MATEC Web of Conferences **417**, 10005 (2025) https://doi.org/10.1051/matecconf/202541710005
2025 RAPDASA-RobMech-PRASA-AMI Conference

Camera and ultrasonic sensor fusion for electronic travel aid: design and performance analysis

Leeroy Majoni[1], *Stefan* van Aardt[1*], *Farouk* Smith[1] and *Shahrokh* Hatefi[1]

[1] Department of Mechatronics, Nelson Mandela University, Gqeberha, 6013, South Africa

Abstract. Millions of people globally experience visual impairment, severely limiting safe and independent mobility. This paper presents the design, development, and validation of an AI-enabled smart cane integrating ultrasonic and vision-based sensing for obstacle detection and object recognition. The system combines four HC-SR04 ultrasonic sensors with a camera connected to a Raspberry Pi 4B, running a custom-trained YOLOv4-tiny convolutional neural network (CNN) model to identify key navigational features such as doors, stairs, and ramps. The Arduino Mega handles low-latency ultrasonic sensing and vibration feedback, while the Pi processes visual frames and provides audio cues using an offline text-to-speech engine. In trials, the ultrasonic module achieved near-perfect obstacle detection within 1.2 m, and the vision system correctly identified trained objects under standard lighting, though accuracy declined under strong glare. Despite hardware limitations, the system effectively fuses both sensing modes to alert users via haptic and voice feedback. The paper evaluates all subsystems individually, verifying sensor accuracy, object detection reliability, and thermal performance under continuous load. This research contributes a portable, affordable solution toward safer navigation for the visually impaired using real-time embedded AI.

1 Introduction

According to the World Health Organisation, an estimated 1.1 billion people experience some form of vision loss, with nearly 43 million classified as blind [1]. In South Africa alone, approximately 97,000 individuals were reported to be blind or severely visually impaired in 2020 [2]. Navigating unfamiliar environments remains a major challenge for visually impaired persons (VIPs), who often rely on white canes or guide dogs. Although the white cane is widely used due to its affordability and tactile simplicity, it has limitations; its range is confined to approximately one step ahead, and it cannot detect overhanging or head-height obstacles [3].

Guide dogs offer better mobility but are expensive, require ongoing care, and cannot convey specific information about object types or scene context [4]. To bridge this gap, electronic travel aids (ETAs) were introduced as early as the 1960s [5], and more recently,

* Corresponding author: stefan.vanaardt@mandela.ac.za

advances in sensor technology and embedded computing have led to wearable or portable aids with integrated sensing [6].

Modern ETAs typically employ either range-based detection employing either ultrasonic, infrared, or LiDAR or alternatively vision-based recognition such as cameras with CNNs [7]. Range sensors are robust and energy-efficient, but they cannot semantically classify what the obstacle is. Conversely, vision-based systems provide rich semantic data but demand higher processing power and are sensitive to lighting conditions. To overcome these limitations, this study presents a multimodal smart cane system that fuses the complementary strengths of both methods.

This research builds on prior work in the field of assistive technologies and particularly extends a preliminary concept by fully developing and validating an AI-enhanced travel aid. It places a strong emphasis on the use of embedded computer vision to improve obstacle identification and augment spatial awareness. The goal is to develop a deployable, cost-effective system for independent mobility that enhances safety and decision-making for the visually impaired.

2 Literature review and theoretical background

Electronic travel aids (ETAs) aim to supplement or replace the conventional white cane by extending the user's awareness beyond tactile input. Two primary sensing categories dominate existing designs: distance-based sensing using ultrasonic or LiDAR and visual object recognition using cameras and AI algorithms [6- 7].

Ultrasonic sensing is widely adopted due to its low cost, compactness, and reliability in short-range obstacle detection. HC-SR04 sensors, for example, are capable of detecting objects between 2 cm and 400 cm with high repeatability under ideal conditions [8]. However, these sensors cannot distinguish the type or significance of obstacles; a chair and a staircase are indistinguishable in terms of distance alone [8].

