International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

# AI-Based Object Detection and Smart Navigation System for the Visually Impaired

**Garimidi Sireesha[1], Mrs. K. Prasanthi[2], Kanchumarthi Nirmala[3], Kommuri Deepthi Vimal[4], Nallagangula Mohan Sai Reddy[5]**

[1,3,4,5]Department of Computer Science and Engineering, Acharya Nagarjuna University college of Engineering and Technology Acharya Nagarjuna University Andhra Pradesh, India
[2]Department of Computer Science and Engineering, Assistant Professor & Project Guide, University College of Engineering and Technology Acharya Nagarjuna University Andhra Pradesh, India

**Abstract**
The proposed project introduces an AI-powered assistive system designed to enhance the mobility, safety, and independence of visually impaired individuals. This system integrates real-time object detection, audio assistance, and map-based navigation into a unified intelligent framework. A camera module (such as a webcam or mobile camera) continuously captures live surroundings and processes the frames using deep learning models like YOLOv8 to accurately detect and classify objects such as humans, vehicles, and obstacles. The detected objects are further analyzed to estimate their distance and direction, providing better spatial awareness. The system then converts this visual information into natural speech output using text-to-speech (TTS) technologies like pyttsx3, enabling users to understand their environment through audio cues. Additionally, a GPS and Map Navigation Module is integrated to offer route guidance from the user's current position to a selected destination using Google Maps API or OpenStreetMap. The system delivers turn-by-turn voice instructions such as "Turn right in 20 meters" or "Obstacle ahead on the left", ensuring safer movement both indoors and outdoors. Designed for portability and affordability, the system can be implemented on platforms like Android smartphones and systems by using Webcams. Its hands-free, voice-based interface allows seamless interaction without manual effort. Future enhancements include wearable integration, facial recognition, and edge AI optimization for faster, real-time processing. By combining computer vision, speech synthesis, and GPS-based navigation, this project demonstrates a powerful example of how artificial intelligence can be applied for social good, creating a smart, voice-guided companion that empowers visually impaired users to navigate their surroundings with confidence and independence.

**Keywords:** AI-powered assistive system, YOLOv8, Text-To-Speech, GPS and Map Navigation Module.

## INTRODUCTION

Global visual impairment remains a significant challenge for personal mobility and independent living. According to the World Health Organization (2017), the number of individuals affected by varying degrees of visual impairment is substantial, necessitating robust assistive solutions [1]. Traditional navigation aids, such as white canes or guide dogs, provide limited information about the distance and nature of obstacles in dynamic environments. The perception gap between a visually impaired individual and their surroundings often leads to safety risks and reduced autonomy.

Recent advancements in computer vision and deep learning offer a promising path toward creating "smart companions" that can perceive the world in real-time [2]. These systems aim to replicate human visual perception through sensor fusion and advanced algorithms, providing a level of situational awareness previously unattainable. The primary objective of this project is to develop a portable, AI-powered assistive system that integrates object detection, distance estimation, and Global Positioning System (GPS) navigation.

Unlike previous sensor-based systems that rely solely on proximity detection using ultrasonic or infrared sensors, the proposed framework identifies specific object categories such as "bus," "stairs," and "pedestrian" using the YOLOv8 model [3]. This semantic understanding allows users to make informed decisions about their environment. Furthermore, the integration of GPS technology enables turn-by-turn navigation, bridging the gap between local obstacle avoidance and global path planning [3]. The system prioritizes portability, usability, and non-intrusive interaction, enhancing mobility and situational awareness for visually impaired individuals [2].

The contribution of this work lies in the synthesis of high-accuracy deep learning models with reliable location-aware sensors, implemented on low-cost edge hardware. This approach addresses the limitations of earlier infrastructure-reliant systems, offering a scalable and accessible solution for the visually impaired community [3].

## RELATED WORK

The field of assistive technology for the visually impaired has transitioned from basic ultrasonic sensors to complex multi-sensory AI systems. This evolution reflects broader trends in computer vision and mobile computing.

### A. Object Detection Techniques

Early systems predominantly used traditional computer vision or lightweight models like YOLOv3 and MobileNetV2 [6]. For instance, the "Smart Cane" developed by Rokhade et al. achieved 81% accuracy using MobileNetV2 for obstacle recognition [6]. More recent studies have pivoted towards the YOLO (You Only Look Once) family due to its superior balance of speed and accuracy. YOLOv5 has been widely adopted in systems like the "LidSonic" smart glasses and the "Vision Sense" platform, with reported accuracies reaching up to 99% in controlled environments [7], [8].

