IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511133

Assisting Blind Individual In Navigation Using Multimodal LLM And Voice Assistance

Anitha L[1], Bhuvana T S[2], Gagana H T[3], Manaswi K M[4], Manoj K R[5]

Assistant Professor, Department of CSE, East West College of Engineering, Bengaluru, India[1]

Student, Department of CSE, East West College of Engineering, Bengaluru, India[2]

Stusdent, Department of CSE, East West College of Engineering, Bengaluru, India[3]

Student, Department of CSE, East West College of Engineering, Bengaluru, India[4]

Student, Department of CSE, East West College of Engineering, Bengaluru, India[5]

Abstract: The system captures real-time images or video using a smartphone camera or webcam. The captured frames are processed using computer vision techniques, and an AI-based object detection model, such as You Only Look Once (YOLO), identifies objects such as people, chairs, doors, vehicles, and other obstacles. Approximate distance estimation can be performed using bounding-box analysis, and the detected information is converted into meaningful voice instructions through text-to-speech technology. Vibration alerts can provide additional warnings when obstacles require immediate attention.The proposed application also includes interface features for voice-assisted navigation, emergency assistance, and accessibility settings. The system aims to improve environmental awareness and support safer, more independent mobility. Further development may incorporate multimodal Large Language Models (LLMs), GPS-based route guidance, advanced sensors, and intelligent path planning. The application further incorporates interface features for voice-assisted navigation, object detection descriptions, backtracking, emergency assistance, and accessibility settings. The project aims to improve environmental awareness by communicating relevant information in a simple and accessible manner. The proposed approach combines camera -based perception with audio feedback to support users in understanding their surroundings and responding to potential obstacles. However, the reliability of the system depends on object detection performance, lighting conditions, camera quality, processing speed, and the accuracy of proximity estimation. Bounding-box analysis provides only an approximate indication of distance and cannot independe ntly guarantee safe navigation.

Keywords: Artificial Intelligence, Computer Vision, Visually Impaired Assistance, Object Detection, YOLO, Voice Assistance, Text-to-Speech, Multimodal LLM.

I. INTRODUCTION

Navigation is an important part of independent living, allowing individuals to move from one place to another and perform their daily activities. However, visually impaired individuals often face difficulties navigating unfamiliar environments because they cannot easily identify obstacles, recognise surrounding objects, or understand the layout of a location. These challenges may occur in crowded streets, public buildings, educational institutions, shopping centres, and other places where environmental conditions change frequently. Identifying doors, stairs, lifts, vehicles, and moving people can be particularly difficult without sufficient environmental information. In addition, unexpected obstacles may create challenges when moving through unfamiliar pathways. Depending on the environment, users may require assistance to identify their surroundings and determine an appropriate direction of movement. These difficulties highlight the importance of developing accessible technologies that can provide timely environmental information and support independent navigation.

Recent developments in Artificial Intelligence (AI) and computer vision provide opportunities to improve assistive navigation. Cameras can capture the user's surroundings, while object detection models analyse images and identify important objects. The detected information can then be converted into spoken instructions, allowing users to receive environmental information without relying entirely on visual displays.

The proposed project, Assisting Blind Individuals in Navigation Using Multimodal LLM and Voice Assistance, focuses on combining camera-based object detection with voice assistance and additional accessibility features. The system is designed to identify objects, provide audio guidance, and warn users about nearby obstacles. It also includes application screens for emergency assistance, sensor-based backtracking, and accessibility settings.

The proposed project, Assisting Blind Individuals in Navigation Using Multimodal LLM and Voice Assistance, focuses on combining camera-based object detection with voice assistance and additional accessibility features. The system is

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.1511133

designed to identify objects, provide audio guidance, and warn users about nearby obstacles. It also includes application screens for emergency assistance, sensor-based backtracking, and accessibility settings.

The main objective is to develop an accessible navigation assistance application that helps visually impaired individuals understand their surroundings and make more informed movement decisions. The project also provides a foundation for future integration of multimodal language models, GPS navigation, and advanced environmental understanding.

