Eurasia Journal of Science and Technology
Print ISSN: 2663-1024
Online ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3168

FIXED-THRESHOLD RGB–POSE FALL DETECTION ACROSS PARTICIPANTS UNDER CONTROLLED IMAGE DEGRADATION

YanFei Yuan*, XiaoNi Liu
College of Cyber Security, Tarim University, Alar 843301, Xinjiang, China.
**Corresponding Author: YanFei Yuan*

Abstract: This study evaluates how fixed-threshold RGB–pose fall detection transfers across participants and responds to controlled image degradation. On UP-Fall, participants are separated into fitting, confirmation and final evaluation groups. A frozen image encoder and a geometric temporal pose classifier are combined by equal probability averaging at threshold 0.5. On final clean trials, the predefined primary seed achieves 94.42% window recall, 1.54% normal-window false-positive rate and 39/39 event hits. On matched trials, recall falls from 90.95% to 87.16% under darkening and 73.49% under strong blur. Across three seeds, mean blur recall is 44.87%, while valid pose geometry covers 19.91% of blurred sampled frames. The evaluation quantifies fixed-threshold tradeoffs and shows why event hits should be reported alongside window recall, seed sensitivity and pose coverage.
Keywords: Fall detection; RGB video; Human pose; Image degradation; Cross-participant evaluation

1 INTRODUCTION

Falls are brief events, so a detector must identify the falling phase while limiting false alarms during ordinary activity. This balance becomes more difficult when images are dark or blurred and when inference is intended for an edge device. An event-level hit alone may hide missed falling windows, delayed decisions, or failures in the visual front end. A useful evaluation should therefore separate participants across fitting and testing, fix the operating point in advance, report window-level errors, and measure whether intermediate pose information remains available.

Research on fall detection spans representation, temporal modeling, fusion, robustness and deployment cost. Human-pose sequences with temporal convolution have been studied for fall detection [1]. MobileNetV3 provides the compact visual architecture used as the image encoder here [2]. Other studies combine audio and video or inertial and visual signals, while research on viewpoint changes and occlusion examines augmentation and inference-time aggregation [3-5]. Edge-oriented studies evaluate lightweight or distilled models on embedded hardware, and graph-based multi-task learning provides another modeling direction [6-7]. Corruption benchmarks offer a general framework for stress testing [8]. These studies inform the evaluation design; differences in datasets, participant splits, labels, thresholds, hardware and metrics preclude direct comparison of their published scores with the present results.

A fixed-threshold RGB–pose pipeline provides a test case for examining cross-participant performance and the joint response of detection and pose availability to image degradation. Pairing clean and degraded trials holds trial membership, windowing and source-time labels constant. Measuring pose coverage alongside detection performance reveals whether event hits conceal missed falling windows, while multiple fixed seeds expose variability across fitted heads.

We evaluate this pipeline on UP-Fall [9]. A frozen MobileNetV3-Small encoder supplies RGB features, and a causal geometric temporal branch supplies pose features; their classifier heads are fitted on development participants. Separate participants are used for confirmation and final evaluation without refitting. Four decision rules are evaluated, with equal probability averaging predefined as the primary comparator. The threshold is fixed at 0.5. We ask: (1) how does the fixed pipeline transfer to unseen participants; (2) how do controlled darkening and blur change recall, false-positive rate and event observability; and (3) how does valid pose coverage change alongside detection performance?

This paper contributes a participant-separated, fixed-threshold evaluation that combines window metrics, event hits, source-time latency and seed sensitivity. A paired degradation analysis relates detection performance to pose availability and includes all four decision rules. A Raspberry Pi 5 CPU replay measures processing cost and checks decision agreement and numerical consistency on a fixed development clip.

2 MATERIALS AND METHODS

The evaluation combines participant-separated data partitions, RGB and pose representations, and a fixed-threshold protocol for assessing detection performance under clean and degraded image conditions.

