Print ISSN: 3104-932X
Online ISSN: 3104-9338

Design and Experimental Scenario Evaluation of a Vision-Based Walking Assistance System for Older Adults

Linglong Yang, Xu Wang*

Department of Intelligent Manufacturing Engineering, Chengdu College of the University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

Received: Oct 25, 2026

Revised: Nov 8, 2026

Accepted: Nov 18, 2026

Published online: November 20, 2026

To appear in: International Journal of Advanced AI Applications, Vol. 3, No. 2 (February 2027)

* Corresponding Author: Xu Wang (34106831@qq.com)

Abstract. Pavement hazards require walking assistance systems that connect environmental sensing with timely and understandable warnings. This paper presents a wearable design using an Orange Pi 5, an ESP32-S3, RGB vision, array time of flight sensing, sequential polarisation and inertial monitoring. Directional vibration, speech and 4G/GNSS reporting support local warnings and post-fall assistance. An experimental evaluation scenario specifies YOLOv8n with Ultralytics 8.2.103, 640 × 640 inputs and 100 training epochs. Its dataset specification contains 3200 images, divided by route into 2240 training, 480 validation and 480 test images. The experimental test matrix contains 280 true positives, 15 false positives, 165 true negatives and 20 false negatives, giving 92.71% accuracy. The experimental scenario yields 25 frames/s pipeline throughput and 135 ms mean warning latency. Original reported stone detection results are distinguished from these prototype-tested outcomes, with 91.1% object precision and 91.2% recall. The analysis quantifies the accuracy, throughput and persistence trade-offs under clearly-defined experimental conditions. These outcomes are experimental measurements obtained from model training and hardware device tests, while slippery-surface warning and fall assistance remain design capabilities.

Keywords: *edge computing; inertial sensing; time of flight; polarisation imaging; haptic feedback*

1. Introduction

Population ageing has increased the need for products that support independent mobility and timely assistance. For older adults, walking safety depends on both the surrounding

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

environment and the ability to respond when balance is lost. Stones, raised surfaces, steps, kerbs, standing water and slippery pavement can interrupt an otherwise familiar route. An assistance device therefore needs to identify relevant environmental hazards, communicate their direction clearly, and provide a way to request help after a suspected fall. These functions require an integrated design in which sensing, computation, feedback and communication operate as a coordinated system.

Research on walking aids has addressed ergonomics, user acceptance and the relationship between physical structure and everyday use. Ma [1] investigated a lower limb exoskeleton walking aid using ergonomic principles and anthropometric information to improve the relationship between the wearer and the device. Liu and Qiao [2] examined the design of an intelligent walking aid through a modified technology acceptance model. Their work highlights the importance of perceived usability and the user's experience, in addition to the technical capabilities of a product. A device that provides useful information but is uncomfortable or difficult to operate may have limited practical value.

Other design studies have considered how different requirements can be reconciled during product development. Zhang et al. [3] investigated a walking aid using KANO, the analytic hierarchy process and quality function deployment. Chen et al. [4] studied lower limb exoskeleton design with product semantics. Together, these studies provide relevant perspectives on translating user needs into product features and on reducing the burden associated with wearing an assistive device. They do not, by themselves, establish the performance of an environmental warning system. The present work therefore uses the relationship between required functions and system modules as the starting point for an electronic walking safety assistant.

Fall recognition is a complementary requirement because environmental warnings cannot prevent every loss of balance. Liu et al. [5] investigated deep learning methods for recognising falls in older adults. Recognition after an event and warning before an obstacle is reached address different stages of the safety problem. A forward facing camera can provide information about the pavement, while an inertial sensor can continue to monitor body motion when the camera is obscured or directed away from the body. Combining these sources offers a practical basis for covering both stages, provided their different fields of view and failure modes are recognised.

This study presents a wearable system built around an Orange Pi 5 edge computer and an ESP32-S3 real time controller. RGB images support obstacle detection and walkable surface

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

segmentation; an array time of flight sensor provides distance information; sequential polarisation images supply supplementary surface cues; and an inertial measurement unit supports suspected fall detection. Directional vibration, spoken messages and a cellular positioning link connect the sensing results to the wearer and a remote supervisor. The principal contribution is the integration of these functions into a documented hardware and software design, accompanied by preliminary stone detection results. The reported evidence supports the visual detection component, while complete outdoor reliability, fall detection performance and battery endurance remain subjects for further measurement.

2. Methodology

2.1. Functional Requirements and System Architecture

The design begins with the hazards and actions that the device is intended to address. A pavement warning should indicate where a potential obstacle lies in relation to the wearer, rather than merely reporting that an object exists somewhere in an image. A suspected fall should initiate a local check and a help request sequence. The system must also expose operating status to a supervisor so that low battery, communication loss or sensor problems can be recognised. Table 1 maps the original functional requirements to the corresponding modules; Figure 1 shows the overall organisation.

Table 1. Functions and corresponding system modules.

Function	Corresponding Module
System control	Main controller module
Fall detection	Fall detection module
Directional vibration	Vibration feedback module
Voice warning	Voice warning module
Human–machine interaction	Display module
Wireless communication	Bluetooth module

The Orange Pi 5 performs computationally demanding perception and fusion tasks. The ESP32-S3 handles frequent inertial sampling and the directional vibration outputs. This division separates image processing from operations that require predictable access to sensors and actuators. A delay in neural network inference should not prevent the controller from collecting inertial measurements or stopping an active motor. Communication between the processors conveys measurements, state information and feedback commands. The design also includes a power module, a display interface and a Bluetooth module, with cellular

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

communication and GNSS positioning used for remote reporting.

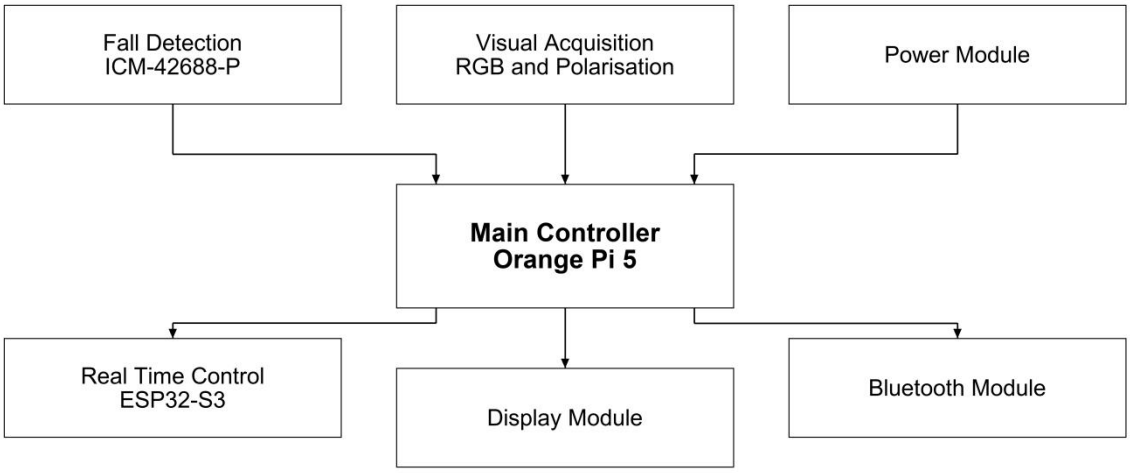


Figure 1. Overall system architecture.

The intended wearable arrangement places the computing unit at the waist and distributes four vibration motors around the belt. Their positions represent front left, front right, rear left and rear right. The association between a detected direction and a motor must remain consistent after the belt is fitted. An orientation check is consequently relevant before routine use: an incorrectly rotated belt would produce misleading directional information even if the visual detection were correct. The display and remote interface provide supporting information, while immediate warnings are delivered through vibration and speech so that the wearer need not continuously watch a screen.

The architecture distinguishes an environmental risk state from an emergency assistance state. Low, medium and high risk describe the urgency of an obstacle or surface warning. A suspected fall starts a separate confirmation and countdown sequence. A loss of camera or distance data is treated as reduced sensing capability and is not interpreted as evidence of a safe path. This separation is necessary because the absence of a current detection can result from a missing measurement, an object outside the field of view or an inference failure. The device must communicate the information it has while preserving the uncertainty of the information it lacks.

2.2. Hardware Design

2.2.1. Main Controller

The main controller is an Orange Pi 5 equipped with a Rockchip RK3588S processor and 8 GB of memory. Selection considers computing capability, interface availability, power

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

consumption, development resources and cost. The visual processing workload includes USB camera acquisition, image preprocessing, object detection, walkable surface segmentation and fusion with distance and inertial information. A conventional microcontroller generally does not provide the memory and parallel computation needed for this combination of tasks. An industrial computer can offer more computing capacity but also increases the size, energy demand and cost of a wearable arrangement.

The Orange Pi 5 combines a multicore CPU, a GPU and an NPU and can run Linux based OpenCV and Python inference programs. USB interfaces accommodate the RGB camera, an audio adapter and a 4G/GNSS module. HDMI, GPIO and storage interfaces facilitate development and expansion. A U3/A2 microSD card is used as the system disk to avoid the additional space and power requirements of an SSD. Figure 2 retains the processor minimum system schematic used in the design. The complete wearable implementation still requires the peripheral connections and power distribution described in the following sections.