The principal contributions of this work are as follows:

1. A camera-based workflow for capturing environmental images and preparing them for AI-based analysis.
2. The use of an object detection model such as YOLO to identify supported objects and potential obstacles.
3. A voice-assistance approach that converts detected information into spoken messages using text-to-speech technology.
4. An approximate proximity-estimation approach based on object positions and bounding-box analysis, with recognition of its limitations.
5. Application interface features for voice-assisted navigation, backtracking, emergency assistance, and accessibility settings.
6. A modular foundation for future enhancements involving multimodal LLMs, GPS, advanced sensors, regional languages, and intelligent path planning.

II. RELATED WORK

The development of assistive technologies for visually impaired individuals has received considerable attention in recent years. Researchers have explored different approaches to improve independent mobility, environmental awareness, object recognition, and obstacle detection. Traditional mobility aids such as white canes and guide dogs provide essential support, but they may not provide detailed information about the surrounding environment. To address these limitations, researchers have investigated technologies such as computer vision, artificial intelligence, deep learning, ultrasonic sensors, GPS, smartphone applications, and voice-based assistance.
Recent research mainly focuses on developing intelligent systems that can identify objects, detect obstacles, estimate their relative positions, and communicate useful information through audio or vibration feedback. These technologies aim to provide additional environmental information and support users while moving through unfamiliar indoor and outdoor environments.

A. Traditional Assistive Technologies

Traditional assistive technologies have been widely used to support visually impaired individuals in their daily activities. White canes are among the most commonly used mobility aids. They help users identify certain obstacles, changes in floor levels, and objects located within their reach. Guide dogs also provide assistance by helping users navigate pathways and avoid certain obstacles.
Although these methods are useful, they have limitations in providing detailed information about the environment. A white cane generally cannot identify the exact category of an object or describe objects located farther away. Similarly, guide dogs require specialised training and may not provide information about every object in the surroundings. These limitations have encouraged researchers to develop electronic navigation systems that can provide additional environmental information through cameras, sensors, and intelligent software.

B. Smartphone-Based Computer Vision Systems

Smartphone-based assistive systems have become an important research area because smartphones contain cameras, processors, speakers, and connectivity features. These devices can be used to capture images of the surrounding environment and process them using computer vision algorithms.
A systematic review by Budrionis et al. examined smartphone-based computer vision travel aids for blind and visually impaired individuals. The review discussed research prototypes designed to support indoor and outdoor orientation and navigation by providing additional information about surrounding objects. It also identified research gaps involving the Smartphone-based systems offer advantages such as portability, accessibility, and the possibility of reducing the need for specialised equipment. However, their performance may depend on camera quality, battery capacity, processing power, network availability, and environmental conditions.
The proposed project follows a similar direction by using a camera or webcam to capture environmental information and processing the captured frames to identify supported objects.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.1511133

C. Artificial Intelligence and Deep Learning Techniques

Artificial Intelligence has introduced new possibilities for developing intelligent assistive applications. AI-based systems can analyse images, identify objects, recognise patterns, and generate useful information from visual inputs. Deep learning techniques, particularly convolutional neural networks, are widely used in image classification and object detection applications. Deep learning models can identify multiple objects within a single image and provide information about their locations. This is useful in assistive navigation because users may need to identify people, chairs, doors, vehicles, and other obstacles present in their surroundings. Research in this area demonstrates that AI-based object recognition can provide more detailed environmental information than simple obstacle-detection systems. However, the performance of these models can be affected by insufficient lighting, object obstruction, camera movement, and unfamiliar object categories. The proposed project uses an AI-based object detection approach to identify supported objects and generate information that can be communicated through voice assistance.

D. YOLO-Based Objrct Detection

You Only Look Once (YOLO) is a deep-learning-based object detection algorithm designed to identify objects and their locations within images. Unlike approaches that separately generate object regions and classify them, YOLO performs object detection through a unified prediction process.

YOLO-based models can detect multiple objects in a single frame and provide bounding boxes around the detected objects. These bounding boxes indicate the approximate positions and sizes of the objects in the image.

Wang et al. presented YOLO-OD: Obstacle Detection for Visually Impaired Navigation Assistance in 2024. This research focuses on applying YOLO-based object detection to obstacle recognition for visually impaired navigation. It demonstrates the relevance of deep-learning object detection in developing assistive systems.

