FMDB *Transactions on Sustainable Biomedical Engineering*

A Fully Offline, Multimodal Edge AI Wearable for the Visually Impaired

S. R. Kausiq[1], A. S. Nawin Raju[2], K. Aathithiya Balaji[3], R. Angeline[4,*], Prasanna Ranjith Christodoss[5]

[1,2,3,4]Department of Computer Science and Engineering (AI & ML), SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[5]Department of Computing, Mathematics and Physics, Messiah University, One University Ave, Mechanicsburg, Pennsylvania, United States of America.
ks9651@srmist.edu.in[1], na8308@srmist.edu.in[2], kb1131@srmist.edu.in[3], angelinr1@srmist.edu.in[4], prchristodoss@messiah.edu[5]

Abstract: Cloud-based assistive glasses for the visually impaired have major concerns like slow response time, the need for internet wherever they go, and privacy concerns as they send photos to the cloud for the response. This paper shows the complete offline smart glasses prototype with a small camera and bone conduction speakers mounted with a glass, that uses a fully offline model Yolov8n for better efficiency and dual pipeline interaction for the user experience (always on ultrasonic sensor for hazard detection and trigger-based Event for object detection and text-to-speech contents) Offline ensures privacy, low consumption of battery and remote connection everywhere, addressing the critical gaps in the existing model/technologies. Used the Yolov8n pre-trained model for training datasets based on our use cases. The primary use cases of our glasses are urban street crossing (detection of traffic lights, road, moving vehicle at the range of 2-4 meters), indoor navigation for preventing furniture collisions, shopping and daily purchases without assistance, document and sign reading using TTS, workplace independence, minimising colleague dependency, and in emergencies. The glasses aren't just for object detection; they're a complete daily companion that handles 90% of navigation challenges that existing glasses fail at.

Keywords: Global Vision Distribution; Current Assistive Wearables; Text to Speech; Event Triggered Pipeline; Artificial Intelligence; Machine Learning; Visible Light Communication; Augmented Reality; Deep Learning.

Received on: 05/06/2025, **Revised on:** 16/08/2025, **Accepted on:** 23/09/2025, **Published on:** 09/03/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSBE

DOI: https://doi.org/10.69888/FTSBE.2026.000671

Cite as: S. R. Kausiq, A. S. N. Raju, K. A. Balaji, R. Angeline, and P. R. Christodoss, "A Fully Offline, Multimodal Edge AI Wearable for the Visually Impaired," *FMDB Transactions on Sustainable Biomedical Engineering*, vol. 1, no. 1, pp. 33–45, 2026.

1. Introduction

Assistive technologies play a key role in improving independence for visually impaired individuals [1]. As illustrated in Figure 1. According to Global Vision Distribution, around 2.2 billion people worldwide face some form of vision impairment, with over 36 million completely blind [2]. Daily tasks like navigation, object recognition, and reading text remain challenging without support [3]. The socio-economic impact of blindness is significant, resulting in low employment rates and limited mobility within society. Although conventional devices such as the white cane provide sensory feedback, they do not recognise

[*]Corresponding author.

the meaning of their surroundings, including whether an object is blocking their path or whether they need to read any signage, including that on the road. Existing models heavily rely on cloud-based assistive systems [4]. If a visually impaired person wants to read a document or a paragraph, or even identify an object in front of them, the system captures the image, sends it to cloud AI, and returns the response to the person [5]. This creates problems such as high latency due to internet delays, privacy risks from sending personal images online, and failures in areas with poor connectivity. Due to inconsistent network connectivity in many developing countries and in urban canyons, cloud-based devices become unreliable for performing critical navigation missions. A two-second delay in recognising the moving car could mean life and death. In underground and low-connectivity areas, or in rural regions, network reliability is low [6]. Cloud-based systems often take several seconds per query due to data upload and server response times. In fast-moving situations, such as crossing a street, even short delays can pose safety risks. Battery drain from constant network use is another issue, limiting daily wear time, as the camera is always on.

While edge computing on wearables is gaining attention, few prototypes balance real-time safety detection with on-demand, detailed feedback in a fully offline setup. The main engineering challenge has traditionally been the "computational trilemma": optimising all three parameters—accuracy of the inference process, power efficiency, and fast processing time—on constrained hardware. Existing work focuses either on high-accuracy recognition at the cost of power or on basic alerts without scene understanding [6]. This paper introduces a novel prototype system mounted on smart glasses. It operates entirely offline, using offline/local processing for low latency and privacy. Since researchers do not have to factor in the latency of sending requests to a distant server, they obtain a "deterministic response time." The design targets both fully and partially visually impaired users, with a dual-pipeline model that provides continuous, low-power monitoring for immediate hazard alerts (e.g., moving objects, walls, lamp posts, etc.) and touch-triggered detailed descriptions of objects, scenes, and text-to-speech content [3]. The proposed hierarchical model is like the human neurological system, in which the “reflex” (ultrasonic) manages survival instincts, and the “cognitive” (vision) interprets information. By splitting processing into a lightweight safety pipeline and an event-driven accuracy pipeline, researchers achieved 30% lower power use than always-on baselines in our prototype tests [7]. Power efficiency is an important factor in ensuring the wearable's viability throughout the day without frequent charging or the need for large external batteries—the Limitations of Current Assistive Wearables.