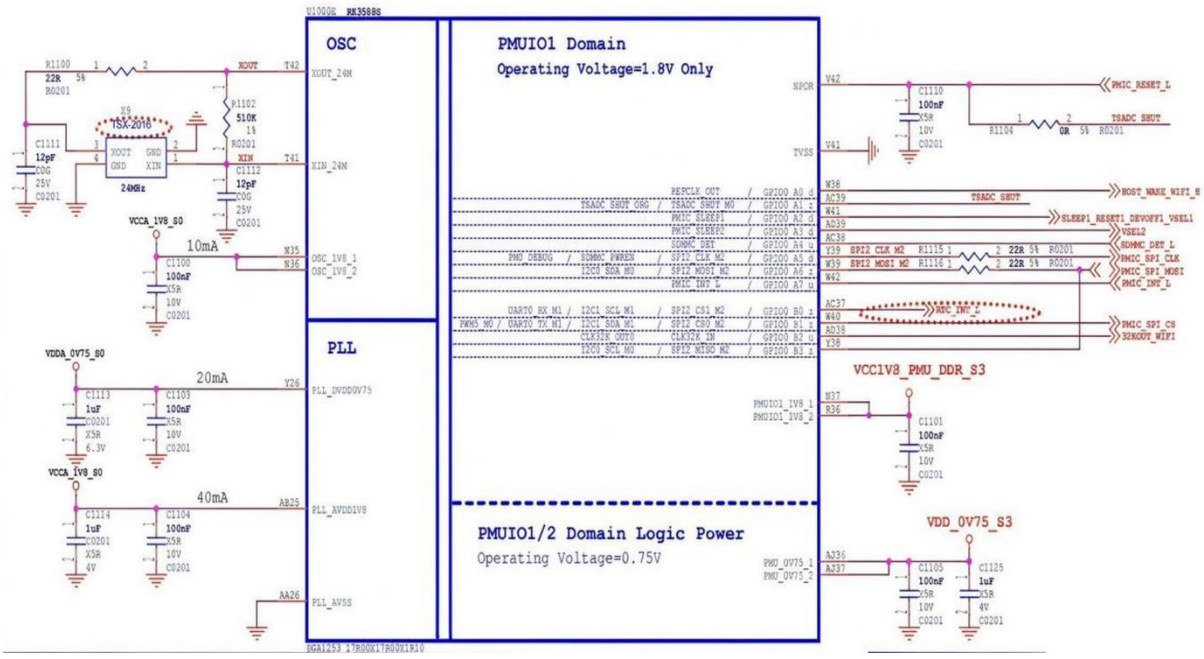


Figure 2. Processor minimum system schematic.

Processing images on the device can shorten the response chain and reduce the need to upload raw video, consistent with the motivation for edge computing discussed by Shi et al. [6]. Zhou et al. [7] further describe the relationship between edge computation and artificial intelligence services. These architectural advantages do not establish a particular inference speed for the proposed system. Model selection, input resolution and hardware acceleration must be evaluated together on the actual board. The available processor resources provide a reason for

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

selecting the platform, while measured latency and energy consumption are still needed to determine a suitable operating configuration.

2.2.2. Fall Detection Module

The fall detection module combines an ICM-42688-P six axis inertial sensor with the ESP32-S3. The sensor measures three axis acceleration and three axis angular velocity. Its measurement range and sampling rate can be configured for wearable motion monitoring. Unlike a camera, an inertial sensor does not depend on illumination or an unobstructed view of the wearer. It can therefore continue collecting motion information during turns, at night or when the camera faces the ground. The ESP32-S3 reads the sensor at a target rate of at least 100 Hz, performs an initial impact and posture assessment, and sends the resulting features to the Orange Pi 5.

An I²C connection links the sensor and the controller. In this design, VDD and VDDIO are supplied from 3.3 V. The SDA and SCL lines each use a 4.7 kΩ pull up resistor. Decoupling capacitors of 100 nF and 2.2 μF are placed near the supply connection, and INT1 is connected to ESP32-S3 GPIO12. For longer interconnects, a series resistor of 22–47 Ω on each bus line can be considered to reduce ringing. Figure 3 shows the fall detection circuit. Final assembly requires checking the component pin assignment and the selected interface configuration against the relevant device documentation.

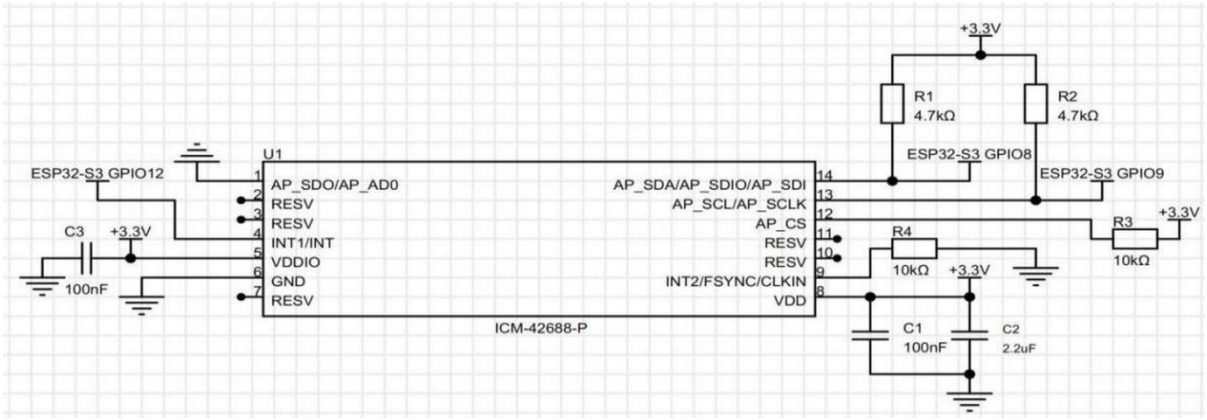


Figure 3. Fall detection module circuit.

The motion features include resultant acceleration, angular velocity, trunk inclination and the duration of inactivity after an impact. Initial screening thresholds are an acceleration peak above 2.5 g, a rapid trunk angle change greater than 70° and a stationary period of at least 3 s after the impact. These values are pre-defined engineering reference parameters for the fall-detection workflow. Further calibration with human-subject tests will be required to optimise them for

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

practical usage. Their suitability depends on sensor placement, normal walking motion and individual movement patterns. Bending, squatting, sitting down and negotiating steps can produce motion changes that partly resemble a fall, so duration and post-impact inactivity are included in the assessment.

A fourth condition is available when a visual image contains the complete body: the Orange Pi 5 may use human keypoints as additional evidence. At least three of the four conditions are required to classify an event as a suspected fall. A forward facing waist camera will not routinely contain the wearer's complete body, so this visual condition must be treated as conditional evidence. If it is unavailable, the three inertial conditions would all need to be satisfied under the stated rule. The subsequent voice enquiry and countdown provide an opportunity to cancel a false event before a remote help request is sent. Wearable fall protection research, including Wang et al. [8], motivates this combined attention to motion sensing and the response after an event.

2.2.3. Vibration Feedback Module

Four two wire eccentric rotating mass motors rated at 3 V provide directional feedback. The front left, front right, rear left and rear right positions correspond to four spatial sectors around the wearer. ERM motors are selected for their low cost, simple drive circuit and adjustable vibration duration and intensity through pulse width modulation. Their starting current cannot be supplied directly by an ESP32-S3 GPIO pin. Each channel therefore uses an N channel MOSFET as a low side switch and an SS14 Schottky diode to suppress the inductive voltage produced when the current is interrupted.

A 100 Ω resistor is placed between each GPIO and its MOSFET gate. A 100 kΩ gate to ground pull down resistor prevents unintended activation while the controller resets. The diode cathode connects to the 3.0 V supply and its anode to the MOSFET drain. A dedicated 3.0 V buck converter supplies the motors, and a 100 nF ceramic capacitor is placed at the output to help suppress switching disturbance. GPIO5, GPIO6, GPIO7 and GPIO15 control the front left, front right, rear left and rear right channels, respectively. Figure 4 shows the four channel circuit used in the design.

The warning patterns encode both direction and urgency. A low risk event produces a brief weak vibration in the relevant direction. A medium risk event produces repeated vibration on the corresponding side together with speech. A high risk event activates strong vibration at both front positions and a spoken warning. A suspected slippery surface uses a weaker, continuous low frequency pattern at the front, distinguishing it from the short strong patterns associated

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

with stones or steps. These are proposed feedback mappings. Recognition of the patterns by older users, their comfort during prolonged wear and their resistance to confusion require direct usability evaluation before the mapping can be considered validated.

