Edge-Deployed Smart Wearable Glasses with Vision-Ultrasonic Late Fusion and Audio Feedback for Assistive Navigation

M Pradeep[1], Aruna Ajjarapu [2], M. Supriya[3], V. Jyothi[4], G. Vijaya Nirmala[5], Radhika Peeriga[6], A. Athiraja[7] and N. Rajeswaran[8, *]

[1] *Department of ECE, Shri Vishnu Engineering College for Women, West Godavari, Bhimaravam, Andhra Pradesh, 534202, India*
[2] *Department of Information Technology, Malla Reddy (MR) Deemed to be University, Maisammaguda, Hyderabad, Telangana, 500100, India*
[3] *Department of CSE (AIML), Geethanjali College of Engineering and Technology, Hyderabad, Telangana, 501301, India*
[4] *Department of ECE, Vardhaman College of Engineering, Shamshabad, Hyderabad, Telangana, 501218, India*
[5] *Department of ECE, CVR College of Engineering, Ibrahimpatnam, Hyderabad, Telangana- 501510, India*
[6] *Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Vijayawada, Andhra Pradesh, 522501, India*
[7] *Department of CSE, Saveetha Engineering College, Sriperumbadur, Chennai, Tamilnadu, 602105, India*
[8] *School of Engineering Technology & AI, MNR University, Fasalwadi, Sangareddy, Telangana, 502294, India*
***Corresponding Author E-mail:** rajeswarann@gmail.com
Received: Mar 24, 2026; **Revised:** Jul 21, 2026; **Accepted:** Aug 13, 2026

Abstract

This study presents a feasibility prototype of edge-deployed wearable glasses that combines camera-based object detection, ultrasonic proximity sensing, and bone-conduction audio feedback for visually impaired users. The revision narrows the contribution to system-level integration and evidence-controlled reporting rather than claiming a new object-detection architecture. The archived dataset records contain 3,250 public-source images and 550 independently captured live frames distributed across six operational classes. An Ultralytics YOLO detector was trained for 50 epochs at 416 × 416 resolution and executed on a Raspberry Pi 4. The retained training plot indicates approximately 0.49 mAP@0.5 and 0.21 mAP@0.5:0.95 at the final epoch; class-wise field results are reported separately because the original "accuracy" formulation was an IoU/Jaccard expression. The reported frame-processing time was 300 ms (approximately 3.3 FPS), which is adequate for static or slowly changing alerts but is not established as safe real-time navigation in dynamic traffic. Seven battery trials yielded 6.2–7.5 h of operation (mean 6.9 h). Preliminary feedback from 10 participants produced a mean overall rating of 4.4/5; because the archived package did not include participant-level responses or an ethics approval identifier, these observations are treated as formative usability evidence rather than clinical validation.

Keywords: Assistive Navigation, Wearable Edge AI, Object Detection, Ultrasonic Proximity Sensing, Late Fusion, Audio Feedback, Raspberry Pi

1. Introduction

At least 2.2 billion people worldwide have near- or distance-vision impairment, and mobility in unfamiliar environments remains a major barrier to independent participation [1]. Conventional aids such as long canes and guide dogs are indispensable, but they do not directly identify object categories or provide semantic scene information. Reviews of wearable mobility aids show that camera-, ultrasonic-, depth-, inertial-, and location-based technologies can complement established aids, although many prototypes remain limited by incomplete safety evaluation, weak user-centred validation, and inconsistent reporting [2–4]. These limitations are especially important for wearable systems because false or delayed alerts can increase cognitive load and create unsafe expectations.

Recent assistive-navigation research increasingly combines computer vision with ranging sensors and auditory or haptic feedback. However, "sensor integration" is often described without clarifying whether measurements are fused at the feature, decision, or output level, and object-specific distances are sometimes reported without a defensible camera-to-range association rule [5],[6]. Edge deployment can reduce cloud dependence and protect image privacy, but it does not by itself establish model optimization or real-time safety. Recent systematic evidence therefore recommends transparent reporting of sensor geometry, latency, energy consumption, failure modes, and evaluations involving visually impaired users [7].

This paper is consequently reframed as an edge-deployed feasibility study. Its specific contributions are: i) a reproducible system description that separates visual detection from ultrasonic sector ranging; ii) a direction-constrained late-fusion rule that withholds class-specific distance announcements when association is ambiguous; iii) corrected reporting of dataset totals, detection metrics, latency, and battery life; and iv) an explicit discussion of limitations associated with approximately 3.3 FPS processing, low-light scenes, reflective surfaces, and incomplete archival records. Section 2 critically reviews related systems, Section 3 presents the revised methodology, Section 4 reports the corrected evaluation and limitations, and Section 5 concludes the study.

2. Related Works

Wearable assistive-navigation research can be organized into five categories: vision-based environment perception, range-sensor obstacle detection, edge-computing deployment, multimodal feedback, and user-centred evaluation. Vision-based travel aids have used RGB, RGB-D, semantic mapping, and object recognition to provide obstacle identity, walkable direction, or scene context [8–10]. These systems offer richer semantic information than range-only aids, but their performance depends on illumination, viewpoint, occlusion, and computational capacity.

Systems that combine localization, crossing-state recognition, or wearable haptic guidance demonstrate that navigation requires more than object classification. Son and Weiland [11] integrated localization, signal recognition, and verbal guidance for crosswalk navigation, while Barontini et al. [12] evaluated wearable haptic obstacle-avoidance cues through a user-centred study. Depth-based methods can improve geometric awareness, but Jafri and Khan [13] emphasized the importance of user-centred design and explicit obstacle-avoidance logic.

Multi-sensor designs can improve coverage when the fusion rule is clearly defined. Bala et al. [14] combined several sensors for a wearable mobility aid, and recent RGB-D/IMU work has coupled obstacle identity, range, and multidimensional feedback [15]. Edge feasibility studies using lightweight YOLO-family detectors on Raspberry Pi platforms demonstrate the value of local processing, but also show that frame rate, model variant, software version, and hardware-specific optimization must be reported rather than inferred [16],[17].

Human-subject evaluation is essential because navigation quality cannot be established from detector outputs alone. Controlled studies such as UNav and teleguidance evaluations report task completion, standardized usability measures, or comparisons with baseline guidance [18],[19]. Across this literature, strong studies disclose participant characteristics, ethics approval or exemption, informed consent, task design, collision or near-collision events, and raw or reconstructable outcome measures.