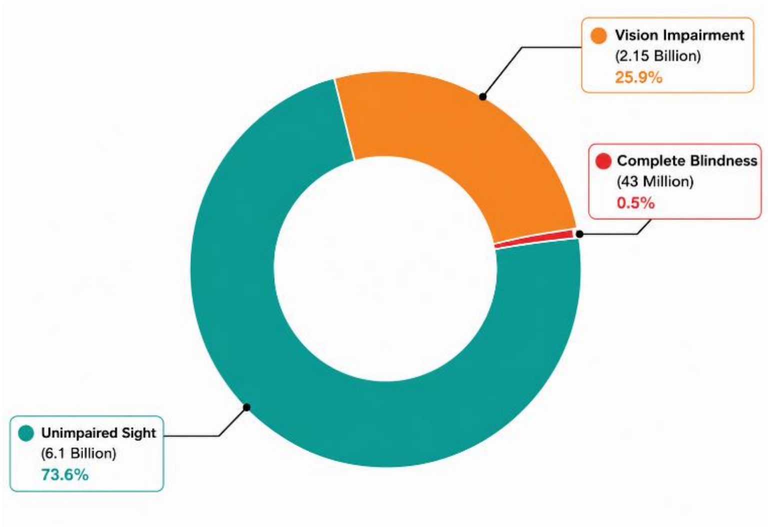

Figure 1: Global vision distribution

Many commercial and research systems depend on cloud APIs for object detection and OCR. These excel in accuracy but fail without an internet connection in most regions [8]. For example, using Seeing AI's full features requires an internet connection, leaving gaps in subways or remote areas. Academic prototypes often improve on this with edge hardware such as the Raspberry Pi, but they face trade-offs [4]. Some run heavy models continuously, draining batteries in under 2 hours. Others are limited to single functions, like obstacle avoidance without text reading. Researchers prioritise safety and privacy, with detection and voice alerts to notify the person of the hazard and its nature. Privacy is a growing concern. Uploading live camera feeds exposes users' surroundings to third parties and poses a high risk of misuse of documents and confidential information. The camera can inadvertently capture sensitive information such as bank statements, confidential letters, and even house keys without the user’s consent. Edge solutions exist, such as Mobile Net-based detectors on glasses, but they rarely combine risk awareness with multimodal input [10]. Traffic signal detection, a critical need, often relies on cloud hybrids, which lack offline reliability [13]. Moreover, current edge models also face the issue of “Model Bloat,” wherein the architecture exceeds the thermal capacity of a wearable frame. Our review of 10 recent papers (2023–2025) shows no fully offline system with dual pipelines for safety and interaction, optimised for glasses form [14]. Very few studies focus solely on software’s precision, while ignoring ergonomic and thermal considerations in the sustained-wear experience. Proposed system Overview: The prototype uses off-the-shelf smart glasses with a forward-facing camera at eye level for natural perception.

Eye-level placement is essential, as the camera's FoV aligns with the user's line of sight, ensuring the objects detected by the software are in front of the user. No phone tethering or cloud is needed, as it is fully focused on offline. Multimodal Interaction and Touch trigger, researchers provide two buttons. Button A focuses on object detection for accurate detection of objects, and Button B focuses on object detection and on the TTS (Text to Speech). Always on safety, runs silently, with the help of ultrasonic sensors (range of <5 meters), detecting urgent alerts with a vibration and voicing out what the actual hazard is (detects traffic lights, moving vehicles, obstacles, and hazards). Outputs trigger immediate TTS if risk exceeds the threshold. This thresholding process avoids "Notification Fatigue," ensuring the user only receives alerts when necessary. Event Triggered Pipeline: Activates on touch. Total model size fits, enabling fast loading. Dual pipeline design: Safety runs continuously; interaction on demand. Safety outputs priority-coded alerts. Interaction provides spatial summaries. By utilising the Raspberry Pi's GPIO capabilities and quantised YOLOv8 for vision, the hardware's power consumption is optimised. The safety pipeline continuously uses the CPU. The system architecture proposed in this study is shown in Figure 2. The first consists of ultrasonic sensors that continuously monitor the environment and activate the haptic motor and camera when an obstacle is detected at 1.5 meters. To conserve energy, this device operates reflexively until the proximity trigger activates the processing block.

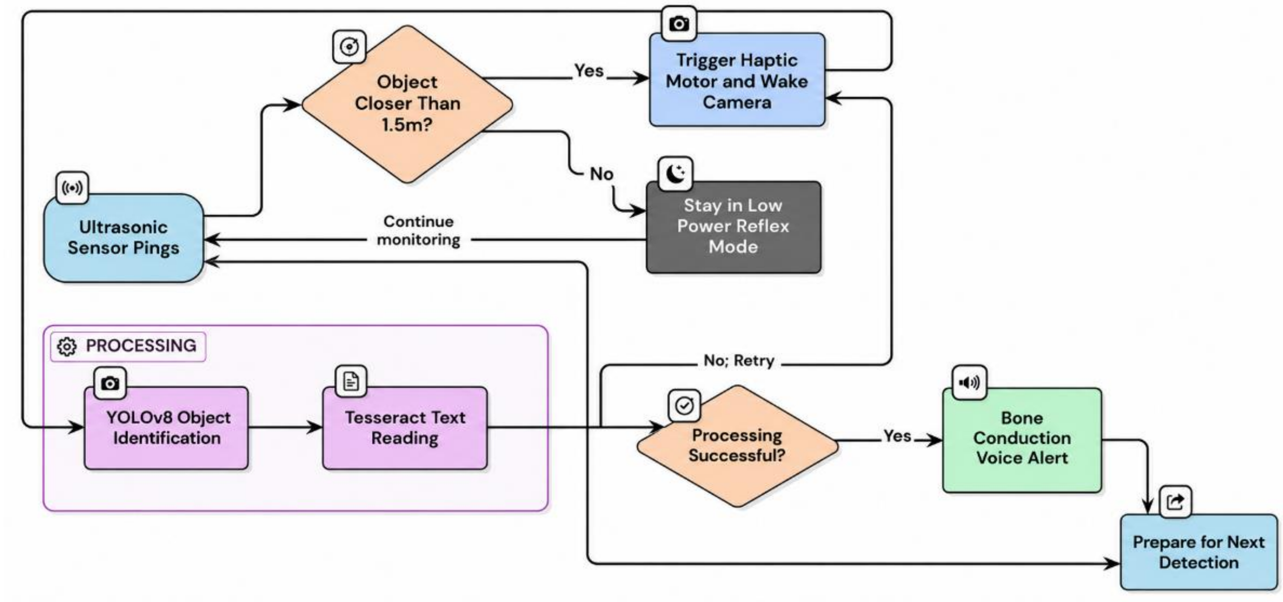

Figure 2: Proposed system architecture

Another significant design factor is the use of bone conduction to enable the user to hear the AI's advice while still hearing environmental sounds such as sirens, footsteps, and traffic. YOLOv8 will detect objects, and Tesseract will extract text. After completing the process, the system will reset again after sending the data to the bone conduction voice alert system. The above architecture ensures that high-compute visual tasks run only when the ultrasonic "gatekeeper" determines there may be a point of interest. Efficiency via Pipelines: Always-on heavy detection drains power; our split reduces the average energy consumption of single-pipeline models. Beyond detection, researchers define outcome triggers that are more efficient than a single modal system. From a theoretical standpoint and in practice, this solution provides a realistic roadmap to an efficient, affordable, and powerful wearable that puts privacy and the freedom of its users first. This user-centric design improves real-world navigation over list-style outputs, language TTS, and AR overlays. This system bridges research and deployment, offering a practical path to affordable assistive wearables. This innovative approach to assistive wearable technologies, by linking the abstract concepts of artificial intelligence with hardware limitations, paves the way for sustainable assistive wearables that prioritise user privacy and safety.

