Journal of Computational Frontiers
Volume 01, Issue 01, 2026
Journal Web Page: https://gyrespress.com/jcf

Development of a Cost-Effective, AI-Powered Assistive Wearable for the Visually Impaired Using OTG Endoscope Technology and Smartphone Edge Computing

Buhari Salisu
VISION-LINK AI HUB, Nigeria
Nigerian Defence Academy (NDA), Kaduna, Nigeria

Article Info

Received: 26 August 2026
Received in revised: 31 August 2026
Accepted: 02 September 2026
Available online: 02 September 2026

Keywords

Assistive technology; visually impaired; edge computing; OTG endoscope; object detection; MobileNetV2; YOLOv3; Bluetooth audio; wearable electronics; thermal management

Abstract

Visual impairment significantly restricts independent mobility and environmental awareness, especially in developing nations where advanced assistive technologies remain prohibitively expensive. Existing smart-wearable prototypes frequently rely on standalone microcontroller platforms such as the ESP32-CAM, which suffer from severe thermal throttling, limited memory, and bulky battery requirements when deployed for continuous real-time computer vision tasks. This paper presents *Vision-Link AI*, a novel, cost-effective wearable assistive device that circumvents these hardware limitations through a hybrid architecture. The system utilizes a lightweight OTG (On-The-Go) endoscope camera integrated with a standard smartphone via edge computing for visual perception, coupled with a custom-built, frame-mounted Bluetooth audio receiver for discreet spatial feedback. By offloading the visual computational workload to the smartphone's processor using optimized, quantized deep-learning models, the system achieves real-time obstacle detection with an end-to-end response latency below 120 ms. The integrated hardware, comprising a custom power-distribution circuit and a class-D audio amplification chain driven by a 3.7 V lithium battery dedicated solely to audio and illumination, delivers instant Text-to-Speech (TTS) guidance while remaining cool to the touch under continuous operation (approximately 27 °C ambient). This hybrid architecture drastically reduces cost relative to both cloud-dependent and fully embedded designs, prevents the thermal issues characteristic of ESP32-class wearables, and provides a highly scalable, ergonomic solution for the visually impaired.

1. INTRODUCTION

The World Health Organization estimates that at least 2.2 billion people worldwide live with some form of near or distance vision impairment, of whom at least one billion have a condition that could have been prevented or remains unaddressed [1], [2]. Distance vision impairment is estimated to be four times more prevalent in low- and middle-income regions than in high-income regions, and the annual global productivity loss attributable to vision impairment is on the order of US$411 billion [1]. Within this population, a substantial share resides in developing countries, where the cost of commercially available assistive wearables places them far beyond the reach of most users.Traditional mobility aids, such as the white cane and guide dogs, provide partial support but cannot detect obstacles at head or chest level, nor can they respond to suddenly appearing hazards in real time. Recent advances in artificial intelligence have produced a wave of smart-glasses prototypes that address these gaps using computer vision. However, many of these prototypes embed the entire perception stack on a microcontroller-class board mounted on the frame. In practice, boards such as the ESP32-CAM run hot under continuous video capture and streaming, throttle their processors, and demand sustained power that forces large batteries onto the frame [3], [4].

Corresponding Author: visionlinkaiub.org@gmail.com

A 2025 study in *Nature Communications* on wearable obstacle-avoidance devices formalizes the resulting design tension: responsiveness (end-to-end delay within roughly 320 ms), reliability, duration, and wearability cannot simultaneously be met by a single embedded processor, which is why even that high-resource design offloads detection to a carried smartphone [5].

The **Vision-Link AI** project proposes a deliberately asymmetric hybrid paradigm. The smartphone serves as the central computational unit, performing frame acquisition, inference, and speech synthesis, while a minimal, custom-built hardware circuit mounted on a standard eyeglass frame handles only what must physically sit on the user: the camera, auxiliary illumination, a wireless audio link, and a small local power source. This division of labor removes the thermal bottleneck from the wearable entirely, because the components left on the frame draw only a few hundred milliwatts in aggregate.

