

OmniFall: From Staged Through Synthetic to Wild, A Unified Multi-Domain Dataset for Robust Fall Detection

David Schneider^{1*}, Zdravko Marinov¹, Moritz Mistol¹, Zeyun Zhong¹, Alexander Jaus¹, Rodi Düger¹, Rafael Baur¹, M. Saquib Sarfraz², and Rainer Stiefelhagen¹

¹ Karlsruhe Institute of Technology

² Mercedes-Benz Tech Innovation

Abstract. Visual fall detection models are usually trained on small, staged datasets. Their real-world utility remains unclear; such data lacks diversity and evaluation protocols differ from paper to paper. We propose OmniFall, a unified benchmark of 15k videos (80 hours) with frame-level annotations in a single 16-class taxonomy. It spans three domains: OF-Staged unifies eight staged datasets with cross-subject and cross-view splits; OF-Synthetic adds 12k videos (17h) with controlled demographic and environmental diversity; and OF-In-the-Wild provides a test-only set of genuine accident videos. We evaluate fine-tuned models as well as much larger zero-shot multimodal LLMs. On in-the-wild fall events, both do comparably well. The clinically critical *fallen* state is where they part: zero-shot models keep confusing *fallen* with *lying*, whereas models fine-tuned on synthetic data with explicit fallen-state scenes do substantially better. We release the unified annotations, the synthetic data, and the in-the-wild test set to foster the development of fall and fallen-state detectors for uncontrolled environments.

Dataset: <https://hf.co/datasets/simplexsigil2/omnifall>

Keywords: Fall detection · Video understanding · Synthetic data

1 Introduction

Falls remain a leading cause of fatal and non-fatal injury in older adults, claiming about 684,000 lives globally and sending more than 37 million people to medical care, yearly [27]. Preventing death hinges on detecting such falls, but even more so on recognizing the sustained *fallen state*, when a person remains on the ground unable to summon help. "Long-lie" episodes increase one-year mortality significantly [13, 20]. Beyond elderly care, falls pose risks in workplaces, public spaces, and during recreational activities, often under diverse, uncontrolled conditions that static indoor datasets cannot represent.

Computer vision-based fall detection could automate monitoring in care environments, yet faces critical limitations: (1) systems target brief falls rather than

* Corresponding author: david.schneider@kit.edu

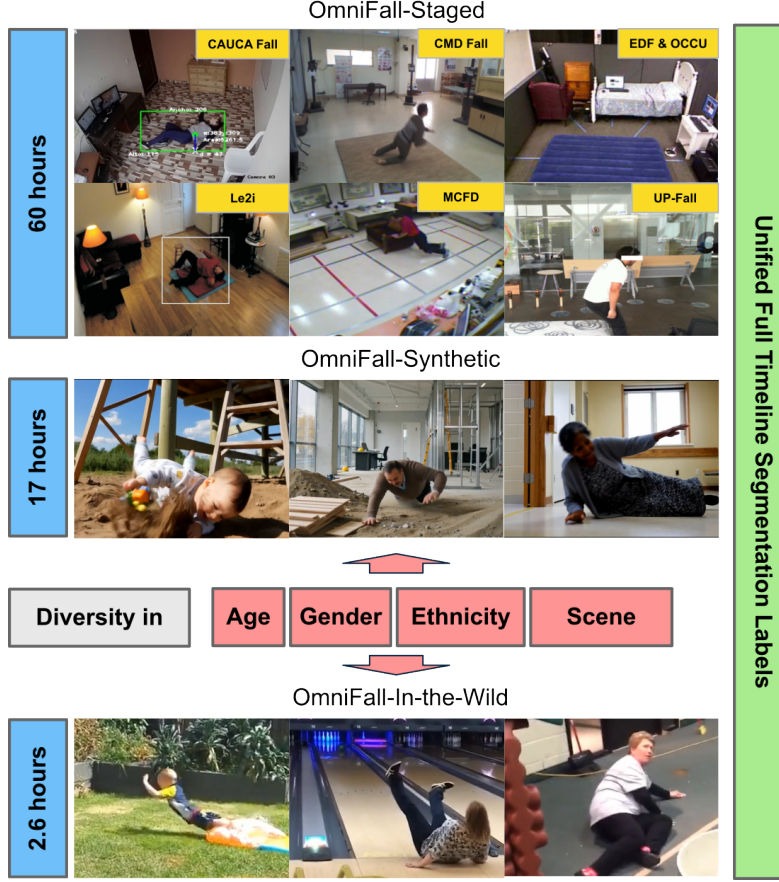


Fig. 1: OmniFall comprises OF-Staged, OF-Synthetic and OF-In-the-Wild with cross-dataset compatible labels.

persistent fallen states; (2) models trained on small, homogeneous datasets [1, 4, 8, 15, 16, 18, 22, 28] generalize poorly across environments; (3) collecting fall data from vulnerable elderly populations raises ethical and practical barriers; (4) no standardized benchmark evaluates genuine accidents or cross-domain generalization; and (5) Multimodal Large Language Models (MLLMs) are powerful generalist vision models, yet underexplored for fall detection.

We introduce *OmniFall*, a unified benchmark comprising three complementary components: *OmniFall-Staged* (*OF-Sta*: staged real-world data), *OmniFall-In-the-Wild* (*OF-ItW*: genuine accident videos), and *OmniFall-Synthetic* (*OF-Syn*: large-scale synthetic data). The benchmark provides 33 hours of densely annotated footage. Five of the eight staged datasets include synchronized multi-view cameras; counting all camera feeds raises the total to 82 hours. This scale enables systematic cross-domain evaluation previously impossible due to annotation heterogeneity.

A clinically motivated 16-class taxonomy (Sec. 3) supports detection of critical fall states even when the fall itself is missed.

Why unify existing staged datasets? Staged datasets cover few environments and subjects with poor generalization to new domains. *OmniFall-Staged* merges eight such datasets [1, 4, 8, 15, 18, 22, 28] spanning 101 subjects and 29 camera views under one annotation scheme. Videos are frame-wise re-annotated allowing for cross-subject and cross-view evaluations across dataset boundaries.

Why the need for in-the-wild evaluation? Staged datasets display healthy actors performing falls in controlled settings with safety equipment, so their generalization to real-world deployment remains unknown. *OmniFall-In-the-Wild (OF-ItW)* addresses this with genuine accident videos from uncontrolled environments, held out as a wild test target whose 1.8 h test portion alone exceeds the total duration of many existing fall detection datasets.

Why add synthetic data? Recording real falls of diverse demographic populations including the elderly and children raises ethical concerns and practical barriers. *OmniFall-Synthetic (OF-Syn)* sidesteps both with generated video offering controlled demographic and environmental variation at no privacy cost. Critically, it explicitly models extended post-fall ground contact, capturing the clinically important *fallen* state that staged recordings underrepresent, since actors there stand up immediately.

Our contributions are threefold. **(1)** We manually re-annotate eight public fall detection datasets frame by frame in a clinically motivated 16-class taxonomy, and define the first standardized staged-to-wild protocol with cross-subject/view splits and a held-out in-the-wild test set of genuine accidents. **(2)** OF-Synthetic adds synthetic-to-wild evaluation and diverse training data with controlled demographic variation. **(3)** We test both fine-tuned models and zero-shot MLLMs and find they fail differently: both reach similar fall-event detection on genuine accidents, but zero-shot MLLMs systematically miss the clinically critical *fallen* state. Fine-tuned models trained on synthetic data, which shows extended post-fall contact, detect the fallen state best across domains. The distinction thus appears to need task-specific training, rather than model scale alone.