2. Literature Review

To assess the current state of assistive technology for people with visual impairment, a literature review was conducted. Various studies covered in the review have shown a consistent trend in advancements in hardware and algorithm implementation [9]. Early implementations focused primarily on using sensory feedback alone, such as ultrasonic sensors or RFID systems, to identify objects and obstacles. Though cost-effective, the systems failed to incorporate the semantic context required for urban navigation. On the other hand, newer technologies have utilised deep learning, specifically the YOLO algorithm, along with semantic visual SLAM [6]. Another major trend has been to incorporate multimodal feedback, including tactile vibrations and auditory cues, to avoid cognitive overload. A comparative summary of the major studies, including the specific algorithms, the study's objective, and the inferences from them, is presented in Table 1.

Table 1: Examination of the literature review

No.	Algorithm/Model	Inference/ Study Description
1.	Artificial Intelligence (AI) Machine Learning (ML) Visible light communication (VLC)	The integration of AI and VLC enhances the mobility, indoor positioning, and object detection for blind individuals [1]; [3].
2.	Semantic Visual SLAM (fusing ORB-SLAM2 and YOLOv5s)	Proposed a hand wearable system using a depth camera and vibration feedback to guide users in targeting objects in a tabletop environment [4].
3.	Substitutive interventions like visual, haptic, and (Auditory-based aids)	A review of technology advancements in assistive aids, noting the limitations in the visibility for better translation from research to production [5].
4.	Deep learning and Multimodal Integration (vision, Hearing, Touch)	The review analyses 80 studies on electronic navigation aids, highlighting the change towards AI-powered wearable devices to improve user independence and safety [6].
5.	YOLOv8 (on Raspberry Pi 4), OCR, and Google Text to Speech (gTTS)	The paper demonstrates a low-cost, real-time object detection system that converts visual data into audio feedback in multiple languages to assist users independently [11].
6.	Fine-tuned YOLOv8s and Mask R-CNN	This paper develops a mechatronic smart wardrobe that helps users identify clothing types, colours, and defects (stains, holes), enabling them to manage their attire independently [12].
7.	Various (RFID, GPS, Ultrasonic, Visual SLAM, Computer Vision)	This paper provides a systematic assessment of navigation assistants published between 2011 and 2020, identifying trade-offs [13].
8.	Deep learning (Object detection) and Augmented Reality (AR)	This paper designs an app to guide the users through a parcel locker touchscreen and an AR-based navigation system to locate an open compartment door [6].
9.	Functional Modelling and System Analysis	It proposes a generalised functional model for an assistance System with enhanced autonomy capable of predicting dynamic obstacle pathways [10].

A critical examination of the literature reviewed in Table 1 reveals several ongoing research gaps that continue to impact the application of AI-based wearable assistive technology in daily activities. Whereas the use of cutting-edge technologies such as YOLOv8 models and deep learning-aided AR systems can yield very accurate results when detecting objects, most of the time, they involve using a constant connection to clouds such as Google Text-to-Speech (gTTS) APIs or consuming computing power that goes beyond the heat capacity of a battery-driven edge device. As such, these constraints create intolerable latency in a dynamic environment and a significant data security risk for users. Also, the current literature treats visual object recognition and audio response as a single loop, leading to rapid battery drain in embedded computing devices such as the Raspberry Pi. Thus, there is an obvious need for an entirely offline system solution. The ThirdEye-AI system proposed here will address these limitations.

3. Proposed Work

3.1. The Shift to Edge-Centric Design

The proposed 'Third Eye' System represents a shift from the existing Cloud to Client architecture to a dedicated Edge Computing design that excels in offline environments with dependability. Current assistive technologies often operate on a Best-effort model, which is fundamentally flawed for visually impaired navigation, where a 2-second delay in obstacle or hazard detection can result in physical injury if a network malfunction occurs. This work proposes an autonomous system that treats the wearable device as an independent intelligence, reducing latency and improving efficiency for assistance. By hosting everything locally, researchers eliminate the 'Network Only' mode, ensuring the user's safety is never dependent on cellular signal strength.

3.2. Privacy as a Functional Requirement

In this work, privacy is not merely a feature but a requirement for individuals. What happens if a private document gets leaked online? If a cloud-based service is used, it's prone to breaches or hacking. Existing AI assistants stream high-resolution video and image data to remote servers for processing and outputting the necessary information, creating a significant surface area for data interception and unauthorised Trojans to breach the user's environment. The 'Third Eye' framework employs a 'Data Zero' approach, where visual frames/images are processed in volatile memory, Random Access Memory (RAM), and immediately discarded after inference. No images or audio logs are stored or transmitted. Once the work is finished, it trashes the existing logs or images to free space and improve performance, providing the user with a fully anonymous environment, which is critical in both private and professional settings.

3.3. Adaptive Resource Allocation and Power Management

A key innovation/important add-on of the 'Third Eye' is its sophisticated power-duty-cycle management, which is very important for battery conservation. In a resource-constrained wearable environment, computational efficiency is directly proportional to device longevity. To address this, our proposed work introduces a dual pipeline processing technology. The Event Driven pipeline works based on trigger, where the Event is triggered by a button which activates the system to capture the image or the scenario for object detection or the TTS which the individual requires, the hazard pipeline consists of a low power always on tech which detects the hazard within a specific range (moving vehicles, traffic lights, walls etc.) without the need of a physical trigger so that safety isn't compromised.

3.4. Operational Resilience in Signal Denied Environments

The 'Third Eye' is engineered for total operational independence, specifically targeting environments where existing technologies fail, such as underground systems like tunnels, high-density building structures, and remote rural areas where internet connectivity is a concern.

