Multimodal Confirmation for Reducing False Fall Alerts: A Simulation Study of the ResQ Decision Policy

Sofia Bouazizi
Walnut Grove High School
Prosper, Texas, USA
Sofia.Bouazizi@gmail.com
First author: high school student

***Abstract*—Camera-based fall detectors can identify that a person is on the floor but cannot reliably determine whether the event is an emergency. ResQ is a proposed privacy-preserving emergency guidance system that pairs on-device pose analysis with a conversational confirmation stage. This paper evaluates that decision policy independently of any unbuilt hardware, using a Monte Carlo simulation of 100,000 monitoring events across 27 operating conditions spanning fall prevalence, pose sensitivity, and pose false-positive rate. In the primary scenario, false emergency alerts fell from 75.5 to 2.9 per 1,000 monitored events, precision rose from 0.378 to 0.932, and F1 rose from 0.537 to 0.864, while recall fell from 0.926 to 0.805. Across all 27 conditions the mean false-alert reduction was 95.9% and the mean recall change was -12.1 percentage points. We also derive the closed-form behavior of the policy and show that the false-alert reduction equals the confirmation reliability exactly and that the recall cost depends only on the human-response parameters, not on the vision model. This decoupling means the perception stage and the interaction stage can be tuned independently, and it identifies user-response reliability rather than detector accuracy as the safety-critical unknown. All numbers reported here are simulation outputs and must not be read as measured device or clinical performance.**

***Index Terms*—fall detection, human pose estimation, multimodal systems, edge AI, decision policy, cost-sensitive classification, simulation**

I. Introduction

Automated fall detection is attractive for people who live alone, because an injured person may be unable to reach a wearable button. A camera-based system, however, faces a context problem. Lying down to rest, stretching on the floor, reaching under furniture, and an injurious fall can all produce similar static poses. A detector that treats every floor posture as an emergency will generate frequent false alarms.

False alarms are not a cosmetic problem. In monitoring systems the negative class dominates, so even a modest false-positive rate produces far more false alerts than true ones. The result is alert fatigue, unnecessary caregiver interruption, and eventual loss of trust in the system.

The public ResQ concept [1] addresses this by adding a conversational check-in between visual detection and emergency escalation. Local MoveNet pose estimation [2] supplies the visual trigger, and a risk-assessment policy decides whether that trigger becomes an alert. Prior work shows that lightweight pose estimation supports real-time fall detection [3], but public fall corpora remain dominated by staged actions and controlled camera views [4]–[6]. A system should therefore be evaluated not only on visual sensitivity but on the downstream policy that converts a visual trigger into an emergency alert.

This paper isolates that policy question. Under a range of plausible pose-detector error rates, how much can a user-confirmation stage reduce false emergency alerts, and what loss in recall does that reduction impose? Framed in data-mining terms, this is a cost-sensitive cascade over a highly imbalanced event stream, where the second stage is a noisy human oracle rather than a second classifier.

The contributions are as follows. First, a reproducible simulation framework that separates decision-policy performance from visual-model accuracy. Second, a closed-form analysis showing that the precision gain of the confirmation stage is determined entirely by confirmation reliability, while the recall cost is determined entirely by human-response behavior, with the two effects independent of the vision model. Third, an explicit statement of the assumptions and safety risks that must be tested before any deployment. This paper does not claim that ResQ has achieved the reported metrics on real data.

II. ResQ System Concept

A. Proposed Architecture

The proposed architecture has four logical components: an on-device visual pose estimator; a temporal fall-candidate rule; a conversational confirmation stage; and a severity and escalation policy. The public prototype description specifies browser-based TensorFlow.js and MoveNet processing, with future hardware comprising a Raspberry Pi camera, a wearable trigger, a mobile robot, solar charging, and low-connectivity communication. Only the decision logic in the last two components is evaluated here.

B. Safety Logic

The pose-only baseline sends an emergency alert whenever the visual stage fires. The ResQ policy instead initiates a check-in. A responsive user who confirms safety cancels the alert. A user who reports distress, or who does not respond before a timeout, is escalated. Non-response therefore escalates rather than suppresses, so that an unconscious user is not silently ignored.