2 Related Work

Many fall detection datasets have been introduced in recent years [1, 2, 4, 6, 8–10, 15, 16, 18, 22, 28], reflecting the growing interest in automated fall detection.

Fall Detection Datasets in OmniFall-Staged. To collect existing public fall detection datasets from prior work, we conducted a systematic search in accordance with PRISMA guidelines (5)-(9) [19]. Of the 98 surveys retrieved, only 8 met all inclusion criteria. These datasets (1)-(8) are listed in Table 1.

(1) The **CMDFall** dataset [22] is a multimodal, multiview collection acquired from seven synchronized Kinect sensors (providing color and depth data) together with two accelerometers. It comprises recordings of 50 subjects performing eight distinct fall types alongside twelve non-fall activities. (2) **UP Fall** [18]

Table 1: Datasets in the OmniFall benchmark. The first eight rows constitute OF-Staged (OF-Sta), followed by OF-In-the-Wild (OF-ItW) and OF-Synthetic (OF-Syn).

Dataset	Year	Vids	Res.	FPS	Sub	Views (Sync.)	$\bar{t}$ (s)	Segments		Dur. (h)	
								SV	MV	SV	MV
MCFD [4]	2010	192	480p	≤ 30	1	8 (✓)	4.3	169	1,352	0.2	1.6
Le2i [8]	2013	190	240p	25	9	6 (✗)	3.0	967		0.8	
EDF [28]	2014	10	240p	30	5	2 (✓)	3.1	254	508	0.2	0.4
OCCU [28]	2014	10	240p	30	5	2 (✗)	3.5	245	484	0.3	0.5
CMDFall [22]	2018	384	480p	20	50	7 (✓)	4.3	6,026	42,143	7.1	49.8
UP Fall [18]	2019	1,118	480p	18	17	2 (✓)	13.6	1,213	2,426	4.6	9.2
CAUCA [15]	2022	100	480p	23	10	1	3.9	258		0.3	
GMDCSA [1]	2024	160	720p	30	4	3 (✗)	2.8	458		0.4	
OF-ItW	2026 [†]	818	var.	30		>818	2.4	4,022		2.7	
OF-Syn	2026	12k	720p	16		12k	3.2	19,228		16.9	

[†] With videos from [12], 2019. **Sub**: subjects, **Sync.**: synchronized views, **SV/MV**: single-view / multi-view, $\bar{t}$: mean segment duration.

offers a multimodal setup from 17 subjects in a laboratory environment. It includes vision data from a pair of synchronized cameras (640×480 at 18 fps), inertial measurements from accelerometers and gyroscopes worn at different body locations covering six activities of daily living (ADL) and five fall types. (3) The **Le2i** dataset [8] contains 143 staged falls and 79 ADL actions recorded from a single fixed camera at 640×480 pixels and 25 fps across four environments and six views. (4) **GMDCSA** [1] delivers high temporal resolution data, capturing 81 fall and 79 ADL clips recorded with a stationary laptop camera at 1280×720 pixels and 30fps in three distinct home settings. Its design emphasizes intra-class variability, with subjects altering clothing and backgrounds between clips. (5) The **EDF** and (6) **OCCU** datasets [28] were acquired in a controlled apartment environment using two Kinect sensors at 320×240 pixels and 30 fps. The EDF set focuses on 40 non-occluded falls by five subjects performing falls from eight directions, as well as 30 non-fall actions. In comparison, OCCU presents a more challenging scenario in which 30 falls become partially occluded near the end, accompanied by a suite of 80 ADL actions. (7) **The Multiple Camera Fall (MCFD)** dataset [3,4] leverages eight synchronized IP cameras installed in a fixed indoor setting to capture 22 fall scenarios, facilitating 3D reconstructions either explicitly or implicitly. (8) The **CAUCA Fall** dataset [15] provides recordings from 10 subjects in an apartment setting under varying lighting conditions. Captured at 720×480 resolution and 23 fps, it includes both color and infrared modes (in low-light situations) with detailed, frame-wise annotations that report not only fall versus non-fall labels but also spatial metrics, such as distance and angle with respect to the camera. **OOPS!** [12] is a 20.7 k-clip (>50 h) video dataset with labels marking when intentional actions turn accidental, supporting benchmarks in intention classification and accident forecasting.

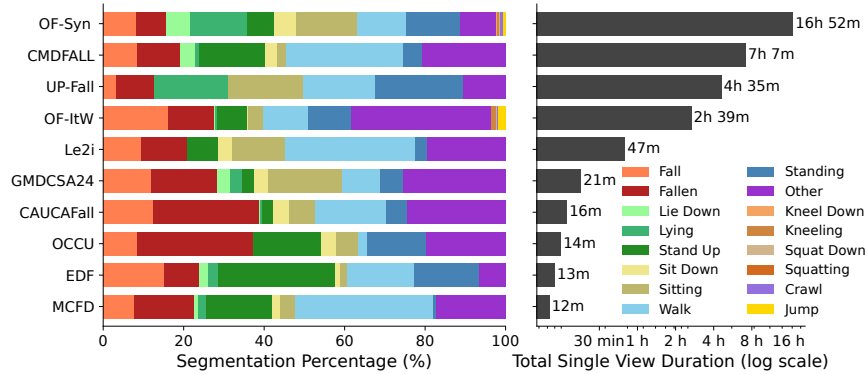


Fig. 2: Label distribution and single-view duration per dataset: eight OF-Sta constituents, OF-ItW, and OF-Syn (33h total).

Other Datasets. We identified several additional fall detection datasets, but excluded them from OF-Staged due to specific limitations. UR Fall [16] was not included due to its limited size. MUVIM [11] offers multimodal data, but all videos are captured from a top-down view. FPDS [17] consists of static images, we only considered full video datasets. While High Quality FSD [6], FallFree [2], SIMPLE [10], and ACT4² [9] align well with our objectives, we were unable to include them due to missing usage licenses or the lack of usage permission.

What Is Missing. These datasets comprise *staged* recordings of healthy participants in controlled environments. They are useful for method development, but none shows how models transfer to *in-the-wild* falls, and their heterogeneous label sets and protocols make cross-dataset comparison hard. Three gaps stand out: (1) *Staging bias*: lab scenes, fixed cameras, and scripted routines dominate. (2) *Annotation heterogeneity*: labels range from binary fall/no-fall to coarse ADL categories. Few offer dense, consistent timelines. To reconcile these schemes, every frame must be re-annotated under a domain-appropriate taxonomy, which no prior work has done at scale. (3) *Lack of standardized OOD evaluation*: no widely-used protocol quantifies staged→wild generalization.

OmniFall (Table 1) addresses all three: we unify eight staged datasets by re-annotating them frameworkwise (OF-Staged), curate an in-the-wild test set from genuine accidents (OF-ItW), and add a diverse synthetic dataset (OF-Syn). The benchmark and its three components are detailed in the following section.