2.1 Data and Split

All trials and degradation derivatives of a participant remain in one partition. Subjects 1–6 provide 18,339 development windows with 610 positives. Subjects 7 and 9 are confirmation participants without refitting. Subjects 10–12 are final participants. The final acquisition contains 99 image archives and 99 CSV files. Six trials containing the undocumented

© By the Author(s) 2026, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

label 20 (Subject 10 and Subject 12, Activity 2, Trials 1–3) are quarantined in full. The remaining 93 eligible trials contain 8,903 retained windows, 208 positive windows and 39 annotated fall events. Timestamp eligibility rejects 2,454 additional candidate windows, including 35 positive windows; the reported results therefore concern retained windows. All 39 final events have an eligible window. The earlier confirmation has one unobservable event, which is retained in its original event denominator in the supplement. Official activity labels 1–5 denote five fall types and are treated as positive at the frame level; labels 6–11 denote negative activities, including lying.
Table 1 summarizes the participant partitions. Figure 1 shows the two representations derived from the same RGB stream and the fixed decision rules.

Table 1 Participant Partitions and Final-Evaluation Sample

Partition	Participants	Use and retained sample
Development	Subjects 1–6	Fit scalers and heads; 18,339 windows, 610 positive windows.
Confirmation	Subjects 7 and 9	Separate confirmation; no refitting.
Final evaluation	Subjects 10–12	93 eligible trials; 8,903 windows, 208 positives, 39 fall events.
Quarantined	Subjects 10 and 12	Activity 2, Trials 1–3: six trials with unknown labels.

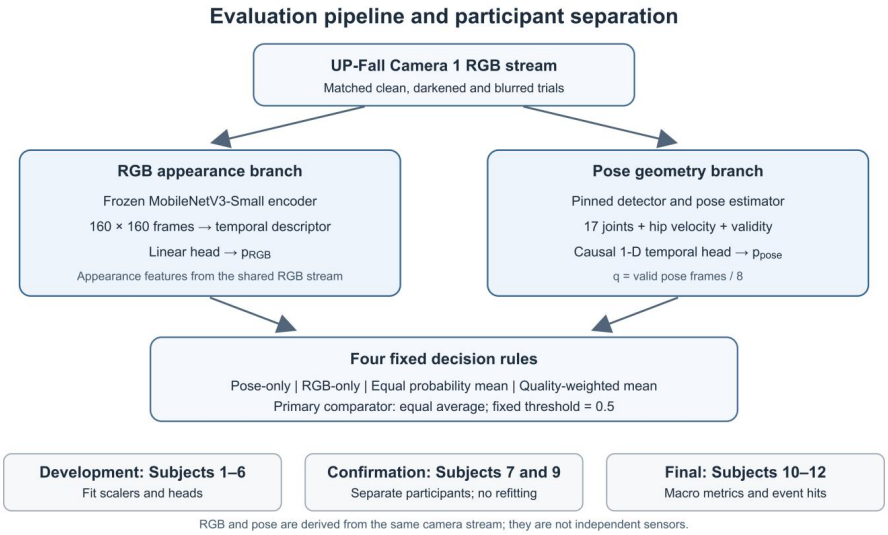


Figure 1 Evaluation Pipeline and Participant Separation. Both branches represent the same Camera 1 RGB stream.

Frames are sampled at 8 Hz using the latest frame not later than the target timestamp. A causal window has eight samples and emits at 4 Hz. A window is rejected when any frame lag is at least 100 ms. Each window receives the official phase label of its last sampled source frame; the same source timestamp defines its emission time for event scoring. No temporal frame padding, temporal interpolation or post-hoc smoothing is used in window construction.