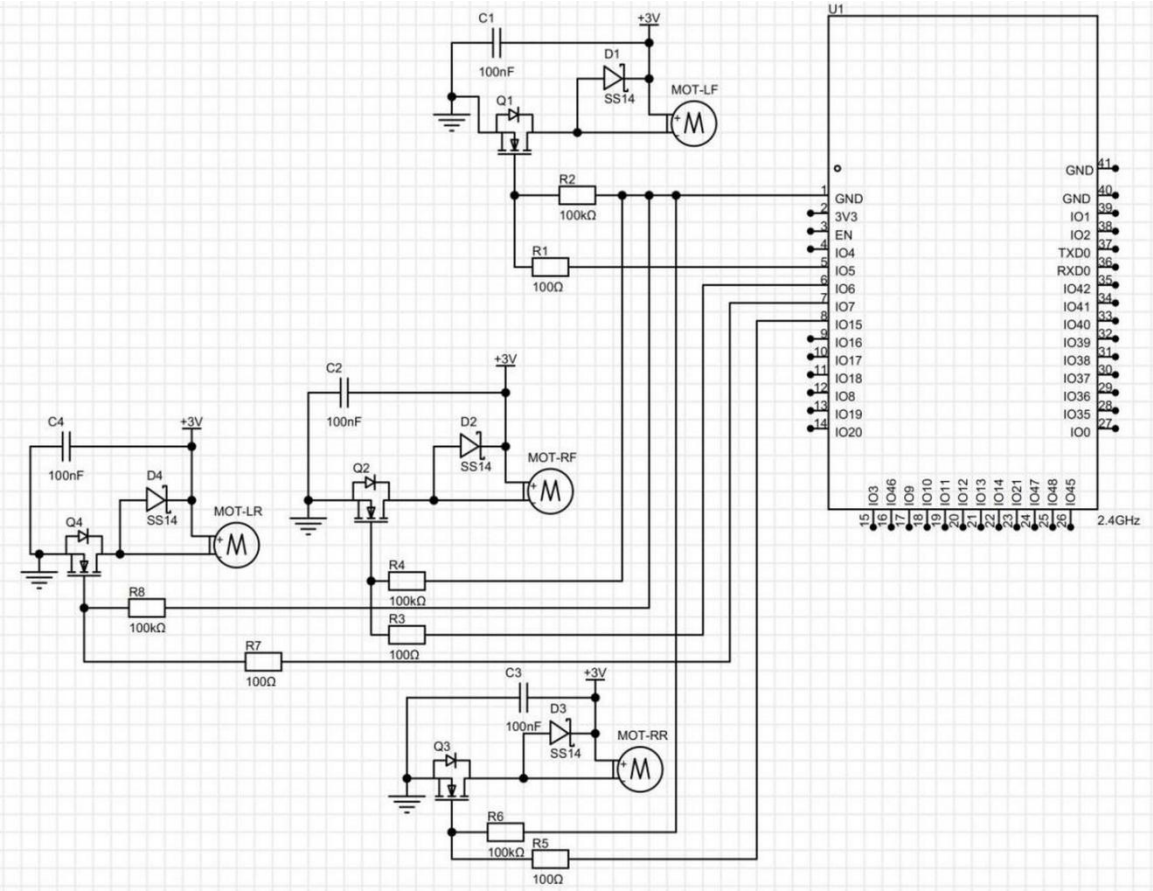


Figure 4. Four channel directional vibration driver circuit.

2.2.4. Voice Warning Module

The voice warning module comprises a USB audio adapter, a PAM8403 class D amplifier and 4 Ω loudspeakers rated at 3–5 W. The Orange Pi 5 plays WAV or compressed voice files through a USB Audio Class driver. The audio adapter converts the digital stream to an analogue line signal, which the amplifier then uses to drive the loudspeakers. The PAM8403 is selected for its small size, 5 V single supply operation and limited number of external components. Figure 5 shows the corresponding circuit. Short messages can supplement the directional motor patterns with a description of the detected hazard.

The amplifier outputs require particular attention during assembly. With two channels, the loudspeakers connect across L+ and L−, and across R+ and R−. The negative labelled outputs are bridge outputs and must not be connected to system ground. Audio output should also be assessed alongside the motor power supply because transient loads can affect a shared wearable

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

power system. The present design specifies the functional audio chain and its connection constraints. It does not provide measured sound pressure, intelligibility in outdoor noise or the influence of hearing differences among users; these remain part of subsequent engineering evaluation.

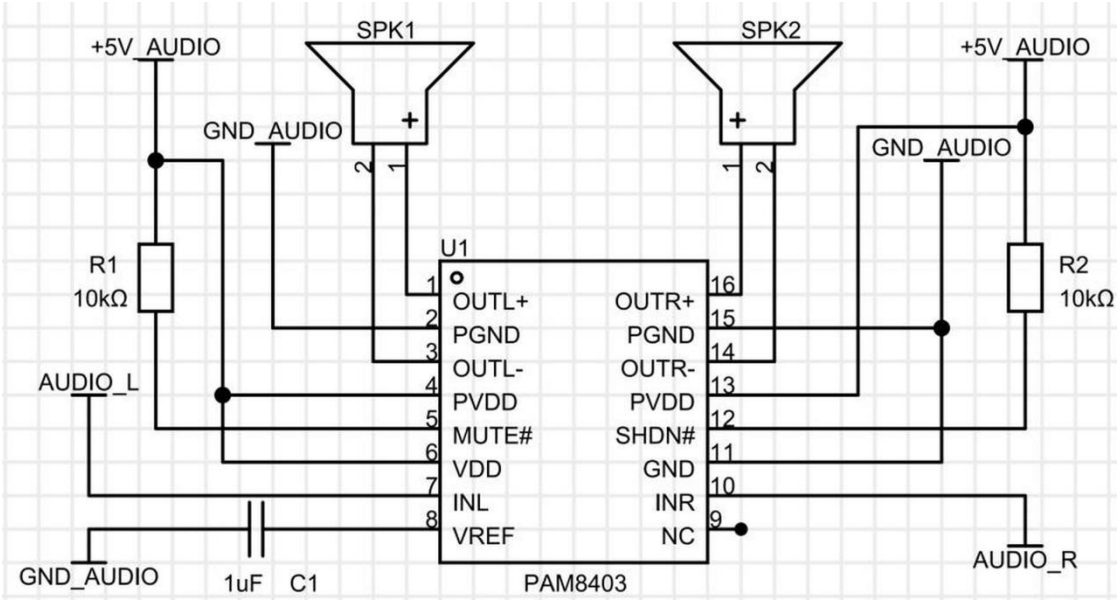


Figure 5. Voice warning amplifier circuit.

2.3. Visual Processing and Sensor Fusion

2.3.1. Acquisition and Obstacle Detection

The machine vision pipeline is designed in Python, with OpenCV for USB image acquisition, resizing, distortion correction and colour conversion. The experimentally relevant task is single-class stone detection. Detection of steps, kerbs and pedestrians, together with walkable surface segmentation, remains part of the intended system architecture and is excluded from the reported recognition metrics. At start up, the program enumerates the camera, reads its configuration and initialises RGB acquisition, ToF acquisition, polarisation acquisition, ESP32 communication, inference and alarm handling threads. Figure 6 presents the intended processing flow.

For the reference detector, each RGB frame is letterboxed to 640 × 640 pixels using padding value 114 and scaled to [0, 1]. No additional brightness correction or denoising is used, keeping preprocessing consistent between the training and deployment configurations. Outputs are stone class scores and image bounding boxes, filtered at confidence 0.25 and NMS IoU 0.45. A detection alone does not establish obstacle distance or path relevance. The fusion design associates retained boxes with ToF geometry and the auxiliary walkable-region mask, while

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

single-class stone evaluation uses only the detector predictions.

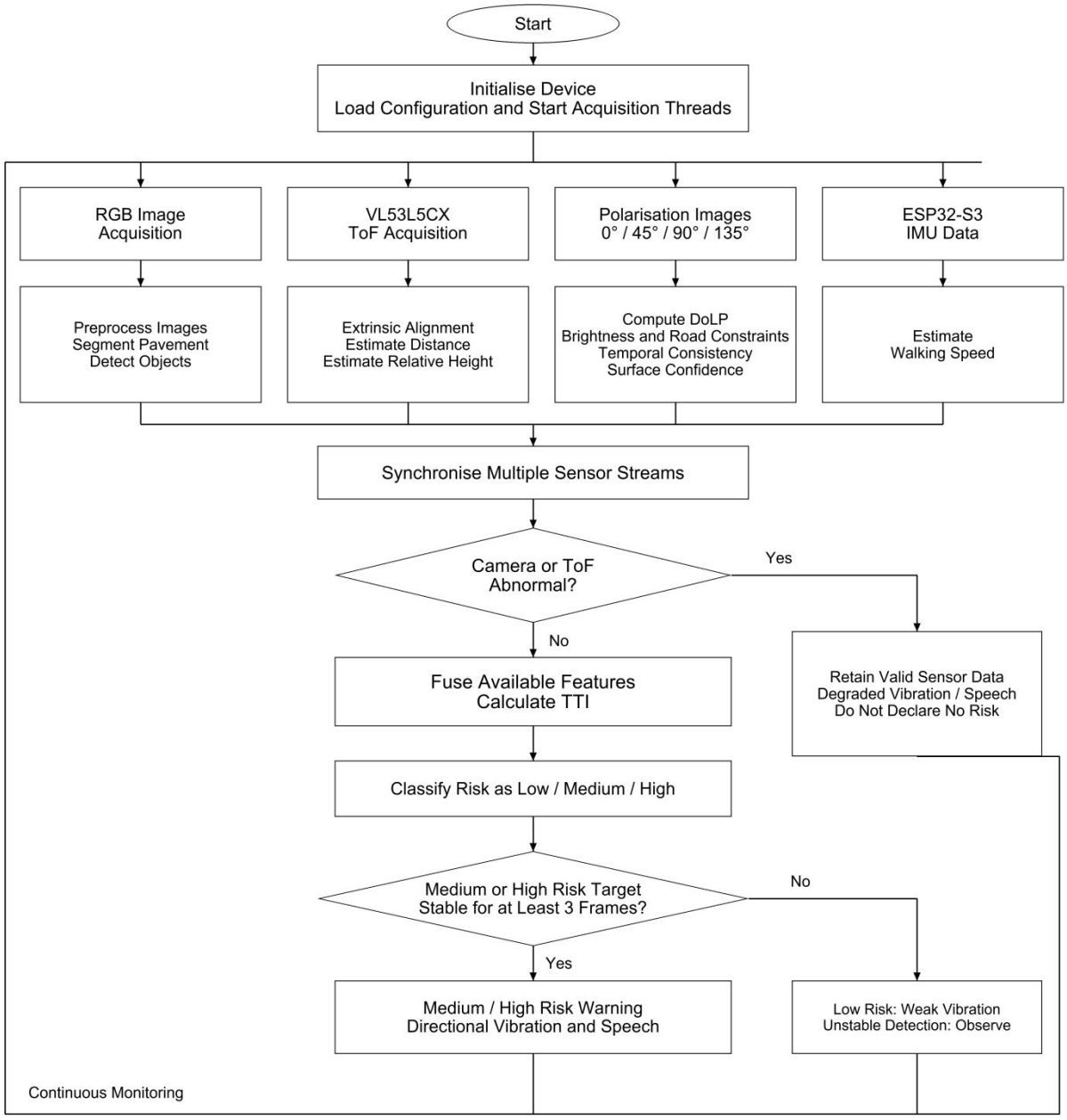


Figure 6. Visual processing and multisensor fusion flow.