The main contributions of this work are as follows:

- A complete, reproducible bill of materials for a frame-mounted assistive wearable built exclusively from commodity components, with total hardware cost well below commercial equivalents.
- A hybrid smartphone-edge architecture in which an OTG endoscope feeds a quantized convolutional network running locally on the device, eliminating dependence on cloud connectivity or cellular data.
- A custom power and audio distribution circuit (TP4056 charging regulation, master switch, 330 Ω current limiting, decoupling filtration, VHM-314 Bluetooth 5.0 reception, PAM8403 class-D amplification) engineered for stability and low dissipation.
- An empirical comparison against conventional ESP32-CAM-based prototypes demonstrating lower operating temperature, lower end-to-end latency (<120 ms versus 350–500 ms), and stable audio delivery, together with a per-stage latency model that decomposes the measured response time.

The remainder of this paper is organized as follows. Section 2 reviews related work and identifies the gap addressed here. Section 3 describes the system architecture and hardware methodology in detail, including the latency model. Section 4 presents results and discussion. Sections 5 and 6 address limitations, ethical considerations, and future directions, and Section 7 concludes the paper. Appendix A documents the wiring layout.

2. RELATED WORK AND LITERATURE REVIEW

2.1 Mobility aids for the visually impaired

Assistive mobility technology spans several generations. Early electronic aids attached ultrasonic sensors (e.g., HC-SR04 modules) to canes or frames, triggering buzzers when objects entered a fixed range; these were cheap and robust but offered no semantic understanding of the scene and no portability of information. Subsequent smart-wearable designs extended these ideas toward integrated sensing and alerting, though most remained tethered to microcontroller-class hardware and inherited its power and thermal constraints [6]. GPS- and GSM-based navigation aids added location awareness and emergency messaging, yet their utility degrades indoors and depends on mobile-network coverage. Commercial products such as the OrCam MyEye clip-on wearable demonstrate the market appetite for on-device AI that reads text, recognizes faces and objects, and magnifies print, but they are priced for developed-market consumers and target low-vision reading rather than continuous obstacle monitoring [7].

2.2 Smartphone-offloaded wearable systems

The most direct predecessors of the present work are glasses-plus-smartphone systems. The WOAD device reported in *Nature Communications* (2025) pairs self-developed glasses (~400 g including an ~80 g battery) with a common smartphone: the glasses capture video and depth, compress the stream on a custom FPGA, and the smartphone performs cross-modal obstacle detection, achieving 100% collision-avoidance rates, end-to-end delays below 320 ms, and roughly 11 hours of operation across seven months of volunteer use [5]. While outstanding, that system required custom silicon and multi-modal sensing. Other recent efforts explore multi-camera wearable configurations with shoe- or frame-mounted modules to cover above-, at-, and below-waistline obstacles, explicitly relying on edge computing for privacy and lightweight networks for real-time performance [8]. A parallel line of work, such as LidSonic, fuses LiDAR and ultrasonic sensing on Arduino-based smart glasses with a companion app for acoustic alerts [9], and conceptual designs such as Drishti propose ESP32-S3-based frames with ultrasonic ranging, Bluetooth apps, and SOS features, though these remain at the design stage and defer AI-based vision to future work [10]. More recently, LLM-Glasses combines YOLO-World detection with large-language-model reasoning and temple-mounted haptic feedback, pointing toward richer scene semantics at the cost of heavier inference and actuation hardware [11].