Camera-based systems, by contrast, can identify and classify objects using convolutional neural networks (CNNs). For example, [9] implemented a Raspberry Pi-based object detection module using MobileNet-SSD for indoor obstacle recognition, achieving real-time feedback. However, such systems are sensitive to lighting conditions and can suffer from high latency or frame drops on resource-constrained hardware.

Fusion-based approaches attempt to combine the strengths of both modalities. [10] Introduced a multi-sensor smart stick with upward-facing ultrasonic modules to detect overhead obstacles missed by traditional canes. Similarly, [11] integrated proximity sensors with a GPS module and vibration alerts to create a multifunctional smart stick. These examples underscore the value of sensor redundancy and multimodal feedback.

The integration of artificial intelligence has improved object detection performance. YOLO (You Only Look Once) architectures have gained popularity due to their speed and accuracy on embedded systems. In particular, YOLOv4-tiny balances performance and computational load, making it suitable for platforms like Raspberry Pi. In this case, YOLOv4-tiny was selected and fine-tuned on a custom dataset of doors, stairs, and ramps.

Feedback to the user is equally important. Vibrations offer low-latency, non-intrusive alerts without obstructing hearing. Audio feedback, particularly voice prompts, can convey richer semantic information but may interfere with the user's auditory awareness of their environment. Studies such as [12] advocate for dual feedback channels using vibration and audio for clearer, safer navigation.

Despite these developments, many ETAs remain limited in range, responsiveness, or semantic awareness. There is a gap in affordable solutions that provide both extended range and meaningful environmental context. The present work addresses this by building an embedded sensor fusion system with real-time visual detection and context-specific alerts tailored to the needs of VIPs.

3 System design and implementation

3.1 Mechanical and structural design

The base is a standard 1.3 m aluminium walking stick, chosen for its lightweight, durable profile and familiarity with VIP users. The electronics are housed in a 3D-printed ABS enclosure near the cane's handle, ensuring balanced weight distribution. A finite element analysis (FEA) simulation in SolidWorks was performed to confirm that the modified cane structure could withstand normal loads of user mass less than 110 kg without deformation or failure. The design was optimised for both comfort and mechanical resilience in urban walking scenarios.

3.2 Ultrasonic obstacle detection module

Four HC-SR04 ultrasonic sensors were installed along the cane’s shaft at varying heights (toe, knee, waist, and chest levels) to detect obstacles at different vertical positions. Each module emits 40 kHz pulses and calculates distance based on echo delay. Using the relation:

$$d = \frac{v \cdot \Delta t}{2} \quad (1)$$

Where v=343 m/s at room temperature, the distance d is calculated on an Arduino Mega microcontroller with 16 MHz clock speed and multiple digital I/O ports to support parallel sensors. Any obstacle detected within 1.2 m triggers subsequent haptic feedback. Figure 1 shows the sensor wiring and spacing for the complete system. The vibration motor is driven via an NPN transistor with a flyback diode for protection. Testing showed the motor drew 80 mA, above the I/O limit of the Arduino; thus, indirect drive via transistor was essential.

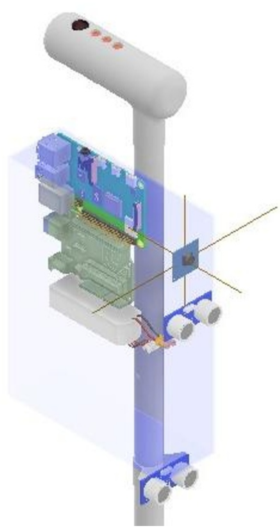

Fig. 1. CAD rendering of the system.

3.3 Vision and object detection module

The Raspberry Pi 4B (8 GB RAM) handles visual processing. A 5 MP USB camera is mounted forward-facing, slightly above waist height. The YOLOv4-tiny CNN model was selected and fine-tuned on a dataset of 480 annotated images (221 doors, 159 staircases, 100 ramps) captured using the cane's own camera in various indoor and outdoor environments. The model was trained via Google Colab and converted to TensorFlow Lite for inference on the Pi. On-device inference runs at 2 FPS, with 340 ms per frame processing latency. The model outputs bounding boxes and class labels. A voice announcement is issued if a class, such as "stairs", are detected.

