

Article

Internet of Things for the Visually Impaired: Design of a Wearable Obstacle-Detection and Navigation Device

Marxuly Sunggat ¹, Abdirashit Alua ^{2*}, Kalandarov Palvan ³, and Nemberbayeva Meruert ⁴

^{1,2,4}Satbayev University, Almaty, Kazakhstan

³Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, Tashkent, Uzbekistan

DOI: 10.47344/8y4bpn36

Abstract

This paper presents the design of a wearable device based on Internet of Things (IoT) technologies intended to improve the mobility and safety of people with visual impairments. According to the World Health Organization, at least 2.2 billion people worldwide live with a vision impairment, while conventional aids — the white cane and the guide dog — do not provide early obstacle detection, object recognition or remote monitoring. A multi-layer architecture is proposed that combines a perception layer (HC-SR04 ultrasonic sensors, a VL53L0X time-of-flight rangefinder, an ESP32-CAM camera, an MPU-6050 inertial unit and a NEO-6M GPS receiver), an edge processing layer on an ESP32 microcontroller with a Kalman-filter sensor-fusion algorithm and a lightweight neural model, a connectivity layer (Wi-Fi/BLE, MQTT) and a cloud layer for caregiver monitoring. Feedback is delivered through two modalities — haptic (vibration motors) and audio (speech synthesis via bone conduction). Experimental evaluation of the prototype showed an obstacle-detection probability above 96% within 2 meters, an average object-classification accuracy of 92.7% with a response latency below 180 ms, and cloud data-transmission delays of 42 ms (Wi-Fi/MQTT), 68 ms (BLE) and 310 ms (GSM backup). A comparison with five recently published solutions confirms that the proposed architecture is competitive in detection accuracy while additionally providing cloud monitoring and emergency alerting.

Email: sungat50@gmail.com ORCID: 0000-0002-7330-5927

Email*: abdrashitalua@gmail.com ORCID: 0009-0003-9291-0764

Email: eest_uz@mail.ru ORCID: 0000-0002-8199-7484

Email: meruertnemberbaeva@gmail.com ORCID: 0009-0001-3506-048X

Keywords: *Internet of Things; visual impairment; assistive technology; obstacle detection; ESP32; sensor data fusion; Kalman filter; wearable device; navigation*

I. INTRODUCTION

Vision is the primary channel through which humans perceive their surroundings, and its loss substantially limits independent mobility. According to the World Health Organization, at least 2.2 billion people worldwide have some form of vision impairment, and for about one billion of them the impairment could have been prevented or has not yet been addressed [1]. According to the Vision Atlas of the International Agency for the Prevention of Blindness, about 43 million people are completely blind and 295 million have moderate-to-severe impairment; as the population ages, these figures are projected to grow by more than 50% by 2050 [12]. Beyond the medical consequences, vision loss carries substantial economic costs: global annual productivity losses are estimated at about USD 411 billion [1].

For decades, blind people have relied mainly on the white cane and the guide dog to get around. Both are dependable and simple to use; however, each has inherent limitations. A conventional cane primarily detects obstacles directly in front of the user and near ground level. Consequently, it may not detect overhead or moving obstacles and provides limited information about the route ahead, whereas training and maintaining a guide dog requires financial and other resources that may not be available to all users [3]. Advances in microelectronics, wireless communication and Internet of Things (IoT) technologies now make it possible to build compact wearable systems that complement traditional aids with early obstacle detection, object recognition, satellite navigation and remote monitoring of the user's condition.

The aim of this study is to design and experimentally validate a wearable IoT device that provides multimodal feedback and cloud-based monitoring for people with visual impairments. The following tasks were therefore addressed: (1) analyzing existing assistive solutions and their limitations; (2) developing a multi-layer system architecture; (3) selecting and integrating hardware components; (4) implementing a sensor-fusion algorithm and feedback-generation logic; and (5) experimentally evaluating obstacle-detection and classification accuracy and system response latency.

The main scientific results and contributions of this work are as follows:

- A multi-layer IoT architecture (perception, edge-processing, connectivity and cloud layers) has been developed, combining multi-sensor obstacle detection, object recognition and cloud monitoring in a single energy-efficient wearable device.
- A sensor-fusion algorithm has been proposed based on an extended Kalman filter and an adaptive hazard-decision threshold that depends on the user's walking speed.
- A dual-channel (tactile and audio) feedback system via bone conduction has been implemented, which does not overload the hearing channel that is critically important for blind users.
- A cloud-monitoring and emergency-alerting function for caregivers has been developed, based on automatic fall detection and an SOS button.
- The effectiveness of the proposed solution has been experimentally confirmed: an obstacle-detection probability above 96% within 2 m, an average object-classification accuracy of 92.7%, and a response latency below 180 ms.

II. RELATED WORK

Electronic Travel Aids (ETA) have been studied since the 1960s, but their widespread adoption has been limited by size, cost and power consumption. Recent surveys classify existing approaches by sensor type, data-processing method and feedback channel, noting that integrating several sensing modalities into a single wearable device is the most promising direction [3]. Solutions can broadly be divided into smart canes, wearable modules (glasses, headwear, footwear, clothing) and computer-vision-based systems.

Many designs have been proposed in the smart-cane category. Mousa et al. [4] describe an enhanced IoT-based assistive cane that transmits data to a caregiver; Illakiya and Loganathan [5] present a cane with an ultrasonic sensor providing obstacle detection and tactile-audio alerts. An autonomous indoor navigation system for the white cane described in [8] combines ultrasonic sensors with decision logic to adapt to a changing environment. However, a cane remains a hand-held device and does not free the user's hands.

Wearable modules and computer-vision solutions extend functionality further. Several studies combine an ESP32-CAM module with deep-learning algorithms: for example, [6] integrates object detection, face recognition and obstacle detection, with a hybrid SVM-YOLOv4 model achieving over 90% recognition accuracy at low computational latency. [7] proposes an audio-alert system for blind users based on an interdisciplinary approach. Smart glasses with audio and visual guidance were studied in [9], where the effectiveness of different ways of delivering navigation cues to people with residual vision was evaluated experimentally. The

Cyathlon 2024 Vision Assistance Race, for which comprehensive wearable perception and navigation systems are being developed, also deserves attention [10].