The reviewed literature shows that a camera, Raspberry Pi, ultrasonic sensor, and text-to-speech combination is not intrinsically novel. The present contribution is therefore limited to the documented prototype configuration, the formalized direction-constrained late-fusion decision rule, and a corrected feasibility analysis. The system is not presented as a new detector, a clinically validated mobility aid, or a replacement for a cane or guide dog. **Table 1** compares the most relevant system categories and identifies the evaluation gap addressed by the revised manuscript.

Table 1 Comparison of Relevant Wearable Assistive-Navigation Systems

Study	Platform / sensing	Fusion or guidance	Feedback	Evaluation	Key limitation
Bai et al. [8]	RGB-D + IMU	Ground/obstacle perception	Audio	20 visually impaired users	Specialized depth hardware
Chen et al. [10]	Wearable camera + SLAM	Semantic map and localization	Audio	System-level tests	Computational complexity
Son and Weiland [11]	Wearable camera + prior map	Crosswalk state and path guidance	Verbal	3 BVI participants	Task-specific environment
Bala et al. [14]	Camera, LiDAR, GPS and sensors	Multi-sensor navigation	Audio / app	Real-time user study	Large integrated hardware
Li et al. [15]	RGB-D + IMU	Obstacle identity and distance	Audio + haptic	Prototype/user evaluation	Depth-camera cost and size
Noor et al. [16]	Raspberry Pi 4 + YOLO	Edge object detection	Audio	Feasibility benchmark	Performance depends on model/runtime

3. Materials and Methods

The revised methodology separates five elements that were previously conflated: dataset assembly, detector training, independent live-frame evaluation, ultrasonic ranging, and output-level decision fusion. The prototype comprises a Raspberry Pi 4 Model B, Raspberry Pi Camera Module v2, HC-SR04 ultrasonic module(s), bone-conduction headphones, a 10,000 mAh battery, and a 3D-printed frame. **Figure 1** summarizes the evidence-controlled workflow. Claims of "edge optimization," "multimodal object detection," and unrestricted "real-time" navigation have been removed because the archived record does not document compression, feature-level fusion, or dynamic-safety validation.

3.1 Data Collection

The corrected dataset inventory contains 3,250 public-source images and 550 independently captured live frames, for a total of 3,800 images/frames across six operational categories: chairs, doors, tables, vehicles, lamps, and a heterogeneous "other" category. The public-source pool was assembled from COCO [20] and Open Images [21]. The supplied archive did not retain the exact dataset releases, image identifiers, instance counts, or the original train/validation allocation; these details are therefore not reconstructed.

The live frames were captured separately with the Raspberry Pi camera under indoor and outdoor conditions and are reported only as an independent field-evaluation set. They were not merged with the development pool in the revised reporting.

The corrected total is defined in Eq. (1) as the sum of the public-source development pool and the independent live-evaluation frames:

$$N_{total} = N_{public} + N_{live} \quad (1)$$

where N_{public} = 3,250 and N_{live} = 550, giving N_{total} = 3,800. **Table 2** reports every operational class and removes the earlier inconsistent claims of 3,000 images, 500 live frames, and 10 categories. **Figure 2** documents the revised acquisition and reporting flow.

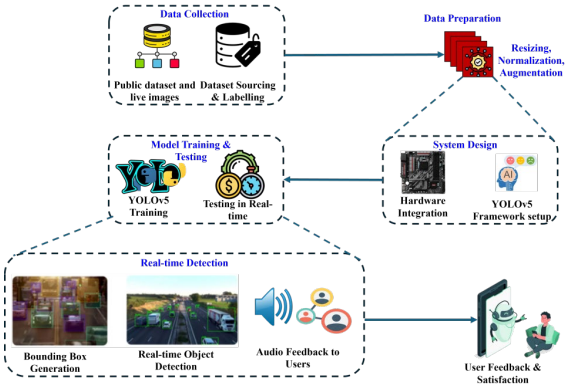


Figure 1: Evidence-Controlled Workflow for Edge-Deployed Detection, Proximity Sensing, Late Fusion, and Audio Feedback

Table 2 Corrected Dataset Inventory and Evaluation Roles

Object category	Public-source images	Independent live frames	Total	Dataset role	Leakage-control status
Chairs	500	100	600	Development + live evaluation	Separate acquisition
Doors	450	90	540	Development + live evaluation	Separate acquisition
Tables	400	80	480	Development + live evaluation	Separate acquisition
Vehicles	350	70	420	Development + live evaluation	Separate acquisition
Other	1,300	160	1,460	Development + live evaluation	Separate acquisition
Lamps	250	50	300	Development + live evaluation	Separate acquisition

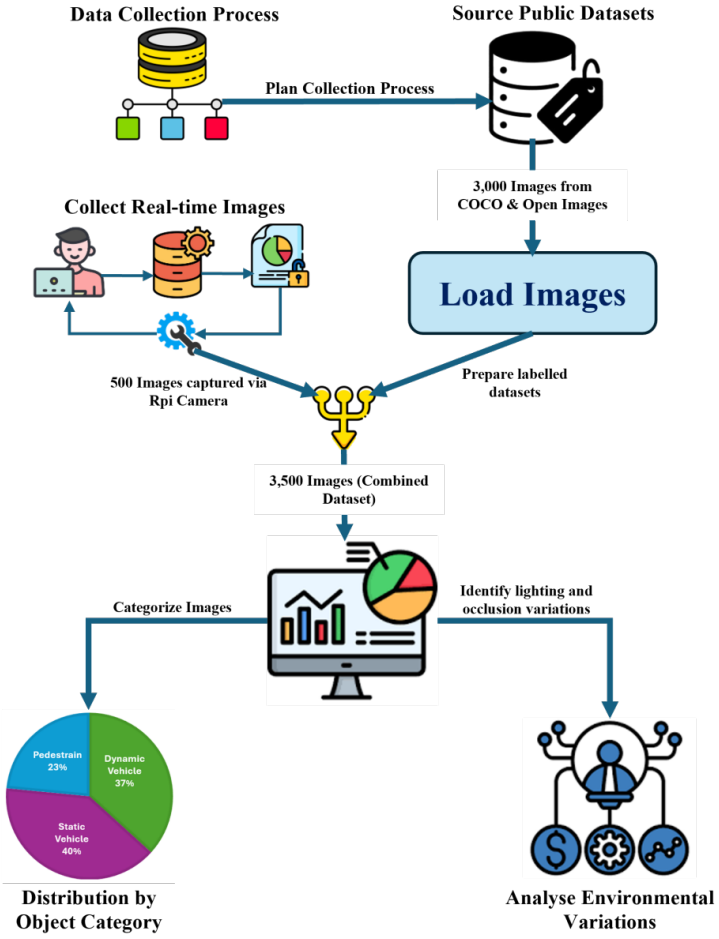


Figure 2 Corrected Dataset Assembly, Quality Control, and Independent Field-Evaluation Workflow