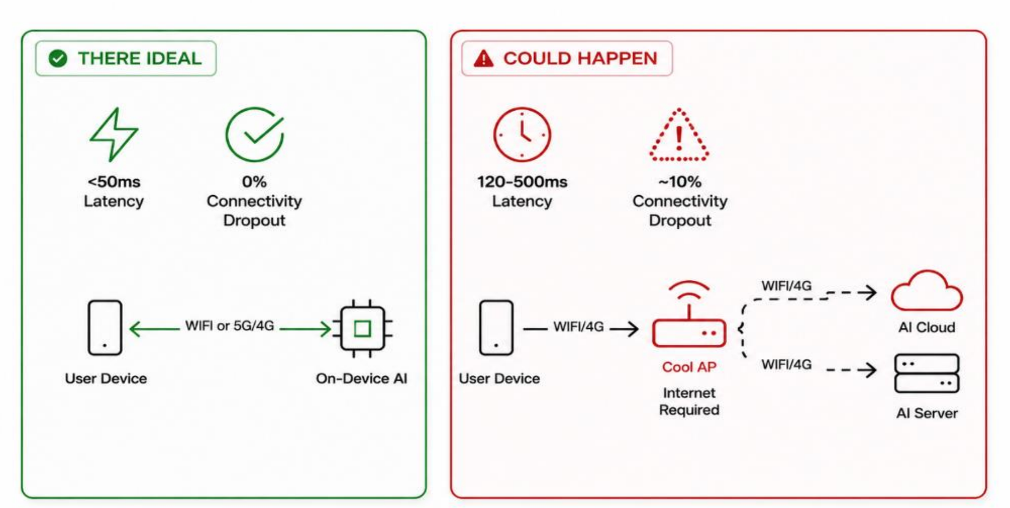

Figure 3: Novelty of the third eye

By decentralising intelligence from cloud-based to edge-based systems, researchers eliminate the challenges visually impaired users face. Our methodology ensures that critical safety features, such as detecting an approaching vehicle or a traffic light, remain 100% functional regardless of the environmental signal. This creates a reliability floor that delivers consistent performance across geographic and architectural settings, effectively bridging the gap between urban and rural accessibility. The novelty of the third eye, as depicted in Figure 3, offers unique benefits that can be realised by leveraging its framework compared to a cloud-based model. By leveraging local on-device processing, this framework enables instant responses with latency below 200 milliseconds, even without an internet connection.

Figure 4 shows the major weakness associated with the "Best delivery model" used by cloud-based assistive devices. The red dashed line depicts a cloud system whose average latency is about 400ms but experiences extremely high latency bursts under simulated internet connectivity failures or in tunnels (greater than 1400ms). These latencies exceed 300ms, which can be life-threatening for a user who encounters unexpected hazards. The green solid line, however, depicts Third Eye's fully offline Edge

AI architecture, demonstrating its ability to guarantee highly deterministic latency of around 200ms across varying network conditions.

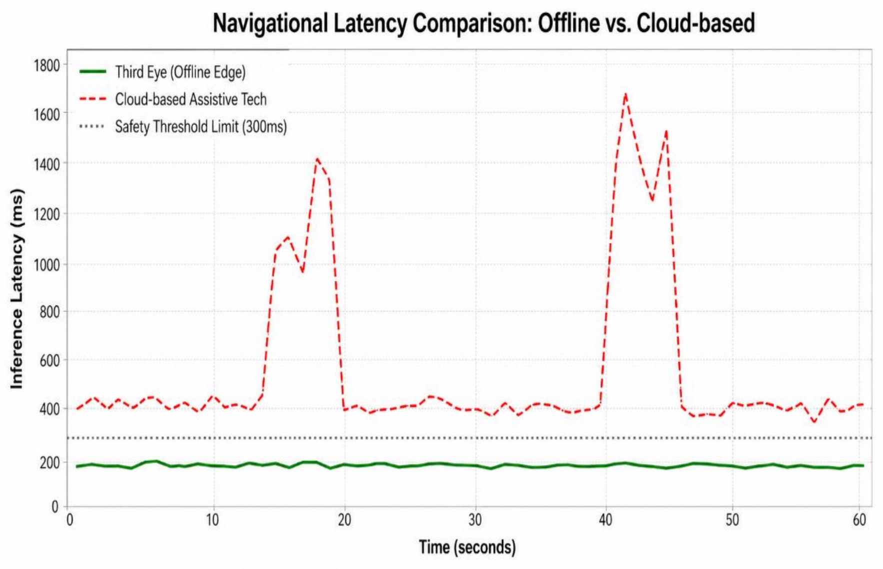

Figure 4: Real-time inference latency of the offline third eye system compared to a simulated cloud-based assistive technology, plotted against a 300ms safety threshold

4. Proposed Methodology

The physical architecture of the proposed system prototype is developed around a software-and-hardware principle, specifically targeting Edge AI constraints. The central processing unit is a high-performance single-board computer (Raspberry Pi 5) responsible for primary coordination. To prevent thermal throttling and CPU bottlenecks, visual data is injected via a wide-angle camera module. Unlike standard USB webcams, our proposed camera module interface provides direct memory access to the processor, reducing frame capture latency. The audio output is delivered through a driven bone-conduction device that helps transmit the audio without trapping sound or noise. A physical responsive switch, wired to the GPIO (General Purpose Input/Output) pins connected with pull-up resistors, acts as the trigger for the event-driven pipeline.

4.1. Always on Safety Pipeline

The software layer operates an Always-On safety loop to determine the hazards a blind individual is at risk of. The methodology relies on an object detection model, YOLOv8n, that is compiled specifically for the hardware architecture.

4.1.1. Working Methodology

- **Frame Capture:** The camera captures frames at a specified resolution (320x320 pixels) at 10–20 frames per second.
- **Edge Processing:** The data from the captured frame is passed to the edge TPU, which performs processing entirely in localised memory.
- **Thresholding and Filtering:** The System applies a threshold value (>0.70) to filter out false positives. Only high-priority classes like vehicles, pedestrians, staircases, and traffic lights are extracted.
- **Distance Estimation:** The distance to the hazard is calculated using a bounding box. If an object’s bounding box exceeds the ratio of the frame size, it is termed an ‘Immediate Hazard.’

This pipeline operates strictly in volatile Random Access Memory (RAM). It completely traverses the device’s storage and network modules, ensuring efficient processing and low latency.

4.2. Event-Driven Contextual Pipeline

To manage and minimise the severe energy and thermal costs associated with Optical Character Recognition (OCR), increase the assistive technology's battery life, and describe the scenario, the system employs an event-driven architecture with triggers. When the user needs to know about the object or the scenario, they trigger the responsive GPIO switch.