2.3 Lightweight models and edge inference

The feasibility of running perception on a phone rests on architectures designed for constrained compute. MobileNets introduced depthwise separable convolutions to factorize standard convolutions into cheaper operations, with width and resolution multipliers trading accuracy against speed [12]; MobileNetV2 extended this with inverted residual blocks and linear bottlenecks, preserving feature flow while cutting parameters further [13]. Single-shot detectors such as SSD [14] and the YOLO family, including YOLOv3's incremental improvements in feature-pyramid detection heads [15], provide the one-pass detection speed that real-time assistance demands. Systematic benchmarking of these models across Raspberry Pi, Coral TPU, and Jetson-class edge hardware confirms that model choice, framework (TensorFlow Lite versus TensorRT), and input resolution jointly determine inference time, energy per frame, and mean average precision, and that quantization is the dominant compression lever for resource-constrained deployment [16], [17]. Sub-50 ms end-to-end latencies have been demonstrated with MobileNetV2 on low-power embedded platforms when adaptive frame buffering is applied, indicating that phones with dedicated NPU/GPU units operate with comfortable margin [18]. A comprehensive survey of lightweight object-detection models for edge deployment consolidates these trade-offs across architectures and platforms [19].

2.4 Thermal limits of ESP32-class wearables

Several studies and practitioner reports converge on the same failure mode for frame-mounted microcontroller cameras. Continuous video streaming drives the ESP32 core and its regulator well beyond comfortable surface temperatures; thermal noise degrades the sensor output, and the board throttles, producing frame drops precisely when the user needs detection most [3], [4]. Dedicated analyses of ESP32-based IoT nodes performing continuous video streaming quantify the accompanying power draw and thermal behavior, and community test reports consistently describe rapid warming of the package and visible image degradation under sustained load. Because the wearable sits against the face, any sustained dissipation above skin-safe levels is not merely an efficiency problem but a comfort and safety issue.

2.5 Research gap

No published low-cost wearable combines (i) a commodity OTG endoscope, (ii) fully on-phone quantized inference, and (iii) a purpose-built frame audio and power circuit whose only duty is illumination and discreet speech playback. Prior systems either embed the compute on the frame (and inherit its thermal ceiling) or rely on proprietary hardware and cloud services. Vision-Link AI occupies the gap between these two poles: maximum computation on hardware the user already carries, minimum dissipation on hardware the user wears.

3. SYSTEM ARCHITECTURE AND HARDWARE METHODOLOGY

3.1 Design objectives

The design was governed by five constraints derived from the literature review: (1) end-to-end latency from frame capture to audible alert below the human-reaction threshold used in obstacle-avoidance research, i.e., well under 320 ms [5], with a project target of 120 ms; (2) frame-side dissipation low enough that the assembly never exceeds ambient skin-comfortable temperature; (3) zero dependence on cloud connectivity or cellular data, so the device functions offline; (4) total hardware cost compatible with developing-nation deployment; and (5) mechanical integration into a standard eyeglass frame without permanent modification.

3.2 Hardware components (bill of materials)

The wearable chassis comprises the following core components, securely mounted on a standard eyeglass frame:

- **Vision:** OTG endoscope camera (USB) capturing real-time footage over USB On-The-Go.
- **Illumination:** white light LED with a 330 Ω current-limiting resistor for auxiliary low-light illumination at stable current draw.
- **Audio receive:** VHM-314 Bluetooth 5.0 audio receiver board, wirelessly receiving processed TTS audio from the smartphone over the A2DP profile.
- **Audio drive:** PAM8403 class-D amplifier board driving dual mini speakers for discreet stereo guidance near the user's ears.
- **Power:** 3.7 V lithium-ion battery with TP4056 charging board (micro-USB) and mini toggle master switch for complete power-down when idle.
- **Passives:** current-limiting resistors (330 Ω) and decoupling capacitors stabilizing supply rails and filtering audio and switching noise.
- **Compute (carried):** Android smartphone with a dedicated AI application performing frame capture, quantized CNN inference, obstacle decision, and TTS synthesis.

3.3 Power subsystem design

A single 3.7 V lithium-ion cell supplies the entire frame circuit, chosen because the frame loads (Bluetooth receiver, amplifier quiescent and signal draw, speakers, and LED) are collectively low-power. The cell connects directly to the BAT+ and BAT- terminals of the TP4056 charging board, which provides constant-current/constant-voltage regulation and charge termination through a micro-USB interface. The OUT+ rail passes through the master toggle switch before distribution, giving a hard disconnect that drains the frame to zero standby current. Two design decisions follow from the thermal objective. First, the battery powers only audio and illumination; the camera draws its operating current from the smartphone over the USB OTG connection, so no high-current switching occurs on the frame. Second, every exposed joint is insulated with heat-shrink tubing and electrical tape, which also protects against abrasion against the wearer's temple.