3 Dataset

Overview. **OmniFall** combines three complementary components, each detailed below (Tab. 1, Fig. 1): **OF-Staged** (eight public staged datasets), **OF-In-the-Wild** (OF-ItW), and **OF-Synthetic** (OF-Syn). Together they total ~33 h

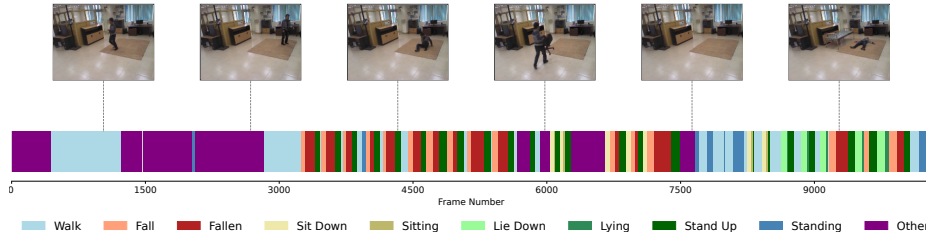


Fig. 3: Our annotations on an example video of CMDFall [22].

of single-view footage and enable *in-distribution* training on staged data and *out-of-distribution* (OOD) testing on wild and synthetic.

Design Principles. Our goal is *comparability across domains*. To that end we: (i) *Standardize labels* by manually (re-)annotating all datasets. (ii) *Densely annotate* full videos (Figure 3), so methods can be evaluated on both classification and timeline segmentation. (iii) *Unify splits* by providing cross-subject (CS) and cross-view (CV) partitions for each staged dataset, mirroring original protocols when available. (iv) *Enable OOD assessment* with a held-out OF-ItW split.

Annotations. Because source datasets make varying assumptions about fall semantics and their original annotations often cover only isolated events rather than full video timelines, we annotate all videos from scratch rather than mapping existing labels, avoiding bias from preexisting schemes and ensuring complete temporal coverage. Our annotations are frame-wise and comprise dynamic actions: *fall*, *sit_down*, *lie_down*, *stand_up*, *walk*, *kneel_down*, *squat_down*, *crawl*, *jump*; six static states: *fallen*, *sitting*, *lying*, *standing*, *kneeling*, *squatting*; plus *other* as a catch-all category for remaining frames. Interviews with medical professionals identified a *fall/fallen* distinction as clinically essential for catching “long-lie” episodes, so our annotation taxonomy models the *fallen* state separately from the fall itself. We further include visually similar actions such as *lie_down* and *lying* since avoiding confusion between lying down and falling has great practical relevance. Additional classes are determined by reviewing the datasets and adding major activities. The annotation process is described in detail in the appendix.

OF-Staged. OF-Staged spans 14 h of recordings across eight datasets (CMD-Fall [22], UP-Fall [18], Le2i [8], GMDCSA24 [1], CAUCA Fall [15], EDF [28], Occu [28] and MCFD [4], Fig. 2), with synchronized views excluded from duration accounting but preserved for CV evaluation. When multi-view recordings are included, CS training includes up to 61.5 h of footage, providing OF-Sta with a training volume advantage over OF-Syn (16.9 h single-view) under cross-subject evaluation. We provide CS/CV splits per dataset and *combined* splits (OF-Staged) that support training and testing across domains.

OF-In-the-Wild. Starting from OOPS [12], we filter clips containing falls, then *manually verify and densely annotate* 818 videos (2.65 h). These contain



Fig. 4: Expression conditioning mitigates Wan 2.2’s smile bias on falls: without (left) and with (right) explicit expression prompts.

1.3k segments in 10/20/70 train/val/test splits with real-world camera motion, occlusion, and clutter. We selected clips with a clearly visible fall, and covered indoor and outdoor scenes under varied lighting. Collecting genuine falls from elderly people is ethically impractical; OF-ItW, with its unscripted accident videos, is a close available proxy. In this work we almost exclusively use OF-ItW as *test-only*, measuring staged→wild generalization.

OF-Synthetic. OF-Syn contributes 12k videos (17 h) generated with Wan 2.2 [25] at 1280×720 and 16 fps, each a five-second clip of a unique individual and scene. Domain-gap analyses (§4.1) show OF-Syn sits closer to OF-ItW in feature space than individual staged datasets. This enables its use for sim→real studies.

We build each clip from a structured prompt with varying axes: who the person is, where they are and what they wear, how the fall unfolds, and where the camera sits. Demographics span six age groups from toddler to elderly and the seven OMB SPD-15 race and ethnicity categories [24], alongside body type and gender, while scenes draw on over a thousand distinct environments. Each fall combines one of 30 mechanisms (e.g., a slip on ice) with a direction and a recovery attempt, and the camera ranges from eye-level to CCTV and top-down views. We fixed specific Wan 2.2 failure modes with prompt engineering, such as a real-time frame-rate cue reducing slow-motion artifacts, and conditioning on expressions such as *startled* or *panicked* suppressing smiling faces during falls (Fig. 4). We screened all 12,000 clips and re-ran generation for clips with unsafe content, implausible motion, or artifacts. Additional details in the appendix.

4 Experiments

Our evaluation comprises three complementary experiments using different methodologies. For domain shift analysis (Sec. 4.1), we extract frozen motion oriented features from I3D [7] and VideoMAE-K400 [21] and more appearance-based features from SigLIP2 [23] to analyze dataset similarities and feature clustering. For action and fall classification (Sec. 4.2), we fine-tune VideoMAE-K400 end-to-end on our benchmark datasets. Timeline segmentation follows established protocols using SigLIP2 features (Sec. 4.3). Additional details in the appendix.

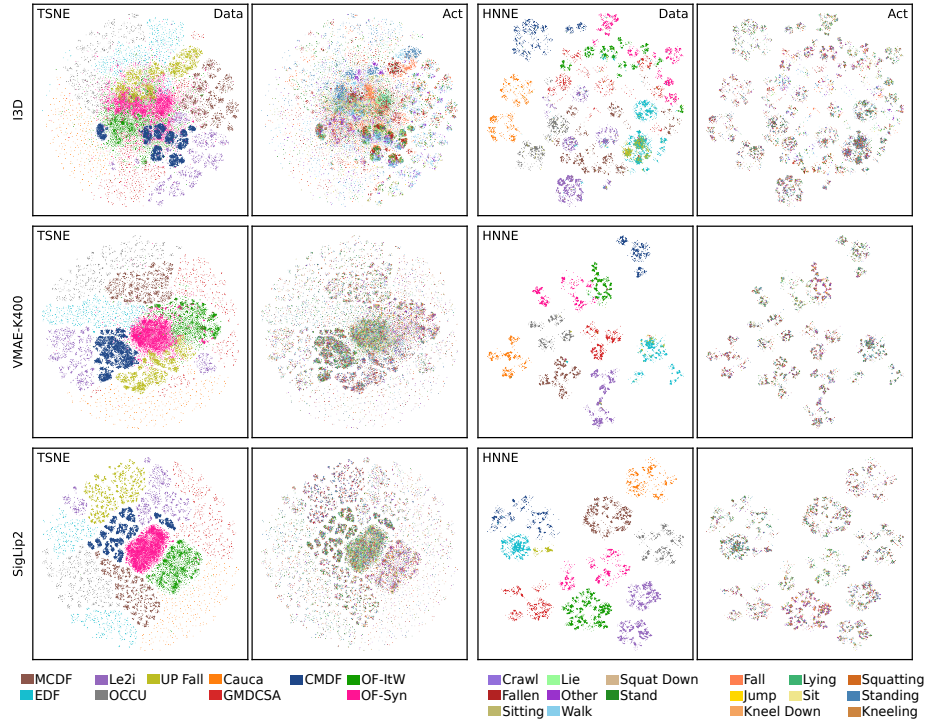


Fig. 5: 2D projections of frozen embeddings (I3D, VideoMAE-K400, SigLIP2), color-coded by dataset (odd columns) and activity (even columns). Rows correspond to feature type; columns show t-SNE and h-NNE clustering. Full-size plots in the appendix.