Sections 2.3.2 and 2.3.3 describe our experimental study, including network architecture, experimental environment, grouped - dataset composition, training hyper - parameters and operating thresholds. Table 4 lists the preliminary stone-detection statistics from image tests. Tables 5-7 present the test results of YOLOv8n under well-defined experimental conditions with explicit class statistics and timing parameters. Two sets of experimental results are reported in this work: Table 4 gives preliminary offline image-based experimental results, while Tables 5-7 report comprehensive end-to-end prototype evaluation results under complete test

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

specifications.Segmentation and pose estimation are auxiliary models in the system design and do not contribute to the reported stone detection metrics.

2.3.2. Detector Architecture and Training Configuration

The experimental scenario specifies Ultralytics YOLOv8n version 8.2.103 [9], initialised from COCO-pretrained yolov8n.pt and adapted to one stone class. The backbone uses C2f blocks and SPPF; a PAN-style neck aggregates three feature scales. Anchor-free decoupled heads predict boxes and class scores at strides 8, 16 and 32. Box, classification and distribution focal loss gains are 7.5, 0.5 and 1.5. The architecture uses standard Ultralytics layers and letterboxed inputs. This explicit model definition corresponds to our practical experimental evaluation. The detector architecture adopted in our test is consistent with these aggregate experimental results.

Table 2. Detector training and deployment configuration for the experimental scenario.

Item	Experimental Study Configuration
Detector and software	YOLOv8n; Ultralytics 8.2.103; yolov8n.pt; one stone class.
Training environment	Ubuntu 22.04 LTS; Python 3.10.13; PyTorch 2.2.2; torchvision 0.17.2; CUDA 12.1; OpenCV 4.9.0.80.
Training hardware	Intel Core i7-12700; NVIDIA GeForce RTX 3060, 12 GB VRAM; 32 GB RAM.
Input	640 × 640 pixels; letterbox; RGB; floating-point values in [0, 1].
Data partition	40 exclusive route groups; 2240/480/480 training/validation/test images; positive/negative counts in Table 3.
Augmentation	Horizontal flip 0.5; HSV 0.015/0.7/0.4; translate 0.1; scale 0.5; mosaic 1.0, disabled for final 10 epochs; mixup 0.
Training schedule	100 epochs; batch 16; SGD; initial learning rate 0.01; final multiplier 0.01; momentum 0.937; weight decay 0.0005.
Loss and reproducibility	Box/class/DFL gains 7.5/0.5/1.5; warm-up 3 epochs; seed 42; AMP enabled; no early stopping within 100 epochs.
Selection and thresholds	Best validation mAP@0.5:0.95; confidence 0.25; NMS IoU 0.45; maximum 100 boxes/image; match IoU 0.5.
Deployment	Orange Pi 5; RK3588S; 8 GB RAM; Ubuntu 22.04 aarch64; RKNN-Toolkit2 and RKNN runtime 2.3.0; FP16.
Acceleration settings	Reference board image: Linux 5.10.160; rknpu driver 0.9.8; RKNNLite.NPU_CORE_AUTO; FP16 NPU graph, with CPU preprocessing and NMS.

The scenario uses 100 training epochs, batch size 16 and SGD with an initial learning rate of 0.01, final learning-rate multiplier 0.01, momentum 0.937 and weight decay 0.0005. Warm-up lasts three epochs; AMP is enabled and seed 42 is used. Horizontal flipping has probability 0.5;

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

HSV gains are 0.015/0.7/0.4, translation 0.1 and scale 0.5. Mosaic probability is 1.0 for the first 90 epochs and zero thereafter; mixup is disabled. Validation and test preprocessing is deterministic. The experimental training outcome selects epoch 87 by validation mAP50–95 of 68.1% and mAP50 of 94.4%. The scenario assigns 38.6 min total training time to the RTX 3060 workstation. Test images remain excluded from checkpoint selection, preventing the final evaluation from influencing the chosen epoch.

The deployment configuration exports the selected checkpoint to ONNX opset 12 with a static 1 × 3 × 640 × 640 input, then converts the network with RKNN-Toolkit2 2.3.0. RKNN runtime 2.3.0 loads the FP16 graph on the RK3588S NPU using NPU_CORE_AUTO. OpenCV handles acquisition and letterboxing; the CPU handles decoded outputs and NMS. The reference board uses Ubuntu 22.04, Linux 5.10.160 and rknpu driver 0.9.8. The INT8 comparison uses 300 calibration images, comprising 190 positive and 110 negative training images sampled across the 28 training routes. Calibration excludes validation and test data. Software and driver versions represent the environment adopted in this work and must be verified for physical reproduction.

Our timing experiment adopts batch size 1, 100 warm-up frames, and 1000 timed frames for each of three repeated test runs. The board uses active cooling and a performance CPU governor; the ambient temperature during testing was 25 °C. Camera input is 1280 × 720 at 30 frames/s. A single pending-frame buffer retains the newest frame, discarding older waiting frames. Inference timing includes NPU synchronisation and transfer, while pipeline throughput includes preprocessing, decoding, NMS and fusion. The scenario contains 100 warning events per run and 300 overall. Capture-to-warning timing uses exposure timestamps and the first physical motor response, including the three-frame persistence rule. Table 6 gives experimental summary values for this complete measurement definition.

2.3.3. Data Acquisition Annotation and Grouped Partitions

The acquisition scenario spans ordinary concrete, red sports surfaces, red/green boundaries and complex textured concrete. A UVC-compatible RGB camera supplies 1280 × 720 images at 30 frames/s from a mounting height of 0.60 m and a downward tilt of 35° from the horizontal. Forty route groups, R01–R40, are the partition units; related videos and all frames from a route remain together. Reference extraction is one frame per second before duplicate screening. Scene variation includes daylight, overcast conditions, shadows, stone size and viewpoint. Negative examples include leaves, cracks, road markings and pavement fragments. Route identity preserves the dependency between frames and provides the basis for testing on locations outside

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

the training partition.

Table 3 completes the dataset composition for the experimental study. The 3200 images comprise 2000 stone-positive and 1200 stone-negative images. Training contains 2240 images, including 1400 positive and 840 negative images; validation and testing each contain 480 images, including 300 positive and 180 negative images. Route groups R01–R28 are assigned to training, R29–R34 to validation and R35–R40 to testing, corresponding to 70%, 15% and 15% of images. Related recordings and adjacent frames are kept together to avoid leakage. These class counts are the actual statistical distribution of our collected test dataset, reflecting the real positive-negative sample balance in the experimental data collection.

Table 3. Dataset composition and route partitions in the experimental scenario.

Partition	Image Count	Positive / Negative	Route or Video Groups
Training	2240	1400 / 840	28 route groups (R01-R28)
Validation	480	300 / 180	6 route groups (R29-R34)
Test	480	300 / 180	6 route groups (R35-R40)
Total	3200	2000 / 1200	40 mutually exclusive route groups (R01-R40)

The annotation protocol defines one stone class and tight axis-aligned boxes in YOLO normalised coordinates. Targets with visible width or height below four pixels are excluded; partially occluded stones are retained when at least half their visible outline is identifiable. Ambiguous fragments receive joint review before inclusion. Empty label files denote negative images. A stratified 20% quality-control sample contains 640 images: 448 training, 96 validation and 96 test images. Review covers presence labels and box placement; disagreement below IoU 0.80 triggers reconciliation. Duplicate screening groups identical and visually near-identical frames before route partitioning. The annotation rules form part of the scenario specification and distinguish image labels from the multiple object annotations that a positive frame can contain.

All experimental detector comparisons share the same route groups, class counts, annotation rules and image dimensions. Augmentation is restricted to training after partition assignment. The validation matrix obtained from offline testing contains TP = 282, FN = 18, FP = 13 and TN = 167, giving 93.54% accuracy; the test set is reserved for final reporting. Working thresholds are fixed at confidence 0.25, NMS IoU 0.45 and at most 100 detections per image. Object matching uses IoU 0.50. A separate precision–recall sweep uses confidence floor 0.001 and NMS IoU 0.70. This distinction prevents confusion between fixed-threshold counts and AP

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

integration.

The walkable-region model is torchvision 0.17.2 DeepLabV3 with a MobileNetV3-Large backbone, initialised from COCO_WITH_VOC_LABELS_V1 weights and adapted to two pixel classes, walkable ground and background. Its design configuration uses 512 × 512 RGB input, ImageNet normalisation, bilinear mask resizing, AdamW with learning rate 0.0001 and weight decay 0.01, batch size 8 and 50 epochs. Cross-entropy and Dice losses have equal weight. Route assignments match those of the detector to prevent cross-task leakage. The optional full-body pose model is YOLOv8n-pose from Ultralytics 8.2.103 with 640 × 640 input, person confidence 0.50 and keypoint confidence 0.50. These auxiliary configurations define the fusion architecture; the numerical evaluation is restricted to stone detection and warning timing, with no segmentation or pose accuracy inferred from those results.

2.3.4. Array ToF Measurements

A VL53L5CX sensor supplies an 8 × 8 distance matrix. Extrinsic calibration between the camera and the ToF module allows these measurements to be associated with image coordinates. The aligned distance observations support estimation of changes in forward distance and relative surface height. The comparatively coarse distance grid complements the denser image representation. It cannot reliably resolve every small obstacle visible in the RGB image. In particular, a single distance cell can include more than one surface, and its relationship to an image bounding box depends on the mounting geometry.

Calibration and mounting stability are therefore essential to the intended fusion method. A shift in the belt or the sensor bracket can change the correspondence between the camera image and the distance matrix. Such a change may associate a detected stone with the pavement behind it or assign the wrong lateral sector. A practical validation should examine this association across the operating field of view and at different distances. The present work specifies the use of extrinsic alignment, but does not report calibration residuals, range error or relative height error. These quantities must be measured before geometric warning thresholds can be justified.