2.2 Representations and Heads

The RGB branch uses a frozen MobileNetV3-Small ImageNet encoder with 160x160 letterboxed images [2]. Spatial mean pooling yields a 576-dimensional vector per frame; its temporal mean, mean absolute adjacent difference and last-frame vector are concatenated into a 1,728-dimensional descriptor. A training-only mean/std scaler feeds a linear classifier.
The pose branch uses the pinned OpenCV Zoo MediaPipe person detector and pose estimator, with implementation lineage described by MediaPipe and BlazePose [10-12]. The saved commit is 47534e27c9851bb1128ccc0102f1145e27f23f98. The highest detector-confidence person is selected without ground-truth regions. Seventeen joints, confidence values, two normalized hip velocities and one validity indicator form 54 dimensions per frame. Landmark confidence is the minimum of the two saved confidence channels. Both shoulders and hips require confidence at least 0.5, and torso length must be at least 1% of image height. Coordinates are hip-centered, torso-normalized and clipped to [-8,8]; low-confidence joint coordinates are zeroed. Valid consecutive hips yield velocity only at positive intervals at most 0.2 seconds; normalized velocities are clipped to [-8,8]. Invalid geometry is retained as a zero vector. Two causal 1-D convolutions (54-32-32, kernel 3, ReLU), each with two positions of left zero-padding, temporal mean pooling and a 32-1 linear layer form the head.
Both heads use weighted binary cross-entropy, AdamW (learning rate 0.003, weight decay 0.0001), batch size 64, gradient clipping at 1 and 25 fixed epochs. The positive loss weight is the development negative/positive count ratio; RGB standard deviations are floored at 1e-4. Seeds are 20261015, 20261016 and 20261017. The primary seed is 20261015 and the threshold is fixed at 0.5.
Four decision rules are retained: pose only, RGB only, equal probability averaging, and a quality-weighted mean with pose weight q and RGB weight (1-q). Here q is the fraction of eight frames with valid pose geometry. No smoothing,

debounce or cooldown is applied. Darkening is floor(uint8 pixel x 0.25). Blur is a 21x21 Gaussian kernel with sigma 5. Degradation is applied before resizing and is compared with the paired clean Trial1 subset.

Let $p_{RGB,t}$ and $p_{pose,t}$ denote the positive-class probabilities for window t. Equations (1)–(3) formalize the two fusion rules and the fixed decision threshold. Here v_t counts frames satisfying the pose-validity criteria defined above, so $0 \leq q_t \leq 1$.

$$p_t^{equal} = \frac{p_{RGB,t}+p_{pose,t}}{2} \tag{1}$$

$$q_t = \frac{v_t}{8}, p_t^{quality} = q_t p_{pose,t} + (1 - q_t) p_{RGB,t} \tag{2}$$

$$\hat{y}_t = 1[p_t \geq 0.5] \tag{3}$$

The indicator $1[\cdot]$ equals one when its condition is true and zero otherwise; p_t is the score produced by the selected decision rule.

2.3 Metrics

Recall is TP/(TP+FN), and normal-window FPR is FP/(FP+TN). Recall, FPR and average precision (AP) are computed separately for Subjects 10–12 and then macro-averaged. Event hits are summed over those participants. An event is hit when at least one positive decision occurs within its official falling interval. Latency is measured in source time from onset to first hit, excluding device computation and transport. Participants are the independent units; overlapping windows and repeated seeds provide within-participant evaluations.

For participant s, TP, FN, FP and TN count retained windows. Equations (4) and (5) define participant-level rates and their macro average; M denotes recall, FPR or AP, and S = {10, 11, 12}.

$$Recall_s = \frac{TP_s}{TP_s+FN_s}, FPR_s = \frac{FP_s}{FP_s+TN_s} \tag{4}$$

$$M_{macro} = \frac{M_{10}+M_{11}+M_{12}}{3} \tag{5}$$

For event e, T_e contains eligible window-emission times within the official falling interval. Equation (6) defines an event hit. Equation (7) defines pose coverage over a frame set F, where g_j indicates valid geometry for frame j; F is either all sampled frames or only falling-phase frames.

$$H_e = 1[\exists t \in T_e : \hat{y}_t = 1] \tag{6}$$

$$C(F) = \frac{\sum_{j \in F} g_j}{|F|} \times 100\% \tag{7}$$

3 RESULTS

3.1 Primary Seed

Table 2 Primary-Seed Performance at the Fixed Threshold, Macro-Averaged across Final Participants