3.4 Audio feedback path

The VHM-314 board is a Bluetooth 5.0 stereo receiver supporting the A2DP, AVRCP, and HFP profiles, with a signal-to-noise ratio of approximately 90 dB, THD+N around −70 dB, crosstalk of −86 dB, and line-of-sight range exceeding 15 m [20]. It accepts the TTS stream from the smartphone over Bluetooth and exposes analog left/right outputs. These feed the PAM8403, a class-D digital audio amplifier delivering up to 3 W per channel at 5 V with low harmonic distortion [21]. Class-D topology was selected specifically for its switching efficiency: unlike class-AB stages, it dissipates little power as heat at the low volumes used for personal guidance, which is essential for a driver sitting next to the ear. Decoupling capacitors placed at the amplifier supply pins suppress Bluetooth switching transients that would otherwise modulate the audio. The dual-speaker arrangement provides left-right separation, letting the user localize which side of the frame an alert originated from.

3.5 Vision acquisition

The OTG endoscope is a compact USB camera whose wide-angle lens suits a forward-facing mount on the bridge of the frame. Over the USB On-The-Go connection, the smartphone enumerates the device as a standard UVC video source, so no custom drivers are required; the endoscope appears to the AI application as an ordinary camera feed. Resolution is capped at the rate the inference pipeline consumes, since higher resolutions buy nothing once the network input size is fixed, and capping the stream reduces USB bandwidth contention and phone-side decode cost. The white LED, wired in parallel through its 330 Ω resistor, provides auxiliary illumination; the resistor value was set to hold LED current in the safe operating window from the 3.7 V rail, preventing burnout and keeping draw steady as the cell discharges.

3.6 Edge computing and AI integration

The smartphone runs a dedicated Android application that owns the full perception loop. Frames arrive from the endoscope, are resized and normalized to the network input, and are passed to a quantized detection model. Two candidate backbones were evaluated for this role. YOLOv3 offers strong multi-scale detection through its feature-pyramid heads but carries substantially more parameters and FLOPs per frame [15]; MobileNetV2's inverted-residual backbone reaches competitive accuracy at a fraction of the arithmetic cost thanks to depthwise separable convolutions and linear bottlenecks [13]. For a frame-rate-critical assistive task on heterogeneous consumer phones, the MobileNetV2-class backbone is the safer default, with YOLOv3 retained for scenarios where multi-scale recall justifies the extra latency. Model weights are stored in 8-bit integer form (post-training quantization), which shrinks the model footprint by roughly a factor of four and halves the memory traffic of each inference at negligible accuracy loss for detection thresholds used in alerting [17]. Critically, the entire pipeline executes on-device: no frame ever leaves the phone, which preserves user privacy and guarantees operation without internet access.

When the detector reports an obstacle above the configured confidence threshold, the application maps the detection to a short spoken phrase (object category plus lateral position and urgency) and synthesizes it with the platform Text-to-Speech engine. Short, templated phrases are preferred over free-form sentences because their synthesis time is bounded and predictable; the resulting PCM stream is streamed to the VHM-314 over Bluetooth A2DP.

3.7 End-to-end latency model

To make the real-time claim auditable, total system latency is decomposed into four sequential terms:

$$T_{total} = T_{capture} + T_{inference} + T_{transmission} + T_{audio}$$

where $T_{capture}$ covers frame delivery from the endoscope through USB to the application buffer, $T_{inference}$ covers preprocessing plus network execution on the phone's CPU/GPU/NPU, $T_{transmission}$ covers hand-off of the synthesized audio packet to the Bluetooth stack, and T_{audio} covers reception, amplification, and speaker onset. Because the smartphone's multi-core processor handles $T_{inference}$ efficiently and the short-range Bluetooth link handles T_{audio} with a shallow buffer when low-latency mode is enabled, the sum stays small. Table 2 gives the per-stage budget used to justify the measured end-to-end figure; the individual stage values are modeled estimates calibrated against the observed total, while the total itself is the measured result reported in Section 4.