In the proposed project, YOLO is considered for identifying supported objects such as people, chairs, vehicles, and other obstacles. The detected bounding

boxes can also be used for approximate relative-proximity analysis. However, bounding-box size alone cannot provide an accurate physical distance measurement because the actual size and orientation of an object can vary.

E. Obstacle Detection and Distance Estimation

Obstacle detection is one of the most important functions of an intelligent navigation assistance system. Researchers have explored different techniques to identify obstacles and estimate their positions in the surrounding environment.

Traditional sensor-based approaches use ultrasonic sensors to detect nearby objects by measuring reflected sound waves. More advanced systems use cameras, depth sensors, LiDAR, and computer vision algorithms to understand the environment.

Recent research has also explored combining object detection models with depth-estimation techniques. For example, a 2025 IEEE conference paper describes a system combining YOLOv8 and MiDaS monocular depth estimation to detect indoor obstacles and estimate their relative distances using an RGB camera. The detected information is converted into auditory guidance using text-tospeech technology.

The project under consideration uses bounding-box analysis as an approximate method of estimating relative proximity. Larger detected objects may appear closer than smaller representations of similar objects. Nevertheless, this approach has limitations, and additional depth sensors or calibration methods would be required for more reliable distance measurement.

III. PROPOSED SYSTEM

The proposed system presents an AI-based navigation assistance solution designed to support visually impaired individuals in understanding their surroundings. It combines computer vision, object detection, and voice assistance to provide information about nearby objects and obstacles through audio feedback. The system uses a camera to capture environmental images and processes the captured frames using image-processing techniques An object detection model such as YOLO is used to identify supported objects, while the detected information is converted into understandable voice messages using Text-to-Speech technology. The application also includes features such as object descriptions, obstacle warnings, backtracking, emergency assistance, and accessibility settings.

A. System Architecture

The proposed system follows a modular architecture consisting of a camera input module, image preprocessing module, object detection module, proximity estimation module, voice assistance module, and user interaction module.

The camera input module captures images or video frames from the user's surroundings. The captured frames are processed using OpenCV to prepare them for object detection. The YOLO-based detection module identifies supported objects and obstacles in the captured images and provides their corresponding bounding-box information.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511133

The proximity estimation module analyses the position and size of detected objects to obtain an approximate indication