Condition	Method	Recall (macro)	FPR (macro)	AP (macro)	Events
clean	pose	96.08%	3.53%	0.8758	39/39
clean	rgb	86.49%	1.54%	0.8072	39/39
clean	equal_average	94.42%	1.54%	0.8776	39/39
clean	quality_fallback	96.54%	3.27%	0.8564	39/39
clean_trial1	pose	95.38%	3.13%	0.8788	13/13
clean_trial1	rgb	86.33%	1.57%	0.8100	13/13
clean_trial1	equal_average	90.95%	1.42%	0.8904	13/13
clean_trial1	quality_fallback	95.38%	2.89%	0.8588	13/13
dark025	pose	77.57%	14.02%	0.3878	13/13
dark025	rgb	87.54%	10.65%	0.5638	13/13
dark025	equal_average	87.16%	4.30%	0.7133	13/13
dark025	quality_fallback	90.03%	14.43%	0.4516	13/13
blur21sigma5	pose	17.97%	9.48%	0.1234	5/13
blur21sigma5	rgb	93.26%	5.52%	0.7964	13/13
blur21sigma5	equal_average	73.49%	1.26%	0.5643	13/13
blur21sigma5	quality_fallback	94.15%	12.03%	0.4789	13/13

The degradation comparison uses paired clean Trial1 trials as the reference. Under blur, equal averaging yields 73.49% recall and 1.26% FPR, compared with 93.26% and 5.52% for RGB alone; the corresponding AP values are 0.5643 and 0.7964. Equal averaging therefore lowers normal-window FPR at this operating point while also reducing recall.

Figure 2 displays the Table 2 operating points with separate recall and FPR scales. The paired clean subset provides the reference for darkening and blur.

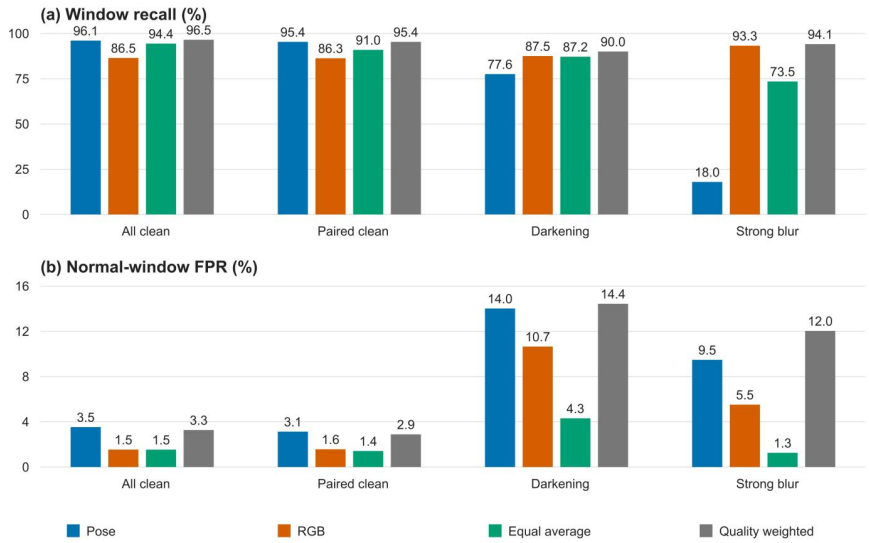


Figure 2 Primary-Seed Recall and Normal-Window FPR, Macro-Averaged over Subjects 10–12. All clean includes 93 trials; paired clean and each degradation include the same 31 Trial1 trials. Bars show operating points at threshold 0.5. The recall and FPR axes use separate percentage scales.