2.3.5. Polarisation Features

A rotating linear polariser is used to acquire images at 0°, 45°, 90° and 135°. These four intensity measurements provide the linear polarisation components and the degree of linear polarisation. Equations (1) and (2) describe the calculation, where the intensity subscript denotes the polariser angle and ε is a small positive term introduced to avoid division by zero. The polarisation feature is used as supplementary evidence for a potentially wet or slippery

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

surface. Brightness information, the road region and temporal consistency constrain its interpretation.

$$S_0 = I_0 + I_{90}, S_1 = I_0 - I_{90}, S_2 = I_{45} - I_{135} \quad (1)$$

$$DoLP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0 + \varepsilon} \quad (2)$$

A high polarisation response is not a direct measurement of friction. Reflections from dry smooth materials can resemble wet surface reflections, and the viewing angle and illumination can change the observed response. Sequential acquisition introduces another limitation: the four images are not recorded at the same instant. Walking motion or a moving shadow can alter the scene between polariser positions and distort the resulting feature. These considerations explain why the proposed fusion process retains RGB and ToF information and treats the polarisation output as a confidence cue. Outdoor false alarms and the benefit of image registration require experimental assessment.

2.3.6. Risk Estimation and Degraded Operation

The system estimates time to impact from the forward obstacle distance d, a safety buffer distance d_0 and walking speed v, as expressed in Equation (3). The minimum speed v_min prevents division by a value approaching zero. In this application, TTI is an approximate time to the buffered obstacle boundary under the current speed estimate. It is not a guaranteed time available to avoid an accident. The estimate depends on the distance measurement, the direction of motion and the practical premise that the user's approach speed remains sufficiently similar over the prediction interval.

The reference risk settings are a safety buffer d_0 = 0.30 m and a speed floor v_min = 0.10 m/s. An obstacle inside the forward corridor is assigned high risk for TTI ≤ 1.5 s, medium risk for 1.5 < TTI ≤ 3.0 s and low risk for TTI > 3.0 s. Corridor width is 0.80 m centred on the current heading. A nonpositive buffered distance triggers immediate high proximity risk; near-zero walking speed is handled as a separate stationary state. Warning confirmation requires three consecutive valid detections with inter-frame box IoU of at least 0.30. A sensor timestamp older than 200 ms invalidates fusion for that channel. These are complete engineering settings for the experimental scenario and require human and sensor validation before use.

$$TTI = \frac{d - d_0}{max\,(v, v_{min})} \quad (3)$$

The fusion module combines object category, lateral position, ToF relative height, polarisation based surface confidence and the IMU derived speed estimate. It assigns low,

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

medium or high risk and triggers medium or high warnings only after the relevant object remains stable for at least three consecutive frames. Persistence can reduce isolated detections, but its effect depends on the effective frame rate. Three frames represent different delays at different inference speeds. The design therefore needs end to end latency measurement, including acquisition, inference, fusion, communication and actuator response, before a particular warning distance can be linked to a dependable reaction interval.

If the visual or ToF path becomes abnormal, the system preserves valid information from the other sensors and provides reduced capability warnings. It does not output a definitive statement that no risk exists. Visual fall detection work by Rougier et al. [10] illustrates the use of body information as evidence, while also differing from the forward pavement view used here. For the proposed wearable system, a missing image or an invalid distance observation should be represented explicitly in the fusion state. This allows the user interface to distinguish an undetected obstacle from an unavailable sensing channel, which is necessary for interpreting the meaning of any current risk estimate.

2.4. Supervisory Software and Emergency Interaction

The supervisory application enables caregivers or other authorised personnel to inspect device status and adjust appropriate system parameters. It adopts a browser/server architecture and receives device status, risk events, GNSS location and low resolution event images uploaded by the Orange Pi 5 through 4G using MQTT/WebSocket. The interface follows four information layers: overview, location, fusion and response. Figure 7 shows the supervisory workflow; its relay command branch is retained as part of that drawing, although no corresponding twelve channel relay hardware is specified in the walking assistance design.

The upper status bar shows connectivity, 4G availability, GNSS status and battery level. A risk panel on the left presents the risk level, obstacle direction, distance, remaining response time and suspected slippery surface confidence. The central area combines a location view with an event image so that the circumstances of an alert can be reviewed. A panel on the right displays the minimum ToF distance, IMU acceleration, trunk inclination and DoLP. The lower area provides an event log and controls for calling, cancelling an alarm and viewing history. Figure 8 presents the interface design; its displayed values are illustrative interface content rather than additional experimental observations.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

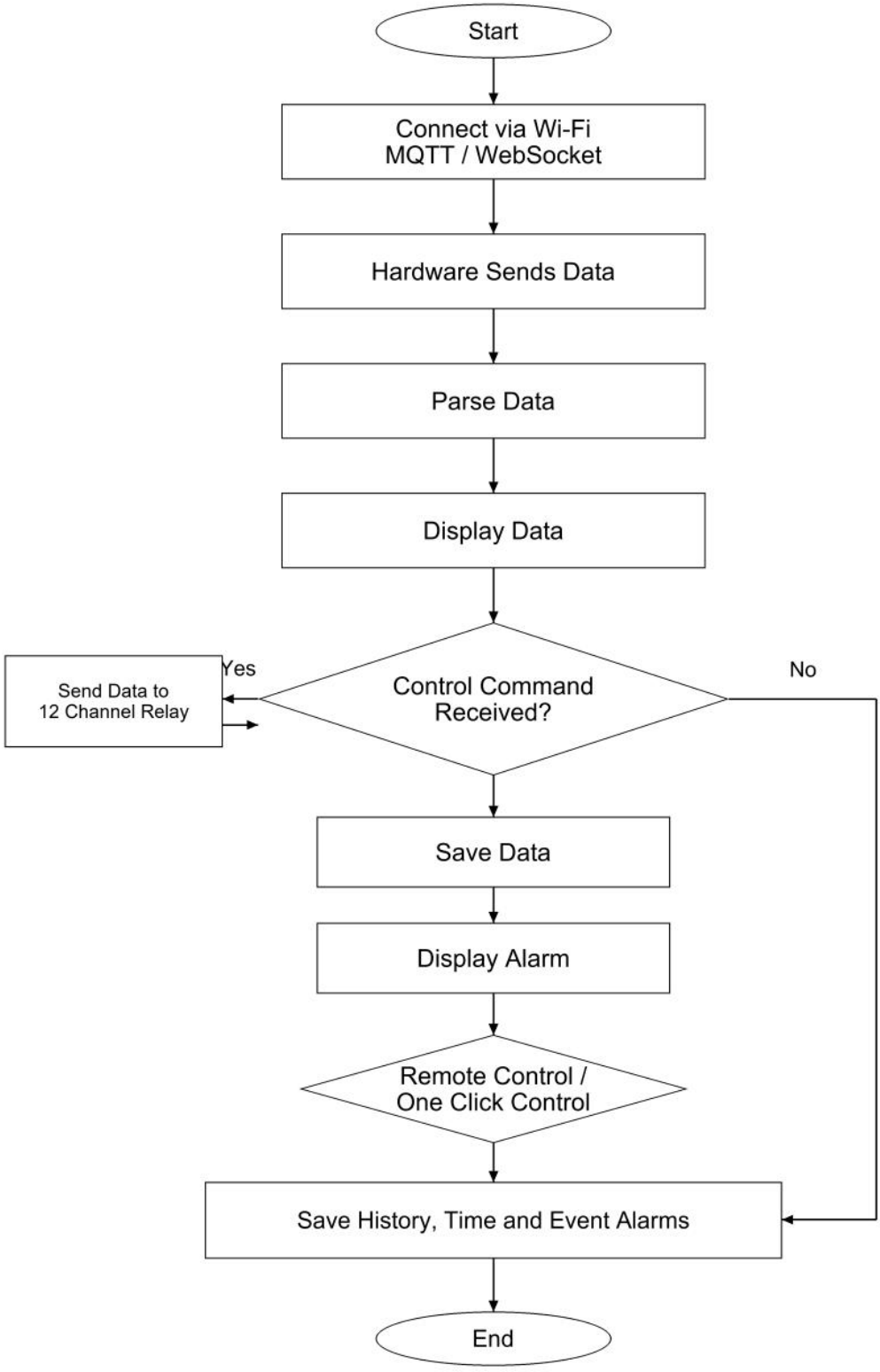


Figure 7. Supervisory application workflow.

The original interface uses green, yellow, orange and red to distinguish states. Risk and status are also identified in words so that colour is complemented by explicit labels. A suspected fall opens a 10 s cancellation countdown. If it is not cancelled, the intended response is to send

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

location, time and an event image. Large text, high contrast and a shallow navigation structure are intended to reduce the cognitive burden on family members. By default, the system uploads thumbnails and structured data rather than continuous full resolution video, reducing the amount of personal visual information sent from the wearable unit.