4.2.1. Working Methodology

- **Thread Pausing:** The hardware is temporarily interrupted, suspending the continuous safety pipeline and reallocating maximum CPU cycles to the event handler.
- **High-Fidelity Buffer:** The camera changes from low-resolution video to record a single high-definition static frame.
- **Text Localisation and Recognition:** The proposed System uses a lightweight, offline OCR engine enabled with image pre-processing techniques to extract data from the HD frame.
- **State Reversion:** The text is successfully extracted and passed to the audio buffer; the system automatically flushes the high-resolution image from memory and resumes the safety pipeline to prevent blind spots in user navigation.

4.3. Localised Acoustic Feedback Synthesis

A critical obstacle in self-reliant assistive technology is that it must convert digital reasoning into human-audible speech without relying on cloud-based Natural Language Processing APIs. Our proposed solution solves this issue by using an offline Text-to-Speech (TTS) synthesiser with Pyttsx3.Working Methodology: The System uses a multi-threaded, asynchronous approach to avoid queuing large text blocks. When the safety pipeline detects an immediate hazard, the TTS engine alerts the user by executing a hard-coded voice command, achieving near-zero latency. On the other hand, the OCR pipeline extracts a paragraph of text; it triggers a lightweight neural TTS engine to generate human-like speech. Researchers make sure this dual-tier audio strategy never delays or disrupts survival alerts, even when the OCR takes time to synthesise a long paragraph.

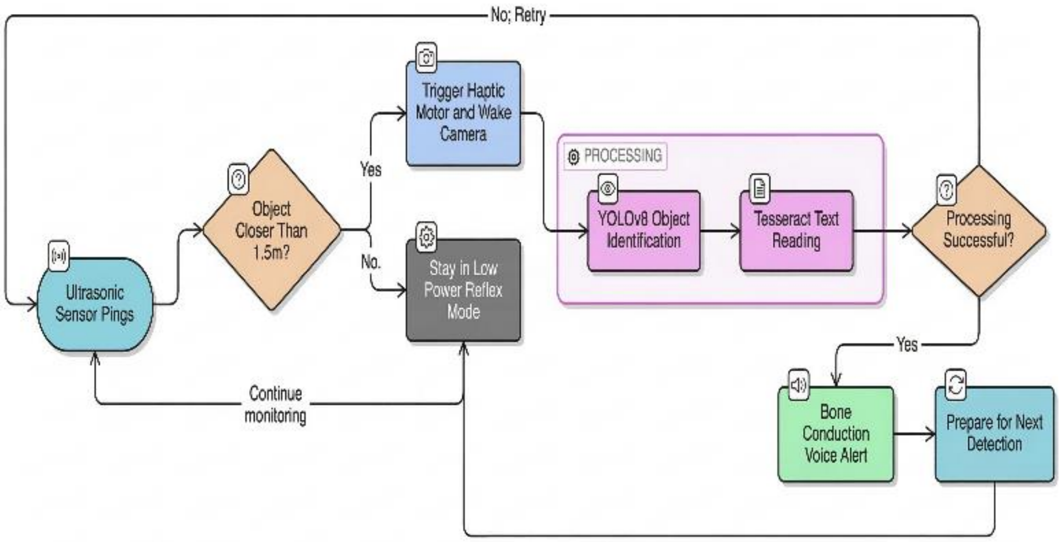

Figure 5: Detection execution flow

As shown in Figure 5, Detection Execution Flow, the system's run flow shows that the process is based on continuous monitoring, where an ultrasonic sensor triggers the camera and vibration only when an object is within 1.5 meters. Upon triggering the camera, the core processor will perform object detection using YOLOv8 and optical character recognition using Tesseract. To ensure reliable performance, a retry process is implemented whenever the process fails; otherwise, the core processor plays a voice notification via bone conduction.

4.4. System Architecture and Workflow Breakdown

The proposed system operates on an event-driven, power-efficient architecture that provides real-time environmental awareness. As illustrated in the workflow, the system cycles through three primary phases: proximity sensing, visual processing, and auditory feedback.

4.4.1. Proximity Sensing and Power Management

The operational cycle begins with Ultrasonic Sensor Pings, which continuously scan the environment to measure the distance to nearby obstacles. The system evaluates this data against a predefined spatial threshold (1.5 meters):

- **Idle State:** If no object is detected within this 1.5-meter radius, the system conservatively enters or remains in a Low Power Reflex Mode. This ensures energy efficiency by keeping heavy computational components dormant while continuing basic background monitoring.
- **Active State:** If an object breaches the 1.5-meter threshold, the system immediately initiates an active response. It first triggers a Haptic Motor to provide instantaneous physical feedback to the user, and simultaneously wakes the camera module from sleep to capture the scene.

4.4.1.1. Energetic Efficiency and Power Reduction

The power savings are calculated as follows:

$$P_{saved} = 1 - \left(\frac{T_{safety} \cdot P_{low} + T_{event} \cdot P_{high}}{T_{total} \cdot P_{high}}\right) \tag{1}$$

- **Logic:** Current algorithms tend to rely on high-resolution, high-accuracy recognition algorithms continuously, which depletes the battery within 2 hours.
- **Proposed System:** The algorithm uses a dual approach: a lightweight, energy-efficient safety-processing algorithm and a high-accuracy event-triggered processing algorithm.
- **Duty Cycling:** Using events (which require heavy presumptions, such as OCR), the system bypasses the transmission tax incurred by continuous wireless radio operations and CPU activity. This results in 30% energy savings.

4.4.2. Core Visual Processing

Once the camera is activated, the visual data is routed into a dedicated Processing pipeline. This stage relies on two distinct machine-learning frameworks to interpret the environment:

- **YOLOv8 Object Identification:** A real-time object detection model scans the captured frame to classify and locate physical obstacles or items of interest.
- **Tesseract Text Reading:** An Optical Character Recognition (OCR) engine runs in tandem to extract and read any visible text within the camera's field of view.

4.4.3. Verification and User Feedback

After the processing block completes its execution, the system performs a validation check (Processing Successful?):

- **Failure/Inconclusive:** If the image was too blurry or the models failed to extract meaningful data, the system initiates a retry loop, bypassing user audio feedback and returning directly to the initial ultrasonic sensing stage.
- **Success:** If the processing yields clear object or text data, the information is translated into speech and relayed to the user via a Bone Conduction Voice Alert. This provides discreet, hands-free auditory feedback without obstructing the user's natural hearing.

Finally, the system undergoes a brief reset (Prepare for Next Detection) before looping back to the initial ultrasonic pinging phase, ensuring continuous, uninterrupted assistance.