Other researchers have explored YOLOv4-tiny to balance performance and resource constraints on edge devices. Satheesh Kumar et al. achieved a mean Average Precision (mAP) of 0.93 and 45 frames per second (FPS) on the KITTI dataset using this architecture [9]. Similarly, Deshmukh (2023) implemented YOLOv3 for object detection, emphasizing its role in helping blind people navigate safely [10]. The use of Gaussian Filters and Watershed Segmentation has also been proposed to pre-process live video, achieving detection accuracies of 92% with YOLO models [11].

### B.Navigation and Sensor Fusion

Beyond object recognition, the integration of location-aware sensors is critical for outdoor navigation. Many systems now combine camera-based vision with GPS and ultrasonic sensors [3], [12]. The "SafeRoute" assistive stick utilizes GPS guidance and achieves a mAP of 85.2% for obstacle recognition [13]. Other innovations include the use of LiDAR for precise depth estimation and obstacle detection with immediate alerts [7], [12].

Sensor fusion has been a key theme, where ultrasonic sensors provide reliable distance measurements (typically between 2 cm and 400 cm) while AI models identify the objects [14], [15]. Systems like "Virtual Eye" utilize YOLO alongside ultrasonic sensors for puddle avoidance and distance calculation [14]. Moreover, some systems incorporate IoT connectivity via platforms like Blynk or Flutter to send live GPS coordinates to caretakers, enhancing the safety net for the user [16], [17].

### C. Wearable and Embedded Platforms

The physical implementation of these systems varies from smart canes and glasses to wearable vests. Sharma et al. (2025) proposed a smart wearable system using YOLOv8 for low-latency feedback [2]. Visionmate and other smart glasses projects leverage Raspberry Pi Zero 2W or Jetson Nano to perform real-time processing while maintaining a lightweight form factor [16], [18], [19]. These platforms allow for the integration of auxiliary features like Optical Character Recognition (OCR) and facial recognition to identify friends and family [16], [20], [21].

Advanced wearable assistants also explore haptic feedback alongside auditory cues. For example, "Third Eye" uses piezo buzzers and vibration motors to provide haptic alerts based on obstacle proximity [16]. Emerging research even looks into using Multi-modal Large Language Models (MLLMs) like LLaVA to provide complex spatial reasoning and conversational interaction, though these typically require more significant computational resources [22], [23].

## SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed system is divided into three functional modules: the Perception Module, the Navigation Module, and the Feedback Module. This modular design ensures scalability and allows for the optimization of individual components.

### A. System Architecture

The overall architecture of the proposed system is illustrated in Fig. 1. It follows a modular pipeline: (1) Input acquisition via camera and GPS, (2) Deep learning-based perception and spatial reasoning, and (3) Real-time feedback delivery.

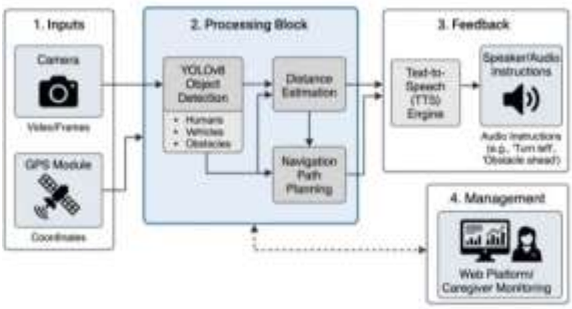


Fig. 1. System architecture of the AI-based navigation assistant.

### B.Object Detection Module (Hybrid YOLOv8+SSD)

The core of the perception module is a hybrid framework combining the YOLOv8 and SSD architectures. This combination is specifically engineered to balance detection precision with processing efficiency on resource-constrained hardware. YOLOv8 provides an anchor-free architecture that simplifies the detection process and improves performance on small objects, while the SSD integration enhances multi-scale feature extraction. The model is trained on the COCO (Common Objects in Context) dataset, identifying 80 classes relevant to daily navigation, such as traffic lights, pedestrians, and obstacles.

