs can provide high-quality predictions on new datasets in milliseconds when GPU-accelerated.	×	0.07
TFMs are pretrained using synthetic data generated from structural causal models (SCMs).	×	0.11
Current publicly available, competitive TFMs have been pretrained on datasets generated from a fixed prior distribution	×	0.05
Fixed priors underrepresent certain regions of the parameter space, potentially degrading performance on real-world data	×	0.03
State-of-the-art TFMs still lag behind tree-based methods on some benchmarks.	×	0.06
The work introduces an optimality gap concept to target regions where the TFM underperforms relative to the best achievable model.	×	0.09
The proposed RTFM algorithm can significantly improve the ranking of TabPFN on several real-world tabular benchmarks with structured data.	×	0.09
The optimality gap can be computed in a matter of seconds when parallelized, given sufficient CPU cores.	×	0.04

## References

- <http://arxiv.org/abs/2409.16382v1>
- <http://arxiv.org/abs/2410.02152v1>

- <http://arxiv.org/abs/2512.03307v1>