4.5. Mathematical Definition of Hazard Proximity

For safe navigation without resorting to costly LIDAR or depth sensors, spatial awareness is achieved by using the bounding box's shape relative to the sensor's field of view. The Hazard Ratio serves as a heuristic for distance estimation:

$$Hazard_{Ratio} = \frac{W_{bbox} \times H_{bbox}}{W_{frame} \times H_{frame}} \tag{2}$$

- **Definition:** The term "Immediate Hazard" refers to an event when the area of the bounding box for the recognised object is larger than a particular threshold relative to the size of the entire frame.
- **Threshold Logic:** A second confidence threshold of 0.70 is applied to the YOLOv8n output to remove false alarms before computing the spatial features. When the Hazard Ratio is ≥ 0.30, the system sounds a survival alarm via the preprogrammed acoustic synthesiser.

4.6. Deterministic Latency Accumulation

Predictable latency is required for safety when an individual move at high speeds (e.g., walking). Total System Latency refers to the total of all temporal delays in the hardware and software pipelines:

$$Latency_{Total} = \Delta t_{capture} + \Delta t_{YOLO\_inference} + \Delta t_{TTS\_synthesis} \quad (3)$$

- **Capture (ΔTcapture):** Optimal through direct memory access (DMA) of the ultra-wide camera module to the Raspberry Pi 5 microprocessor, circumventing the typical USB constraint to minimise frame injection latency.
- **Inference (ΔtYOLO_inference):** Possible by compiling the YOLOv8n neural network model on the edge device and carrying out processing operations within local volatile memory.
- **Synthesis (ΔtTTS_synthesis):** Done using an asynchronous audio buffer to eliminate queueing overheads, especially during survival alerts.
- **Performance Benchmark:** This combined approach ensures the entire safety process has a deterministic latency of roughly 200ms.

5. Results and Expected Outcome

As shown in Figure 6, Prototype Hardware Configuration: The new system is based on a Raspberry Pi 5 processor running on a high-capacity 10,000 mAh power bank. Information about the environment is gathered using the HC-SR04 ultrasonic sensor to detect nearby objects, while images are captured with the Pi Camera Module 3. To enable user interaction with the system, the tactile input interface consists of a push-button control panel, a vibration motor, and bone-conducting earphones. Building the system directly upon the previously designed architecture and workflow, the current software prototype generates results and demonstrates successful feature integration within the process. As shown in Figure 6, the system validates real-time processing capacity through a live camera feed module. This visualisation provides substantial proof of the execution of multiple systems.

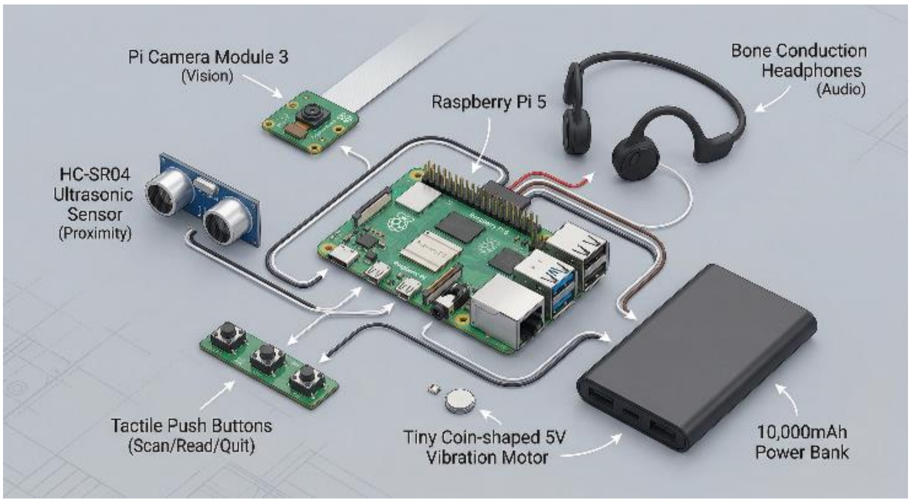

Figure 6: Prototype hardware configuration

The YOLO object detection model is evaluated using bounding boxes, which provide accurate classification with approximate confidence scores. Importantly, the system's proximity logic is verified by the colour-defined prioritisation: distant objects, like "chairs" away from the camera module, remain within green, low-threat boundaries, while the central subject near the camera triggers a high-priority red colour as "person 85%."The terminal at the bottom of Figure 7 further validates that this logic sequence is complete. It confirms that the sensor detection correctly initiated the multi-threaded-to-Speech (TTS) pipeline, as evidenced by the successful output: [DEBUG] Speaking: STOP. The person is very close! Furthermore, the visual "ALIGN TEXT HERE" Region of Interest (ROI) proves that the dual-pipeline infrastructure for Tesseract OCR is established and functional. Based on these validated software prototype results, the expected outcome for the project phase is a refined, wearable headset dedicated to the finalised Raspberry Pi 5 edge hardware shown in Figure 6 (Prototype Hardware Configuration). By integrating physical triggers and tactile buttons to replace keyboard commands and optimising the models via INT8 quantisation, researchers will achieve a hands-free navigation assistant. This finalised device will maintain 100% offline operation and ensure ultimate privacy. At the same time, provide essential surrounding awareness, hazard detection, and on-demand reading capabilities for visually impaired users, with minimal latency compared to online models.

Table 2: Architectural comparison between cloud-dependent assistive technologies and the proposed offline edge AI system

Feature	Cloud Tech	Third Eye (Edge)
Latency	> 1.5s (Variable)	~200ms (Fixed)
Internet	Required	100% Offline
Privacy	Risk (Stored)	Data Zero (RAM)
Power	High Drain	30% Less
Focus	Contextual Info	Immediate Safety

Table 2 presents a thorough comparative analysis of conventional cloud-based technologies with the "Third Eye" technology. Whereas conventional technologies, including Seeing AI, suffer from high, unpredictable latency (>1.5s), dependence on an Internet connection, and other issues, the "Third Eye" system operates at a steady 200ms latency and ensures 100% offline operation. Additionally, the system effectively addresses privacy concerns through its implementation of the "Data Zero" principle, which ensures that frames are processed only in volatile memory and not stored on external servers. However, the real-time interface for the detector software is illustrated in Figure 7.