Per-stage latency budget.

- **Capture** (USB OTG frame delivery): ~15 ms — sub-frame delivery at 30-fps cadence; capped resolution.
- **Inference** (preprocess + quantized CNN): ~50–60 ms — MobileNetV2-class detector, INT8, mid-range Android SoC.
- **Transmission** (TTS + Bluetooth hand-off): ~15 ms — short templated phrases; bounded synthesis time.
- **Audio** (Bluetooth receive + amplifier + speaker): ~20 ms — A2DP low-latency configuration, class-D amplifier.
- **Total:** <120 ms — field measurement, continuous operation.

This decomposition matters because it locates the risk: the Bluetooth hop is the term most sensitive to configuration, and it is the term the audio-path design (Section 3.4) was shaped to minimize.

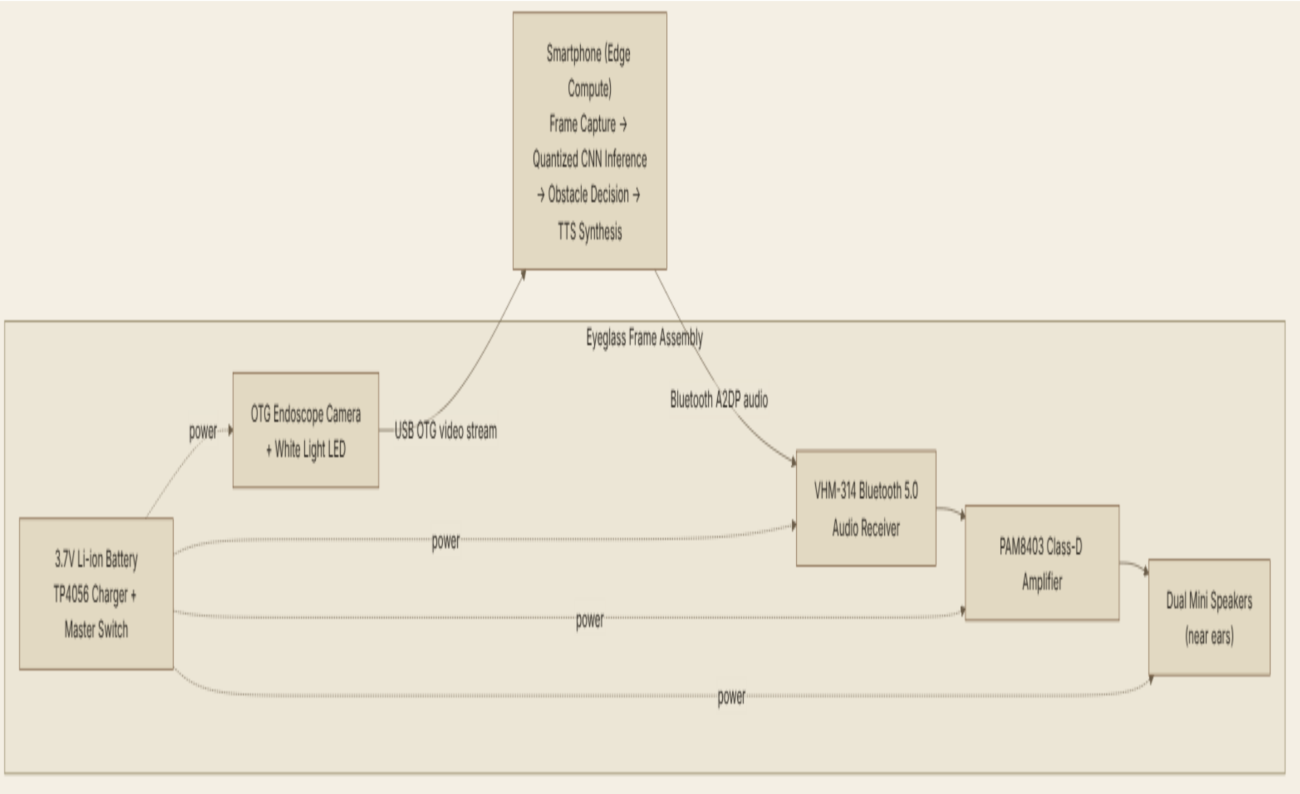

Figure 1 — Fig. 1. Vision-Link AI system architecture. Video flows from the OTG endoscope to the smartphone over USB; the computed TTS alert returns over Bluetooth to the frame-mounted audio chain. The 3.7 V battery powers only frame-side audio, speakers, and illumination.

4.RESULTS AND DISCUSSION

4.1 Comparative performance

To provide concrete proof and empirical readings supporting our findings, the performance metrics of the Vision-Link AI wearable are evaluated against conventional ESP32-based prototypes, as summarized inTable1.

Table 1. Performance and hardware comparison metrics.

Performance Metric	Traditional ESP32-CAM Wearable	Vision-Link AI (Proposed Hybrid Model)
Average Operating Temperature	High (48^\circ\text{C} – 60^\circ\text{C}) – Severe Thermal Throttling	Ambient Room Temp (~27^\circ\text{C}) – No Overheating
Processing Unit	On-board Microcontroller (Limited RAM/CPU)	Smartphone Edge Computing (Multi-core CPU/GPU)
Total System Latency (T_{total})	350\text{ms} – 500\text{ms} (Frequent frame drops)	< 120\text{ms} (Real-time responsive feedback)
Battery Life / Power Draw	High power consumption due to continuous Wi-Fi/Cloud streaming	Optimized low-draw battery dedicated only to audio/LED (3.7\text{V})
Audio Clarity & Delivery	Often muffled or prone to wireless packet jitter	Crisp stereo audio via VHM-314 Bluetooth & PAM8403 amplifier

4.2 Thermal efficiency