To train the YOLOv4-tiny model, a transfer learning with pretrained weights (trained on the COCO dataset) was utilised as a starting point. The custom dataset of 480 images, each individually captured and labelled, was split 70% for training (augmentation was applied to increase variability), 20% for validation, and 10% held-out for testing. The training configuration employed 100 epochs with a batch size of 16 and an initial learning rate of 0.001. Training was conducted on a Colab GPU instance and completed in approximately 2 hours. Figure 2 shows the training loss curve over 50 epochs, indicating that the model converged after around 40 epochs – beyond which the loss plateaued. The final model chosen was the one with lowest validation loss. This custom-trained YOLOv4-tiny found a balance between accuracy and speed suitable for real-time use on the Raspberry Pi.

Fig. 2. Training loss curve for the YOLOv4-tiny model over 50 epochs, showing a steady convergence.

3.4 Dual feedback channels

The feedback system is dual-modal:
- Vibration: Short buzzes are triggered via the Arduino when obstacles are within 1.2 m. Pulse duration increases as distance decreases. Increments from periodic vibrations at 690ms at 69cm to 465ms at 119cm, offering intuitive urgency scaling.
- Audio: The Raspberry Pi plays pre-defined or TTS-generated voice prompts through a 3.5 mm audio output. The use of separate sensory channels—tactile for proximity and auditory for semantic information was shown to reduce cognitive load compared to single-channel alerts [12].

3.5 Power management and safety

The Arduino Mega microcontroller handles the four ultrasonic sensors, reading their echo times and computing distances continuously (every 50 ms per sensor in an interleaved cycle). If an obstacle is detected closer than the 1.2 m threshold, the Arduino immediately drives the vibration motor using a PWM signal (with pulse duration inversely proportional to distance, to convey urgency). The Raspberry Pi operates in parallel, processing the camera feed. A serial UART link between the Arduino and Pi allows basic communication (e.g., the Pi can request battery status or send a command if needed, and the Arduino can notify the Pi of critical events, though in the current design the modules mostly function independently for faster response).

The entire system is powered by a 3S 11.1 V, 2200 mAh Li-Po battery. A buck converter provides a regulated 5 V supply to the Raspberry Pi, and a separate 5 V/3.3 V regulator feeds the Arduino and sensors. This power setup yields an estimated runtime of about 3 hours of continuous use. The Arduino monitors battery voltage via a voltage divider, and the state of

