

# Synthetic vs. Real Gesture Classification: Accuracy Gaps Across Pre-Trained Video Models

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the accuracy gap between synthetic feature-based k-nearest neighbors classification and fine-tuning on real gesture datasets change when comparing different pre-trained video models (e.g., 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: An Evaluation of Large Pre-Trained Models for Gesture Recognition using Synthetic Videos. Research question: How does the accuracy gap between synthetic feature-based k-nearest neighbors classification and fine-tuning on real gesture datasets change when comparing different pre-trained video models (e.g., TimeSformer, ViViT, MViT) at varying scales (e.g., small, medium, large)?.

## 2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

## 3 Results

10 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

## 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/2404.17929v1>
- <http://arxiv.org/abs/2410.02152v1>
- <http://arxiv.org/abs/2110.06500v2>