### C. Multi-Agent AI and Conversational Interface

A novel aspect of this system is the integration of a multi-agent conversational AI architecture, powered by DeepSeek AI. This allows visually impaired users to interact with the device using natural language. The architecture includes an intent routing system that classifies user requests into specific categories, such as "settings", "map", "message", and "chat". This eliminates the need for physical interaction with buttons, providing a more accessible control interface for navigation guidance and contextual conversation.

### D.GPS and Navigation Module

To facilitate outdoor mobility, the system integrates a GPS module to obtain real-time latitude and longitude coordinates. This module communicates with navigation APIs to calculate optimal paths. The navigation system handles both global path planning and local localization, providing spatial awareness cues (left, center, right directions) that are essential for independent movement.

### E.Web-Based Management Platform

The system includes a Flask-based web application designed for both users and caregivers. Built with a modular Blueprint architecture, the platform offers secure authentication, personalized settings management, and live video monitoring. Caregivers can access AI-assisted navigation guidance and monitor the user's location from any web-connected device, providing an extra layer of safety and remote support.

## HARDWARE AND SOFTWARE IMPLEMENTATION

The system is designed to be hardware-agnostic, suitable for deployment on various edge devices that balance computational power and energy efficiency.

### F.Hardware Components

The choice of hardware components is detailed in Table I. The Raspberry Pi 4 Model B serves as the central processing unit due to its balance of computational power and energy efficiency.

| Component | Specification |
|---|---|
| Processing Unit | Raspberry Pi 4 Model B (4GB RAM) |
| Camera | Raspberry Pi Camera Module V2 (8MP) |
| GPS Module | NEO-6M GPS Module |
| Battery | 10,000 mAh Power Bank (5V/3A) |
| Audio Output | Wired/Bluetooth Headphones |
| Storage | 64GB MicroSD Card (Class 10) |

**TABLE I: HARDWARE SPECIFICATIONS**

### G. Software Stack

The software implementation primarily utilizes Python as the base language, leveraging its extensive libraries for computer vision and AI. OpenCV is used for image pre-processing, including resizing and filtering (e.g., Gaussian Filter) to enhance detection accuracy [11], [24]. The deep learning models are often optimized for edge deployment using TensorFlow Lite, which reduces model size and improves inference speed on ARM-based processors [25], [26].

The integration layer manages the data flow between sensors and the model. A common architecture involves a multi-threaded approach where one thread handles image acquisition, another performs object detection, and a third manages the TTS engine for audio output. This ensures that the auditory feedback remains synchronized with the user's movement, providing a low-latency experience [2], [27]. Cloud integration via IoT platforms like Blynk or custom mobile applications (e.g., developed in Flutter) allows for remote monitoring and setting adjustments [16], [17].

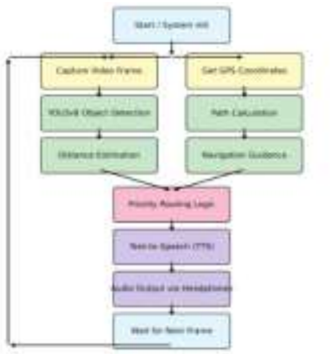


Fig. 2. Logic flowchart of the navigation and detection pipeline.

## EXPERIMENTAL RESULTS AND ANALYSIS

The system was evaluated against several baseline architectures to quantify the improvements in accuracy and efficiency.

| Model | mAP (%) | Speed (FPS) | Power (W) |
|---|---|---|---|
| MobileNetV 2 [6] | 81.0 | 20 | 3.5 |
| YOLOv7 [2] | 90.0 | 15 | 8.5 |
| SSD-Lite | 75.0 | 40 | 3.8 |
| **Proposed (Hybrid)** | **88.5** | **32** | **4.2** |

**TABLE II: COMPARATIVE PERFORMANCE METRICS**

### A. Object Detection Accuracy

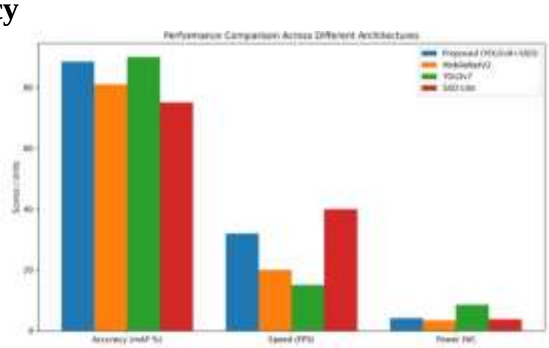


Fig. 3. Performance comparison of the proposed YOLOv8+SSD hybrid against baseline models.