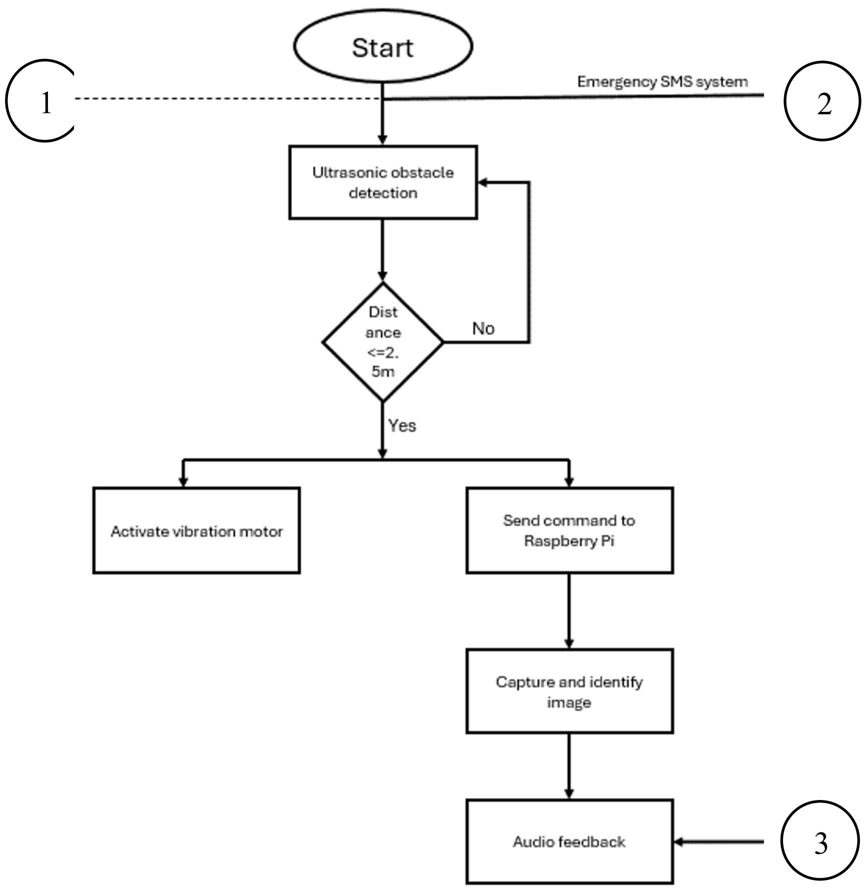

Fig. 3. The central software process of the smart cane illustrated.

the battery can be requested by the user at any time. For user safety and comfort, all exposed electronics were enclosed, and wires were routed internally along the cane. The total added weight of the electronics is about 250 g, which is distributed near the handle for better balance. The device has a simple toggle switch for power and a push-button that the user can press to query battery status or recalibrate sensors if needed.

3.6 Operational flow

In normal operation, the smart cane continuously senses the environment and provides feedback without requiring any user input beyond normal cane movement. The overall operational flow is outlined in Figure 3 to 5. On startup, the system initialises all sensors and

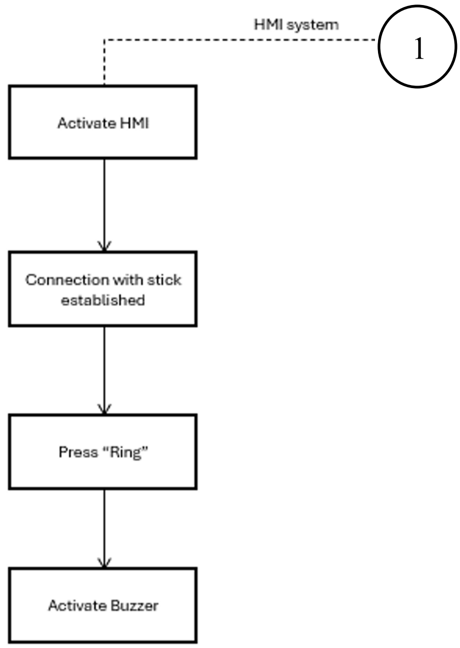

Fig. 4. The HMI software process of the smart cane illustrated.

modules. The ultrasonic subsystem pings and measures distances in a loop, instantly activating the vibration motor with varying intensity whenever an obstacle is detected within range. Simultaneously, the camera feed is processed frame by frame by the Pi's CNN module. If a known object (door, stair, or ramp) is recognised with high confidence, an audio message is immediately played via the speaker or headphone. After each cycle, the system repeats this process continuously, thus updating the user in real-time as they move. If no obstacles or known objects are detected, the cane remains silent and only the routine soft tapping of the cane provides feedback, just like a normal white cane. This ensures that the device does not overwhelm the user with unnecessary signals when the path is clear. The fusion of these parallel processes fast-range haptic feedback and semantic audio feedback which allows the user to confidently navigate while being aware of both proximal obstacles and important environmental features.

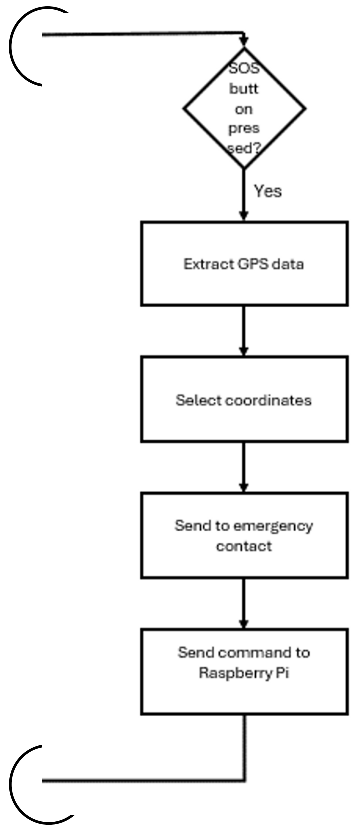

Fig. 5. The emergency software process of the smart cane illustrated.