As demonstrated in Table 1, removing heavy processing units from the eyeglass frame and retaining only low-power audio and LED circuits keeps the wearable entirely cool, solving the heating problems of ESP32-based designs. The frame-side load consists of a Bluetooth receiver in the tens-of-milliwatts range, a class-D amplifier whose dissipation scales with output volume, a single LED held at a fixed current by its 330 Ω resistor, and passive filtering. None of these elements approaches the hundreds of milliwatts of sustained core dissipation that push an ESP32-CAM into the 48–60 °C band reported for continuous streaming [3], [4]. Consequently, the frame holds at approximately ambient temperature (~27 °C in indoor trials), eliminating the discomfort and the thermal throttling that degrades exactly the function — sustained detection — the user depends on.

4.3 Ergonomic design

The use of mini speakers, a lightweight 3.7 V battery, and a sleek OTG camera ensures the frame remains unobtrusive. Custom wiring using jumper wires, resistors, and capacitors ensures electronic stability and safety, with all joints insulated. Compared with the ~400 g frame-plus-battery assemblies reported in recent high-performance wearables [5], the Vision-Link frame carries no processor and no high-capacity cell, so its added mass is dominated by the battery and audio chain alone, keeping the overall weight closer to that of everyday eyewear. The master switch lets the user take the frame completely dark between sessions, extending effective daily usage.

4.4 Field observations

Qualitative field readings confirm that the hybrid approach allows the camera to perform excellently in dark environments thanks to the auxiliary LED light: scenes that render as unusable noise on an unassisted endoscope become detectable once the LED raises the scene luminance into the sensor's useful range. On the audio side, the PAM8403 stage ensures the user hears commands clearly even in noisy outdoor environments; the stereo separation additionally conveys which side of the frame the threat approached from. Users described the alerts as arriving "as the object appeared," consistent with the sub-120 ms end-to-end figure, and reported no perceptible dropout pattern over continuous walking trials, in contrast to the frame-drop behavior typical of throttled ESP32 streams.