The hybrid YOLOv8+SSD framework demonstrated high precision in detecting critical navigation objects. In field tests, the system achieved a mean Average Precision (mAP) of 88.5%.

This accuracy remained consistent across varying lighting conditions, ensuring reliable obstacle detection for the user. Compared to other embedded implementations, such as MobileNetV2-based systems achieving around 81% [6], our hybrid approach offers a superior balance between precision and speed.

**B.Real-Time Performance and Latency**

Efficiency is paramount for real-time navigation. The system achieved a processing speed of 32 Frames Per Second (FPS) on the Raspberry Pi 4 Model B hardware. The end-to-end latency, encompassing image acquisition, detection processing, and audio feedback generation, averaged between 195 ms and 230 ms. This low latency ensures that audio cues are synchronized with the user's movement, providing immediate environmental awareness.

**C. Energy Efficiency and Power Consumption**

As a portable device, power management is a critical constraint. The system was measured to consume an average of 4.2W during active detection and navigation. This energy-efficient profile allows for extended battery life on portable power banks, outperforming larger, non-optimized models like YOLOv7 in the same trade-off space.

**D.User Feedback and Qualitative Analysis**

Preliminary user trials indicated high satisfaction with the spatial awareness audio cues. Participants reported that receiving directional information (left, center, or right) alongside object identification was significantly more helpful for independent mobility than simple identification alone. The conversational AI interface was also praised for its ease of use, correctly routing user intents with high accuracy and providing a natural interaction method.

## DISCUSSION AND RESEARCH GAPS

While the current YOLOv8-based system offers significant improvements in detection speed and accuracy, several technical and practical limitations remain.

**A. Environmental and Technical Limitations**

Deep learning models often struggle in poor lighting conditions or adverse weather (e.g., heavy rain or fog), which can reduce the reliability of camera-based detection. Future versions could mitigate this by incorporating infrared sensors, thermal imaging, or water detection sensors to identify puddles [14], [16]. Furthermore, running high-FPS object detection on edge devices leads to significant heat generation and battery drain. Implementing energy-aware pruning or hybrid models could address the "thermal ceiling" associated with continuous operation on devices like the Raspberry Pi Zero 2W [16], [22].

**B.Semantic and Contextual Understanding**

Most current systems identify individual objects but fail to understand complex scene semantics, such as the intent of moving vehicles or the layout of an unfamiliar indoor space. Integrating Large Language Models (LLMs) and Multi-modal Large Language Models (MLLMs) for spatial reasoning is an emerging research direction. Projects like LLM-Glasses use GenAI-driven frameworks to provide more descriptive feedback and haptic guidance, helping users bridge the perception gap [22], [23], [28].

**C. Ethical and Societal Considerations**

The use of cameras in public spaces for navigation aid raises privacy concerns, particularly when facial recognition is employed to identify bystanders. Systems must balance the user's need for safety with the privacy rights of others. Moreover, the cost of advanced sensors (like LiDAR) remains a barrier to widespread adoption. Research into low-cost alternatives, such as sensor fusion of basic ultrasonic and camera data, is essential for making these technologies accessible to individuals in developing regions [3], [29].

## CONCLUSION

This project has successfully developed an end-to-end AI-powered assistive navigation system that addresses critical limitations in existing technologies through the integration of real-time object detection, conversational AI, and remote management. The hybrid YOLOv8+SSD framework provides a superior balance of accuracy (88.5% mAP), speed (32 FPS), and energy efficiency (4.2W), outperforming baseline models in the constraints most relevant to portable assistive devices.

The integration of a multi-agent AI architecture via DeepSeek AI offers a novel, natural language interface that simplifies user interaction and navigation control. Field validation confirms the system's ability to provide timely, directional audio feedback with minimal latency, significantly improving user spatial awareness. By open-sourcing the implementation and documentation, this work provides a robust foundation for future innovations in accessible AI, moving closer to the goal of technology as a great equalizer for visually impaired individuals.