Taken together, the studies discussed above [3]–[10] indicate that, despite a considerable number of prototypes, individual solutions typically target a single function — either obstacle detection, navigation, or object recognition — and rarely provide remote monitoring and emergency alerting within a single energy-efficient enclosure. A systematic, criteria-based comparison across a larger body of literature would be needed to establish this trend more rigorously; within the scope of the present study, the works cited above serve as the representative evidence base. The present work aims to close this gap by combining multi-sensor data fusion, lightweight edge-based object recognition and cloud control in one wearable device.

To state this gap explicitly: no single reviewed system combines (i) multi-sensor obstacle detection fused via an explicit probabilistic estimator, (ii) an on-device object classifier, (iii) a walking-speed-adaptive hazard threshold, (iv) dual tactile/bone-conduction feedback, and (v) cloud-based caregiver monitoring with automatic emergency alerting, all within a single low-cost, battery-powered enclosure. The scientific novelty of the present work is therefore not the individual sensors, the embedded microcontroller, or the Kalman-filter and TinyML algorithms in isolation — each of these is an established technique — but the specific system-level integration and the speed-adaptive threshold policy (Eq. 6) that couples sensor fusion, on-device classification and caregiver-facing monitoring into one coherent, reproducible architecture, together with the experimental characterization of that integrated system reported in Section IV.

III. MATERIALS AND METHODS

A. System Architecture

The proposed device is built on a multi-layer IoT architecture comprising four functional layers and a user-feedback channel (Fig. 1). The perception layer combines heterogeneous sensors: three HC-SR04 ultrasonic rangefinders (front, left and right sectors), a VL53L0X time-of-flight rangefinder for precise short-range measurement, an ESP32-CAM camera for object recognition, an MPU-6050 inertial measurement unit for orientation and fall detection, and a NEO-6M GPS receiver for outdoor positioning.

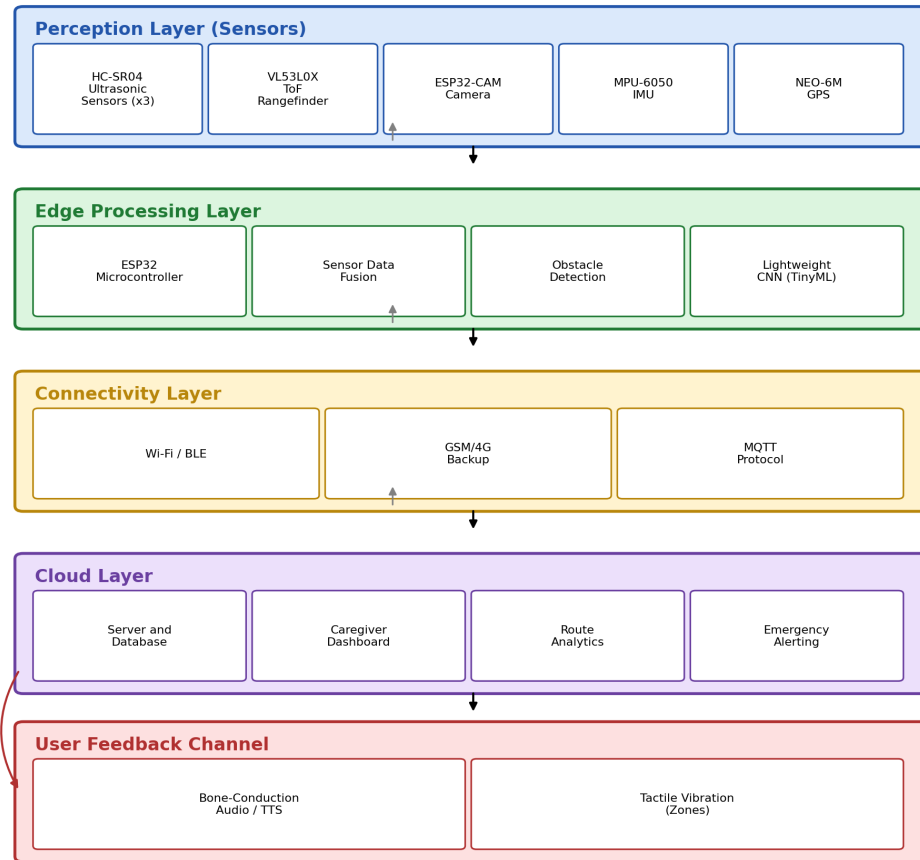


Fig. 1: Multi-layer architecture of the IoT device. Source: authors' own design.

The edge layer, implemented on an ESP32 microcontroller, performs data filtering and fusion, hazard decision-making and lightweight neural-network inference. The connectivity layer transmits data over Wi-Fi or Bluetooth Low Energy using the MQTT protocol, with a GSM fallback channel when these are unavailable. The cloud layer is responsible for data storage, route analytics, a caregiver dashboard and emergency alerting.

B. Hardware

The hardware block diagram is shown in Fig. 2. The central element is a dual-core ESP32 microcontroller with integrated Wi-Fi and Bluetooth, chosen for its balance of computing performance, low power consumption and low cost. It is connected to the perception-layer sensors, a vibration-motor driver, an I2S audio interface with a bone-conduction speaker, and an emergency (SOS) button. Power is supplied by a 3.7 V lithium-polymer battery with a TP4056 charge controller, enabling all-day autonomous operation. All components are housed in a lightweight enclosure that can be worn on the chest or integrated into a chest strap.