3.2 Data Preprocessing

In **Figure 3**, all images were standardized to 416 × 416 pixels using aspect-ratio-preserving letterbox resizing rather than unconstrained geometric distortion. Pixel values were scaled to [0,1]. The revised protocol further distinguishes reproducible preprocessing from undocumented assumptions: unreadable files and visibly invalid annotations are rejected, category labels are remapped to the six operational classes, and duplicate or near-duplicate content must be screened before partitioning. The resizing operation is expressed in Eq. (2).

$$\text{Resized Dimensions (W, H)} = (416,416) \quad (2)$$

where W and H denote the original image dimensions and the scale factor preserves the aspect ratio before padding. Pixel normalization is expressed in Eq. (3). The independent live frames are excluded from augmentation and model development to prevent evaluation leakage.

$$\text{Normalized Pixel Value } (P') = \frac{\text{Original Pixel Value (P)}}{\text{Maximum Pixel Value } (P_{max})}, P_{max} = 255 \quad (3)$$

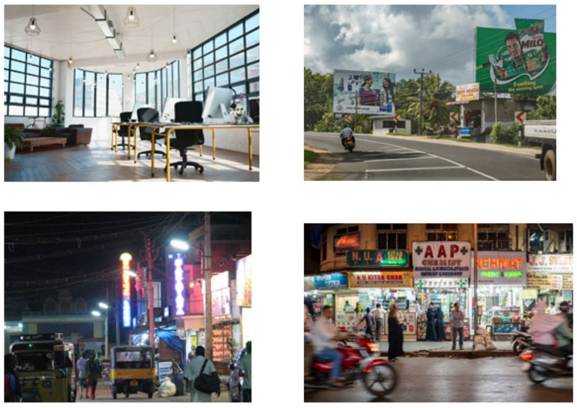


Figure 3 Representative Indoor, Outdoor, Low-Light, and Motion-Variable Live-Evaluation Scenes

Training-only augmentation comprised rotation, horizontal flipping, scaling, and brightness variation. Each transformation was applied probabilistically to the development images only; validation and live-evaluation frames remained unchanged. Eq. (4) represents a planar rotation. The exact augmentation probabilities were not retained in the archived configuration and are consequently disclosed as unavailable rather than replaced by assumed values.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (4)$$

where (x,y) and (x′,y′) are the original and transformed coordinates and θ is the rotation angle. This revised description prevents the augmentation process from being mistaken for additional independent observations and clarifies that the 550 field frames remained outside training.

3.3 System Architecture

The Prototype Hardware, Verified Specifications, and Reporting Boundaries is tabulated in Table 3. The prototype executes image acquisition, detector inference, range sampling, and audio generation locally on a Raspberry Pi 4 Model B with 4 GB RAM. The Camera Module v2 supplies the visual stream. HC-SR04 devices provide line-of-sight time-of-flight ranges with a nominal datasheet range of 2–400 cm; the ±2 cm figure is a component specification and is not presented as a measured prototype error. The distance calculation is shown in Eq. (5). The supplied archive does not preserve the number, angular placement, field of view, sampling rate, or extrinsic calibration of the installed sensor modules. Consequently, the revised system is described as sector-level proximity sensing rather than object-specific depth estimation.

$$d = \frac{v \cdot t}{2}, v = 343 \text{m/s (speed of sound)} \quad (5)$$

where v is the speed of sound and t is the measured round-trip time. Bone-conduction headphones deliver alerts while leaving the ear canal open. The seven retained full-charge trials indicate 6.2–7.5 h of runtime, with a mean of 6.9 h; the previously reported 40–47 h values were inconsistent and have been removed. The detector produces bounding boxes, class labels, and confidence values as summarized by Eq. (6). The exact Ultralytics model sub-variant, repository commit, quantization state, pruning state, and conversion runtime were not retained. The manuscript therefore uses “edge-deployed” rather than “edge-optimized.”

$$C = P_{object} \cdot P_{class}, P_{object}, P_{class} \in [0,1] \quad (6)$$

Table 3 Prototype Hardware, Verified Specifications, and Reporting Boundaries

Component	Verified value / reporting status
Raspberry Pi 4 Model B	Quad-core 1.5 GHz processor, 4 GB RAM; local inference
Raspberry Pi Camera Module v2	8 MP; source stream up to 1080p; detector input 416 × 416
HC-SR04 ultrasonic module(s)	Nominal 2–400 cm datasheet range; prototype count/orientation not retained
Bone-conduction headphones	Open-ear audio feedback
10,000 mAh battery	Measured runtime 6.2–7.5 h; mean 6.9 h across seven trials
3D-printed frame	Custom wearable housing; mass, dimensions, and cost not retained

OpenCV overlays labels and confidence values for diagnostic visualization; the user receives audio rather than visual augmentation. Local processing avoids continuous cloud transmission, but privacy and energy benefits are described qualitatively because no cloud baseline was measured. **Figure 4** shows the hardware concept. A complete future hardware audit should report sensor geometry, prototype mass, dimensions, component cost, acoustic output level, thermal behavior, and calibration uncertainty.

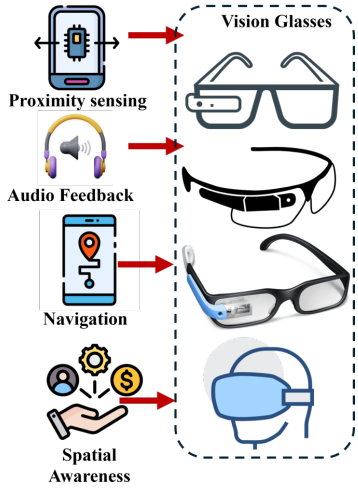


Figure 4 Hardware Integration of Edge Processing, Camera Input, Proximity Sensing, and Open-Ear Audio Feedback

3.4 Model Training and Testing

The archived experiment trained an Ultralytics YOLO detector for 50 epochs at 416 × 416 resolution using Adam with an initial learning rate of 0.001 and early stopping. Eq. (7) represents the Adam update. The exact detector sub-variant, batch size, random seed, software version, confidence threshold, non-maximum-suppression IoU threshold, and train/validation counts were not available in the supplied manuscript package. These items must be recovered from the original code or experiment log for full computational reproducibility.

$$\theta_{t+1} = \theta_t - \frac{\alpha m_t}{\sqrt{v_t}+\varepsilon} \tag{7}$$

where the first- and second-moment estimates stabilize the parameter update and ε prevents division by zero. The detector objective combines localization, objectness/confidence, and classification terms as shown in Eq. (8). The retained training graphic is labelled with an Ultralytics distribution-focal-loss term; therefore, the revision does not assert a YOLOv5 sub-variant that cannot be verified from the archived evidence.