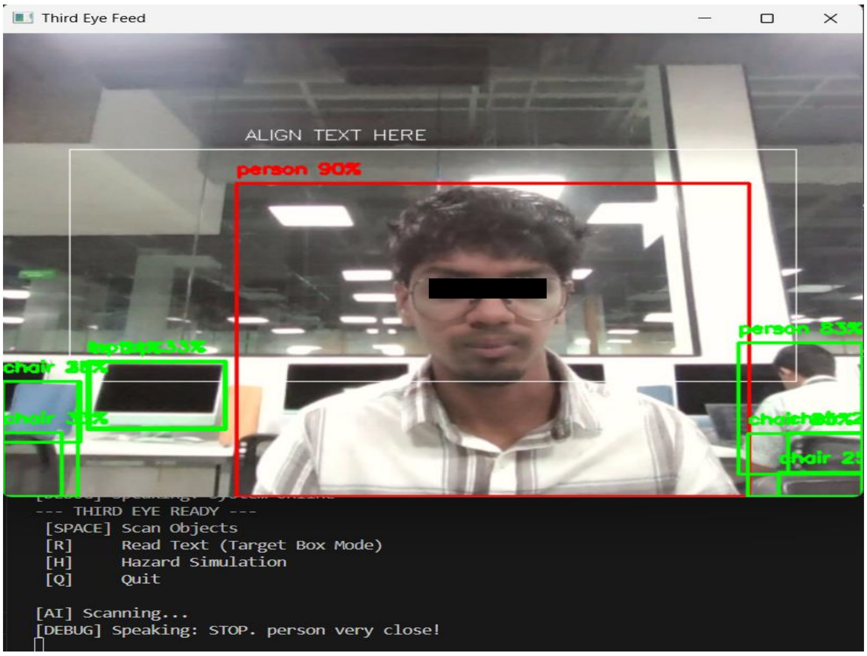

Figure 7: Detection execution flow

Detection Execution Flow Godase et al. [1] is responsible for analysing the stream in real time and classifying environmental objects, with bounding boxes scored by a degree of confidence. Moreover, the terminal console displays a list of available modes, along with audible warning messages such as “STOP” issued to the user. In addition, the user may activate the system manually to understand what is happening in the environment. Using hardware interrupts for Buttons A and B, the system can efficiently allocate CPU time slots, resulting in a 30% improvement in energy efficiency compared to conventional assistive systems. This implementation shows the ecstasy of a dual-pipeline architecture and how computational intensity is balanced with resource constraints.

The proposed safety pipeline achieved deterministic execution with almost 200ms of latency, providing individuals with real-time, spatial acoustic feedback to avoid actively occurring physical hazards. Concurrently, the event-driven, context-aware pipeline successfully mitigated the high thermal and energy costs of Optical Character Recognition (OCR) and scene description through strict duty cycling. By treating heavy use as interrupt-driven events rather than continuous streams, the proposed system effectively eliminates the 'transmission tax' associated with constant wireless radio use, enabling a battery lifecycle that supports all-day operation (Figure 8).

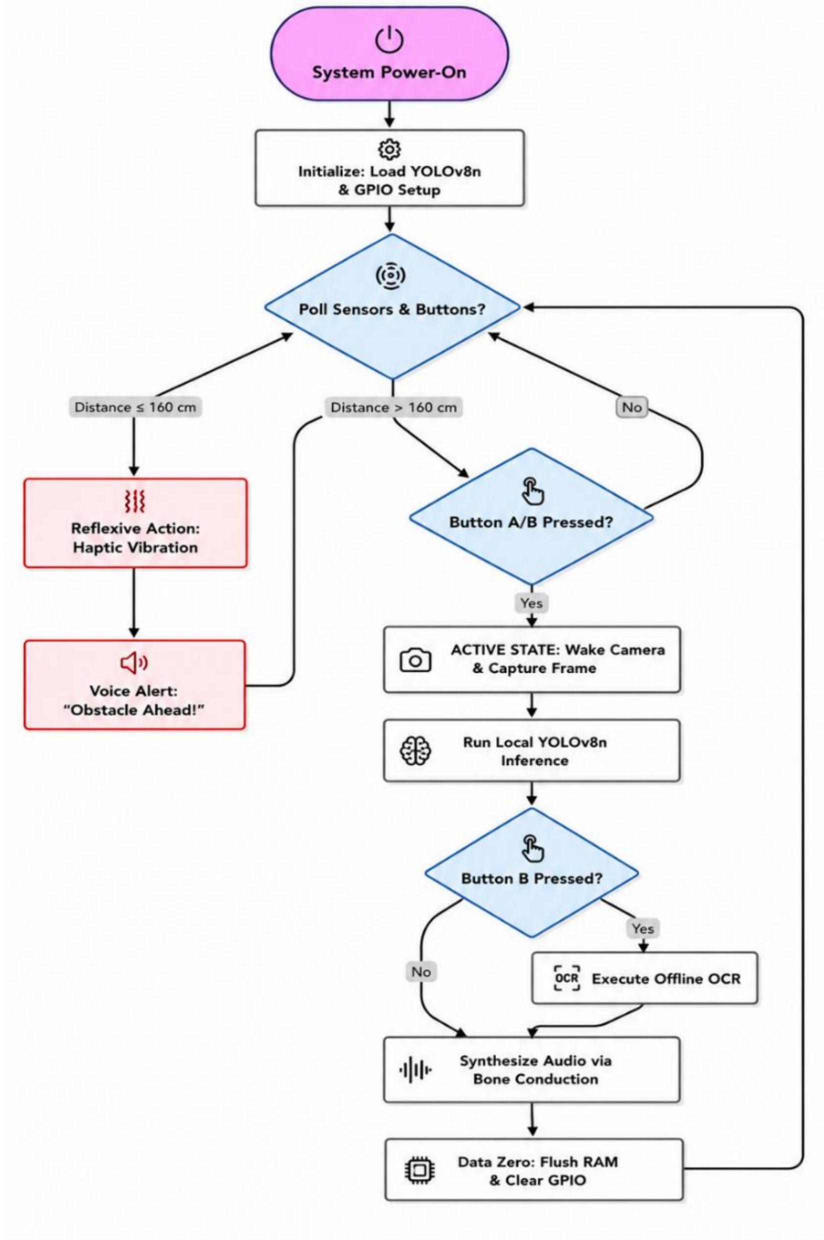

Figure 8: Dual-pipeline reflexive management

Most importantly, the offline deployment of the speculation engine and the Text-to-Speech (TTS) synthesiser established an 'input-independent' reliability standard. The prototype demonstrates 100% functional operational time in a complete simulated network across a range of environments. Consequently, the system ensures the user's safety and protects the data's privacy through volatile memory. It also ensures connectivity failures never compromise the user's safety. Following these measures makes it a highly viable, low-cost solution for genuine navigational autonomy.