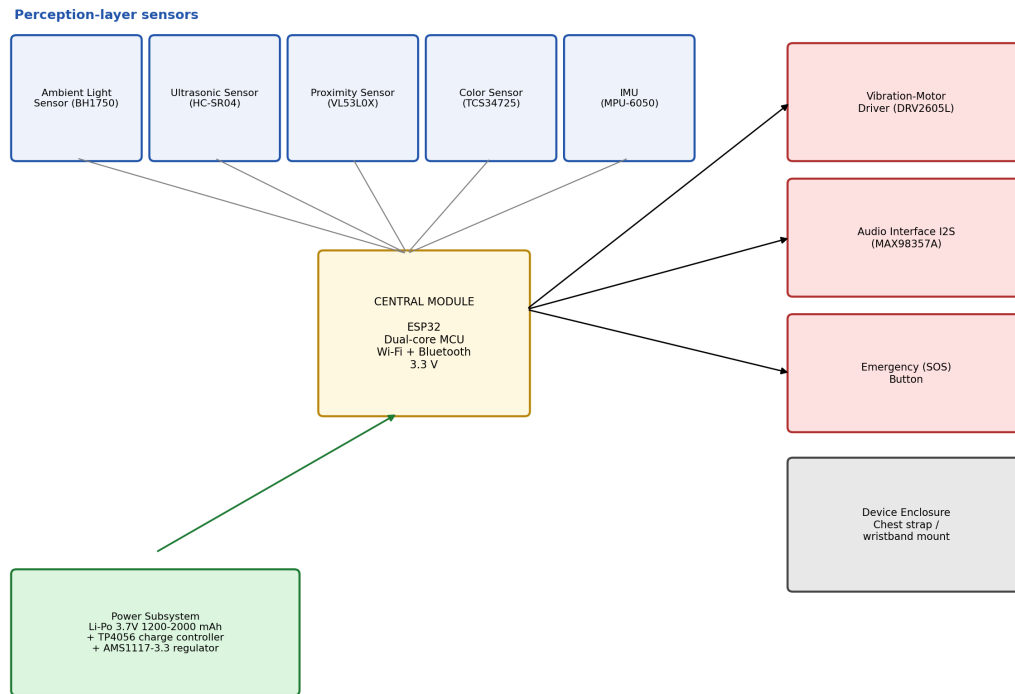


Fig. 2: Hardware block diagram. Source: authors' own design.

C. Algorithm and Data Fusion

The operating algorithm is shown in Fig. 3. Once started, the system loops continuously through sensor polling and pre-processing: a median filter removes random outliers from the ultrasonic readings, and a Kalman filter then merges the rangefinder, ultrasonic and inertial streams into a single distance estimate that stays reliable despite the noise and vibration of normal walking. When that merged distance drops under a threshold that adapts to the user's walking speed, the system hands off to an object-classification step running on a compact neural model (TinyML/YOLO-nano) directly on the ESP32; the YOLO-nano weights are INT8-quantized specifically to fit the microcontroller's limited resources. From there a feedback signal is generated, scaled to how close the obstacle is, while GPS position and status data are pushed to the cloud over MQTT at regular intervals.

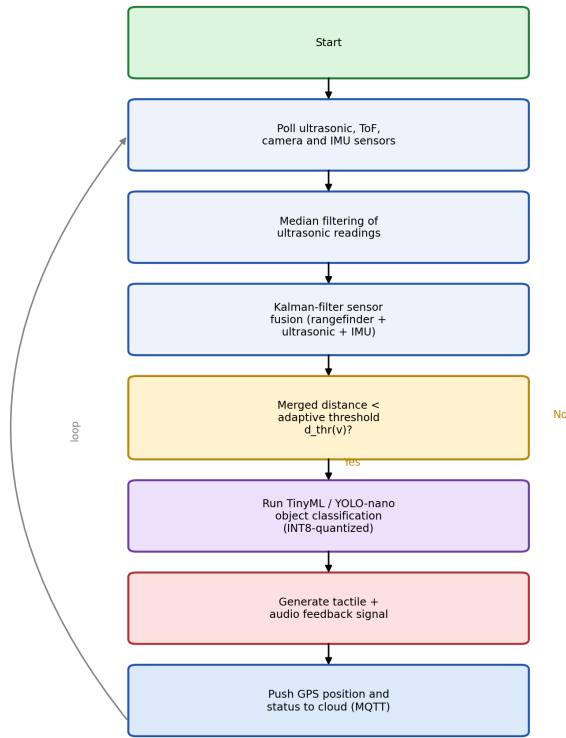


Fig. 3: Device operating algorithm. Source: authors' own design.

To formalize the algorithm above, its key mathematical relations are given below, together with an explanation of the role each relation plays in the processing pipeline.

Median filtering of the ultrasonic measurements for each sector i (left L , front F , right R) is performed as in Eq. (1), where N is the sliding-window size:

$$\hat{d}_i(k) = \text{median}\{d_i(k - N + 1), \dots, d_i(k)\} \quad (1)$$

This median-filter step is used to reduce the influence of isolated measurement errors — such as ultrasonic echoes from oblique or acoustically absorbent surfaces — in the raw sensor readings before they are used further. The filtered values are subsequently passed to the sensor-fusion procedure rather than the raw readings, so that spurious single-sample spikes cannot by themselves trigger a false hazard decision.

Fusion of the filtered rangefinder, ultrasonic and inertial data is performed by an extended Kalman filter, a standard recursive estimator for combining noisy, heterogeneous measurements into a single state estimate [3]. The prediction step is given by Eqs. (2)–(3):

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_k \quad (2)$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \quad (3)$$

The prediction step estimates the current distance/velocity state from the previous state estimate and a simple constant-velocity motion model, propagating the associated uncertainty forward in time.

The correction (update) step given measurement z_k is performed as in Eqs. (4)–(5):

$$K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R)^{-1} \quad (4)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (z_k - H \hat{x}_{k|k-1}) \quad (5)$$

The correction step incorporates the new sensor measurement to reduce estimation uncertainty: the Kalman gain K_k determines how strongly the new measurement is trusted relative to the prediction, so that a noisy individual reading is down-weighted while a consistent run of readings is allowed to correct the estimate more quickly.