III. Methods

A. Study Design

A Monte Carlo simulation generated $N = 100{,}000$ independent monitoring events for each of 27 operating conditions. The conditions were the Cartesian product of fall prevalence $p \in \{0.01, 0.05, 0.10\}$, pose sensitivity $s \in \{0.88, 0.93, 0.97\}$, and pose-stage false-positive rate $f \in \{0.03, 0.08, 0.15\}$. These are sensitivity-analysis inputs chosen to bracket published pose-based detector performance. They are not measurements of the current ResQ prototype.

B. Confirmation Model

Three human-response parameters govern the confirmation stage. For a true fall reaching confirmation, a fraction $u = 0.12$ of users were modeled as unresponsive and therefore escalated automatically. Among responsive true-fall users, distress was reported with probability $d = 0.85$. For non-fall events incorrectly flagged by the pose stage, the simulated user correctly confirmed safety with probability $c = 0.96$; the remaining 4% represented non-response, misunderstanding, communication failure, or a precautionary distress report.

C. Compared Policies

Policy 1, pose-only, escalated every pose-positive event. Policy 2, ResQ confirmation, escalated a pose-positive event only when a true-fall user was unresponsive or reported distress, or when a non-fall user failed to confirm safety. Events missed by the pose stage could not be recovered by either policy.

D. Analytic Model

Because confirmation is conditional on the pose trigger, the policy admits a closed form that is useful for interpreting the simulation. For the pose-only policy, recall is s and the expected false alerts per 1,000 events are $1000(1-p)f$. For the ResQ policy,

$$R_{\text{ResQ}} = s\left[u + (1-u)\,d\right], \tag{1}$$

$$\text{FA}_{\text{ResQ}} = 1000\,(1-p)\,f\,(1-c). \tag{2}$$

Dividing (2) by the pose-only expression gives a false-alert reduction of exactly c, independent of p, s, and f. Likewise, (1) shows that the fraction of detected falls retained, $u+(1-u)d$, depends only on the human-response parameters. We return to this decoupling in Section IV-C.

E. Outcomes and Reproducibility

We computed precision, recall, specificity, F1 score, and false emergency alerts per 1,000 monitoring events. A fixed pseudorandom seed (20260718) was used. The simulation script, environment file, parameter grid, and complete cell-level results accompany this paper.

TABLE I
Primary Simulation Scenario ($p = 0.05$, $s = 0.93$, $f = 0.08$)

Policy	Prec.	Recall	Spec.	F1	FA/1,000
Pose-only	0.378	0.926	0.921	0.537	75.48
ResQ confirmation	0.932	0.805	0.997	0.864	2.90

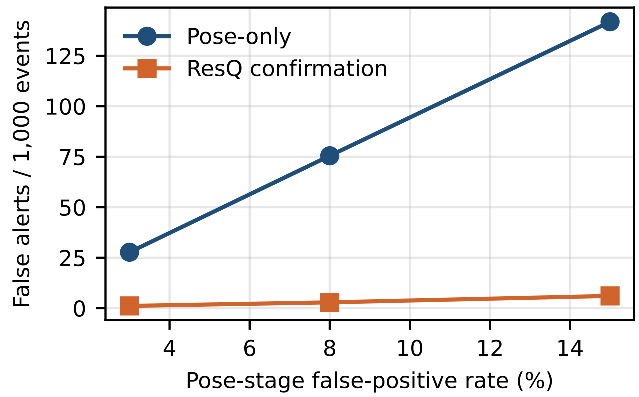

Fig. 1. False emergency alerts per 1,000 events at 5% fall prevalence and 0.93 pose sensitivity.

IV. Results

A. Primary Scenario

The primary scenario used $p = 0.05$, $s = 0.93$, and $f = 0.08$. The pose-only policy produced 75.5 false emergency alerts per 1,000 events; ResQ confirmation reduced this to 2.9. Precision rose from 0.378 to 0.932 and F1 from 0.537 to 0.864. The benefit came with lower recall, 0.805 against 0.926 for pose-only. Table I summarizes these results and Fig. 1 shows the false-alert gap widening as the pose stage degrades.