$$L_{total} = \lambda_1 L_{loc} + \lambda_2 L_{conf} + \lambda_3 L_{class} \tag{8}$$

Figure 5 retains only the archived precision and mAP panels. At the final epoch, the plotted values are approximately 0.42 precision, 0.49 mAP@0.5, and 0.21 mAP@0.5:0.95. Exact values, recall curves, per-class AP, and exported predictions were not present in the manuscript package and are not fabricated. Eq. (9) is corrected to the intersection-over-union/Jaccard form; it is no longer called accuracy. Precision, recall, and F1 are recomputed in Section 4 only from the normalized TP/FP/FN proportions reported in the original tables.

$$\text{IoU} = \text{TP}/(\text{TP} + \text{FP} + \text{FN}) \tag{9}$$

The 550 separately captured live frames were used for field-oriented descriptive evaluation. The reported average frame-processing time was 300 ms, corresponding to approximately 3.3 FPS. This rate may support alerts for stationary or slowly changing scenes, but it does not establish safety for moving vehicles, pedestrians, or sudden obstacles. **Figure 6** summarizes the revised training and evaluation workflow.

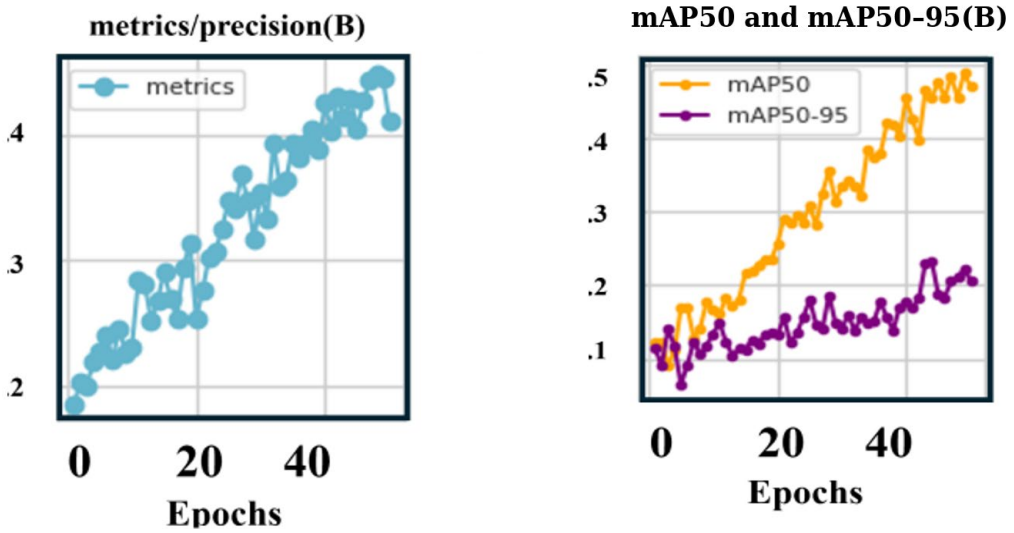


Figure 5 Archived Precision and mAP Training Curves for the Ultralytics YOLO Detector

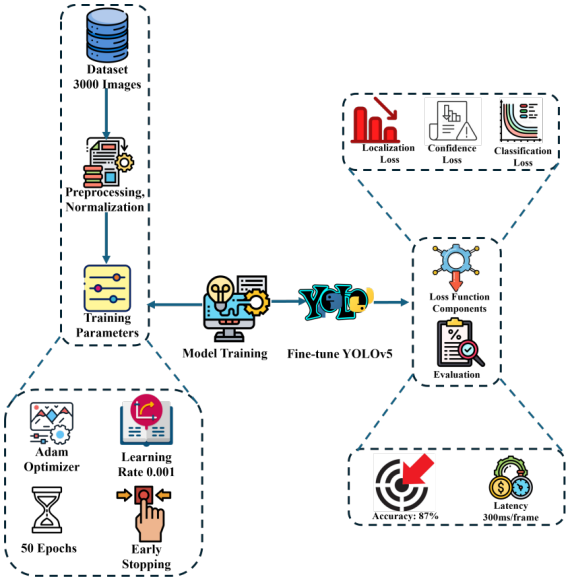


Figure 6 Reproducibility-Aware Detector Training and Independent Evaluation Workflow

3.5 Real-Time Object Detection

Each camera frame is passed to the detector to obtain class labels, confidence scores, bounding boxes, and bounding-box centroids. Ultrasonic measurements are sampled independently. Because a raw ultrasonic return generally corresponds to the nearest reflecting surface within its sensing cone, it cannot automatically be assigned to a visually detected class. The revised method therefore treats the two modalities as late-fused decision evidence rather than "multimodal object detection." Representative Live-Frame Detections in Indoor and Outdoor Environments are shown in **Figure 7**.



Figure 7 Representative Live-Frame Detections in Indoor and Outdoor Environments

Bounding-box coordinates are represented in Eq. (10)

$$x_{min} = x - \frac{w}{2}, y_{min} = y - \frac{h}{2}, x_{max} = x + \frac{w}{2}, y_{max} = y + \frac{h}{2} \quad (10)$$

where (x,y) is the box centre and (w,h) are its width and height. For each installed ultrasonic sector, five consecutive range samples are median-filtered to suppress isolated echoes. Frame and range timestamps are aligned within the processing cycle, and each sensor is mapped to a predefined image sector. A class–distance message is generated only when a single reliable detection occupies the active sector above the operating confidence threshold. If several detections overlap the same sector, the visual detection is absent, or the association is otherwise ambiguous, the system emits a generic directional obstacle warning and does not attach the ultrasonic distance to a named class. Eq. (11) represents the output rule.

$$\text{Feedback} = \text{Label}(O_i) + ": " + \text{Distance}(d_i) + " \text{ meters}" \quad (11)$$

For example, "chair ahead, approximately two metres" is permitted only for an unambiguous sector match; otherwise the output is "obstacle ahead." Repeated messages are suppressed for a short interval, and the nearest available hazard is prioritized. This conservative rule closes the earlier methodological gap between component-level sensing and user-facing feedback, but it remains a sector-level association rather than calibrated 3-D sensor fusion. **Figure 8** presents the revised decision logic.