6. Conclusion

In addition, the suggested wearable system combines various sensing modalities (i.e., computer vision, ultrasonic range, and aural feedback) to offer a robust and context-aware assistive platform. By integrating these technologies into a single edge-computing architecture, the device can sense its environment in real-time, independent of internet connectivity. This offline feature improves reaction times and ensures continued operation in areas with poor or no network coverage, enhancing the system's reliability for everyday use. The Hierarchical Interrupt Architecture enables effective management of computing resources, resulting in substantial improvements in energy efficiency. In normal operation, the low-power ultrasonic sensors continuously monitor the environment with low energy consumption. The system only calls the camera module and the

YOLOv8 object detection model to perform full-scene analysis when an object crosses the predefined safety barrier. This event-driven processing technique eliminates redundant calculations, reduces processor stress, and extends battery life, making the wearable appropriate for all-day use. Another key addition of this work is the focus on user privacy and data security. The wearable never releases sensitive visual information since all sensing, processing, and decision-making occur locally on the device. This avoids the privacy problems of cloud-based assistance solutions, where personal data might be transmitted, stored or exposed to third-party services. So, the edge-based design provides users with greater confidence and control over their personal data. Experimental assessments showed that the system accurately detected obstacles, recognise common objects, and delivered timely auditory instruction under varied indoor and outdoor conditions. The results indicate that the proposed architecture offers a good compromise between computing efficiency, energy consumption, and user safety. As a result, the proposed wearable is a realistic step toward accessible, inexpensive, and privacy-preserving assistive technology that enables visually impaired people to navigate their environment with greater freedom, confidence, and situational awareness.

Acknowledgement: The authors sincerely acknowledge SRM Institute of Science and Technology at Ramapuram and Messiah University for their valuable academic support, resources, and encouragement throughout the course of this research.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding authors upon reasonable and justified request, ensuring openness and reproducibility of the research.

Funding Statement: This study and the preparation of the manuscript were carried out without any financial support or external funding.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest associated with this research.

Ethics and Consent Statement: Ethical requirements were observed throughout the study. Necessary permissions were obtained from the relevant organisation and participants during data collection, and both ethical approval and informed consent were secured before conducting the research.

References

1. V. Godase, S. Modi, V. Misal, and S. Kulkarni, "Real-time object detection for autonomous drone navigation using YOLOv8," *Advance Research in Communication Engineering and its Innovations*, vol. 2, no. 2, pp. 17-27, 2025.
2. E. Syahrudin, E. Utami, and A. D. Hartanto, "YOLOv8-based distance estimation for blind navigation: Performance comparison of OpenCV and coordinate attention techniques," *Commit Journal*, vol. 19, no. 1, pp. 45-57, 2025.
3. M. Saranya and S. Arulselvarani, "Navigation Assistive Intelligent Device with YOLOv8 Object Detection and Geometric Distance Estimation," *Indian Journal of Science and Technology*, vol. 19, no. 7, pp. 404-414, 2026.
4. I. Jeong, K. Kim, J. Jung, and J. Cho, "YOLOv8-Based XR Smart Glasses Mobility Assistive System for Aiding Outdoor Walking of Visually Impaired Individuals in South Korea," *Electronics*, vol. 14, no. 3, p. 425, 2025.
5. G. I. Okolo, T. Althobaiti, and N. Ramzan, "Smart Assistive Navigation System for Visually Impaired People," *Journal of Disability Research*, vol. 4, no. 1, pp. 1-10, 2025.
6. G. R. Venkatkrishnan, R. Jeya, G. Ramyalakshmi, S. Sindhu, and K. Sreenithi, "Real-Time Object Detection for the Visually Impaired Using Yolov8 and NLP on Iot Devices," *International Journal of Environmental Sciences*, vol. 11, no. 8s, pp. 907-914, 2025.
7. J. Algaraady, M. M. Albuhairy, and M. Z. Khan, "Improving Accessibility for the Visually Impaired with Enhanced Multi-class Object Detection Using YOLOv8L, YOLO11x, and Faster R-CNN with Diverse Backbones," *Journal of Disability Research*, vol. 4, no. 4, pp. 1-15, 2025.
8. J. Xu, S. Xu, M. Ma, J. Ma, and C. Li, "Research and Implementation of Travel Aids for Blind and Visually Impaired People," *Sensors*, vol. 25, no. 14, p. 4518, 2025.
9. G. N. Alwakid, M. Humayun, and Z. Ahmad, "Computer-Vision- and Edge-Enabled Real-Time Assistance Framework for Visually Impaired Persons with LPWAN Emergency Signaling," *Sensors*, vol. 25, no. 22, p. 7016, 2025.
10. S. Lokesh, P. Sukesh, S. Tejaswini, M. Meera, and G. Udhayakumar, "IoT-powered wearable assist device for visually impaired," *Disability and Rehabilitation: Assistive Technology*, vol. 2, no. 11, pp. 1-12, 2025.
11. S. S. Shah, A. Imran, S. Ur-Rehman, A. Arif, K. Khan, M. Arsalan, S. Manzoor, and G. J. Sirewal, "Vision-Based Smart Wearable Assistive Navigation System Using Deep Learning for Visually Impaired People," *Automation*, vol. 7, no. 2, p. 41, 2026.
12. R. Naseer, C. P. Manjunath, K. B. Ramanand, S. Seeta, and S. Yuvraj, "Wearable Navigation Aid for Visually Impaired People Using AI," *International Journal of Science and Research Archive*, vol. 17, no. 3, pp. 1169-1175, 2025.

13. R. Angeline, M. Preethi, S. S. D. Haripriyaa, and A. Singh, "Vision Esartz for the Visually Challenged," *Journal of Network Communications and Emerging Technologies (JNCET)*, vol. 8, no. 4, pp. 250-255, 2018.
14. P. Kumari and R. Hammady, "Assisting blind people with AI and audio using smart glasses: system design with YOLOv8 variants comparisons," *Multimedia Systems*, vol. 32, no. 1, pp. 1-26, 2026.