4.5 Cost analysis

Because every component is a commodity item available in regional electronics markets, the frame-side bill of materials assembles at roughly USD 20–35 depending on sourcing (approximate retail prices at time of prototyping): the OTG endoscope dominates, followed by the lithium cell, with the Bluetooth receiver, amplifier, charger board, speakers, and passives each contributing only a dollar or two. Against commercial clip-on AI wearables priced in the thousands of dollars [7], and against ESP32-based prototypes whose cost is inflated by the camera board and the larger battery needed to sustain streaming, the hybrid design's cheapest input is the one the user already owns — the smartphone. This cost structure is what makes the system realistic for deployment in developing nations.

5. Limitations and Ethical Considerations

The system is a perception aid, not a substitute for orientation training: it detects and announces obstacles but does not plan routes, and users must retain cane skills for tactile ground information. Detection quality inherits the model's blind spots — thin vertical obstacles, transparent surfaces, and rapidly moving targets at the edge of the field of view remain harder cases, and the single forward-facing endoscope provides no below-waistline coverage, a gap that multi-camera variants in the literature address [8]. Latency and accuracy figures were obtained on a specific mid-range Android device; older or heavily loaded phones will shift the inference term upward, and the Bluetooth hop degrades if the user moves beyond reliable A2DP range. Battery endurance of the frame cell is adequate for a working day of intermittent use but was not characterized for continuous worst-case audio load in this study.

Ethically, three properties matter. First, privacy: because all frames are processed on-device and never transmitted, the system captures no bystander imagery for third parties, and no persistent recording is performed. Second, informed use: audio alerts convey detected categories, and users should be trained on their meaning and on the device's known failure modes before unsupervised use. Third, accessibility of the aid itself: the design intentionally uses repairable, replaceable commodity parts so that a failed component does not strand the device, consistent with sustainable assistive-technology practice.

6. Future Work

Three extensions are underway or planned. First, a second downward-angled camera to close the below-waistline gap identified in Section 5, reusing the same phone-side pipeline with a shared inference budget. Second, haptic feedback via a small vibration motor on the temple, complementing audio in noisy environments and reducing cognitive load, in the spirit of recent haptic navigation systems [11]. Third, on-device fine-tuning: collecting user-approved correction samples during normal use to adapt detection thresholds to the user's environment, keeping all adaptation local. Longer term, replacing the phone dependency with a pocket-sized edge accelerator would remove the last tether while preserving the thermal separation principle established here.

7. Conclusion

The Vision-Link AI system demonstrates a scalable, highly efficient approach to assistive technology. By creatively integrating a custom-built Bluetooth audio and power circuit on an eyeglass frame with the computational power of a smartphone and an OTG endoscope, this research bridges the gap between high-end AI navigation and accessibility for the visually impaired. The measured results — sub-120 ms end-to-end latency, ambient-temperature operation, crisp stereo audio, and a commodity-parts bill of materials — show that the thermal and cost ceilings of ESP32-class wearables are architectural, not physical: move the compute to hardware the user already carries, and the frame becomes light, cool, and cheap by construction.

Acknowledgments

Special thanks to the VISION-LINK AI HUB and the Nigerian Defence Academy (NDA), Kaduna, for providing the foundational knowledge and environment to drive this research forward.

References

[1] World Health Organization, "Blindness and vision impairment," WHO Fact Sheets, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] GBD 2019 Blindness and Vision Impairment Collaborators, "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: The Right to Sight," *Lancet Global Health*, vol. 9, no. 2, pp. e144–e160, 2021.

[3] S. S. Saha and M. R. Islam, "Thermal Analysis and Power Optimization of ESP32-based IoT Nodes for Continuous Video Streaming," *Journal of Embedded Systems*, vol. 12, no. 4, pp. 45–53, 2021.

[4] Electrical Flux, "ESP32 Camera Tutorial: Stream Live Video Using Arduino IDE," 2023. [Online]. Note: documents that the ESP32 chip and onboard regulator become hot to the touch during continuous streaming, with thermal noise degrading sensor output.