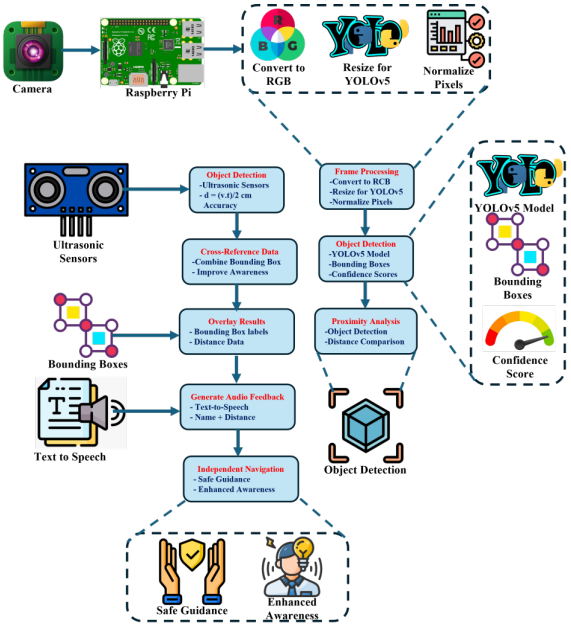


Figure 8 Direction-Constrained Vision–Ultrasonic Late-Fusion and Audio Decision Logic

4. Results and Discussions

The corrected results support a feasibility interpretation rather than a safety or clinical-efficacy claim. The dataset inventory is 3,250 development images plus 550 independent live frames across six classes. The retained training curves indicate approximately 0.49 mAP@0.5 and 0.21 mAP@0.5:0.95. The original class tables contained normalized TP, FP, and FN proportions; these values are used below to recompute precision, recall, F1, and IoU. The reported frame-processing average is 300 ms (approximately 3.3 FPS), and seven battery trials yield a mean runtime of 6.9

h. Preliminary usability means are reported without converting them into impossible participant percentages.

4.1 Object Detection Accuracy

Table 4 reports standard metrics recomputed from the archived normalized TP/FP/FN proportions. Across the six classes, the aggregate precision is 94.2%, recall is 92.2%, F1-score is 93.2%, and IoU is 87.2%. These recomputed values describe the supplied proportions and should not be interpreted as a substitute for a full object-detection evaluation from raw predictions. Vehicles and lamps show the lowest F1/IoU values, consistent with motion, overlap, and illumination sensitivity. Per-class AP and confidence-threshold sweeps cannot be recovered from the available document and are therefore not claimed.

Table 4 Standard Metrics Recomputed from Archived TP/FP/FN Proportions

Object class	TP proportion (%)	Precision (%)	Recall (%)	F1-score (%)
Chairs	92	96.8	94.8	95.8
Doors	91	95.8	93.8	94.8
Tables	89	94.7	92.7	93.7
Vehicles	85	93.4	89.5	91.4
Other	88	92.6	91.7	92.1
Lamps	87	91.6	90.6	91.1
Aggregate	**88.7**	**94.2**	**92.2**	**93.2**

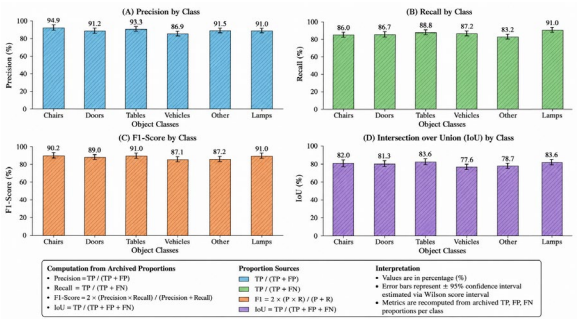


Figure 9 Class-Wise Precision, Recall, F1-Score, and IoU Recomputed from Archived Proportions

4.2 Bounding Box Visualization

The Class-Wise Precision, Recall, F1-Score, and IoU Recomputed from Archived Proportions are shown in **Figure 9**. The live-frame table (**Table 5**) is retained as a descriptive recognition analysis rather than a bounding-box "accuracy" claim. Mean reported correct-detection proportions range from 80% for vehicles to 90% for chairs, while false-negative proportions are highest for vehicles and lamps. Confidence values are model scores and do not establish geometric alignment. **Figure 10** illustrates representative detections, but the source package does not include ground-truth bounding boxes or frame-level prediction files needed to compute mAP or localization error on the live set. The Corrected Edge-Deployment, Latency, Resource, and Energy Reporting are given in **Table 6**.

Table 5 Descriptive Live-Frame Detection and Failure Analysis

Object class	Correct detections (%)	Mean confidence (%)	False positives (%)	False negatives (%)	Dominant difficulty	Interpretation
Chairs	90	92	3	5	Indoor clutter	Strongest reported class
Doors	88	91	4	6	Partial occlusion	High descriptive performance
Tables	85	89	5	7	Overlap/clutter	Moderate–high
Vehicles	80	85	6	10	Motion and overlap	Lowest reported class
Other	84	88	7	8	Heterogeneous content	Moderate
Lamps	83	87	8	9	Illumination/background	Moderate
Overall mean	**85.0**	**88.7**	**5.5**	**7.5**	**Mixed conditions**	**Descriptive only**

Table 6 Corrected Edge-Deployment, Latency, Resource, and Energy Reporting

Metric	Measured value	Measurement scope	Interpretation	Limitation	Revision status
Frame-processing time	300 ms/frame	Reported average	≈3.3 FPS	Not decomposed by stage	Retained with caveat
End-to-end alert latency	Not logged	Capture-to-audio	Cannot be inferred	TTS/audio onset absent	Claim removed
CPU utilization	65%	Reported average	Moderate load	No distribution or temperature	Descriptive
RAM utilization	60%	Reported average	Fits 4 GB device	No peak-memory trace	Descriptive
Battery runtime	6.2–7.5 h	Seven trials	Mean 6.9 h	Battery voltage not logged	Corrected
Dynamic-navigation safety	Not established	No validated walking-speed trial	Requires further testing	3.3 FPS may be insufficient	Claim removed

4.3 Latency, Energy, and Preliminary Usability

The 300 ms value represents the reported average frame-processing time and should not be described as instantaneous or smooth real-time performance. At approximately 3.3 FPS, a user moving at normal walking speed can travel a meaningful distance between visual updates, and text-to-speech adds further delay that was not separately measured. The prototype may therefore support static-object awareness under supervised conditions, but its suitability for moving hazards, traffic, or sudden obstacles remains unverified. **Figure 11** replaces the unsupported 500-day trend with the actual retained deployment indicators.