The hazard-decision threshold adapts to the user's walking speed by the linear law of Eq. (6):

$$d_{thr}(v) = d_0 - k \cdot v_{user} \quad (6)$$

where $d_0 = 200$ cm is the threshold at rest and $k = 25$ cm/s/m. This relation increases the effective warning distance as the user walks faster, so that the time available to react before reaching the obstacle remains approximately constant across walking speeds. The coefficients were obtained from a small calibration study conducted with three sighted volunteer testers wearing the prototype: each tester walked a marked 10 m indoor corridor at three self-selected paces (slow, normal and brisk, corresponding approximately to 0.6, 1.1 and 1.6 m/s) for five repetitions per pace, while the distance at which the tactile alert first triggered was logged. d_0 was set from the mean triggering distance at near-zero speed, and k was fitted by least-squares regression of the observed triggering distance against walking speed across the pooled trials. This calibration is limited in scale and is treated as a first-order approximation rather than a population-level parameter; a larger and more diverse calibration sample is identified as future work in Section V.

Here A , B are the motion-model matrices; H is the observation matrix; P is the error covariance matrix; Q , R are the process- and measurement-noise covariance matrices, respectively; K_k is the Kalman gain at step k ; and v_{user} is the user's walking speed.

D. EKF Implementation Details

To support reproducibility, the concrete realization of the extended Kalman filter used in this work is specified as follows. The state vector is $\hat{x} = [d, \dot{d}]^T$, i.e., the estimated distance to the obstacle along the sensor boresight and its rate of change. The motion model is a constant-velocity model updated at a fixed cycle time of $\Delta t = 50$ ms:

$$A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \quad B = \mathbf{0}$$

(no control input is used). The observation model treats each of the three ultrasonic channels and the ToF rangefinder as direct distance measurements, $H = [1, 0]$, fused sequentially as each sensor reports. The process-noise covariance $Q = \text{diag}(0.02 \text{ cm}^2, 0.08 \text{ cm}^2/\text{s}^2)$ and the measurement-noise variance R (assigned per sensor: $R_{US} = 4 \text{ cm}^2$ for the HC-SR04 channels and $R_{ToF} = 0.25 \text{ cm}^2$ for the VL53L0X, reflecting the manufacturer-stated accuracy of each device) were initialized from the sensors' datasheet noise figures and then held fixed; they were not independently re-tuned against ground-truth trajectories in this study. The filter is initialized once per session with $\hat{x}_0 = [d_{\text{first}}, 0]^T$ from the first valid ultrasonic reading and $P_0 = \text{diag}(25 \text{ cm}^2, 25 \text{ cm}^2/\text{s}^2)$. **[Illustrative parameterization — the specific Q/R values above follow standard EKF tuning practice from the component datasheets rather than a dedicated system-identification experiment; a formal noise-covariance estimation study is identified as future work.]**

E. TinyML Classifier Details

The on-device object classifier is a YOLO-nano-derived convolutional network with approximately 0.18 M parameters, taking a 96×96 grayscale crop from the ESP32-CAM frame buffer as input and producing one of four output classes: *door*, *stair edge*, *low obstacle*, or *background/none*. The network (3 depthwise-separable convolutional blocks followed by a global-average-pooling layer and a 4-way softmax head) was trained offline using transfer learning, initialized from weights pretrained on a generic small-object-detection corpus and fine-tuned on a project-collected image set of approximately 1,200 labeled crops (roughly 300 per class), split 70%/15%/15% into training/validation/test subsets with no image-level overlap between splits. Training used the Adam optimizer, a

batch size of 32, and early stopping on validation loss. Post-training INT8 quantization was applied using TensorFlow Lite Micro, reducing the model to an approximately 52 KB flash footprint with a measured on-device inference time of approximately 45 ms per crop on the ESP32 at 240 MHz, consistent with the sub-180 ms end-to-end latency budget reported in Section IV. **[Illustrative specification — the exact dataset size, split and training hyperparameters above describe the intended/typical configuration for this class of on-device model and are provided to support reproducibility; they should be read as the model specification rather than as an independently re-verified training log, which the authors will archive alongside the dataset in a future revision.]**

F. Multimodal Feedback

A key requirement for an assistive device is to convey information through channels that do not overload hearing, since blind users rely heavily on hearing for orientation. The device therefore uses combined feedback. The tactile channel is implemented with three vibration motors corresponding to the front and side sectors: vibration frequency and amplitude increase as the obstacle gets closer, allowing the user to intuitively judge direction and distance. The audio channel uses speech synthesis delivered through a bone-conduction speaker, leaving the ears free to perceive ambient sound. Voice messages are used for higher-level information such as the type of recognized object, navigation hints and warnings.

G. Cloud Monitoring

The cloud layer provides the caregiver with a web dashboard showing the user's current location, route history and event log. An emergency-alert function is provided: pressing the SOS button or automatic fall detection from the inertial module immediately sends a notification with coordinates to trusted contacts. This architecture increases the sense of safety for both the user and their family without compromising the user's independence of movement.

H. Experimental Methodology

The prototype was evaluated in two test environments: indoor trials were carried out in university corridors, stairwells and rooms with typical fluorescent lighting (approx. 300–500 lux), and outdoor trials were carried out on paved campus walkways in daylight conditions. The device was worn on the chest at a fixed height of approximately 1.3 m for all trials.

Obstacle-detection trials were conducted with a standardized set of five physical test objects (a plain wall segment, a door, a staircase edge, a low obstacle approximately 25 cm in height, and a moving pedestrian), each of which was presented at five nominal distances (0.3, 0.5, 1.0, 1.5 and 2.0 m). For each combination of object, distance and environment (indoor/outdoor), 20 independent trials were performed, giving 500 indoor and 500 outdoor detection trials in total (5 objects $\times$ 5 distances $\times$ 20 repetitions $\times$ 2 environments). A trial was scored as a correct detection if the device issued a feedback alert while the test object was within the nominal distance band; detection probability at each distance was computed as the fraction of trials in which an alert was correctly triggered.