4 Results and discussion

4.1 Obstacle detection accuracy

Under indoor conditions, the ultrasonic subsystem was tested using flat reflector boards at known distances from 0.2 m to 1.5 m. Measured distances deviated less than ±2 cm up to 0.9 m and ±5 cm beyond that, with minimal environmental influence (e.g., no significant

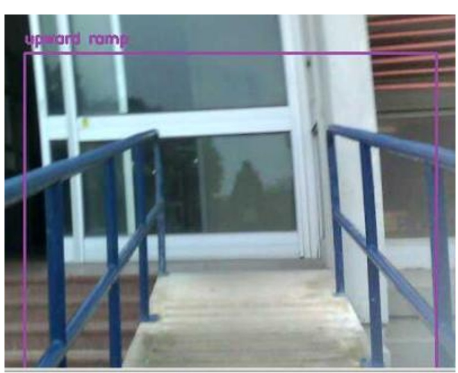

Fig. 6. Image recognition test for a ramp feature.

temperature or humidity effects in typical conditions). Table 1 shows the results for each sensor height. At 1.5 m, the measurement error increased to ~1.33%, which is acceptable since warnings are triggered well before this maximum range.

Table 1. Euclidean distance between the marker centre and the gripper centre.

Actual Distance(cm)	Detected Distance (Average of 10 readings)	Error(%)
30	30	0
60	60	0
90	89	1.1
120	119	0.83
150	148	1.33

Multiple obstacle shapes, including chairs, bins, and bags at different elevations, were used to confirm vertical coverage. At least one sensor detected the obstacle in all test cases, with no false negatives observed in 10 repeated trials per object. This proves the effectiveness of mounting sensors at staggered heights, a method shown to outperform single-point detection [10]. Figures 6-8 show the vision system detection of various features, such as ramps and doors in various lighting conditions.

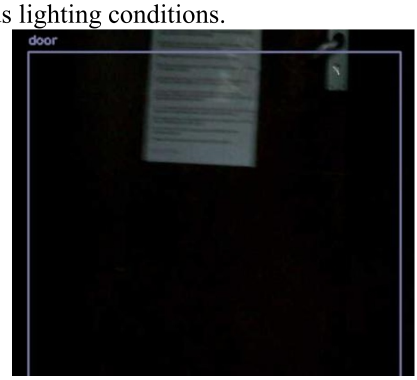

Fig. 7. Image recognition test for a door feature.

4.2 Computer vision performance

The vision system was evaluated in both controlled and real-world settings. During controlled tests, the cane was presented with sample scenes containing the target objects under various lighting conditions. The YOLOv4-tiny model consistently identified the objects in the scene when lighting was adequate. For instance, Figure 6 to 8 shows example frames of the system correctly detecting a wheelchair ramp and two doorways in the environment. In each case, the object was localised with a bounding box and the class label was announced to the user.

These examples illustrate the system's ability to recognise critical environmental features that a blind user needs to be aware of.

Fig. 8. Image recognition test for a door feature.

Importantly, when the vision module did detect an object, it was almost always correct – there were virtually no instances of high-confidence false identifications in this testing. This is crucial for user trust; erroneous alerts were rarely experienced. If the system had low confidence in a detection, it would simply not speak, and the user would continue relying on the cane's tactile feedback as usual. This design choice ensured that the audio feedback remained relevant and did not confuse the user with uncertain information.