3.2 Seed Sensitivity and Front-End Coverage

Table 3 Equal-Average Recall by Seed and the Three-Seed Mean

Condition	Seed 20261015	Seed 20261016	Seed 20261017	Three-seed mean
clean recall	94.42%	88.39%	91.55%	91.45%
clean Trial1 recall	90.95%	88.80%	88.80%	89.52%
dark025 recall	87.16%	58.12%	71.34%	72.20%
blur recall	73.49%	25.03%	36.10%	44.87%

Valid pose geometry covers 78.82% of clean Trial1 sampled frames, 56.68% under darkening and 19.91% under blur. Among falling-phase frames, coverage falls from 96.49% in clean Trial1 to 14.04% under blur. Each seed evaluates the same 13 events in each degraded condition (Table 3).

Figure 3 reports overall pose coverage and falling-phase coverage separately, showing how the availability of valid geometry changes across the paired conditions (Table 4).

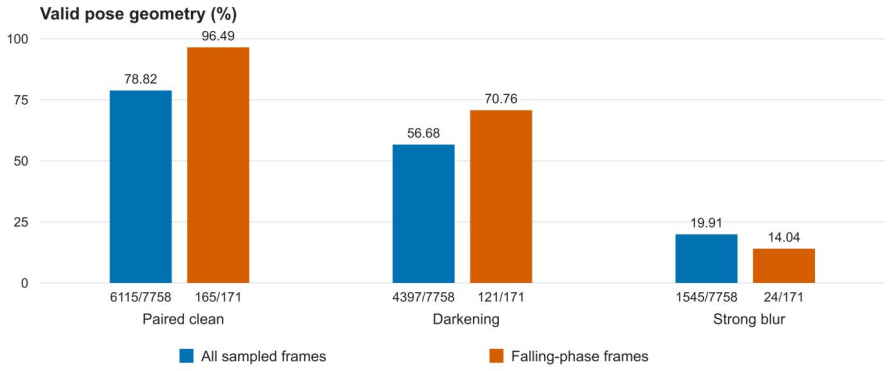


Figure 3 Pose Coverage in Paired Trial1 Conditions, Computed Using Eq. (7). Table 4 gives valid/total frame counts and rates. Frame counts define the coverage denominators; participants remain the independent evaluation units.

Table 4 Pose-Geometry Coverage by Paired Condition

Condition	All sampled frames with valid geometry	Falling-phase frames with valid geometry
Paired clean	6,115/7,758 (78.82%)	165/171 (96.49%)
Darkening	4,397/7,758 (56.68%)	121/171 (70.76%)
Strong blur	1,545/7,758 (19.91%)	24/171 (14.04%)

3.3 Raspberry Pi CPU Replay

The frozen primary pipeline was replayed on a Raspberry Pi 5 with 8 GB RAM (ARM64, CPU only, two threads). Three fresh processes each used the same development trial S01A01T1, with 80 selected PNG frames and 37 output windows from 9.88 s of source video. Full selected-frame processing took 9.269–9.295 s, corresponding to 8.61–8.63

frames/s. The highest sampled temperature was 50.15 °C, with no observed power/throttling flag. Peak process RSS was 421.5 MiB including measurement capture. The three runs measure processing repeatability on one development clip (Table 5).

Table 5 Repeated Pi 5 Replay Runs on the Same Development Clip

Run	Selected frames/s	Processing seconds	Development decisions matching PC
1	8.607	9.295	37/37
2	8.631	9.269	37/37
3	8.614	9.287	37/37

All primary decisions matched the PC reference, and same-device feature/probability arrays were bit-identical across runs. Cross-platform pose-feature and probability differences exceeded the original strict tolerances; the assertions and error records were retained. Raw pose coordinates differed by up to 0.234 pixels, while frozen-head inference on saved Pi features reproduced the device probabilities. Maximum fused-probability difference was 0.000389, and the nearest Pi fusion score was 0.000077 from 0.5. These measurements establish decision agreement for the replayed clip and quantify the remaining numerical differences. The timed workload includes selected-PNG reads and processing, excluding live capture, unselected-frame decoding and real-time queues.