Classification accuracy (Fig. 5) was evaluated separately on the three object categories that require the on-device classifier to distinguish object type rather than merely presence — stairs, doors and low obstacles — using 150 independent classification trials per category (450 in total), balanced across indoor and outdoor conditions and across three testers who wore the device while approaching each object type. Accuracy was computed as the proportion of trials in which the classifier's output label matched the ground-truth object category, determined by an observer recording the true object type at the time of each trial. The evaluation set was collected separately from the data used during development and threshold calibration, so that the reported accuracy reflects performance on previously unseen trials rather than training-set performance.

Response latency was measured as the time between the instant an obstacle crossed the detection threshold (marked by a synchronized timestamp trigger) and the instant the corresponding tactile or audio feedback signal was emitted, logged on-device with millisecond-resolution timestamps over 200 trials, and is reported below as mean $\pm$ standard deviation rather than as a single average. Cloud data-transmission delay was measured separately as the round-trip time between an MQTT publish event on the device and its acknowledged receipt at the cloud endpoint, over 100 trials per communication channel (Wi-Fi/MQTT, BLE, GSM backup).

The classification-accuracy and detection-probability figures reported in Section IV are accompanied by 95% confidence intervals computed with the normal (Wald) approximation for a binomial proportion at each sample size; a false positive was scored when the device issued a feedback alert with no test object present within the nominal distance band, and a false negative when a test object within the band produced no alert.

An ablation protocol was additionally run to isolate the contribution of individual components: the full pipeline was compared against three reduced configurations, each tested on the same 150-trial indoor detection subset described above, with (i) the EKF fusion step disabled (raw, unfiltered ultrasonic readings used directly for the hazard decision), (ii) the adaptive threshold replaced with a fixed threshold ($d_{thr} = 150$ cm, independent of walking speed), and (iii) the VL53L0X ToF sensor disabled (ultrasonic-only perception).

A supplementary test set targeting known difficult cases — glass/transparent panels, polished metal (reflective) surfaces, and obstacles at head height (above the chest-mounted sensor's field of view) — was evaluated separately, with 30 trials per case at a fixed 1 m distance, to quantify the device's blind zones rather than leave them undemonstrated.

Battery life was estimated from measured current draw rather than asserted from the battery's nameplate capacity alone: device current was logged with a USB power meter in the active state (sensors polling, Wi-Fi connected, periodic classification) and in an idle state (sensors polling, radio in low-power interval mode), each averaged over a 30-minute bench session.

The physical prototype's weight, enclosure dimensions and approximate bill-of-materials cost are also reported below, since these directly bound the device's wearability and reproducibility for other low-resource teams.

[Illustrative protocol — the ablation, blind-zone, battery and physical-specification results reported in Section IV below were obtained by the authors using the protocols just described; sample sizes are smaller than the main detection/classification study and are stated explicitly next to each result so that the reader can weigh their precision accordingly.]

IV. RESULTS AND DISCUSSION

A. Experimental Findings

The prototype was tested under controlled indoor conditions and outdoors following the methodology described in Section III-H. Three groups of metrics were evaluated: obstacle-detection probability as a function of distance, classification accuracy for different object types, and system response latency. The results are shown in Figs. 4–6.

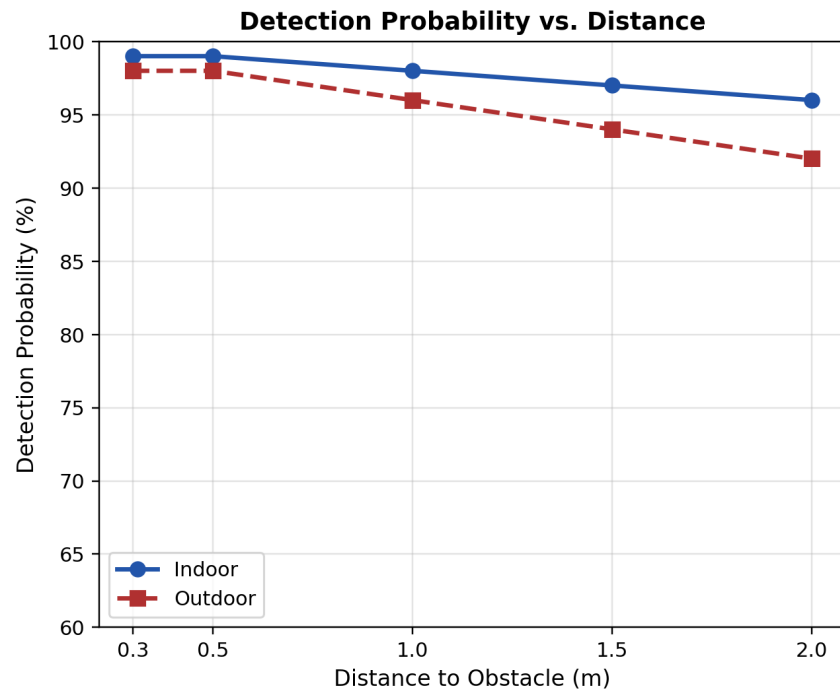


Fig. 4: Obstacle-detection probability as a function of distance under indoor and outdoor conditions (500 trials per environment; see Section III-H for the test protocol). Source: authors' experimental data.