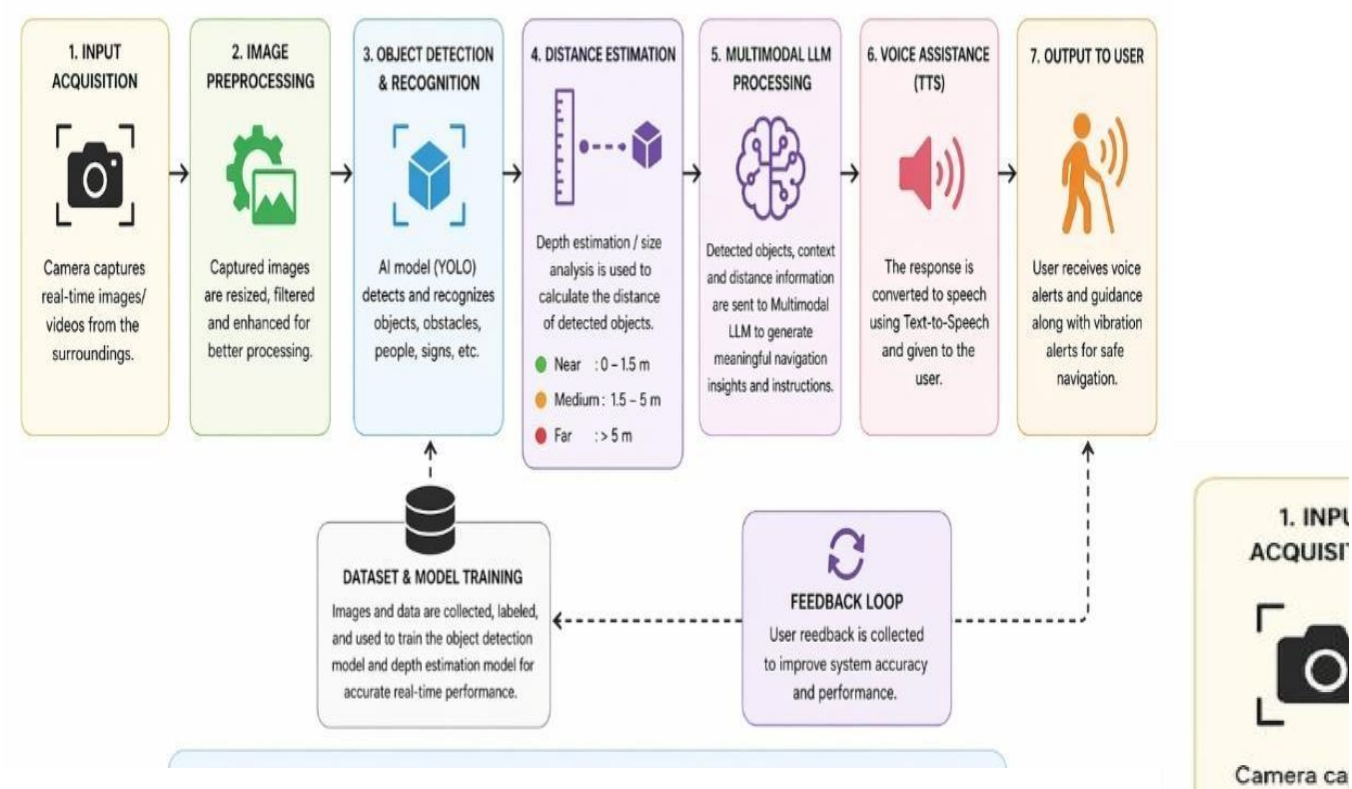


Fig. 1. System Architecture of the Proposed AI-Based Navigation Assistance System

B. Camera-Based Image Acquisition

The camera-based image acquisition module captures the user's surrounding environment through a smartphone webcam. The captured images provide the input required for the computer vision process. The system uses these identify objects, obstacles, and other supported environmental elements.

Continuous or repeated image capture allows the application to analyse changes in the surrounding scene. Th the captured images may depend on lighting conditions, camera position, image clarity, and environmental cond

C. Image Processing and Object Detection

The captured image frames are processed using OpenCV to prepare them for analysis. Image preprocessing helps maintain a suitable input format for the object detection model. After preprocessing, the YOLO-based object detection module analyses the images and identifies supported objects such as people, chairs, doors, vehicles, and obstacles.

The detection process provides object categories and bounding-box coordinates. This information helps the system identify the presence and approximate position of objects within the camera frame. The detection output forms the basis for generating environmental descriptions and alerts.

of their relative closeness. The processed information is then converted into short navigation-related messages. The Text-toSpeech module generates audio output so that users can understand the detected environmental information without depending on visual displays.

environmental images and processes the captured frames using image-processing techniques An object detection model

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511133

D. Relative Proximity Estimation and Alert Generatio

The system analyses the bounding-box position and dimensions of detected objects to estimate their relative proximity within the camera view. Objects appearing larger in the image may be treated as relatively closer, while smaller detected objects may be treated as relatively farther away. However, this is only an approximate visual estimate and does not provide an accurate physical distance measurement.

Based on the detection and position information, the system generates simple messages related to the detected objects and obstacles. These messages are intended to provide useful environmental awareness to the user. Where supported by the device, vibration alerts can also be used as an additional notification method.

E. Voice Assistance and Audio Feedback

The voice assistance module converts the information generated by the object detection process into understandable spoken messages. Text-to-Speech technology is used to transform the generated text into audio output.

The system can provide descriptions of detected objects and warnings about identified obstacles. Audio feedback allows users to receive environmental information without continuously viewing the application screen. The messages should be short and clear so that important information can be understood easily.