B. Sensitivity Analysis

Across all 27 conditions, the mean reduction in false emergency alerts was 95.95% (range 94.91%–96.36%) and the mean recall change was -12.12 percentage points (range -14.34 to -10.59). Fig. 2 shows that every ResQ operating point lies above the pose-only cloud in precision and to its left in recall, and that the two groups do not overlap.

C. Analytic Decomposition

The observed spread around 95.95% is Monte Carlo sampling noise, not a real dependence on the operating condition. By (2), the false-alert reduction equals the assumed confirmation reliability $c = 0.96$ exactly, for every cell in the grid. This is an important qualification: the headline precision improvement is an arithmetic consequence of the confirmation assumption and is not evidence that confirmation works. The simulation neither confirms nor tests c.

What the analysis does establish is a decoupling. The precision gain is governed only by c, and the recall cost is governed only by u and d; neither depends on pose sensitivity or pose false-positive rate. Two design consequences follow. First, the

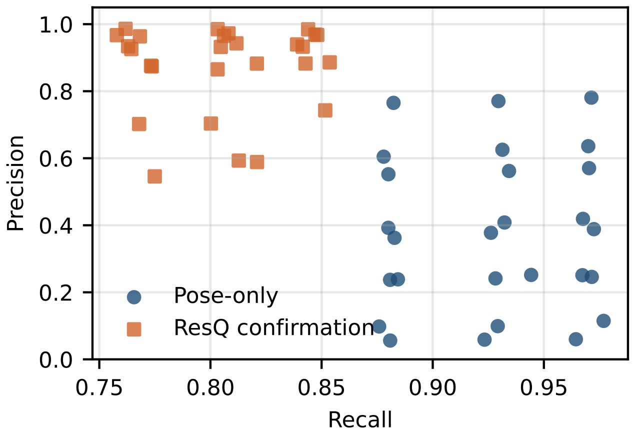

Fig. 2. Precision–recall trade-off across all 27 simulated operating conditions.

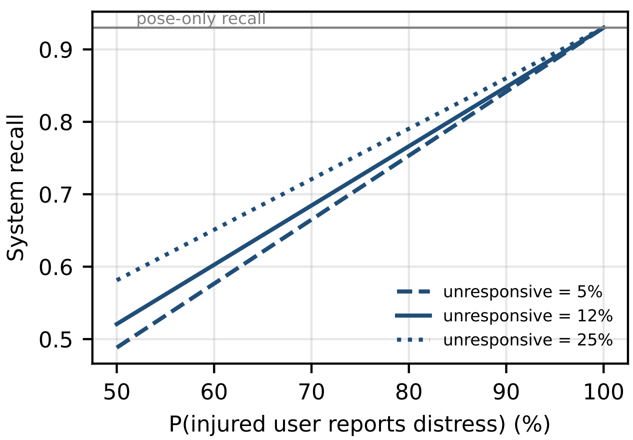

Fig. 3. System recall as a function of the probability that an injured user reports distress, for three unresponsive fractions ($s = 0.93$). Non-response escalation sets the floor.

perception stage and the interaction stage can be engineered and evaluated independently, since improving the pose model does not change the policy's relative effect. Second, the safety-critical unknown is human-response reliability, not detector accuracy, and it is the parameter that a validation study must measure first.

Fig. 3 makes the exposure concrete. Recall degrades linearly in d, and only the unresponsive fraction u provides a floor. If injured users report distress reliably ($d \geq 0.95$), recall stays within 4 points of pose-only. If they do so only 60% of the time, recall falls below 0.60 at $u = 0.05$: worse than no confirmation stage at all, and dangerously so. A policy that depends on a single spoken answer sits on the steep part of this curve.

V. Discussion

A. Main Finding

A conversational confirmation layer can substantially improve the precision of an emergency-alert system whose visual detector produces frequent false positives, and it does so uniformly across the operating conditions examined. Confirmation is not free, however. True falls can be incorrectly canceled when a responsive but injured user fails to report distress, misunderstands the prompt, or answers inaccurately. Under the assumed parameters, roughly one detected fall in eight is lost this way.