4.1 Domain Shift Analysis

Feature-Space Visualization. Figure 5 visualizes frozen embeddings from I3D, VideoMAE-K400, and SigLIP2 via t-SNE and h-NNE projections (5,000 samples per domain, five-frame temporal means). Across all backbones, domains form largely disjoint clusters, confirming a strong dataset shift between staged, in-the-wild, and synthetic videos. The separation is most pronounced for SigLIP2 and clearly visible for VideoMAE and I3D as well.

Figure 6 shows pairwise Fréchet Video Distances across three feature types. CMDfFall demonstrates high similarity (darker colors) to most other staged datasets, consistent with its broad coverage of fall types and ADL categories. OF-Syn and OF-ItW exhibit strong mutual similarity, likely because both contain a large proportion of outdoor activities absent from any staged dataset. This domain overlap partly explains OF-Syn’s transfer advantage in Sec. 4.2. It also illustrates a practical strength of synthetic generation: diverse environments can be covered at negligible cost, whereas staged recordings are typically confined to laboratory settings. Conversely, OF-Staged shows substantial dissimilarity to OF-ItW on VMAE-K400, corresponding to the OOD performance gaps in

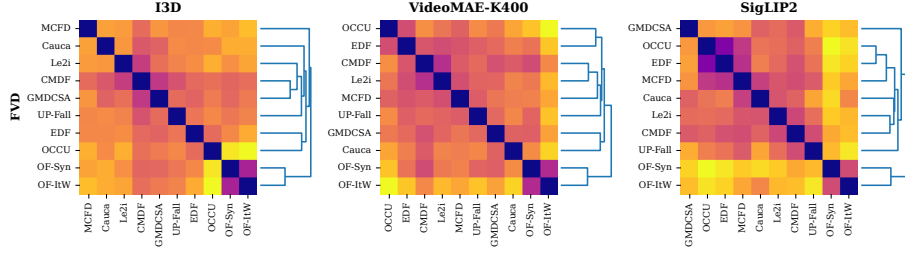


Fig. 6: Pairwise Fréchet Video Distances calculated on features from I3D, VideoMAE-K400 and SigLIP2, visualised as heat-maps. Darker colors indicate higher similarity (smaller distances) between datasets. Numerical values are listed in the appendix.

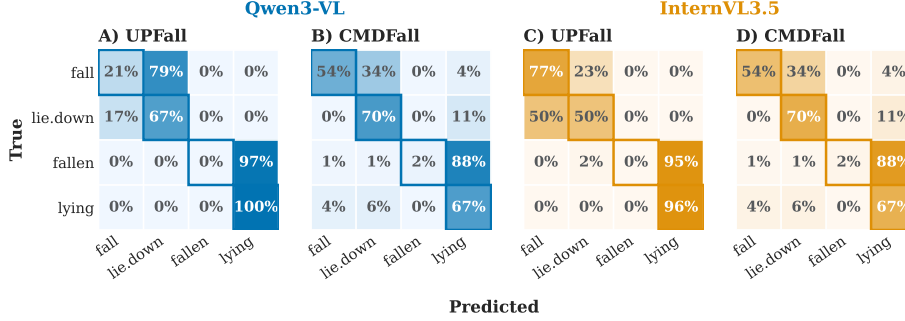


Fig. 7: Zero-shot confusion matrices on UP-Fall and CMDFall for Qwen3-VL-8B and InternVL3.5-8B, restricted to the semantically adjacent classes *fall*, *lie down*, *fallen*, and *lying*. Both models confuse *fallen* with *lying* and, to varying degrees, *fall* with *lie down*.

Section 4.2. We encourage cross-referencing these distance-based visualizations with the dataset examples provided in the supplementary material.

4.2 Action and Fall Classification

Fine-Tuning Setup. We use VMAE-K400 [21] (87M parameters). As a first probe of out-of-domain generalization, we fine-tune on individual staged datasets and test each both in-domain and on OF-ItW (Tab. 2). We then study four training configurations in more detail: (1) CMDFall alone (7h), the largest pre-existing staged dataset, using cross-subject (CS) splits; (2) OF-Staged (14h, 8 datasets) with CS splits; (3) OF-Synthetic with random (Ra) splits (17h); (4) OF-Staged (CS) + OF-Synthetic (Ra) combined. Single-view durations are comparable (OF-Sta: 13.8h, OF-Syn: 16.9h), but CS splits leverage multi-view recordings, giving OF-Sta up to 61.5h of training footage.

For metrics, we report accuracy and macro-F1 for the 16-class task, alongside sensitivity (Se), specificity (Sp), and F1-score for the critical binary subtasks *Fall* and *Fallen*. Macro-F1 scores tend to be lower due to the natural class imbalance

Table 2: Per-dataset **in-domain** (ID) vs. **in-the-wild** (ItW) generalization. A VideoMAE-K400 model is trained on each staged dataset and tested on its own test split (ID) and on OF-ItW. **Sensitivity**, **Specificity**, **F1** score.

Test	Train dataset	Fall			Fallen		
		Se	Sp	F1	Se	Sp	F1
ID	CMDFall	91.2	99.3	93.0	92.0	98.7	90.9
	Up-Fall	57.9	99.8	72.9	40.4	99.8	57.1
	Le2i	100	100	100	100	99.5	97.7
	GMDCSA24	76.5	93.4	74.3	76.5	100	86.7
	EDF	100	100	100	65.0	100	78.8
	OCCU	100	97.7	94.1	100	100	100
	CaucaFall	100	94.7	90.0	100	100	100
OF-ItW	CMDFall	51.1	86.9	52.3	10.0	98.7	16.6
	Up-Fall	9.5	99.6	16.9	1.6	99.8	3.2
	Le2i	56.8	94.8	58.8	50.5	86.7	33.1
	GMDCSA24	37.8	96.7	47.0	12.3	97.7	17.7
	EDF	15.6	98.6	24.8	18.9	94.8	21.0
	OCCU	8.7	98.4	14.5	1.5	98.9	2.7
	CaucaFall	8.0	98.9	13.8	11.2	97.7	16.2

Each row reports one VideoMAE-K400 [21] model trained on that dataset’s training split. ID: same dataset’s held-out test split; ItW: the OF-ItW test set.

across 16 categories. Tables 3 and 4 provide detailed results; best and second-best values per column are highlighted in green and orange.

Per-Dataset Generalization. Trained on a single staged dataset, a model reaches near-ceiling in-domain F1 but collapses on OF-ItW (Tab. 2): Le2i falls from 100 to 58.8 Fall F1, and Up-Fall from 72.9 to 16.9, with *fallen* detection lower still (Le2i: 97.7 to 33.1). No staged dataset displays good OOD generalization.