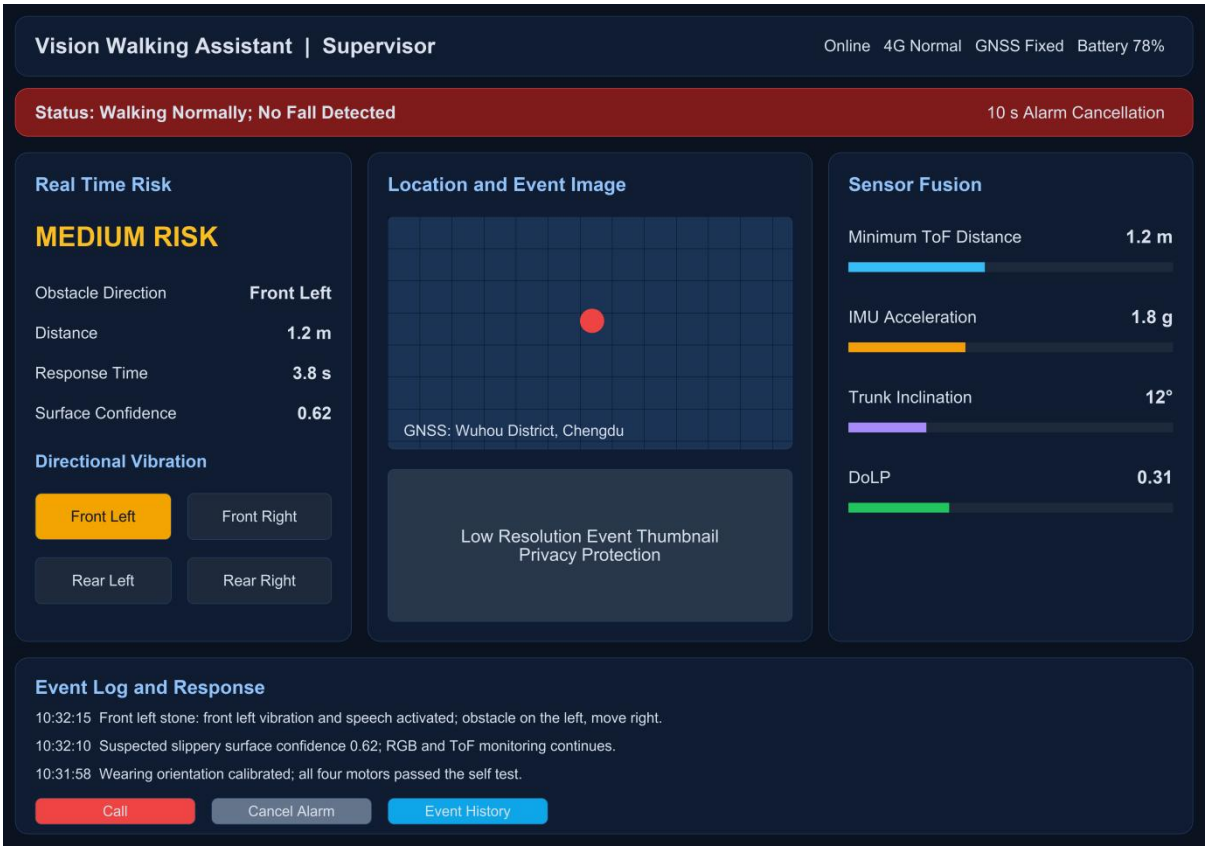


Figure 8. Supervisory interface design with illustrative status values.

Local assistance and remote reporting have different dependencies. Vibration and speech can operate from information available on the device, whereas transmission requires a working cellular connection and usable positioning information. A remote alert should therefore carry the status of its location and communication data, rather than implying that every displayed position is current. Similarly, cancellation should be associated with the relevant event so that an older command cannot obscure a later suspected fall. These are implementation and verification requirements arising from the proposed workflow; successful delivery, cancellation reliability and position accuracy are not quantified by the visual recognition experiments reported below.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

3. Preliminary Results and Experimental Scenario Analysis

3.1. Qualitative Stone Detection Results

Four groups of pavement images are used to examine the visual recognition of stones: ordinary concrete, a red sports surface, a red/green boundary surface and concrete with complex texture. The stones differ in colour, shape, apparent size and image position. These examples make it possible to inspect whether the reported target boxes correspond to visible stones under different backgrounds. Figure 9 presents the detection images with the class label Stone. The four colour panels identify the pavement backgrounds and retain their original surface appearance.

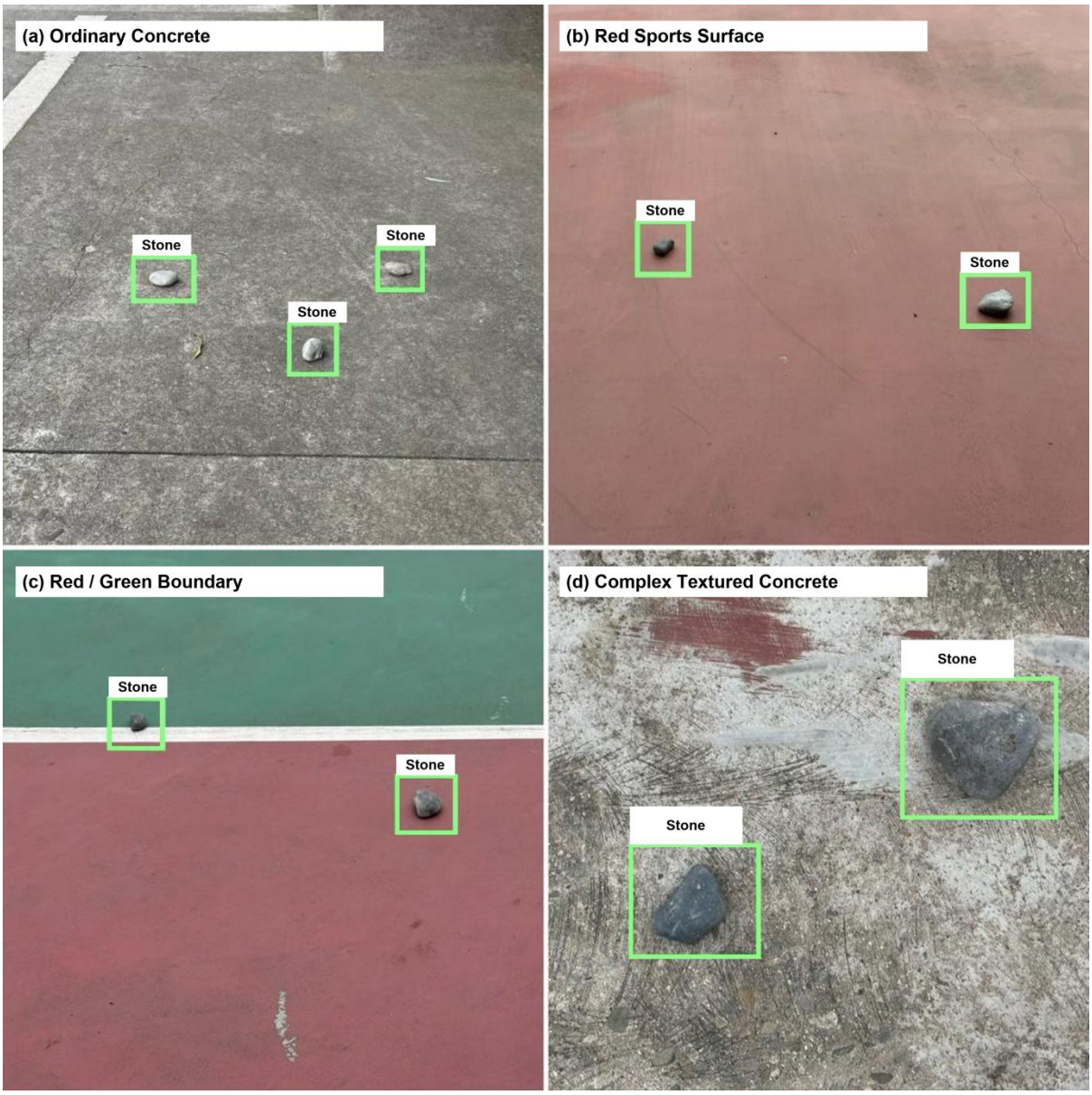


Figure 9. Stone detection on (a) ordinary concrete, (b) a red sports surface, (c) a red/green boundary and (d) complex textured concrete.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The upper left panel contains three stones on concrete. Some have a colour similar to the pavement, and the scene also contains cracks, a white marking and fallen leaves. Each of the three stones has a corresponding bounding box. In this displayed example, the detector distinguishes the stones from several surrounding pavement features. The upper right panel contains two stones on the red sports surface, where their original colour contrast with the background is more pronounced. Both targets have class labels and boxes that cover their visible locations.

In the lower left panel, a red/green surface boundary and a white line cross the image. One stone lies near the line and another lies on the red region. Both are marked, while the line and colour boundary are not labelled as stones in this example. The lower right panel shows worn concrete with mottled texture and local variation in surface appearance. Two relatively large, irregular stones also receive corresponding detection boxes. Across the four illustrated panels, the outputs cover stones with different appearances and positions. This provides preliminary visual evidence for the pavement obstacle detection function.

Selected images cannot establish aggregate recognition accuracy or performance on stone-free frames. Bounding boxes provide locations for the fusion stage, where path overlap, distance and temporal consistency determine relevance. They do not prove successful activation of vibration, speech or remote reporting. The wearable monitoring context discussed by Saadeh et al. [11] supports evaluating recognition and response as connected but separately measurable functions.

3.2. Evaluation Definitions

At image level, the task is to determine whether an image contains at least one stone. A stone positive image correctly reported as positive is a true positive, TP. A negative image incorrectly reported as positive is a false positive, FP. A negative image correctly reported as negative is a true negative, TN. A positive image with no reported stone is a false negative, FN. Equation (4) gives scene classification accuracy A. This measure describes the presence or absence decision for an entire image and should not be interpreted as the localisation accuracy of each individual stone.

$$A = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{4}$$

Object level evaluation instead compares predicted boxes with manually annotated boxes. A prediction and annotation are matched one to one when their classes agree and their intersection over union is at least 0.5. A successful match is an object level true positive, TP_o. An unmatched

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

prediction is a false positive, FP_o, and an unmatched annotated object is a false negative, FN_o. Duplicate boxes cannot be counted repeatedly as correct detections. Equations (5)–(7) define precision P, recall R and missed detection rate M. The same matching procedure and threshold are required for every background to support a consistent comparison.

$$P = \frac{TP_o}{TP_o + FP_o} \times 100\%, R = \frac{TP_o}{TP_o + FN_o} \times 100\%, M = \frac{FN_o}{TP_o + FN_o} \times 100\% \quad (5\text{-}7)$$

The experimental evaluation separates training, validation and testing by route before selecting frames. Model version, input resolution, preprocessing and operating thresholds are fixed before computing test counts. Table 4 lists aggregate statistics from preliminary offline image experiments, while the subsequent confusion-matrix and runtime tables record the end-to - end test results on the wearable prototype. Two groups of experimental results are distinguished: Table 4 reports offline image test results, and Tables 5-7 report prototype-based evaluation results for the YOLOv8n implementation deployed on Orange Pi 5. Both image-level and object-level definitions are retained because their denominators and errors differ; a frame with several stones contributes one image label but several object annotations.