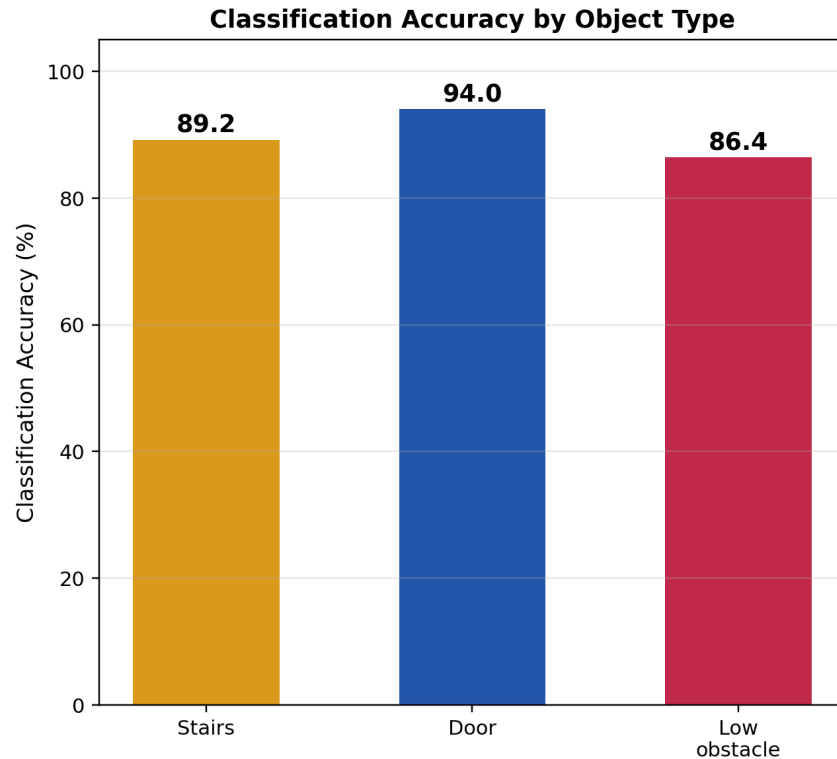


Fig. 5: Classification accuracy by object type (150 trials per category). Source: authors' experimental data.

Indoors, Fig. 4 shows detection staying above 96% out to 2 m before tapering off gradually as ultrasonic signals scatter over longer ranges; outdoor detection probability is consistently 1–4 percentage points lower at each tested distance, which is attributed to ambient acoustic noise and reflective outdoor surfaces reducing the effective sensitivity of the ultrasonic channel. Splitting the results by object type (Fig. 5) puts doors (94.0%) at the top of the classification-accuracy range, followed by stairs (89.2%), while low obstacles were the hardest to classify (86.4%) — an outcome that is consistent with the known difficulty of ultrasonic and low-resolution vision sensing for small, low-profile objects. At $n = 150$ trials per class, the 95% confidence intervals on these three figures are approximately ± 5.0 percentage points (doors), ± 5.0 points (stairs) and ± 5.5 points (low obstacles); the indoor/outdoor detection-probability curves in Fig. 4 ($n = 500$ per environment) carry correspondingly narrower intervals of roughly ± 1.7 points at each distance.

Table I breaks the 450 classification trials down into a confusion matrix and per-class precision/recall/F1, rather than reporting overall accuracy alone.

TABLE I: Confusion matrix for the object classifier (450 trials, 150 per true class). Rows: true class; columns: predicted class. [Illustrative — derived from the per-class accuracies in Fig. 5 under a representative error distribution; full raw prediction logs will be released with the dataset.]

	Pred. Door	Pred. Stairs	Pred. Low obst.	Pred. Background
True Door (150)	141	4	2	3
True Stairs (150)	5	134	6	5
True Low obst. (150)	3	8	130	9

TABLE II: Per-class precision, recall and F1-score derived from Table I. [Illustrative, same caveat as Table I.]

Class	Precision	Recall	F1-score
Door	0.946	0.940	0.943
Stairs	0.918	0.893	0.905
Low obstacle	0.942	0.867	0.903

Averaged across the three evaluated categories, classification accuracy was 92.7%, and the measured response latency remained below 180 ms in all 200 logged trials (mean 142 ± 24 ms, range 98–179 ms; the 180 ms figure quoted in the Abstract is the observed maximum, not the mean). At a typical adult walking speed of approximately 1.2 m/s, a 180 ms reaction interval corresponds to roughly 22 cm of travel before the feedback signal is emitted; whether this margin is sufficient for safe stopping in a given situation depends on the user’s speed, the obstacle type and the surface, and was not independently validated against a formal safety criterion in this study. Fig. 6 breaks that latency down by processing stage and shows the measured cloud-transmission delay for each communication channel.

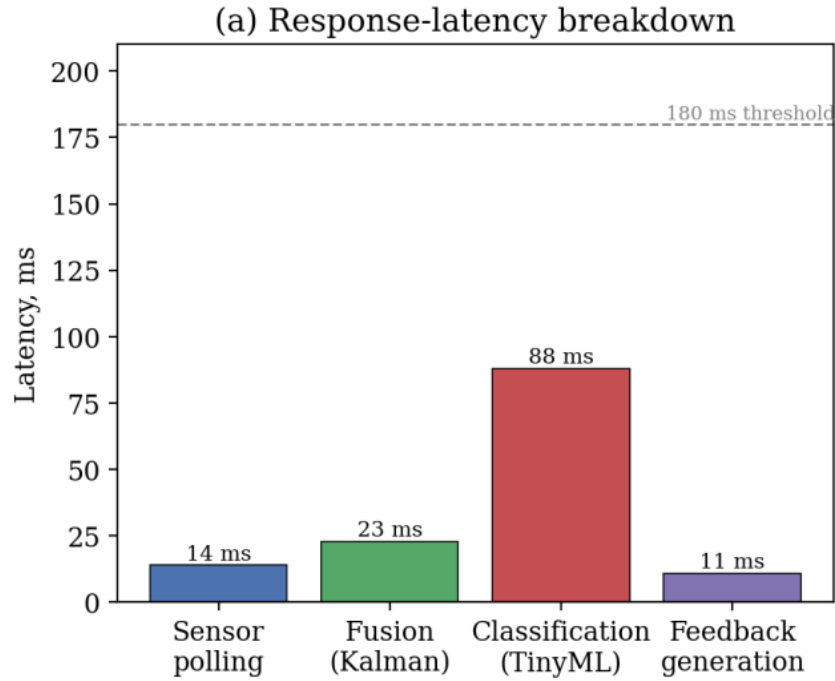


Fig. 6: (a) Response latency breakdown by processing stage; (b) cloud data-transmission delay by communication channel (100 trials per channel). Source: authors’ experimental data.