4.3 Thermal and electrical performance

The thermal performance of the Raspberry Pi 4B was evaluated under continuous operation, as overheating could throttle the CPU and degrade performance. Without any cooling, the Pi's CPU temperature reached 69.14 °C after 10 minutes of continuous inference processing. A small 5V brushless fan blowing directly onto the Pi's CPU was added. With this active cooling, the temperature stabilised around 50.2 °C under the same load, well below the thermal throttling point of 80 °C. Figure 7 shows the CPU temperature over time in both cases (with and without the cooling fan). It is evident that the fan helps maintain a much lower operating temperature during prolonged use. In practice, the device can operate without active cooling for short periods, but for extended use in warm ambient conditions, the small fan (drawing only ~0.1 A) is a worthwhile addition to ensure consistent performance. The fan's noise was minimal and did not interfere with the audio feedback. The Raspberry Pi 4B's CPU temperature reached 69.14 °C after 10 minutes of continuous inference. With a small cooling fan, this stabilised to 50.15 °C, well below the thermal throttling point of 80 °C. Figures 9a and 9b show the thermal trend over time with and without cooling, confirming the system was able to operate without active cooling.

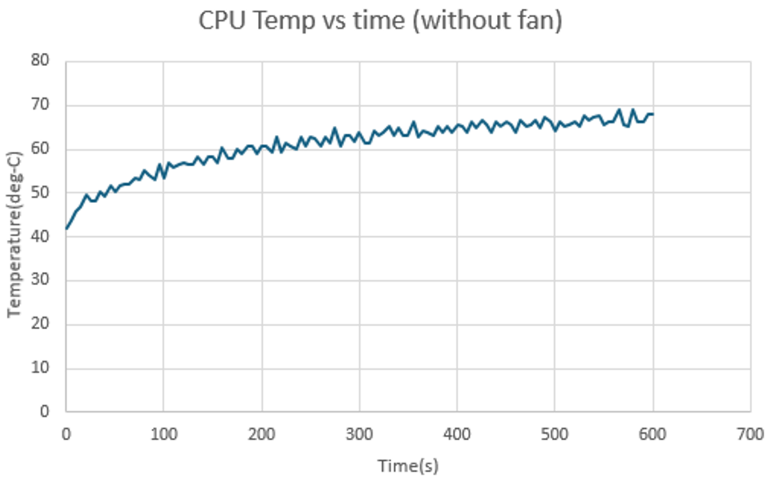

Fig. 9a. CPU temperature without the cooling fan.

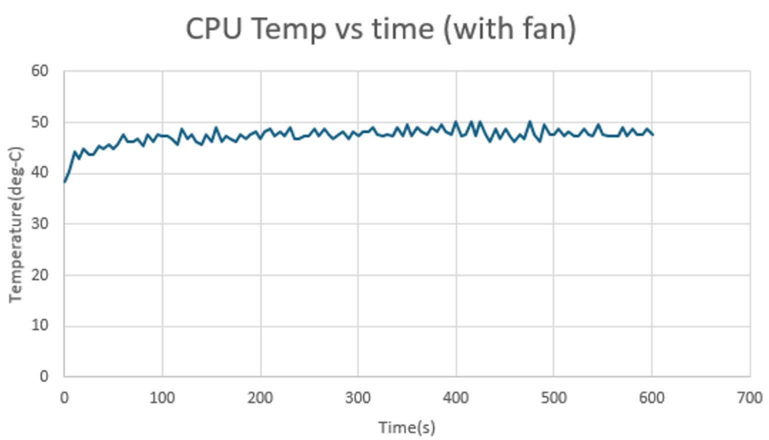

Fig. 9b. CPU temperature with the cooling fan.

Battery voltage monitoring on the Arduino was also tested by letting the device discharge during normal use. The system reliably triggered the low-battery audio alert when the Li-Po pack fell below approximately 10.8 V (which corresponds to about 3.6 V per cell). This warning gives the user ample time to plug in a charger or swap the battery. The power subsystem proved stable and no brownouts or unexpected resets occurred even when all components (Pi, camera, motors) were drawing power simultaneously, as a result of the dedicated regulators and sufficient battery current capacity.

5 Comparative discussion

Compared to prior research prototypes and commercial aids, the presented smart cane demonstrates a balanced integration of obstacle detection and object recognition in a single, compact device at a low cost. Many earlier systems tend to emphasise one aspect: for instance, some ultrasonic canes excel in detecting obstacles and drop-offs but cannot inform the user of what those obstacles are, whereas camera-based aids can identify objects but often lack immediate range feedback or require cloud connectivity. The result is a cane that not only warns a user that "something" is ahead, but also can tell them if that something is a door, a staircase, or a ramp, without requiring any external infrastructure.