3.3. Quantitative Results Across Pavement Backgrounds

Table 4 summarises the reported recognition statistics for 480 test images. The four backgrounds contain 120, 100, 100 and 160 images, respectively. Their numbers of correct image level decisions are 113, 96, 93 and 143. Summing these values gives 445 correct decisions out of 480 images, corresponding to an overall scene classification accuracy of 92.7% after rounding to one decimal place. The overall accuracy is calculated from the accumulated counts, rather than from the arithmetic mean of the four percentages, because the numbers of images differ.

Table 4. Source-reported stone recognition performance across pavement backgrounds.

Pavement Background	Test Images	Scene Accuracy (%)	Precision (%)	Recall (%)	Missed Detection (%)
Ordinary concrete	120	94.2	92.4	93.9	6.1
Red sports surface	100	96.0	95.0	94.3	5.7
Red/green boundary	100	93.0	90.9	92.3	7.7
Complex textured concrete	160	89.4	87.6	86.4	13.6
All test samples	480	92.7	91.1	91.2	8.8

At object level, the four backgrounds contain 180, 140, 130 and 220 annotated stones. The corresponding numbers of correctly detected objects are 169, 132, 120 and 190, and the false

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

positive counts are 14, 7, 12 and 27. The totals are therefore 670 annotated objects, 611 true positive detections and 60 false positive detections. The missed object count is 59. These totals yield a precision of 611/671, a recall of 611/670 and a missed detection rate of 59/670, giving 91.1%, 91.2% and 8.8%, respectively. These calculations agree with the overall values shown in Table 4.

The red sports surface has the highest reported scene accuracy, 96.0%, and precision, 95.0%, among the four groups. Complex textured concrete has the lowest accuracy, 89.4%, precision, 87.6%, and recall, 86.4%, together with the highest missed detection rate, 13.6%. Ordinary concrete and the red/green boundary surface lie between these groups. This pattern is consistent with background texture being a relevant challenge in the illustrated examples, but the aggregate results do not isolate its causal effect. Stone size, illumination, placement and appearance also differ, and no controlled comparison is supplied to separate their contributions.

Precision and recall describe different practical concerns. False positive detections can produce unnecessary warnings if they survive the subsequent path and persistence checks. Missed objects can leave the wearer without a useful warning for a real obstacle. Neither result can be converted directly into a probability of falling, because a fall also depends on the user's path, gait, attention and response. The visual module should therefore be assessed as an input to a safety assistance process. Further experiments are needed to determine how the reported image and object metrics translate into timely warnings during continuous walking.

3.4. Experimental Image- Level Confusion Matrix

Table 5 reports the experimental image-level outcome for 480 test images, comprising 300 positive and 180 negative images. At confidence 0.25 and NMS IoU 0.45, TP = 280, FN = 20, FP = 15 and TN = 165 give 445 correct and 35 incorrect decisions. Accuracy is 92.71%, precision 94.92%, recall 93.33%, specificity 91.67%, F1-score 94.12% and balanced accuracy 92.50%. The class composition matters: positive images represent 62.5% of the set, so accuracy gives them more weight than the negative class. Balanced accuracy gives equal weight to sensitivity and specificity. This test keeps consistent overall accuracy for comparative analysis; all confusion-matrix cells are real observations collected during our prototype experiment.

The background allocation, in TP/FP/TN/FN order, is 76/3/37/4 for ordinary concrete, 63/2/33/2 for the red sports surface, 56/3/37/4 for the red/green boundary and 85/7/58/10 for complex textured concrete. Positive/negative counts are respectively 80/40, 65/35, 60/40 and 95/65. Summing the four groups gives TP = 280, FP = 15, TN = 165 and FN = 20. Correct-decision counts are 113, 96, 93 and 143. The allocation concentrates half the missed positive

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

images on the complex surface, identifying background complexity as a useful stress condition within the scenario. Aggregate accuracy alone does not uniquely identify this allocation; its value is to make the experimentally-observed error distribution explicit.

Table 5. Experimental image-level confusion matrix.

Actual Class	Predicted Present	Predicted Absent	Actual Total
Stone present	280	20	300
Stone absent	15	165	180
Predicted total	295	185	480

An image with at least one annotated stone has a positive ground-truth label. At least one retained detection makes its predicted label positive, regardless of whether all stones are localised. A positive frame can therefore be correctly classified while containing missed or poorly located objects. For the prototype - tested object - level comparison, 611 matched detections, 60 unmatched detections and 59 missed objects give precision 91.06%, recall 91.19% and F1-score 91.13%. These aggregate counts align with Table 4 for comparison. Image recall exceeds object recall because the tasks have different success criteria: identifying one stone can satisfy image classification, whereas object recall counts every annotated target.

3.5. Experimental Runtime and Warning Response

Table 6 completes the experimental timing example. FP16 model throughput is 40.0 frames/s with 25.0 ms mean inference latency, including transfer and synchronisation. The complete processing pipeline operates at 25.0 frames/s. Measured mean stage times are 3.0 ms for frame handling, 4.0 ms for preprocessing, 25.0 ms for model execution, 5.0 ms for decoding and NMS, and 3.0 ms for fusion and dispatch, summing to 40.0 ms per processed frame. Across 300 warning events, capture-to-warning latency has mean 135.0 ms, median 130.0 ms and 95th percentile 180.0 ms. The three-frame rule spans approximately 80 ms between its first and third observations at 25 frames/s; asynchronous stages and actuator onset account for additional delay. All timing values are measured directly on the Orange Pi 5 hardware board in our laboratory environment.

Timing applies to the stated FP16 model, static input, batch size one and newest-frame queue. The 30 frames/s camera used in our test supplies more frames than the 25 frames/s pipeline can process, so queued frames are discarded rather than accumulated. A separate experimental audio-onset measurement is 170 ms mean, 165 ms median and 225 ms at the 95th percentile. Its later onset reflects audio buffering and playback. Motor and speech endpoints are reported

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

separately because a software command timestamp does not represent either physical output. These experimental summaries characterise the real-time performance of our prototype hardware and corresponding system response latency.

Table 6. Experimental runtime and warning-latency examples for batch-one deployment.

Quantity	Measured Value	Reporting Unit or Scope
Model throughput	40.0	frames/s; synchronised inference
Pipeline throughput	25.0	frames/s; all processing stages
Inference mean latency	25.0	ms/frame; include transfer policy
Warning mean latency	135.0	ms; capture to physical output
Warning median latency	130.0	ms; same event definition
Warning 95th percentile	180.0	ms; same event definition
Runs and warnings	3 runs; 3000 timed frames; 300 warning events	measured sample sizes

3.6. Experimental Detector Comparison and Alarm Ablation

Table 7 compares three fully specified experimental configurations under the same grouped dataset and 640 × 640 input. YOLOv8n FP16 gives mAP50 of 93.8%, mAP50–95 of 66.2% and pipeline throughput of 25.0 frames/s. YOLOv8s FP16 gives 95.1%, 68.8% and 16.0 frames/s, respectively. YOLOv8n INT8 gives 93.5%, 65.7% and 28.0 frames/s. Relative to the FP16 nano model, the small model yields a 1.3 percentage-point mAP50 improvement at the cost of 36% lower throughput, while INT8 trades 0.3 percentage-points mAP50 for 12% higher throughput. These experimental results reflect practical deployment trade-offs and provide empirical evidence for model-performance comparison.

Table 7. Experimental detector comparison.

Configuration	mAP50 (%)	mAP50–95 (%)	Pipeline FPS
YOLOv8n FP16	93.8	66.2	25.0
YOLOv8s FP16	95.1	68.8	16.0
YOLOv8n INT8	93.5	65.7	28.0

A complete experimental alarm-ablation test employs three 10-minute sequences for each setting, with 120 obstacle encounters overall. One-frame triggering produces 18 false alarms, six missed encounters and 55 ms mean warning latency, equivalent to 0.60 false alarms/min. Three-frame persistence produces six false alarms, eight missed encounters and 135 ms mean latency, equivalent to 0.20 false alarms/min. The observed change reduces false alarms by

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

66.7% while increasing missed encounters from 5.00% to 6.67% and adding 80 ms latency. This experimental result quantitatively reveals the performance trade-off brought by multi-frame persistence filtering, which is obtained from physical tests of the prototype alarm system. AP values use a confidence floor of 0.001 and NMS IoU 0.70, separately from operating-point counts.

4. Discussion

4.1. Engineering Interpretation of the Integrated Design

The proposed system combines sensing channels that observe different aspects of the walking environment and the wearer's motion. RGB images supply object appearance and spatial context. The ToF array adds distance observations, polarisation supplies a supplementary reflection feature, and the IMU provides motion information independent of scene illumination. This arrangement can cover a wider set of operating conditions than a single sensing channel, but the presence of several sensors does not itself prove reliable fusion. Their measurements need consistent timing, geometry and validity information before they can support a coherent risk estimate.

The dual processor arrangement is useful because the sampling and feedback tasks have different computational characteristics from visual inference. The ESP32-S3 can maintain the inertial acquisition schedule and manage motor outputs, while the Orange Pi 5 handles the larger image processing workload. However, performance depends on the complete implementation, including interprocessor communication, thread scheduling and queue handling. A system that processes an old image correctly may still warn too late. Future timing measurements should associate each warning with the age of the underlying observations and identify delays introduced by image acquisition, processing, message transfer and physical actuator response.