4 DISCUSSION AND LIMITATIONS

Event hits, window recall and pose coverage reveal different aspects of performance under image degradation. Across three seeds, equal-average mean recall/FPR are 91.45%/1.03% for full clean, 72.20%/2.41% for darkening and 44.87%/0.60% for blur. The primary seed retains all degraded event hits despite lower window recall. Performance under strong blur varies across fitted heads, and reduced pose coverage accompanies these changes. The present design measures this association without isolating its causal contribution. Window FPR measures the proportion of normal windows classified as positive; estimating false alarms per hour requires an alarm-level evaluation. Full per-participant, per-method and per-seed records, lying negatives and source-time latencies are supplied in the supplementary tables.
The final evaluation covers three healthy young volunteers performing simulated falls in one controlled room. Both representations use the same Camera 1 RGB stream. Generalization to older adults, clinical settings, other cameras and backgrounds requires corresponding evaluation data. The six trials containing undocumented labels and the timestamp-based window exclusions define the analyzed sample. Device measurements cover one development clip; sustained live-camera performance and end-to-end latency require separate measurement.

5 CONCLUSION

On controlled UP-Fall trials from three final participants, the fixed RGB–pose baseline achieves 91.45% mean window recall across three seeds on clean data. The primary seed hits all 13 events in each degraded condition, while the three-seed mean recall under strong blur is 44.87%. These results show why event hits should be interpreted together with window recall, pose availability and seed sensitivity. The Pi replay characterizes CPU processing and numerical agreement on a fixed development clip.

DATA AVAILABILITY

UP-Fall images and labels are obtained from the original dataset at https://sites.google.com/up.edu.mx/har-up/. This study is a secondary analysis of public data. The dataset publication reports research-committee approval at Universidad Panamericana and voluntary participation [9]. The supplementary material provides the fixed protocol, final trial membership and all saved per-participant, per-method and per-seed metric records. Raw media and third-party pretrained weights are excluded from redistribution. The device supplement supplies three short Pi replay runs, original numerical-failure records and their independent audit.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This paper is supported by the Tarim University Euphrates Poplar Talent Introduction Scientific Research Foundation Project (TDZKSS202543).

REFERENCES

[1] Qi L, Sun H. Multistage fall detection framework via 3D pose sequences and TCN integration. Scientific Reports, 2025, 15(1): 27832.
[2] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1314-1324.

[3] Jiao S, Li G, Zhang G, et al. Multimodal fall detection for solitary individuals based on audio-video decision fusion processing. Heliyon, 2024, 10(8): e29596.
[4] Rehouma H, Boukadoum M. Fall detection by deep learning-based bimodal movement and pose sensing with late fusion. Sensors, 2025, 25(19): 6035.
[5] Zobi M, Bolzani L, Tabii Y, et al. Robust 3D Skeletal Joint Fall Detection in Occluded and Rotated Views Using Data Augmentation and Inference–Time Aggregation. Sensors, 2025, 25(21): 6783.
[6] Pham Q N, Ninh D K. Lightweight fall detection on edge devices via knowledge distillation. The University of Danang – Journal of Science and Technology, 2025: 74-78.
[7] Zhao X, Jin P. Graph Topology-Guided Multi-Task Learning for Fall Detection. Sensors, 2026, 26(17): 5529.
[8] Hendrycks D, Dietterich T. Benchmarking neural network robustness to common corruptions and perturbations. arXiv preprint arXiv:1903.12261, 2019.
[9] Martínez-Villaseñor L, Ponce H, Brieva J, et al. UP-fall detection dataset: A multimodal approach. Sensors, 2019, 19(9): 1988.
[10] OpenCV Zoo. Person detection and pose estimation MediaPipe model implementations, commit 47534e27c9851bb1128ccc0102f1145e27f23f98, 2026. https://github.com/opencv/opencv_zoo/tree/47534e27c9851bb1128ccc0102f1145e27f23f98/models.
[11] Lugaresi C, Tang J, Nash H, et al. Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172, 2019.
[12] Bazarevsky V, Grishchenko I, Raveendran K, et al. Blazepose: On-device real-time body pose tracking. arXiv preprint arXiv:2006.10204, 2020.