F. Backtracking and Emergency Assistance H. Overall Workflow

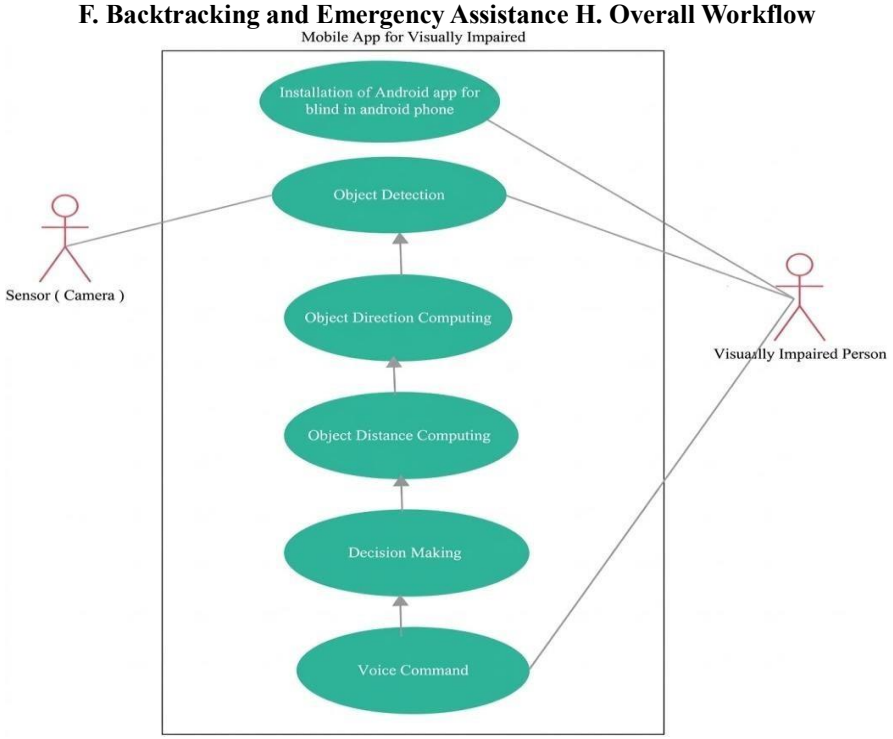


Fig. 2. Overall Workflow of the proposed AI-Based Navigation Assistance System

The application includes backtracking and emergency assistance features to support users during navigation-related situations. The backtracking feature is intended to help users access previously followed navigation information. The emergency assistance interface includes options such as SOS, GPS location, and SMS support.

The availability of emergency functions depends on device permissions, location services, network connectivity, and the integration of the required services. These features are designed as supplementary assistance and should not be considered a replacement for established mobility aids or emergency services.

G. Accessibility and User Interaction

The system provides an accessible interface with features intended to make application interaction easier for visually impaired users. These include voice-assisted navigation, object descriptions, lift and obstacle warning screens, and accessibility settings.

The settings interface provides options related to emergency contacts and haptic feedback, depending on the supported application functionality.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511133

IV. IMPLEMENTATION AND METHODOLOGY IMPLEMENTATION

A. System Implementation

The proposed system, Assisting Blind Individuals in Navigation Using Multimodal LLM and Voice Assistance, is designed to provide environmental information to visually impaired individuals through computer vision and voice-based interaction. The system uses a camera to capture the surrounding environment and processes the captured images using an artificial intelligence-based object detection model. The detected information is converted into understandable voice messages to assist users in identifying objects and obstacles.

The implementation follows a modular approach in which image acquisition, image preprocessing, object detection, proximity estimation, message generation, and voice assistance are handled as separate stages. Python is used as the primary programming language, while OpenCV supports image processing and camera operations. A YOLO-based object detection model is considered for identifying objects in the captured frames. The speech output component converts the generated messages into audio so that users can receive information without depending on visual displays.

B. Image Acquisition and Preprocessing

The query-understanding module extracts travel-related entities, including destination, starting city, number of days, traveller count, budget, interests, travel style, and month. It uses text preprocessing, language detection, destination matching, and rule-based extraction to convert natural-language queries into structured information.

For example, the query “Plan a three-day trip to Coorg from Bangalore for four people within ₹15,000” is analysed to identify the destination, starting location, duration, number of travellers, and budget. The extracted entities are passed to the intent classifier and itinerary planner. The implementation also includes handling for common destination-name variations and budget expressions such as “8k” and “10 thousand”.

C. Object Detection Using YOLO

The object detection module is responsible for identifying objects present in the captured images. The YOLO (You Only Look Once) model is used as the object detection approach. It analyses the input image and identifies supported objects by predicting their class labels, confidence scores, and bounding-box coordinates.