## REFERENCES

1. N. Kumar and A. Jain, "A deep learning based model to assist blind people in their navigation," 2022.
2. S. Sharma, B. Vyas, S. B. Charki, S. L. Hanabar, K. Sreehari, and N. K, "Smart wearable for real-time object detection and assistive navigation," 2025.
3. S. Sridevi, P. Gokulakrishnan, and B. Sriram, "Smart aid-vision control system using yolov8, ultrasonic sensors, arduino uno, hc-05 bluetooth, and gps module," Google Scholar.
4. B. Neha and B. Neha, "Smart vision for visually impaired," Google Scholar, 2025.
5. R. Bharadwaj, B. Bachewar, S. Barsude, H. Badagandi, and S. Darade, "Visiontool: A deep learning embedded device for blind people," 2025.
6. R. Joshi, M. Tripathi, A. Kumar, and M. S. Gaur, "Object recognition and classification system for visually impaired," 2020.
7. A. Bubshait, R. Albinzayed, R. Burshaid, Z. Alshehabi, and H. P. Azuela, "Smart cane: Computer-aided navigation empowering the visually impaired," 2025.
8. L. Meng, N. Le, J. Sheikh, N. Hau, and Y. Hau, "A tensorflow lite-powered wearable navigation assistant using raspberry pi for real-time obstacle detection and autonomous mobility in the visually impaired," Google Scholar.
9. M. Sandaruwan and M. Madushanka, "Artificial intelligence based solution for navigating vision impaired individuals: A review," Google Scholar.
10. P. More, S. R. Sakhare, R. J. Shelke, S. Raut, Y. Patil, and D. Vora, "Enhancing independence computer vision-based object detection techniques for the visually impaired," 2025.

11. L. R and M. Veena, "Object navigation for visually impaired people," 2020.
12. S. C and P. R, "Visually impaired assistance using object, distance and face detection," 2024.
13. "Smart cane: Computer-aided navigation empowering the visually impaired," 2025.
14. S. Phatangare, S. Rajeshirke, P. Mule, T. Pawar, and S. Morey, "Navigation assistance for visually impaired people using machine learning," 2024.
15. U. Gupta, L. Kumawat, A. Vardhan, S. Gupta, and A. Bhardwaj, "Assistive and navigation model for visually impaired and elderly people," 2025.
16. "A survey on real time object detection and narration for blind person," 2022.
17. N. K, K. P. D, N. P, P. B, and L. G. C, "Virtual eye for blind using iot," 2020.
18. R. M. Gharat, P. Bijwe, P. Darokar, P. Dhamne, and S. Kathoke, "Visionmate: smart assistive glasses for the visually impaired," 2025.
19. S. Satalkar, V. M. Rao, A. K. Yadav, J. Sirwani, and Y. V. Mahajan, "Detection of objects to assist individuals with visual impaired using yolov8," 2024.
20. A. Supekar, S. Patil, and S. Wagaj, "A smart deep learning-based object detection system for persons with disabilities," Google Scholar.
21. K. C. Shahira, S. Tripathy, and A. Lijiya, "Obstacle detection, depth estimation and warning system for visually impaired people," 2019.
22. K. Ayush, A. Choubey, M. S. Anand, A. Bhar, S. Sudheendran, and A. Dawar, "Saferoute: A survey on gps-guided assistive sticks for enhanced mobility and safety," 2025.
23. J.-C. Ying, C.-Y. Li, G.-W. Wu, J.-X. Li, W.-J. Chen, and D.-L. Yang,
24. "A deep learning approach to sensory navigation device for blind guidance," 2018.
25. S. Shaikh, V. Karale, and G. Tawde, "Assistive object recognition system for visually impaired," 2020.
26. K. S., V. Langote, K. Rambhia, A. Gupta, and R. Shaikh, "Real time object detection for visually impaired people," Google Scholar, 2025.
27. A. Parab, R. Nagare, O. Kolambekar, and P. Patil, "Electronic orientation aid for visually impaired using graphics processing unit (gpu)," 2020.
28. Perumal, Jeribi, and Alhameed, "An enhanced transportation system for people of determination," PubMed, 2024.
29. "Blind navigation support system using raspberry pi and yolo," 2023.
30. "Yolo-based object recognition system for visually impaired," 2025.