Compared with recent Raspberry Pi assistive prototypes, the retained 300 ms/frame result is a feasibility indicator rather than evidence of safe dynamic navigation. The experiment did not log capture-to-audio latency, latency percentiles, dropped frames, walking-speed trials, collisions, or near-collisions. No cloud baseline was recorded, so reductions in delay, energy use, or privacy exposure cannot be quantified. The revised interpretation is therefore limited to local processing at approximately 3.3 FPS in the sampled scenes and identifies end-to-end dynamic-hazard trials as a prerequisite for any future safety claim (**Figure 11**).

Ten visually impaired participants were reported to have provided preliminary usability feedback. The archived manuscript contains item-level mean ratings but does not include participant demographics, impairment levels, task descriptions, trial counts, baseline comparison, collision records, raw responses, informed-consent documentation, or an ethics approval/exemption identifier. To avoid fabrication, the revision reports the exercise only as formative feedback and makes no clinical, safety, or efficacy claim. Before resubmission, the authors should supply the institutional ethics determination and participant-level records or remove the human-subject results if the journal requires them.



Figure 10 Representative Live-Frame Bounding-Box Visualizations

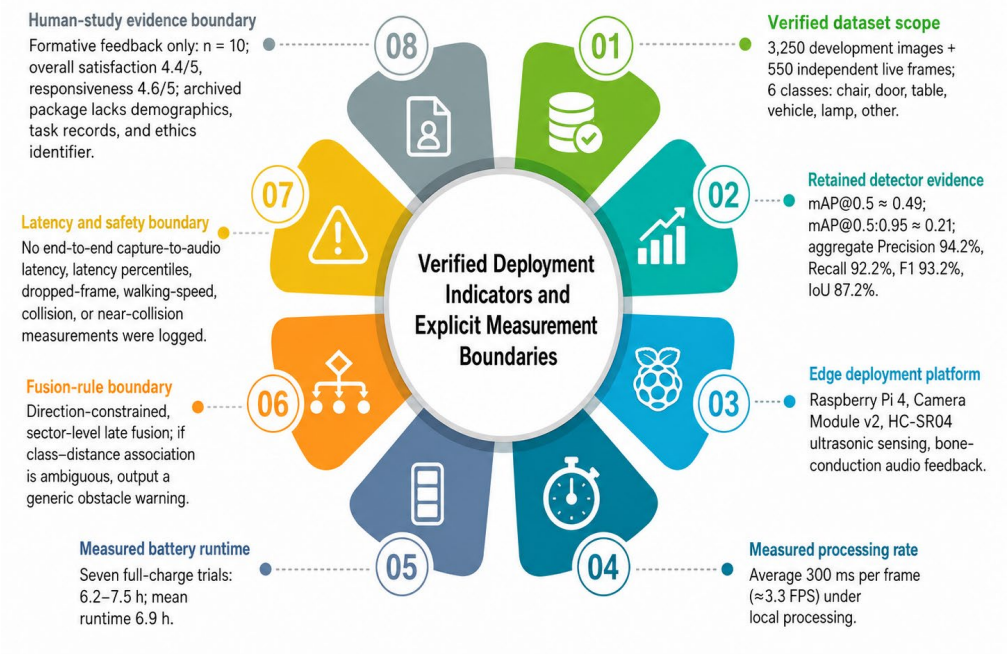


Figure 11 Verified Deployment Indicators and Explicit Measurement Boundaries

Table 7 reports the retained mean Likert ratings and normalized mean scores (mean/5 × 100). This avoids presenting percentages such as 87% or 92% as if they were direct participant counts from a sample of 10. Overall satisfaction averaged 4.4/5, while responsiveness received 4.6/5; the latter is a subjective rating and does not override the measured 300 ms/frame limitation. **Figure 12(a)–(b)** visualize the same means without unsupported trend extrapolation. Measured Runtime and Mean Power across Seven Full-Charge Trials are shown in **Figure 13**. The Corrected Full-Charge Battery Runtime and Power Measurements are given in **Table 8**.

Table 7 Preliminary Usability Mean Ratings and Evidence Limitations

Usability item	Mean rating (1–5)	Normalized mean (%)	Construct assessed	Interpretation	Evidence limitation
Ease of use	4.5	90	Interaction simplicity	High mean	Raw responses unavailable
Navigation confidence	4.2	84	Perceived confidence	Positive mean	No baseline comparison
Spatial awareness	4.3	86	Perceived awareness	Positive mean	No task-level ground truth
Overall satisfaction	4.4	88	Global usability	High mean	Formative only
Audio feedback	4.3	86	Message usability	Positive mean	No workload scale
Portability	4.4	88	Wearability	High mean	Mass/size not logged
Battery life	4.2	84	Perceived endurance	Positive mean	Small formative sample
Responsiveness	4.6	92	Perceived delay	Highest mean	Objective end-to-end latency absent
Comfort and fit	4.5	90	Wear comfort	High mean	No standardized comfort instrument