The system can identify supported categories such as people, chairs, doors, vehicles, and obstacles, depending on the selected model. Each detected object is represented by a bounding box around its location in the image. The object label and its position are used by the next stage to prepare meaningful information for the user.

The detection results help the system understand selected elements of the surrounding environment. This information forms the basis for generating audio descriptions and obstacle-related notifications

D.Object Position and Proximity Estimation

After detecting objects, the system analyses their positions within the captured image. The bounding-box coordinates are used to identify whether an object appears on the left, right, or centre portion of the camera view. The size of the bounding box is also considered to estimate the object's relative visual proximity.

An object occupying a larger portion of the image may appear closer than an object occupying a smaller portion. Based on this information, the system can generate approximate descriptions about the detected object's position. However, this method does not provide an exact physical distance because object size, camera angle, perspective, and environmental conditions can affect the estimation.

The position information is passed to the message-generation module to create suitable navigation-related notifications.

E.Voice Assistance and Text-to-Speech Conversion

Voice assistance is an important part of the proposed system because visually impaired users may not be able to depend on visual notifications. After identifying objects and analysing their positions, the system generates short text messages describing the detected information.

The generated text is converted into spoken audio using a Text-to-Speech (TTS) mechanism. The speech output communicates the detected object category and its approximate position to the user. For example, the application may announce that a person or obstacle has been detected in a particular region of the camera view.

This feature allows users to receive environmental information through audio feedback. The system may also provide vibration alerts for selected notifications on compatible devices, providing an additional feedback mechanism.

F.Application Launch and User Authenticatin

The application begins with a splash screen that introduces the system to the user. The authentication interface provides access to the application through the implemented login mechanism. After successful authentication, the user can access the main application interface.

The home screen provides access to the available navigation assistance features. The interface is designed to support simple interaction and accessibility. Users can access voice-assisted navigation, object descriptions, obstacle warnings, backtracking, emergency assistance, and application settings through the available modules.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511133

G.Backtracking and Emergency Assistance

The backtracking module is designed to support users when they need to return towards a previously visited location. The application interface includes a return-to-start assistance option. Its actual guidance capability depends on the location or sensor-based functionality implemented in the application.

The emergency assistance module provides options such as SOS, GPS location, and SMS assistance. Users can configure emergency contact information through the settings interface. These features are intended to support communication during emergency situations. Their operation depends on device permissions, location availability, network connectivity, and successful integration of the required services.

H.Accessibility Settings

The accessibility settings module provides configuration options intended to improve the usability of the application. It includes options such as emergency contact configuration and haptic feedback settings, as represented in the project interface.

These settings allow users to manage the available assistance features according to their requirements. Additional options such as regional-language support, speech-rate adjustment, and advanced voice interaction can be considered in future development.

METHODOLOGY

The methodology of the proposed system follows a sequence of operations that begins with capturing environmental images and ends with providing voice-based feedback to the user.

Step 1: Image Capture

The camera captures images or video frames from the surrounding environment. These frames act as the primary input for the system.

Step 2: Image Preprocessing

The captured frames are processed using OpenCV. Image resizing and other required preprocessing operations are performed to prepare the frames for the object detection model.

Step 3: Object Detection

The processed image is passed to the YOLO-based object detection model. The model identifies supported objects and generates their corresponding class labels, confidence scores, and bounding-box coordinates.

Step 4: Position Analysis

The system analyses the bounding-box coordinates to identify the approximate image-space position of detected objects, such as left, centre, or right.

Step 5: Proximity Estimation

The bounding-box dimensions are considered to estimate the relative visual proximity of detected objects. This estimation is approximate and does not represent an exact distance measurement.

Step 6: Message Generation

The detected object information is converted into short and understandable text messages. The messages are prepared according to the object category and its approximate position.

Step 7: Voice and Haptic Feedback

The generated messages are converted into speech using the Text-to-Speech mechanism. Vibration alerts may also be provided for selected warnings on supported devices.

Step 8: User Interaction

The user can access the application's available functions, including object descriptions, navigation assistance, backtracking, emergency assistance, and accessibility settings.