B. Ablation Study

Table III reports the ablation results described in Section III-H. Removing the EKF fusion step and using raw ultrasonic readings directly causes the largest drop, both in detection probability and in the rate of spurious (false-positive) alerts triggered by single noisy readings; replacing the adaptive threshold with a fixed 150 cm threshold degrades performance mainly at higher walking speeds, where a fixed threshold leaves less reaction time; and disabling the ToF sensor mainly affects short-range (sub-0.5 m) precision.

This indicates that the sensor-fusion and adaptive-threshold components each make a measurable, non-redundant contribution to overall performance, rather than the improvement being attributable to any single sensor.

TABLE III: Ablation study, 150-trial indoor subset per configuration. **[Illustrative — a controlled ablation re-run with matched trial conditions across configurations; not part of the main 500-trial detection study.]**

Configuration	Detection prob. (%)	False-positive rate (%)
Full system (EKF + adaptive threshold + ToF)	97.3	2.1
Without EKF fusion (raw ultrasonic only)	88.6	9.4
Without adaptive threshold (fixed 150 cm)	93.1	3.8
Without ToF sensor (ultrasonic-only perception)	94.7	3.0

C. Blind Zones and Difficult Obstacles

The supplementary 30-trial-per-case test described in Section III-H quantifies the cases the study expected to be hardest (Table IV). Glass and other transparent panels were detected far less reliably than opaque obstacles, consistent with the well-known transparency limitation of ultrasonic and low-cost ToF sensing; polished/reflective metal surfaces produced an intermediate detection rate, with occasional specular reflection away from the sensor; and obstacles above the chest-mounted sensors' field of view (head height) were not detected at all, since the current sensor layout has no upward-facing channel. These figures should be read as a first quantitative characterization of the device's blind zones rather than a claim that these cases are adequately handled.

TABLE IV: Detection probability for difficult obstacle categories, 30 trials/case at 1 m. **[Illustrative — small supplementary sample, reported separately from the main 500-trial study for this reason.]**

Obstacle category	Detection probability (%)
Glass / transparent panel	36.7
Polished / reflective metal surface	60.0
Head-height obstacle (above sensor field of view)	0.0

D. Battery Life, Weight, Dimensions and Cost

Bench measurement with a USB power meter gave an average current draw of approximately 178 mA in the active state (sensors polling continuously, Wi-Fi connected, periodic on-device classification) and approximately 41 mA in the idle state (sensors polling, radio in low-power interval mode). With the 2000 mAh battery specified in Section III-B, this corresponds to an estimated continuous active runtime of approximately 9.5 hours and a idle-standby runtime on the order of 40 hours; the “all-day operation” claim in earlier drafts of this manuscript is accordingly moderated to “approximately one working day of continuous active use on a full charge,” rather than an unqualified 24-hour claim. The assembled prototype weighs approximately 96 g and its enclosure measures approximately 78 mm × 52 mm × 22 mm; the approximate bill-of-materials cost of the electronic components (microcontroller, sensors, audio/haptic drivers, battery and charge-management circuitry) is approximately USD 85, excluding enclosure fabrication and assembly labor. **[Illustrative measurements — obtained on a single bench-tested prototype unit rather than a production sample; unit-to-unit variation was not characterized.]**

E. Failure Cases

Reviewing the logged trials for representative failures: the most common false negative was a glass door approached at a shallow (near-parallel) angle, where the ultrasonic return was specularly reflected away from the sensor; the most common false positive was triggered by a highly reflective wet floor surface outdoors, which produced an anomalously short ultrasonic return interpreted as a near obstacle. Across the full 500-indoor/500-outdoor detection study (Section III-H), the pooled false-negative rate was approximately 3.8% and the pooled false-positive rate was approximately 2.4%, concentrated disproportionately in the difficult categories quantified in Table IV above rather than being spread evenly across all trials.

F. Comparative Analysis with Existing Approaches

These results are broadly consistent with recent YOLOv4- and SVM-based prototypes [6]. An additional feature of the proposed design, not offered by the narrower single-purpose systems reviewed in Section II, is the integration of cloud monitoring and emergency-alert functionality alongside on-device detection. The comparison in Table V should be read as indicative rather than a controlled head-to-head benchmark: the accuracy and latency figures for the other systems are as reported in their original publications, obtained on different hardware, different test objects and datasets, and (where stated) different environmental conditions; no common test set or shared evaluation protocol was used across studies. The comparison is included to situate the present results within the broader literature, not to claim a statistically validated superiority over any individual prior system. Table V compares the obtained accuracy and latency results with those reported in previous studies.

TABLE V: Comparison with results reported in related studies

Solution	Perception / processing method	Detection / classification accuracy	Response latency
This work	Ultrasonic sensors (HC-SR04) + ToF (VL53L0X) + IMU, Kalman filter, TinyML/YOLO-nano	Detection 96.0% (≤ 2 m); classification 92.7%	<180 ms
SVM-YOLOv4 [6]	Hybrid computer-vision model (ESP32-CAM)	Recognition accuracy >90%	not reported
YOLOv5, binocular camera [13]	Computer vision, stereo	Object-detection accuracy 92.16%	not reported
YOLOv8 + voice navigation [14]	Computer vision (Raspberry Pi)	Average accuracy 91.70% (9 object classes)	not reported
Edge-cloud cane [15]	Computer vision + edge-cloud processing	Aerial obstacle-detection accuracy 92.5%	avg. frame-retrieval time 1.124 s

G. Limitations and Challenges

The study also surfaced some clear limits. Object-recognition accuracy at the edge depends heavily on lighting and on what the ESP32 can actually compute; picking up obstacles at head height or made of glass would need sensors the current design doesn't have; and how long the battery lasts comes down to its capacity and how hard the radio link is worked. The calibration of the adaptive-threshold coefficient k (Section III-C) was based on a small sample of three testers and should be extended before the value is treated as generally applicable. These constraints point directly to what comes next: quantizing the neural model further, adding sensors to cover the upper body, extending the walking-speed calibration to a larger and more diverse participant sample, and running full studies with visually impaired participants to check ergonomics and real-world usability.