Training on CMDFall. We examine CMDFall in more detail, the largest and most diverse staged dataset (Tab. 3). On cross-subject splits it performs strongly in-domain (83.3% BAcc, Fall F1: 93.0%), yet still generalizes only weakly to OF-ItW (Fall F1: 52.3%, Fallen F1: 16.6%). Even the largest staged dataset thus lacks the diversity needed for real-world deployment, particularly for the clinically critical *fallen* state. We next ask whether unifying datasets and adding controlled synthetic diversity closes this gap.

Training on OF-Staged. Training on the unified OF-Staged superset substantially improves out-of-distribution performance. Despite similar data volume to CMDFall alone (61.5h vs 49.8h multi-view), the diversity of eight datasets with varied environments, subjects, and viewpoints improves OF-ItW results to 61.2% Fall F1 and 39.4% Fallen F1, in-distribution performance remains strong (CMDFall CS BAcc: 82.9%). This suggests dataset diversity rather than quantity drives better real-world generalization for both fall events and fallen states.

Table 3: Evaluation of zero-shot MLLMs and differently fine-tuned VMAE-K400 models on OF-Staged (CS test splits). **Bal.** Accuracy, **Sensitivity**, **Specificity**.

Test Model	Train	16-class			Fall Δ			Fallen Δ			Fall $\cup$ Fallen Δ		
		BAcc	Acc	F1	Se	Sp	F1	Se	Sp	F1	Se	Sp	F1
[22]	Qwen ZS	50.6	47.8	27.6	53.9	98.3	65.4	2.4	100	4.8	30.6	98.1	44.9
	IntVL ZS	52.4	53.6	30.6	84.5	96.6	81.7	1.8	99.9	3.6	47.3	96.2	59.4
	VMAE CMDF	83.3	85.4	83.3	91.2	99.3	93.0	92.0	98.7	90.9	95.6	98.9	96.0
	VMAE Sta	82.9	85.4	82.9	91.8	99.1	93.0	92.6	98.9	92.0	95.8	98.9	96.2
	VMAE Syn	45.8	50.3	34.9	61.0	96.4	66.1	27.6	96.0	34.6	52.4	93.3	60.5
CMDf	VMAE St+Sy	83.0	85.4	83.3	92.6	98.9	92.6	90.9	98.9	91.0	96.3	98.8	96.3
[18]	Qwen ZS	44.7	31.8	20.0	21.1	90.9	27.4	0.0	100	0.0	10.5	87.4	16.6
	IntVL ZS	47.6	41.8	22.9	77.2	86.8	68.8	0.0	100	0.0	38.6	81.6	47.6
	VMAE CMDF	55.6	52.9	41.7	12.3	98.8	21.1	50.9	96.6	62.4	31.6	93.5	45.1
	VMAE Sta	92.5	92.7	83.0	95.6	98.8	95.6	87.7	97.3	88.9	92.1	94.9	92.7
	VMAE Syn	48.0	52.9	31.3	63.2	99.3	76.2	2.6	99.5	5.0	33.8	99.0	50.0
UP-Fall	VMAE St+Sy	92.6	92.0	82.4	96.5	99.3	96.9	79.8	98.0	85.4	88.2	96.3	91.4
[8]	Qwen ZS	48.0	54.7	28.8	9.1	99.4	16.0	0.0	100	0.0	4.7	99.4	8.7
	IntVL ZS	48.0	55.2	28.1	27.3	98.3	38.7	0.0	100	0.0	14.0	98.1	23.1
	VMAE CMDF	65.9	71.9	52.7	77.3	97.8	79.1	85.7	99.5	90.0	86.0	98.1	89.2
	VMAE Sta	81.4	89.7	72.9	100	100	100	100	100	100	100	100	100
	VMAE Syn	53.1	57.1	39.5	59.1	100	74.3	9.5	98.9	16.0	39.5	100	56.7
Le2i	VMAE St+Sy	83.2	89.7	82.4	100	99.4	97.8	100	100	100	100	99.4	98.9
[1]	Qwen ZS	62.4	48.4	36.1	58.8	94.7	64.5	17.6	100	30.0	38.2	93.2	51.0
	IntVL ZS	55.6	44.1	33.5	58.8	92.1	60.6	0.0	100	0.0	29.4	89.8	40.0
	VMAE CMDF	59.5	54.8	50.4	64.7	89.5	61.1	52.9	100	69.2	61.8	88.1	67.7
	VMAE Sta	64.5	73.1	66.8	88.2	93.4	81.1	88.2	97.4	88.2	91.2	89.8	87.3
	VMAE Syn	64.5	60.2	50.3	52.9	97.4	64.3	23.5	100	38.1	41.2	98.3	57.1
GMDCSA	VMAE St+Sy	80.3	76.3	75.0	64.7	94.7	68.8	82.4	97.4	84.8	76.5	91.5	80.0
[28]	Qwen ZS	51.5	38.3	20.9	0.0	100	0.0	0.0	100	0.0	0.0	100	0.0
	IntVL ZS	41.5	34.4	20.7	25.0	97.3	34.8	0.0	100	0.0	11.1	96.7	18.6
	VMAE CMDF	52.6	58.6	43.7	6.2	96.4	9.5	75.0	96.3	76.9	50.0	93.5	60.0
	VMAE Sta	56.6	75.8	57.2	82.4	95.5	77.8	100	95.3	89.4	100	92.2	91.6
	VMAE Syn	33.7	43.8	24.2	35.3	92.8	38.7	47.6	93.5	52.6	47.4	85.6	52.2
EDF	VMAE St+Sy	55.1	74.2	54.2	82.4	95.5	77.8	100	95.3	89.4	100	92.2	91.6
[28]	Qwen ZS	55.1	53.5	39.5	31.2	100	47.6	0.0	100	0.0	15.6	100	27.0
	IntVL ZS	44.8	44.6	30.2	12.5	100	22.2	0.0	100	0.0	6.2	100	11.8
	VMAE CMDF	44.2	38.6	29.4	0.0	100	0.0	6.2	97.6	10.5	3.1	97.1	5.7
	VMAE Sta	86.0	90.1	84.6	100	97.6	94.1	100	97.6	94.1	100	94.2	94.1
	VMAE Syn	54.1	53.5	30.5	37.5	100	54.5	0.0	100	0.0	18.8	100	31.6
OCCU	VMAE St+Sy	84.4	91.1	81.9	100	97.6	94.1	100	98.8	97.0	100	95.7	95.5
[15]	Qwen ZS	56.3	55.3	36.2	100	92.1	85.7	11.1	100	20.0	55.6	89.7	64.5
	IntVL ZS	38.1	42.6	24.5	100	86.8	78.3	0.0	100	0.0	50.0	82.8	56.2
	VMAE CMDF	62.4	74.5	61.3	66.7	100	80.0	100	94.7	90.0	88.9	96.6	91.4
	VMAE Sta	79.5	80.9	77.5	88.9	100	94.1	100	100	100	94.4	100	97.1
	VMAE Syn	60.1	55.3	40.9	77.8	94.7	77.8	22.2	100	36.4	50.0	93.1	62.1
CaucaF.	VMAE St+Sy	76.8	80.9	76.2	100	100	100	100	97.4	94.7	100	96.6	97.3

Qwen: Qwen3-VL-8B [5], IntVL: InternVL3.5-8B, VMAE: VideoMAE-K400 [21].