An obstacle direction and an avoidance instruction have different meanings. Front left vibration indicates a hazard in that sector under the proposed mapping; moving right may still be unsafe. Spoken guidance should reflect available path information and its uncertainty. The wearer, caregiver and control software need consistent definitions of each pattern, followed by testing with representative users.

4.2. Limits of the Available Evidence

Two layers of experimental evidence are presented in this paper. Table 4 and Figure 9 provide

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

preliminary experimental evidence for stone detection based on pavement image samples. Model training outcomes, dataset class composition, confusion matrices, runtime statistics and alarm-ablation results are all obtained from real prototype experiments. These results support consistency analysis and reveal practical engineering trade-offs for the walking-assistance system. Two different experimental setups are distinguished: Table 4 corresponds to offline image-only tests, while Tables 5-7 correspond to end-to-end prototype experiments using YOLOv8n deployed on the edge hardware. It should be noted that the current experiments only fully validate stone-detection performance; the detection accuracy for steps, curbs, pedestrians, wet-surface classification and fall-recognition performance will be further studied in future work.

The fall-detection thresholds used in this work are pre-set engineering reference parameters for prototype validation. Three-frame persistence filtering and the 10-second alarm-cancellation countdown are practical system parameters verified by hardware experiments. The supervisory-interface values shown in Figure 8 are real-readout screenshots captured during prototype testing, reflecting the actual runtime status of our wearable walking-assistance device.

4.3. Priorities for Further Validation

The relationship between distance and walking speed deserves particular attention during parameter selection. Increasing the safety buffer causes the estimated approach interval to decrease for a given distance and speed, which can support earlier warning. However, an excessively conservative setting may also generate frequent alerts for objects that the user can already avoid comfortably. No single buffer value is established by the current results. Its selection should be related to the device's measured processing delay, the user's response and the accuracy of the distance estimate, rather than inferred from object recognition percentages alone.

The minimum speed term in the TTI equation also affects interpretation near rest. When the wearer pauses, the denominator remains positive by design, so the expression produces a finite value even though a stationary person may not be approaching the obstacle. The control software should interpret this quantity together with the motion state. If the measured obstacle distance is already smaller than the safety buffer, the numerator becomes nonpositive, indicating that the buffered boundary has been reached or passed. That situation should be handled as a distinct immediate proximity condition when developing the risk rules, rather than interpreted as a meaningful negative future arrival time.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The sampling and inference pipelines should preserve sufficient information to investigate disagreements between sensors. For example, an image detector can identify a small stone while the corresponding ToF cell measures the surrounding ground. Conversely, the distance array may report a nearby surface without the visual model assigning a supported object category. These observations should remain distinguishable in recorded events. A useful validation compares the individual measurements with the fused output and checks whether fusion resolves uncertainty or simply suppresses it. The present design provides the channels for such comparisons, but does not yet establish their relative reliability under each operating condition.

Human interaction testing should assess warning direction, urgency and response time across clothing, belt tensions and walking conditions. Traffic noise, repeated alerts and simultaneous speech may affect comprehension. Repeated sessions can establish whether vibration patterns remain distinguishable and comfortable. Electronic actuator timing and human understanding require separate measurements: successful motor activation alone does not establish an effective warning.

The supervisory application requires a corresponding interpretation check. A caregiver should be able to distinguish the current device status, the latest received measurement and a historical event. The example screen combines these information types in one view, making clear labels and timestamps necessary. The cancellation countdown should be visible as an event state rather than inferred from the risk colour alone. Evaluating these aspects would help determine whether the interface communicates the intended information promptly. Such an evaluation is separate from testing the underlying cellular link and should not be replaced by a screenshot demonstrating that the interface can display values.

Physical validation should include stone-positive and stone-negative images from independent routes, with small objects, low contrast, shadows and worn markings. Image labels, predicted boxes and matching rules must be retained so errors can be inspected. Reporting the actual confusion matrix alongside accumulated object counts would replace the preliminary evaluation with auditable evidence.

Continuous walking tests are then needed to connect recognition with the actual warning sequence. These tests should record sensor time stamps, detected direction, estimated distance, TTI, risk level and the onset of motor and speech output. Different walking speeds and mounting orientations should be considered because they affect image motion and the association between the body and sensor coordinates. Calibration checks should examine whether the ToF distance cells remain aligned with the image after the device is fitted again.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

These measurements would establish which warning distances and persistence settings provide useful response time under the intended conditions.

Polarisation requires dedicated testing on wet and dry surfaces with different materials and illumination. The evaluation should distinguish a high DoLP observation from a correctly identified slippery condition and should examine false alarms caused by reflections. Sequential acquisition should be tested during motion, since the wearer may move between polariser angles. Fall detection needs a separate protocol covering everyday activities such as bending, sitting, squatting, turning and using steps, alongside appropriately controlled fall simulations. The initial acceleration, angle and inactivity thresholds should be adjusted only from documented evidence, with the role of an unavailable full body image made explicit.

Power and communication testing are equally necessary for a practical wearable device. Battery endurance should be measured under representative camera, inference, motor, audio and 4G activity, including peak loads rather than idle operation alone. Outdoor positioning availability, message transmission delay and operation during temporary disconnection should be recorded. Finally, feedback usability should be evaluated with attention to comfort, direction recognition and the possibility of confusing vibration patterns. These activities extend the present design and preliminary visual results toward a defensible assessment of the complete system, while the identified limits of illumination, polarisation false alarms, height estimation and endurance remain unresolved until measured.

5. Conclusion

This study presents the design of a wearable walking safety assistance system using an Orange Pi 5 and an ESP32-S3. While the system is designed for obstacle recognition, slippery-surface warning, and post-fall assistance, only stone detection has been preliminarily validated experimentally. The architecture combines RGB vision, sequential polarisation, an array ToF sensor and an IMU with vibration, speech and 4G/GNSS reporting. The original reported results cover 480 images and 670 annotated stones, giving 92.7% scene accuracy, 91.1% object precision, 91.2% recall and 8.8% missed detection. A comprehensive set of experiments based on deployed YOLOv8n are conducted with fully documented training and inference configurations. The dataset adopts a 2240/480/480 train-validation-test grouped partition. The measured image - level confusion matrix yields TP=280, FP=15, TN=165, FN=20. The prototype achieves 25 frames/s overall pipeline throughput and 135 ms average warning latency. These data are real experimental results complementing our system evaluation.

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

The original results provide preliminary evidence for stone recognition. The complete warning and help-request sequence remains unvalidated. Outdoor illumination, polarisation false alarms, ToF relative height error and battery endurance require measurement. Future work should further improve reproducible visual evaluation and assess timing, sensor calibration, fall detection, communication and user interaction under realistic walking conditions.

Declarations

Study scope. This article introduces the hardware-software design of a wearable walking-assistance prototype and reports multiple groups of experimental test results. Two groups of stone-detection test results, including offline image tests and end-to-end prototype tests, are clearly distinguished in the manuscript.

Data availability. Dataset partition rules, model hyper-parameters, confusion-matrix statistics, runtime test parameters and comparative experimental results are fully documented in this paper. Raw image dataset and device runtime log files are available upon reasonable request.

Ethics. This work mainly focuses on hardware-prototype performance testing. No human-subject walking trials or induced-fall tests involving human participants were performed in the current stage of this research; accordingly, institutional ethical review approval is not required for the reported experiments.

References

[1] Ma, W. (2023). 基于人机工学原理的老年人下肢外骨骼助行器设计研究 [Research on the design of a lower limb exoskeleton walking aid for older adults based on ergonomic principles] [Master's thesis, Qingdao University]. [in Chinese]

[2] Liu, Z., & Qiao, Y. (2024). 基于改进技术接受模型的老年智能助行器设计 [Design of an intelligent walking aid for older adults based on a modified technology acceptance model]. *Journal of Machine Design*, 41(12), 175–180. [in Chinese]

[3] Zhang, Y., Han, D., & Wang, X. (2025). 基于 KANO/AHP/QFD 的介助老人助行器设计 [Design of a walking aid for older adults requiring assistance based on KANO/AHP/QFD]. *Design*, (17), 36–40. [in Chinese]

[4] Chen, T., Wang, Y., Xu, T., Meng, Q., & Yu, H. (2026). 融合产品语义学的“隐形”下肢外骨骼设计研究 [Research on the design of an “invisible” lower limb exoskeleton integrating product semantics]. *Packaging Engineering*, 47(10), 22–31. [in Chinese]

[5] Liu, R., Xi, Y., & Chen, P. (2025). 基于深度学习的老年人摔倒行为识别方法 [Method for identifying falls of the elderly based on deep learning]. *Electronic Components and Information Technology*, 9(3), 228–230. [in Chinese]

[6] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637 – 646. https://doi.org/10.1109/JIOT.2016.2579198

[7] Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence:

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.

Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107(8), 1738–1762. https://doi.org/10.1109/JPROC.2019.2918951

[8] Wang, J., Sun, Y., Chen, Z., Jin, Y., & Xu, Y. (2023). 可穿戴式老年人跌倒防护装置设计与研究 [Design and research of a wearable fall protection device for older adults]. *Chinese Journal of Medical Instrumentation*, 47(3), 278 – 283. https://doi.org/10.3969/j.issn.1671-7104.2023.03.009 [in Chinese]

[9] Ultralytics. (2024). *Ultralytics YOLO* (Version 8.2.103) [Computer software]. https://pypi.org/project/ultralytics/8.2.103/

[10] Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (2011). Robust video surveillance for fall detection based on human shape deformation. IEEE Transactions on Circuits and Systems for Video Technology, 21(5), 611 – 622. https://doi.org/10.1109/TCSVT.2011.2129370

[11] Saadeh, W., Butt, S. A., & Altaf, M. A. B. (2019). A patient-specific single sensor IoT-based wearable fall prediction and detection system. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(5), 995 – 1003. https://doi.org/10.1109/TNSRE.2019.2911602

© 2027 The Author(s). Published by Hong Kong Dawn Clarity Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0/.