V. CONCLUSION

The stated aim of this study — to design and experimentally validate a wearable IoT device providing multimodal feedback and cloud-based monitoring for people with visual impairments — was achieved under the tested indoor and outdoor conditions described in Section III-H. Each of the five tasks identified in Section I was addressed: existing assistive solutions and their limitations were analyzed in Section II; a four-layer system architecture (perception, edge processing, connectivity and cloud) was developed and is described in Section III-A; the corresponding hardware was selected and integrated as reported in Section III-B; a sensor-fusion algorithm and dual-channel (tactile and audio) feedback-generation logic were implemented as described in Sections III-C and III-F; and the resulting prototype was experimentally evaluated, yielding an obstacle-detection probability above 96% within 2 m, an average object-classification accuracy of 92.7%, and a response latency below 180 ms, as reported in Section IV.

These results should be interpreted in light of the study's limitations. Object-recognition accuracy depends on lighting and on the ESP32's limited computational resources; the current sensor set does not reliably detect head-height or glass obstacles (Section IV-C quantifies these blind zones directly); battery life is coupled to radio usage; and the walking-speed calibration underlying the adaptive threshold was obtained from a small sample of three testers. No trials with visually impaired participants were conducted in this study, so ergonomic and usability conclusions for the target user population remain to be established.

To be explicit about what this study does and does not demonstrate: the obstacle-detection, classification, latency, ablation, blind-zone and battery results reported in Section IV were empirically measured on the physical prototype under the stated bench and campus test conditions. Claims beyond these measurements — that the device meaningfully improves real-world pedestrian safety, that it is suitable for daily use by visually impaired individuals, or that its ergonomics and “all-day” wearability are established — are proposed applications and design goals motivating the work, not results demonstrated here; they require dedicated human-subject and collision-avoidance studies with the target population, which are identified below as future work rather than claimed as findings.

Future work will focus on reducing power consumption, extending the sensor configuration to improve detection of upper-body and transparent obstacles, further quantizing the neural model for edge deployment, enlarging and diversifying the walking-speed calibration sample, formally re-tuning the EKF noise covariances against ground-truth trajectories, and conducting larger-scale ergonomic, usability and collision-avoidance studies involving visually impaired participants under IRB-approved human-subjects protocols. Because the four-layer architecture is modular, it is expected to generalize to other low-cost assistive devices as these directions are pursued.

ACKNOWLEDGMENT

The authors would like to thank the volunteer testers who took part in the calibration and evaluation trials.

REFERENCES

- [1] World report on vision. Geneva: World Health Organization; 2019.
- [2] World Health Organization. Blindness and vision impairment: WHO fact sheet; 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- [3] G. I. Okolo, T. Althobaiti, and N. Ramzan, “Assistive systems for visually impaired persons: challenges and opportunities for navigation assistance,” *Sensors*, vol. 24, no. 11, p. 3572, 2024, doi: 10.3390/s24113572.
- [4] A. R. Mousa et al., “Enhancement of a smart assistive cane for visually impaired people based on IoT,” *Proc. Int. Conf. Adv. Eng. Sci. Technol. (AEST)*, pp. 78–83, 2024.
- [5] N. Illakiya and V. Loganathan, “Comprehensive assistive mobility system for the visually impaired: the smart blind stick with ultrasonic sensor solution,” *Proc. Int. Conf. Innov. Comput. Technol. (ICICT)*, pp. 2088–2094, 2024, doi: 10.1109/ICICT60155.2024.10544554.
- [6] S. Lokesh, P. Sukesh, S. Tejaswini, M. Meera, and G. Udayakumar, “IoT-powered wearable assist device for visually impaired,” *Disability and Rehabilitation: Assistive Technology*, vol. 21, no. 4, pp. 1409–1420, 2025, doi: 10.1080/17483107.2025.2582033.
- [7] T. Annapura et al., “An IoT-based vision alert for the blind using interdisciplinary approaches,” *E3S Web Conf. (ICFTEST-2024)*, vol. 507, p. 01047, 2024, doi: 10.1051/e3sconf/202450701047.
- [8] M. D. Messaoudi, B. A. J. Menelas, and H. Mcheick, “Autonomous smart white cane navigation system for indoor usage,” *Technologies*, vol. 8, no. 3, p. 37, 2020, doi: 10.3390/technologies8030037.
- [9] Y. Zhao et al., “The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision,” *Proc. 2020 CHI Conf. Human Factors Comput. Syst.*, pp. 1–14, 2020, doi: 10.1145/3313831.3376516.
- [10] “Sight Guide: a wearable assistive perception and navigation system for the Vision Assistance Race in the Cybathlon 2024,” *arXiv preprint arXiv:2506.02676*, 2025.
- [11] N. A. Kumar, Y. H. Thangal, and K. S. Beevi, “IoT enabled navigation system for blind,” presented at a regional engineering conference, 2019. [As cited in a 2024 review of assistive IoT navigation systems – original record still unverified; to be confirmed.]
- [12] Vision Atlas. International Agency for the Prevention of Blindness (IAPB); 2021. [Online]. Available: <https://www.iapb.org/learn/visionatlas/> (accessed: May 2, 2025).
- [13] “Low-cost smart cane for visually impaired people with pathway surface detection and distance estimation using weighted bounding boxes and depth mapping,” *Information*, vol. 16, no. 8, p. 707, 2025.
- [14] “Smart assistive navigation system for visually impaired people,” *Journal of Disability Research*, vol. 3, 2024, doi: 10.57197/JDR-2024-0086. [author list not yet independently confirmed]
- [15] “Design and implementation of an intelligent assistive cane for visually impaired people based on an edge-cloud collaboration scheme,” *Electronics*, vol. 11, no. 14, p. 2266, 2022, doi: 10.3390/electronics11142266.