ZS: zero-shot, CMDf: CMDFall CS, Sta: OF-Sta CS, Syn: OF-Syn Ra, St+Sy: OF-Sta CS + OF-Syn Ra. Shaded rows: out-of-distribution (test dataset absent from training).

Table 4: Evaluation of zero-shot MLLMs and differently fine-tuned VMAE-K400 models on OF-Syn and OF-ItW. **Bal.** Accuracy, **Sens** Sensitivity, **Spec** Specificity.

Test Model	Train	16-class			Fall Δ			Fallen Δ			Fall $\cup$ Fallen Δ		
		BAcc	Acc	F1	Se	Sp	F1	Se	Sp	F1	Se	Sp	F1
OF-Syn	Qwen ZS	55.9	67.6	48.4	54.2	96.4	59.3	15.3	98.6	23.2	43.1	95.7	53.2
	IntVL ZS	50.3	64.9	44.3	52.4	97.2	59.8	4.7	99.5	8.4	36.5	97.3	49.2
	VMAE CMDF	32.4	51.9	31.0	33.0	96.5	40.9	10.0	97.4	14.2	27.3	94.1	35.7
	VMAE Sta	32.0	50.9	31.2	46.2	97.3	55.1	32.7	94.1	32.2	43.1	90.9	47.3
	VMAE Syn	59.0	81.4	60.5	96.2	95.1	81.6	67.3	97.0	66.4	90.1	92.6	81.1
	VMAE St+Sy	60.5	81.2	62.4	96.2	95.8	83.6	68.7	96.7	66.0	89.5	92.7	81.0
OF-ItW	Qwen ZS	36.4	45.4	26.3	84.6	80.8	67.8	19.6	98.1	29.2	69.8	78.5	66.1
	IntVL ZS	34.9	43.8	26.0	87.7	79.4	68.0	19.6	99.1	31.0	74.4	79.5	69.5
	VMAE CMDF	16.0	35.3	14.7	51.1	86.9	52.3	10.0	98.7	16.6	42.9	85.7	50.3
	VMAE Sta	20.5	41.4	18.5	77.6	77.6	61.2	33.4	95.4	39.4	74.1	73.3	65.7
	VMAE Syn	20.6	46.7	21.2	72.0	84.6	64.2	35.0	96.1	42.3	69.5	81.8	67.9
	VMAE St+Sy	21.4	47.6	21.9	72.9	85.4	65.6	33.1	96.3	41.0	68.2	82.4	67.5
VMAE St+Sy+ItW		25.8	56.5	26.7	79.3	91.3	76.0	57.2	92.8	53.6	80.9	84.7	76.9

Qwen: Qwen3-VL-8B [5], IntVL: InternVL3.5-8B, VMAE: VideoMAE-K400 [21].

ZS: zero-shot, CMDF: CMDFall CS, Sta: OF-Sta CS, Syn: OF-Syn Ra, St+Sy: OF-Sta CS + OF-Syn Ra Shaded rows: out-of-distribution (test dataset absent from training).

Training on OF-Synthetic. Training only on synthetic OF-Syn achieves the best single-source OF-In-the-Wild performance on both Fall F1 (64.2%) and Fallen F1 (42.3%), above the OF-Staged superset (61.2% and 39.4%). It does so despite comparable single-view duration (16.9h vs. 13.8h) and no access to multi-view recordings (OF-Sta CS: up to 61.5h). The Fallen F1 improvement reflects OF-Syn’s explicit modeling of extended post-fall ground contact, as discussed in Sec. 1. This matches the domain shift analysis (Sec. 4.1), where OF-Syn sits closer to OF-ItW in feature space than staged datasets.

Training on OF-Staged+OF-Synthetic. Combined training on OF-Sta(CS) and OF-Syn(Ra) yields the best cross-domain Fall F1 on OF-ItW (Table 4), reaching 65.6% with 83.0% BAcc on CMDFall CS. Improvement over OF-Syn alone is marginal (+1.4 F1 points) suggesting that our synthetic data captures most of the variation needed for real-world generalization. *fallen* detection is strongest when OF-Syn is the sole training source (Fallen F1: 42.3% vs. 41.0%), possibly due to an underrepresentation of the *fallen* class in staged datasets.

Upper-Bound on OF-ItW. To contextualize the cross-domain results, we fine-tune VMAE-K400 on the combined OF-Staged, OF-Synthetic, and OF-ItW training splits (Table 4, last row). This configuration achieves Fall F1 of 76.0 and Fallen F1 of 53.6 (Fall $\cup$ Fallen F1: 76.9). Compared to the best cross-domain configuration (St+Sy: 65.6 / 41.0 / 67.5), substantial headroom remains: even a modest amount of in-the-wild training data improves both *fall* and *fallen* detection considerably.

Zero-Shot MLLMs. As zero-shot baselines, we prompt two 8B-parameter multimodal LLMs (MLLMs), Qwen3-VL-8B [5] and InternVL3.5-8B [26], to classify video clips into one of the 16 classes without any task-specific training (Tables 3 and 4). On OF-ItW, both models achieve high *fall* F1 scores (67.8 and 68.0, respectively), yet *fallen* detection remains poor (F1: 29.2 and 31.0), indicating that detecting sustained post-fall states requires capabilities that large-scale pre-training alone does not provide. On staged datasets, both MLLMs score substantially below the fine-tuned models above. Setup details, prompt templates, and ablations across additional model scales appear in the appendix.

Fallen Class Confusion. Figure 7 illustrates systematic confusion between semantically adjacent classes in zero-shot MLLM predictions. Both models classify nearly all *fallen* instances as *lying* (95–97% on UP-Fall; <3% correct on CMDFall), consistent with these being static ground-contact states that likely require fine-grained temporal reasoning to distinguish. In a care setting, this confusion would result in missed long-lie alerts, as the system would label a person unable to get up as merely lying down. Beyond classification accuracy, 8B-parameter MLLMs are computationally heavy, which limits their use on edge devices or for real-time monitoring, unlike the smaller, task-specific fine-tuned models above.

Fine-Tuned VMAE vs. Zero-Shot MLLMs. On *fall* event detection, both paradigms perform comparably on OF-ItW: VMAE (St+Sy, ~87M parameters) achieves Fall F1 of 65.6, close to Qwen3-VL’s 67.8 and InternVL3.5’s 68.0 (8B parameters each). On the clinically critical *fallen* state, however, the paradigms diverge: the best fine-tuned configuration (VMAE trained on OF-Syn) reaches Fallen F1 of 42.3, substantially outperforming both MLLMs (29.2 and 31.0). This indicates that *fallen*-state recognition requires task-specific training, a gap that large-scale pre-training alone does not bridge. At a fraction of the computational cost and model size, the fine-tuned models are also more suitable for on-site deployment in care environments.

Summary. Across all configurations, *fallen*-state detection is the primary bottleneck: even the best configuration with in-domain training data reaches only 53.6 % Fallen F1 (Table 4), compared to 76.0 % Fall F1 under the same conditions. Cross-demographic generalization experiments (cross-age and cross-BMI on OF-Syn) and cross-view results are in the appendix.