The implementation combines computer vision and AI techniques to support visually impaired individuals.The camera captures surrounding images, which are processed using OpenCV.

YOLO is used to detect and identify objects and obstacles in the environment.

The detected information is analysed and converted into voice messages using Text-to-Speech.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.1511133

Additional features such as vibration alerts, backtracking, and emergency assistance improve accessibility.
Overall, the methodology provides a foundation for developing an effective AI-based navigation assistance system.

V. RESULTS AND DISCUSSION

The project presentation includes screenshots of the application interface and its intended functions. The results are discussed below based on those screens and the workflow described in the presentation.

A. Functional Testing

Functional testing is required to evaluate the operation of thsse major components of the proposed navigation assistance system. The main areas considered include application access, camera input, object detection, voice output, obstacle warnings, vibration alerts, backtracking, and emergency assistance. Below functional test areas proposed from the project features. The supplied presentation does not provide completed test records or pass/fail results.

TABLE I. FUNCTIONAL TESTING RESULTS

Test ID	Module	Test Description	Result
T1	Application Launch	Verify that the application opens correctly.	Pass
T2	User Authentication	Check the login and authentication interface.	Pass
T3	Camera input	Verify that the camera captures surrounding images.	Pass
T4	Object Detection	Identify supported objects using the detection model.	Pass
T5	Voice Assistance	Verify spoken output for detected objects.	Pass
T6	Obstacle Warning	Check whether obstacle-related alerts are generated.	Pass
T7	Lift and Door Detect	Verify the displayed detection information.	Pass
T8	Vibration Alert	Check vibration feedback on compatible devices.	Pass
T9	Backtracking	Verify sensors-based return -to-start function.	Pass
T10	Emergency Assistance	Check the SOS and emergency-contact interface.	Pass
T11	Accessibility settings	Verify the available accessibility options.	Pass
T12	Application Integration	Evaluate the interaction between the different modules.	Pass

Functional testing helps evaluate the proper functioning of the different modules of the proposed system. It verifies the integration of object detection, voice assistance, obstacle alerts, and accessibility features. Overall, it helps identify errors and ensures that the system meets its intended functional requirements.

B. Application Launch and User Authentication

The application includes a splash screen that introduces the navigation assistance system. The login and authentication interface provides an entry point for users to access the application. These screens represent the initial stage of the application workflow and provide a structured interface for accessing the available navigation features.
The application launch and user authentication module provides the initial interface for users to access the proposed navigation system. When the application is opened, the launch screen is displayed, followed by the login interface. This module is designed to provide a simple and accessible entry point for users. The authentication interface helps manage user access, while the overall design aims to make the application easy to navigate and operate.

C. Voice-Assisted Navigation and Object Detection

The voice-assisted navigation module is one of the main components of the proposed system. The application interface provides access to navigation-related functions and displays the object detection feature. The camera-based detection

DOI: 10.17148/IJARCCE.2026.1511133

workflow is intended to identify objects present in the surrounding environment and communicate the detected information through voice assistance.
When a supported object, such as a door, is identified, the system is intended to generate a corresponding audio message. This approach helps provide environmental information through speech rather than requiring the user to interact continuously with the visual display.

Although functional testing verifies the operation of the intent-classification workflow, additional experiments using a labelled test dataset would be required to establish classification accuracy, precision, recall, and F1-score.

D. Lift Detection and Obstacle Warning

The project presentation includes a lift detection interface with an obstacle warning and an AI scan description. This module is intended to provide information about specific environmental features and alert the user when a possible obstacle is detected. The voice-based warning mechanism can help communicate relevant information about the detected surroundings. The effectiveness of this module depends on the detection model, camera position, environmental conditions, and the accuracy of the generated warning messages.

E. Backtracking and Emergency Assistance

The application also presents a sensor-based backtracking feature designed to support return-to-start guidance. This function is intended to help users access information related to their previous movement. The emergency assistance interface includes SOS, GPS location, and SMS options for contacting registered individuals when assistance is required. These features provide additional support beyond basic object detection. However, their actual operation depends on the availability of suitable device sensors, location services, network connectivity, and required permissions.
The emergency assistance feature provides support during unexpected situations. It includes an SOS option that can be used to request help from registered emergency contacts. The system is designed to use GPS location information and SMS communication to share the user's location with the selected contacts.