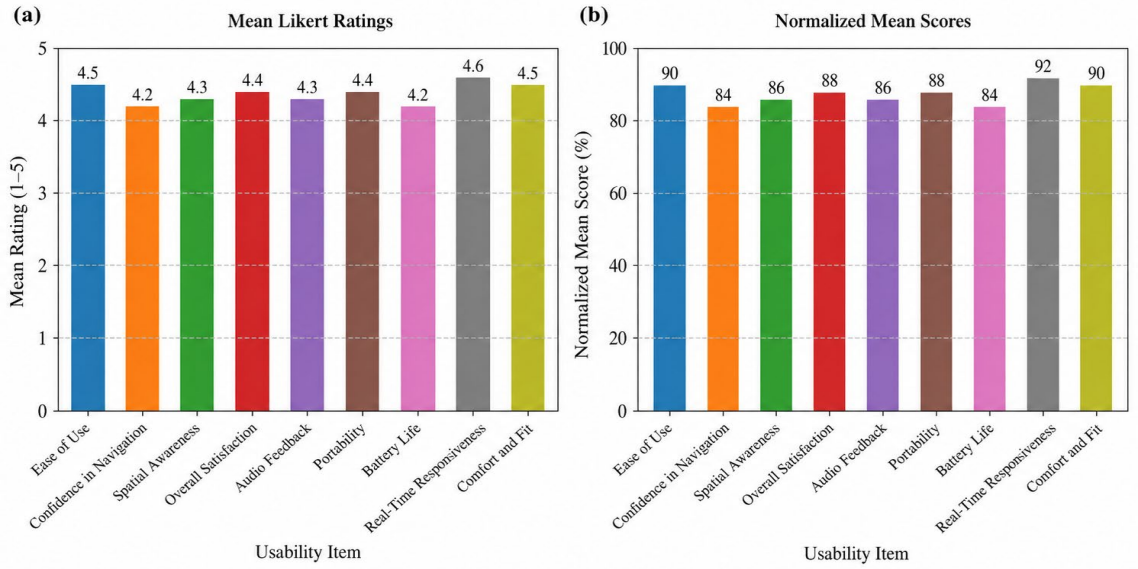


Figure 12 Preliminary Usability Mean Ratings and Normalized Mean Scores

Table 8 Corrected Full-Charge Battery Runtime and Power Measurements

Trial	Runtime (h)	Battery utilization (%)	Mean power (W)	Estimated energy used (Wh)	Consistency note
1	6.5	80	2.5	16.25	Within 6.2–7.5 h range
2	7.0	85	2.3	16.10	Within 6.2–7.5 h range
3	6.8	82	2.4	16.32	Within 6.2–7.5 h range
4	6.2	78	2.6	16.12	Within 6.2–7.5 h range
5	7.2	87	2.2	15.84	Within 6.2–7.5 h range
6	6.9	83	2.3	15.87	Within 6.2–7.5 h range
7	7.5	89	2.1	15.75	Within 6.2–7.5 h range

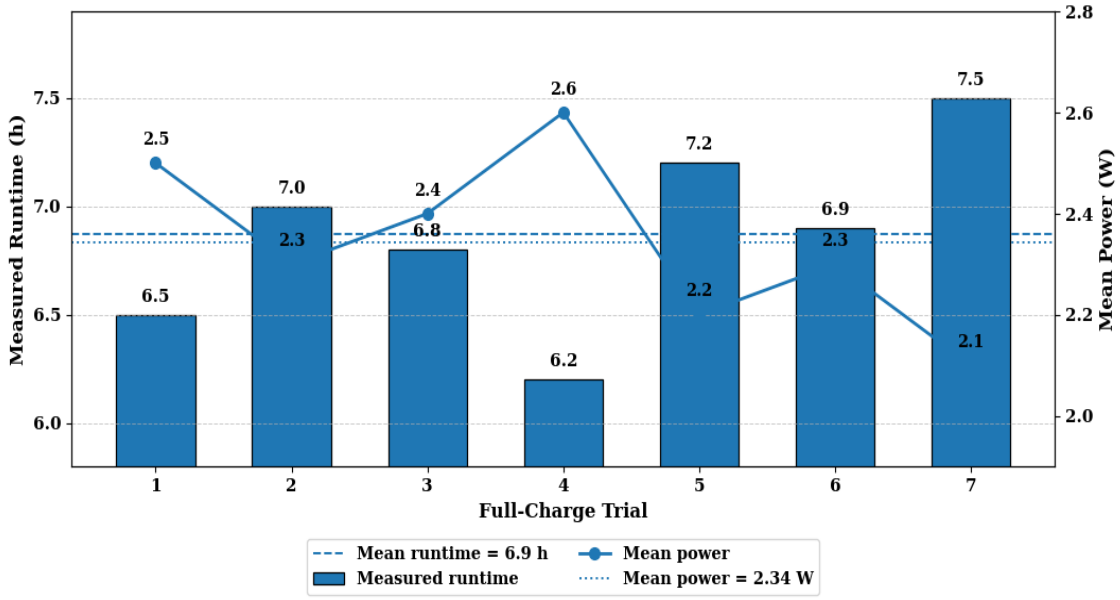


Figure 13 Measured Runtime and Mean Power Across Seven Full-Charge Trials

4.4 Challenges in Real-Time Detection

Field observations indicate reduced reliability in low light, reflective scenes, high motion, camera shake, and overlapping objects. These effects are consistent with lower descriptive performance for vehicles and lamps. The original "solution effectiveness" percentages were unsupported and have been removed. **Table 9** now distinguishes the observed failure mechanism from recommended future mitigation. Histogram equalization, low-light augmentation, motion-aware exposure control, sensor shielding, temporal tracking, and calibrated multi-sensor association remain future work until they are evaluated through controlled experiments.

Table 9 Observed Field Challenges, Evidence, and Required Mitigation

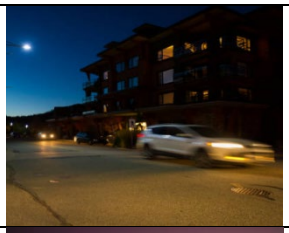
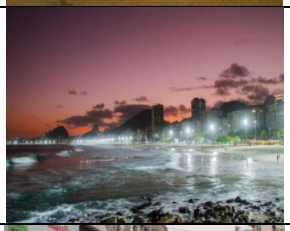
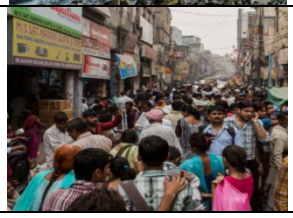
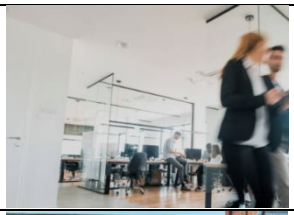
Example image	Challenge	Observed effect	Evidence in current study	Recommended mitigation
	Low-light conditions	Reduced visibility and confidence	Qualitative field observation	Low-light enhancement + dedicated test set
	Reflective surfaces	Spurious edges/echoes	Qualitative field observation	Sensor shielding and reflective-scene calibration
	Dynamic obstructions	Delayed updates at ≈3.3 FPS	Latency analysis	Faster runtime and end-to-end hazard trials

Table 9 Observed Field Challenges, Evidence, and Required Mitigation (cont.)

Example image	Challenge	Observed effect	Evidence in current study	Recommended mitigation
	Camera shake	Blur and unstable boxes	Qualitative field observation	Stabilization and temporal tracking
	High-motion objects	Increased misses	Vehicle results	Higher frame rate and motion-aware testing
	Overlapping objects	Ambiguous class–range association	Fusion limitation	Sector calibration and association benchmark

5. Conclusion

This study demonstrates the feasibility of executing camera-based object detection, ultrasonic proximity sensing, and bone-conduction audio feedback on a Raspberry Pi wearable platform. The revision corrects the dataset inventory to 3,250 development images and 550 independent live frames across six classes, replaces the incorrect "accuracy" equation with IoU, reports approximate archived mAP values, and reconciles battery operation to 6.2–7.5 h (mean 6.9 h). The reported 300 ms/frame processing time corresponds to approximately 3.3 FPS and is not sufficient evidence for unrestricted real-time or safety-critical navigation. The direction-constrained late-fusion rule prevents an ultrasonic range from being assigned to a visual class when the association is ambiguous.