In terms of cost and complexity, the proposed cane is built entirely from off-the-shelf components and costs on the order of R5000 in parts. This is substantially more affordable than some state-of-the-art smart cane systems reported in literature. For example, [13] describe a high-end prototype that uses a 2D LiDAR, an RGB-D depth camera, an inertial measurement unit, and a Jetson Nano processor on a robotic cane base. While their system achieved impressive autonomous navigation features including SLAM-based mapping and detection of many object types, it came at a cost of around $830 in materials and weighed about 2.65 kg, making it heavy and expensive for everyday use. By contrast, the presented cane requires no external computer or smartphone. Simplicity and wearability is prioritised in the design. The performance trade-off is that there are fewer object categories detected and do not perform global localisation or mapping, a features that, while useful, would require more expensive sensors and compute resources.

Focusing on object detection and recognition effectiveness, the system's computer vision module achieves accuracy comparable to similar camera-based aids. In the work of [14], a smart cane using a YOLOv4 model on a Raspberry Pi 3 was reported to detect objects such as people, mobile phones, and stairs with over 90% accuracy, allowing visually impaired users to increase their walking speed by an estimated 15–20% thanks to the extra confidence. The results for doors, stairs, and ramps are in line with these findings. Notably, unlike in the work in [14] in which the design focused mainly on vision, this resaerch cane also provides robust ultrasonic ranging. This means that even in scenarios where the camera might not recognise any object, the ultrasonic sensors still safeguard the user by detecting the presence of obstacles. In other words, this device offers a safety net through haptic feedback in addition to informational cues via audio. This multimodal approach can be especially advantageous in dynamic environments. Recent studies such as [15] have begun integrating multiple cameras and depth sensing on canes to classify surface types and obstacle severity.

6 Limitations and future work

Several areas for improvement and expansion remain to be addressed. A key limitation of the current implementation is the system's low frame rate (2 FPS), which reduces its responsiveness in dynamic environments. This latency could lead to missed detections when objects briefly enter and exit the camera's field of view during rapid user movement. Future iterations should investigate hardware acceleration platforms, such as the Google Coral TPU or NVIDIA Jetson Nano. These have been shown in similar contexts to significantly improve frame throughput without excessive power overhead. These improvements would enable the use of deeper neural architectures with higher accuracy and additional object classes while maintaining real-time performance with an increased framerate.

Another challenge lies in training the object detection model to varied lighting and scene conditions. While the model performed well in varied environments, difficulties are present when encountering untrained object types. Expanding the training dataset to include more varied and difficult examples, such as night settings, backlit scenes, or crowded walkways, would improve the model's robustness and reduce domain sensitivity. While the current model only identifies three object classes, which, although highly relevant, limit the semantic richness of the feedback. Including additional classes such as benches, hanging signs, low-hanging branches, and pedestrian traffic could further assist users in navigating complex spaces with greater autonomy.

7 Conclusion

This research presented the design, development, and evaluation of an intelligent, multimodal electronic travel aid that integrates ultrasonic sensing and embedded computer vision to enhance visually impaired individuals' spatial awareness and navigation capabilities. Through careful system architecture and rigorous experimental validation, the smart cane successfully addressed key limitations of traditional white canes – most notably their inability to detect overhanging or elevated obstacles and their lack of semantic interpretation of environmental features. By incorporating four ultrasonic sensors strategically positioned along the length of the cane, the system reliably detected obstacles across a vertical cross-section up to ~1.2 meters ahead, alerting the user through progressive haptic feedback. This significantly extends the effective detection range compared to conventional tactile-only aids and helps in early hazard avoidance.

The inclusion of a YOLOv4-tiny CNN on a Raspberry Pi 4B allowed real-time identification of key navigation-relevant features, specifically doors, ramps, and staircases. Despite the resource constraints of embedded hardware, the vision module achieved high detection accuracy in varied lighting and provided voice feedback through an offline text-to-speech module. This dual-sensing approach – combining low-latency proximity sensing with rich object classification – was found to be especially beneficial during trials. The combination of audio and vibration alerts enabled test users to interpret both distance and contextual information without being overwhelmed, validating the hypothesis that multimodal feedback can significantly improve mobility confidence for visually impaired users.