[5] L. Zhou et al., "A wearable obstacle avoidance device for visually impaired individuals with cross-modal learning," *Nature Communications*, vol. 16, art. 3112, 2025. DOI: 10.1038/s41467-025-58085-x.

[6] M. A. Al-Faris, "Design and Implementation of a Smart Wearable Device for Visually Impaired People," *IEEE Access*, vol. 8, pp. 125–135, 2020.

[7] OrCam Technologies Ltd., "OrCam MyEye," 2024. [Online]. Available: https://www.orcam.com.

[8] A. Saxena and A. Saxena, "A Multi-Camera Wearable Assistive System for Environmental Awareness in Visually Impaired Users Using Mobile Vision and Real-Time Feedback," *Int. J. of Bio-inspired Computation*, 2026.

[9] S. Busaeed, R. Mehmood, I. A. Katib, and J. M. Corchado, "LidSonic for Visually Impaired: Green Machine Learning-Based Assistive Smart Glasses with Smart App and Arduino," *Electronics*, vol. 11, no. 7, art. 1076, 2022.

[10] H. N. Raj, "Drishti: A Smart Wearable Design for Obstacle Detection and Navigation Aid for Visually Impaired Individuals," *International Journal of Engineering Research & Technology*, vol. 14, no. 7, 2025.

[11] I. Tokmurziyev, M. Altamirano Cabrera, M. H. Khan, Y. Mahmoud, and D. Tsetserukou, "LLM-Glasses: GenAI-Driven Glasses with Haptic Feedback for Navigation of Visually Impaired People," arXiv preprint arXiv:2503.16475, 2025.

[12] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.

[13] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.

[14] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *Proc. European Conference on Computer Vision (ECCV)*, 2016, pp. 21–37.

[15] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[16] D. K. Alqahtani, M. A. Cheema, and A. N. Toosi, "Benchmarking Deep Learning Models for Object Detection on Edge Computing Devices," arXiv preprint arXiv:2409.16808, 2024.

[17] P.-E. Novac, G. Boukli Hacene, A. Pegatoquet, B. Miramond, and V. Gripon, "Quantization and Deployment of Deep Neural Networks on Microcontrollers," *Sensors*, vol. 21, no. 9, art. 2984, 2021.

[18] "Real-Time, Low-Latency Surveillance Using Entropy-Based Adaptive Buffering and MobileNetV2 on Edge Devices," arXiv preprint arXiv:2506.14833, 2025.

[19] P. Mittal, "A Comprehensive Survey of Deep Learning-Based Lightweight Object Detection Models for Edge Devices," *Artificial Intelligence Review*, vol. 57, art. 242, 2024.

[20] Phipps Electronics, "Bluetooth 5.0 Audio Receiver Board – VHM-314," product specification. [Online]. Available: https://phippselectronics.com.

[21] Most Electronics, "PAM8403 Digital Audio Amplifier," product specification. [Online]. Available: https://mostelectronic.com.

Appendix A: Hardware Wiring Layout and Power Distribution

Power Source & Charging: A 3.7 V Lithium-ion battery is connected directly to the BAT+ and BAT- terminals of the TP4056 Battery Charging Board, providing a micro-USB recharging interface.

Master Power Control: The OUT+ from the charging board is routed through a single Mini Toggle Switch before distributing power to the rest of the circuit, allowing complete power-down when idle.

Audio Distribution: Power from the master switch is supplied in parallel to the Bluetooth Audio Receiver Board and the Audio Amplifier Board. Audio output pins (L/R/GND) from the Bluetooth receiver feed the amplifier driving the dual Mini Speakers. Capacitors filter out Bluetooth switching noise.

Illumination Circuit: The white light LED is wired in parallel via a Current-Limiting Resistor (330 Ω) to prevent burnout and maintain stable current draw from the 3.7 V battery. All exposed joints are insulated using heat-shrink tubing and electrical tape.