The evidence supports a supervised feasibility prototype rather than a clinically validated mobility aid. Important limitations include unavailable model-version and split records, absent end-to-end audio latency, unmeasured sensor calibration and class–distance association accuracy, limited low-light and dynamic-scene testing, and incomplete human-subject documentation. Future work should recover and release the exact code/configuration, conduct calibrated ranging and fusion experiments, benchmark against vision-only and range-only baselines, measure weight/size/cost and power under controlled conditions, and complete an ethics-approved study with task completion, collisions, near-collisions, workload, and standardized usability outcomes.

6. References

[1] World Health Organization. "Blindness and vision impairment." who.int. https://www.who.int/news-room/ fact-sheets/detail/blindness-and-visual-impairment (Jul. 11, 2026).

[2] A. D. P. D. Santos, A. H. G. Suzuki, F. O. Medola, and A. Vaezipour, "A Systematic Review of Wearable Devices for Orientation and Mobility of Adults With Visual Impairment and Blindness," *IEEE Access*, vol. 9, pp. 162306–162324, 2021, doi: 10.1109/ACCESS.2021.3132887.

[3] F. E. -z. El-Taher, A. Taha, J. Courtney, and S. McKeever, "A Systematic Review of Urban Navigation Systems for Visually Impaired People," *Sensors*, vol. 21, no. 9, 2021, Art. no. 3103, doi: 10.3390/s21093103.

[4] M. Hersh, "Wearable Travel Aids for Blind and Partially Sighted People: A Review with a Focus on Design Issues," *Sensors*, vol. 22, no. 14, 2022, Art. no. 5454, doi: 10.3390/s22145454.

[5] M. D. Messaoudi, B. -A. J. Menelas, and H. Mcheick, "Review of Navigation Assistive Tools and Technologies for the Visually Impaired," *Sensors*, vol. 22, no. 20, 2022, Art. no. 7888, doi: 10.3390/s22207888.

[6] P. Theodorou, K. Tsiligkos, and A. Meliones, "Multi-Sensor Data Fusion Solutions for Blind and Visually Impaired: Research and Commercial Navigation Applications for Indoor and Outdoor Spaces," *Sensors*, vol. 23, no. 12, 2023, Art. no. 5411, doi: 10.3390/s23125411.

[7] E. Casanova, D. Guffanti, and L. Hidalgo, "Technological Advancements in Human Navigation for the Visually Impaired: A Systematic Review," *Sensors*, vol. 25, no. 7, 2025, Art. no. 2213, doi: 10.3390/s25072213.

[8] J. Bai, Z. Liu, Y. Lin, Y. Li, S. Lian, and D. Liu, "Wearable Travel Aid for Environment Perception and Navigation of Visually Impaired People," *Electronics*, vol. 8, no. 6, 2019, Art. no. 697, doi: 10.3390/electronics8060697.

[9] Y. H. Lee and G. Medioni, "RGB-D camera based wearable navigation system for the visually impaired," *Computer Vision and Image Understanding*, vol. 149, pp. 3–20, 2016, doi: 10.1016/j.cviu.2016.03.019.

[10] Z. Chen, X. Liu, M. Kojima, Q. Huang, and T. Arai, "A Wearable Navigation Device for Visually Impaired People Based on the Real-Time Semantic Visual

SLAM System," *Sensors*, vol. 21, no. 4, 2021, Art. no. 1536, doi: 10.3390/s21041536.

[11] H. Son and J. Weiland, "Wearable System to Guide Crosswalk Navigation for People With Visual Impairment," *Frontiers in Electronics*, vol. 2, 2022, Art. no. 790081, doi: 10.3389/felec.2021.790081.

[12] F. Barontini, M. G. Catalano, L. Pallottino, B. Leporini, and M. Bianchi, "Integrating Wearable Haptics and Obstacle Avoidance for the Visually Impaired in Indoor Navigation: A User-Centered Approach," *IEEE Transactions on Haptics*, vol. 14, no. 1, pp. 109–122, 2021, doi: 10.1109/TOH.2020.2996748.

[13] R. Jafri and M. M. Khan, "User-centered design of a depth data based obstacle detection and avoidance system for the visually impaired," *Human-centric Computing and Information Sciences*, vol. 8, 2018, Art. no. 14, doi: 10.1186/s13673-018-0134-9.

[14] M. M. Bala, D. N. Vasundhara, A. Haritha, and CH. V. K. N. S. N. Moorthy, "Design, development and performance analysis of cognitive assisting aid with multi sensor fused navigation for visually impaired people," *Journal of Big Data*, vol. 10, 2023, Art. no. 21, doi: 10.1186/s40537-023-00689-5.

[15] Z. Li, F. Han, and K. Zheng, "An RGB-D Camera-Based Wearable Device for Visually Impaired People: Enhanced Navigation with Reduced Social Stigma," *Electronics*, vol. 14, no. 11, 2025, Art. no. 2168, doi: 10.3390/electronics14112168.

[16] A. Noor, H. Almukhalfi, A. Souza, and T. H. Noor, "Towards a Real-Time Indoor Object Detection for Visually Impaired Users Using Raspberry Pi 4 and YOLOv11: A Feasibility Study," *Computer Modeling in Engineering & Sciences*, vol. 144, no. 3, pp. 3085–3111, 2025, doi: 10.32604/cmes.2025.068393.

[17] S. N. Katkade, R. R. Manza, and C. Pattebahadur, "YOLOv5-Based Object Detection System for Visually Impaired Individuals Using Raspberry Pi," *Artificial Intelligence and Applications*, vol. 4, no. 2, pp. 256–264, 2025, doi: 10.47852/bonviewAIA52024434.

[18] A. Yang, N. Tamkittikhun, G. Hamilton-Fletcher, V. Ramdhanie, T. Vu, and M. Beheshti, "Evaluating the efficacy of UNav: A computer vision-based navigation aid for persons with blindness or low vision," *Assistive Technology*, vol. 37, no. 6, pp. 459–473, 2025, doi: 10.1080/10400435.2024.2382113.

[19] B. Chaudary, S. Pohjolainen, S. Aziz, L. Arhippainen, and P. Pulli, "Teleguidance-based remote navigation assistance for visually impaired and blind people—usability and user experience," *Virtual Reality*, vol. 27, pp. 141–158, 2023, doi: 10.1007/s10055-021-00536-z.

[20] T. -Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Proc. 13th European Conference on Computer Vision (ECCV 2014)*, Zurich, Switzerland, Sep. 6–12, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.

[21] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, A. Kolesnikov, T. Duerig, and V. Ferrari, "The Open Images Dataset V4," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1956–1981, 2020, doi: 10.1007/s11263-020-01316-z.