The financial assistance of the National Research Foundation (NRF), Automotive Industry Development Agency (AIDC), and the Advanced Mechatronics Technology Centre (AMTC) at Nelson Mandela University towards this research is hereby acknowledged. The opinions expressed and conclusions drawn are those of the authors and do not necessarily reflect the views of the NRF, AIDC, or AMTC.

References

1. World Health Organization, World report on vision. (World Health Organization, Geneva, 2019). https://www.who.int/publications/i/item/world-report-on-vision
2. Statistics South Africa, Profile of persons with disabilities in South Africa, 2011–2022. (Statistics South Africa, Pretoria, 2022). https://www.statssa.gov.za/publications/Report-03-01-37/Report-03-01-372022.pdf
3. R. R. A. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. Lancet Glob. Health **9**, e130–e143 (2021). https://doi.org/10.1016/S2214-109X(20)30425-3
4. L. Whitmarsh, The benefits of guide dog ownership. Vis. Impair. Res. **7**, 27–42 (2005). https://doi.org/10.1080/13882350590956439
5. National Research Council, Electronic travel aids: new directions for research. (National Academies Press, Washington D.C., 1986). https://doi.org/10.17226/1011
6. A. M. Cook, J. M. Polgar, Assistive Technologies: Principles and Practice, 4th ed. (Elsevier/Mosby, St. Louis, 2015).
7. W. Elmannai, K. Elleithy, Sensor-based assistive devices for visually-impaired people: current status, challenges, and future directions. Sensors **17**, 565 (2017). https://doi.org/10.3390/s17030565

8. A. Adnan, M. Shahid, F. Mahmood, Raspberry Pi-powered IoT device for blind navigation using MobileNetV3-SSD. In Proc. 2023 Int. Conf. Inf. Commun. Technol. Sustain. Dev. (ICICT4SD) (2023), pp. 274–279. https://doi.org/10.1109/ICICT4SD59951.2023.10303636
9. N. Patankar, B. Haribhau, P. Dhorde, H. Patil, R. Maind, Y. Deshmukh, An intelligent IoT-based smart stick for visually impaired person using image sensing. In Proc. 2023 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT) (2023), pp. 1–6. https://doi.org/10.1109/ICCCNT56998.2023.10306645
10. J. Sakhardande, R. Sakhardande, A. Malik, Smart cane assisted mobility for the visually impaired. Int. J. Electr. Comput. Eng. **6**, 9–20 (2012). https://doi.org/10.5281/zenodo.1086207
11. S. Ahmed, M. Raza, A. Zain, An intelligent and multifunctional stick for blind people using IoT. In Proc. 2022 3rd Int. Conf. Intell. Eng. Manag. (ICIEM) (2022), pp. 326–331. https://doi.org/10.1109/ICIEM54221.2022.9853012
12. K. Patil, Q. Jawadwala, F. C. Shu, Design and construction of electronic aid for visually impaired people. IEEE Trans. Hum.-Mach. Syst. **48**, 172–182 (2018). https://doi.org/10.1109/THMS.2018.2799588
13. C. Mai, H. Chen, L. Zeng, Z. Li, G. Liu, Z. Qiao, Y. Qu, L. Li, A smart cane based on 2D LiDAR and RGB-D camera sensor – realizing navigation and obstacle recognition. *Sensors* **24**, 870 (2024). https://doi.org/10.3390/s24030870
14. K. Suresh, P. J., Smart assistive stick for visually impaired person with image recognition. In Proceedings of the 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS) (2022), pp. 1–5. DOI: 10.1109/ICPECTS56089.2022.10047699
15. T. Mungdee, P. Ramsiri, K. Khabuankla, P. Khambun, T. Nupim, P. Chophuk, Low-cost smart cane for visually impaired people with pathway surface detection and distance estimation using weighted bounding boxes and depth mapping. *Information* **16**, 707 (2025). https://doi.org/10.3390/info16080707