B. Safety Implications

A deployable policy should not depend on a single spoken answer. Safer escalation can combine repeated prompts, confidence-aware speech recognition, observed immobility, rapid posture change, wearable-button input, and a hard maximum timeout. Each of these raises the effective floor in Fig. 3. The system should communicate uncertainty and should not attempt to diagnose injuries. In a high-risk situation, failure to obtain a reliable response must increase, not decrease, the escalation level.

C. Privacy and Accessibility

On-device pose estimation reduces the need to transmit raw video, but local processing alone does not eliminate privacy risk. A real implementation requires clear data-retention rules, user-visible recording status, secure logs, and the ability to disable or reposition the camera. The confirmation interaction must also accommodate hearing, speech, language, cognitive, and connectivity differences through multimodal prompts and caregiver-configurable fallbacks.

D. Limitations

This study simulates downstream decision logic only. It does not run MoveNet on labeled fall videos, measure browser or Raspberry Pi latency, evaluate speech recognition, test the wearable trigger, or involve patients or caregivers. The response probabilities are stated explicitly but are not empirically estimated, and, as Section IV-C shows, the headline result is a direct function of one of them. Independence between monitoring events is also a simplification; real falls cluster in time and by individual. The numerical results must not be reported as ResQ device accuracy or clinical effectiveness.

VI. Empirical Validation Plan

The next experiment should evaluate the complete software pipeline on at least two labeled RGB fall datasets [5], [6] using subject-independent splits. MoveNet keypoints would be extracted frame by frame, followed by a temporal fall-candidate rule based on body orientation, vertical descent, and post-fall immobility. The preregistered comparison should include pose-only escalation, pose plus fixed confirmation, and confidence-aware escalation. Primary outcomes should be event-level sensitivity, specificity, precision, F1, false alerts per hour, and time-to-alert. Thresholds must be selected on a validation split, with final results reported once on a held-out test set.

Because the present analysis identifies c, u, and d as the parameters that dominate the policy, a separate human-factors study is required to estimate them rather than assume them. Such work should use scripted, non-dangerous scenarios, informed consent, adult supervision, and institutional ethics review where applicable. No participant should be asked to perform a hazardous fall.

VII. Conclusion

ResQ's multimodal design addresses a real weakness of pose-only fall detection: a visual trigger does not establish emergency context. In a reproducible simulation, confirmation reduced false alerts by about 96% and raised F1 from 0.537 to 0.864 under the stated assumptions, while lowering recall by about 12 percentage points. The accompanying analytic result shows that the precision gain follows directly from the assumed confirmation reliability, and that the recall cost is set entirely by human-response behavior and is independent of the vision model. The practical implication is that the next round of work should measure human response, not tune the detector. This paper is a simulation study and a foundation for, not a substitute for, benchmark and user testing.

Acknowledgment

The author thanks Salma Mohamed Elhady for contributions to the original ResQ project concept and physical-design materials.

Data and Code Availability

The simulation script, parameter grid, random seed, environment file, and complete cell-level result table accompany this paper.

References

[1] S. Bouazizi, "ResQ AI," project description, 2026. [Online]. Available: https://sofiabouazizi.org/resq.html
[2] R. Votel and N. Li, "Next-generation pose detection with MoveNet and TensorFlow.js," Google Research / TensorFlow Blog, 2021.
[3] E. Alam, A. Sufian, P. Dutta, and M. Leo, "Real-time human fall detection using a lightweight pose estimation technique," arXiv:2401.01587, 2024.
[4] A. Sucerquia, J. D. López, and J. F. Vargas-Bonilla, "SisFall: A fall and movement dataset," *Sensors*, vol. 17, no. 1, art. 198, 2017.
[5] S. Denkovski *et al.*, "Multi visual modality fall detection dataset," arXiv:2206.12740, 2022.
[6] D. Schneider *et al.*, "OmniFall: A unified staged-to-wild benchmark for human fall detection," arXiv:2505.19889, 2025.