4.3 Timeline Segmentation

With dense labels and multi-domain coverage, OmniFall supports OOD timeline segmentation studies. To demonstrate, we train ASQuery [14] at 10 fps using frozen SigLIP2 [23] features. We consider three training sets: OF-Staged, OF-Synthetic, and their union OF-StaUOF-Syn, and evaluate on in-distribution staged test splits and out-of-distribution OF-ItW. We report F1@IoU (10/25/50), normalized Edit, and frame accuracy (Table 5).

Table 5: ASQuery [14] timeline segmentation with SigLIP2 features.

Test	Train	F1@10	F1@25	F1@50	Edit	Acc
OF-Sta	Sta	82.13	80.64	72.70	81.90	80.19
	St+Sy	82.48	80.99	73.24	82.51	79.88
OF-Syn	Syn	84.17	83.32	78.11	83.71	82.71
	St+Sy	83.10	82.24	76.90	82.41	81.15
OF-ItW	Sta	39.15	33.95	20.26	36.42	43.32
	Syn	53.92	40.56	27.60	49.97	59.39
	St+Sy	52.34	45.66	28.61	48.07	59.83

Sta: OF-Sta CS, Syn: OF-Syn Ra, St+Sy: OF-Sta CS + OF-Syn Ra. Shaded rows: out-of-distribution.

Training Data Comparison. All configurations achieve strong in-distribution performance (F1@50: 72–78%), but OF-ItW scores drop substantially. Training on OF-Synthetic alone achieves F1@50 of 27.60% on OF-ItW, surpassing OF-Staged (20.26%) and approaching OF-Sta $\cup$ OF-Syn (28.61%). This mirrors the classification results (Section 4.2): synthetic data with controlled diversity bridges the domain gap to in-the-wild settings more effectively than staged data alone. OF-Sta $\cup$ OF-Syn is the most robust overall: it combines staged motion structure with synthetic appearance diversity.

Overall, the staged $\rightarrow$ wild transfer is hard: models that score highly in controlled settings lose precision on OF-ItW, where appearance, motion, and camera factors vary widely.

5 Risks and Limitations of Synthetic Data Generation

OF-Synthetic addresses representative gaps concerning demographics and realism of staged fall data. At the same time, the use of generative video models can introduce new biases. While we manually verified the overall quality of the generation process, we cannot guarantee that all attributes are faithfully representative of global populations, particularly for demographic groups underrepresented in the generator’s training data. In order to avoid personal biases regarding ethnicity definitions, we follow the U.S. OMB ethnicity reporting standards [24] which might under-represent certain ethnicities. Furthermore, bound by limitations of the generation model, we limited video generation to a binary gender representation model. We request researchers using our data to be aware of such potential biases and limitations and to consider countermeasures in training and analysis for potential biases of resulting models. Despite these limitations, we believe that our synthetic dataset improves fall detection for all demographic groups.

6 Conclusion

OmniFall is a unified benchmark spanning ~ 80 hours of densely annotated video across staged, in-the-wild, and synthetic domains, with framewise 16-class timeline annotations for both classification and timeline segmentation.

Experiments produce three findings. First, smaller fine-tuned vision models and zero-shot multimodal LLMs fail differently: both reach comparable fall-event detection on genuine accidents, yet diverge on the clinically critical *fallen* state, where zero-shot models systematically confuse *fallen* with *lying*. Fine-tuned models trained on synthetic data with post-fall modeling address this gap at a fraction of the computational cost, and are better suited to on-site deployment.

Second, training data diversity drives cross-domain generalization more effectively than volume alone: unifying eight staged datasets substantially improves in-the-wild performance over the largest single dataset at comparable training hours, and synthetic data with controlled diversity improves further. Timeline segmentation experiments confirm this pattern: synthetic-trained models outperform staged-only models on out-of-distribution data.

Third, *fallen*-state detection is the main bottleneck: even with in-domain training data, *fallen* recognition performs significantly worse than fall-event detection. Since missed *fallen* states lead to missed long-lie alerts, closing this gap is the most important challenge for clinical deployment. OmniFall provides the standardized protocols and multi-domain data to drive this progress.

Acknowledgements

This work has been supported by the Carl Zeiss Foundation through the JuBot project as well as by funding from the pilot program Core-Informatics of the Helmholtz Association (HGF). The authors acknowledge support by the state of Baden-Württemberg through bwHPC. Experiments were performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

References

1. Alam, E., Sufian, A., Dutta, P., Leo, M., Hameed, I.A.: GMDCSA-24: A dataset for human fall detection in videos. *Data in Brief* **57**, 110892 (2024). <https://doi.org/10.1016/j.dib.2024.110892>
2. Alzahrani, M.S., Jarraya, S.K., Salamah, M.A., Ben-Abdallah, H.: Fallfree: Multiple fall scenario dataset of cane users for monitoring applications using kinect. In: 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). pp. 327–333. IEEE (2017). <https://doi.org/10.1109/SITIS.2017.61>
3. Auvinet, E., Multon, F., Saint-Arnaud, A., Rousseau, J., Meunier, J.: Fall detection with multiple cameras: An occlusion-resistant method based on 3-d silhouette vertical distribution. *IEEE Transactions on Information Technology in Biomedicine* **15**(2), 290–300 (2011). <https://doi.org/10.1109/TITB.2010.2087385>
4. Auvinet, E., Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Multiple cameras fall data set. Technical Report 1350, DIRO, Université de Montréal (2010), <https://web.archive.org/web/20250731101305/https://www.iro.umontreal.ca/~labimage/Dataset/>
5. Bai, S., Cai, Y., Chen, R., Chen, K., Chen, X., Cheng, Z., Deng, L., Ding, W., Gao, C., Ge, C., Ge, W., Guo, Z., Huang, Q., Huang, J., Huang, F., Hui, B., Jiang, S., Li, Z., Li, M., Li, M., Li, K., Lin, Z., Lin, J., Liu, X., Liu, J., Liu, C., Liu, Y., Liu, D., Liu, S., Lu, D., Luo, R., Lv, C., Men, R., Meng, L., Ren, X., Ren, X., Song, S., Sun, Y., Tang, J., Tu, J., Wan, J., Wang, P., Wang, P., Wang, Q., Wang, Y., Xie, T., Xu, Y., Xu, H., Xu, J., Yang, Z., Yang, M., Yang, J., Yang, A., Yu, B., Zhang, F., Zhang, H., Zhang, X., Zheng, B., Zhong, H., Zhou, J., Zhou, F., Zhou, J., Zhu, Y., Zhu, K.: Qwen3-VL technical report (2025). <https://doi.org/10.48550/arXiv.2511.21631>
6. Baldewijns, G., Debar, G., Mertes, G., Vanrumste, B., Croonenborghs, T.: Bridging the gap between real-life data and simulated data by providing a highly realistic fall dataset for evaluating camera-based fall detection algorithms. *Healthcare technology letters* **3**(1), 6–11 (2016). <https://doi.org/10.1049/htl.2015.0047>
7. Carreira, J., Zisserman, A.: Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6299–6308 (2017). <https://doi.org/10.1109/CVPR.2017.502>
8. Charfi, I., Miteran, J., Dubois, J., Atri, M., Tourki, R.: Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification. *Journal of Electronic Imaging* **22**(4), 041106 (2013). <https://doi.org/10.1117/1.JEI.22.4.041106>, publisher: SPIE
9. Cheng, Z., Qin, L., Ye, Y., Huang, Q., Tian, Q.: Human daily action analysis with multi-view and color-depth data. In: Computer Vision–ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7–13, 2012, Proceedings, Part II 12. pp. 52–61. Springer (2012). https://doi.org/10.1007/978-3-642-33868-7_6
10. Chua, J.L., Chang, Y.C., Lim, W.K.: A simple vision-based fall detection technique for indoor video surveillance. *Signal, Image and Video Processing* **9**, 623–633 (2015). <https://doi.org/10.1007/s11760-013-0493-7>
11. Denkovski, S., Khan, S.S., Malamis, B., Moon, S.Y., Ye, B., Mihailidis, A.: Multi visual modality fall detection dataset. *IEEE Access* **10**, 106422–106435 (2022). <https://doi.org/10.1109/ACCESS.2022.3211939>, conference Name: IEEE Access

12. Epstein, D., Chen, B., Vondrick, C.: Oops! predicting unintentional action in video. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 916–926 (2020). <https://doi.org/10.1109/CVPR42600.2020.00100>
13. Fleming, J., Brayne, C.: Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. *Bmj* **337** (2008). <https://doi.org/10.1136/bmj.a2227>
14. Gan, Z., Jin, L., Nie, L., Wang, Z., Zhou, L., Li, L., Wang, Z., Li, J., Xing, J., Zhao, J.: Asquery: A query-based model for action segmentation. 2024 IEEE International Conference on Multimedia and Expo (ICME) pp. i–vi (2024). <https://doi.org/10.1109/ICME57554.2024.10687535>
15. Guerrero, J.C.E., España, E.M., Añasco, M.M., Lopera, J.E.P.: Dataset for human fall recognition in an uncontrolled environment. *Data in Brief* **45**, 108610 (2022). <https://doi.org/10.1016/j.dib.2022.108610>
16. Kwolek, B., Kepski, M.: Human fall detection on embedded platform using depth maps and wireless accelerometer. *Computer Methods and Programs in Biomedicine* **117**(3), 489–501 (2014). <https://doi.org/10.1016/j.cmpb.2014.09.005>
17. Maldonado-Bascon, S., Iglesias-Iglesias, C., Martín-Martín, P., Lafuente-Arroyo, S.: Fallen people detection capabilities using assistive robot. *Electronics* **8**(9), 915 (2019). <https://doi.org/10.3390/electronics8090915>
18. Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., Peñafort-Asturiano, C.: UP-fall detection dataset: A multimodal approach. *Sensors* **19**(9), 1988 (2019). <https://doi.org/10.3390/s19091988>, number: 9 Publisher: Multidisciplinary Digital Publishing Institute
19. Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al.: The prisma 2020 statement: an updated guideline for reporting systematic reviews. *bmj* **372** (2021). <https://doi.org/10.1136/bmj.n71>
20. Tinetti, M.E., Liu, W.L., Claus, E.B.: Predictors and prognosis of inability to get up after falls among elderly persons. *Jama* **269**(1), 65–70 (1993). <https://doi.org/10.1001/jama.1993.03500010075035>
21. Tong, Z., Song, Y., Wang, J., Wang, L.: Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training. *Advances in neural information processing systems* **35**, 10078–10093 (2022), https://proceedings.neurips.cc/paper_files/paper/2022/file/416f9cb3276121c42eebb86352a4354a-Paper-Conference.pdf
22. Tran, T.H., Le, T.L., Pham, D.T., Hoang, V.N., Khong, V.M., Tran, Q.T., Nguyen, T.S., Pham, C.: A multi-modal multi-view dataset for human fall analysis and preliminary investigation on modality. In: 2018 24th International Conference on Pattern Recognition (ICPR). pp. 1947–1952 (2018). <https://doi.org/10.1109/ICPR.2018.8546308>, ISSN: 1051-4651
23. Tschannen, M., Gritsenko, A., Wang, X., Naeem, M.F., Alabdulmohsin, I., Parthasarathy, N., Evans, T., Beyer, L., Xia, Y., Mustafa, B., et al.: Siglip 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features. *arXiv preprint arXiv:2502.14786* (2025). <https://doi.org/10.48550/arXiv.2502.14786>
24. U.S. Office of Management and Budget: Statistical policy directive no. 15: Standards for maintaining, collecting, and presenting federal data on race and ethnicity. <https://web.archive.org/web/20260625143817/https://www.federalregister.gov/documents/2024/03/29/2024-06469/revisions-to-ombs-statistical-policy->

- `directive-no-15-standards-for-maintaining-collecting-and` (2024), published March 28, 2024
25. Wan, T., Wang, A., Ai, B., Wen, B., Mao, C., Xie, C.W., Chen, D., Yu, F., Zhao, H., Yang, J., Zeng, J., Wang, J., Zhang, J., Zhou, J., Wang, J., Chen, J., Zhu, K., Zhao, K., Yan, K., Huang, L., Feng, M., Zhang, N., Li, P., Wu, P., Chu, R., Feng, R., Zhang, S., Sun, S., Fang, T., Wang, T., Gui, T., Weng, T., Shen, T., Lin, W., Wang, W., Wang, W., Zhou, W., Wang, W., Shen, W., Yu, W., Shi, X., Huang, X., Xu, X., Kou, Y., Lv, Y., Li, Y., Liu, Y., Wang, Y., Zhang, Y., Huang, Y., Li, Y., Wu, Y., Liu, Y., Pan, Y., Zheng, Y., Hong, Y., Shi, Y., Feng, Y., Jiang, Z., Han, Z., Wu, Z.F., Liu, Z.: Wan: Open and advanced large-scale video generative models. arXiv preprint arXiv:2503.20314 (2025). <https://doi.org/10.48550/arXiv.2503.20314>
 26. Wang, W., Gao, Z., Gu, L., Pu, H., Cui, L., Wei, X., Liu, Z., Jing, L., Ye, S., Shao, J., Wang, Z., Chen, Z., Zhang, H., Yang, G., Wang, H., Wei, Q., Yin, J., Li, W., Cui, E., Chen, G., et al.: InternVL3.5: Advancing open-source multimodal models in versatility, reasoning, and efficiency (Aug 2025). <https://doi.org/10.48550/arXiv.2508.18265>, arXiv:2508.18265 [cs]
 27. World Health Organization: Falls. <https://web.archive.org/web/20250423060144/https://www.who.int/news-room/fact-sheets/detail/falls/> (2021), fact sheet, last updated 26 April 2021. Estimates 684 000 fatal falls and 37.3 million falls requiring medical attention each year
 28. Zhang, Z., Conly, C., Athitsos, V.: Evaluating depth-based computer vision methods for fall detection under occlusions. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., McMahan, R., Jerald, J., Zhang, H., Drucker, S.M., Kambhamettu, C., El Choubassi, M., Deng, Z., Carlson, M. (eds.) *Advances in Visual Computing*, vol. 8888, pp. 196–207. Springer International Publishing (2014). https://doi.org/10.1007/978-3-319-14364-4_19, series Title: Lecture Notes in Computer Science