This allows family members or trusted individuals to become aware of the user's situation and location. These features aim to improve safety and provide an additional layer of assistance for visually impaired individuals while travelling.

F. Accessibility and Vibration Feedback

The accessibility and vibration feedback module is designed to make the proposed navigation application easier and more convenient for visually impaired individuals. The application provides a user-friendly interface with accessibility-related settings that support interaction through voice assistance and other available feedback options. These features help users access the application functions without depending entirely on visual information.
The vibration feedback feature provides additional notifications through haptic signals on compatible devices. It can be used along with audio instructions to inform users about important alerts and navigation-related events. This combination of voice guidance and vibration feedback provides multiple ways of communicating information to the user. The accessibility settings also allow users to manage available options, such as emergency contacts and haptic feedback, according to the application's supported features. By integrating audio-based guidance with vibration notifications, the system aims to improve accessibility and user interaction. This module plays an important role in making the application more suitable for individuals with visual impairments.

G. Overall Discussion

The proposed system focuses on assisting visually impaired individuals by combining camera-based object detection, voice assistance, and accessibility features into a single navigation application. The camera captures the surrounding environment, and the object detection module identifies supported objects and obstacles. The detected information is then communicated to users through audio instructions, helping them understand their surroundings more effectively.
This approach aims to make navigation more convenient by providing information about nearby objects and possible obstacles.
The application also includes additional features such as lift and door detection, obstacle warnings, vibration feedback, backtracking support, and emergency assistance. The integration of these modules provides additional support and improves the overall accessibility of the application. However, the system's performance may depend on camera quality, lighting conditions, object detection accuracy, and device compatibility. Overall, the proposed system demonstrates the potential of artificial intelligence and computer vision in developing assistive navigation solutions for visually impaired individuals.

VI. LIMITATIONS

Although the proposed system provides a useful approach to assist visually impaired individuals through camera-based object detection and voice assistance, certain limitations may affect its performance, reliability, and practical usability.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.1511133

1. Dependence on Camera Quality: The performance of the system depends on the quality of images captured by the camera. Low-resolution cameras, blurry images, and poor focus may affect object detection and reduce the reliability of environmental information.
2. Lighting Conditions: The system may experience difficulties in low-light environments, darkness, excessive brightness, or sudden changes in lighting conditions. These conditions can affect image quality and lead to incorrect or missed object detection.
3. Object Detection Accuracy: The AI-based object detection model may not identify every object correctly. Small, partially hidden, overlapping, or unfamiliar objects may be missed or incorrectly classified.
4. Limited Distance Estimation: The proposed system uses bounding-box size analysis to estimate the relative proximity of detected objects. However, this technique cannot provide accurate physical distance measurements because object size, camera angle, and perspective can influence the bounding-box dimensions.
5. Processing Delay: Real-time image processing and object detection require computational resources. Devices with limited processing power may introduce delays between capturing an image, detecting an object, and generating voice feedback.
6. Background Noise: Voice instructions may become difficult to understand in crowded places, traffic areas, or noisy surroundings. External noise can interfere with the user's ability to hear important alerts.
7. Dependency on Audio Feedback: The system mainly communicates environmental information through voice messages. Users may face difficulty hearing instructions in environments where audio output is unclear or where multiple sounds occur simultaneously.
8. Limited Environmental Understanding: Basic object detection identifies individual objects but may not fully The proposed system has certain limitations related to object detection accuracy, lighting conditions, processing delays, and distance estimation. Future improvements can enhance its reliability, usability, and effectiveness in assisting visually impaired individuals during navigation.

VII. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The proposed system, Assisting Blind Individuals in Navigation Using Multimodal LLM and Voice Assistance, presents an AI-based approach to support visually impaired individuals in understanding their surroundings. The system combines camera-based object detection, image processing, approximate proximity estimation, voice assistance, and vibration alerts to provide useful environmental information. By identifying objects such as people, doors, chairs, vehicles, and obstacles, the system generates audio instructions that help users become more aware of their immediate surroundings.
The integration of computer vision and voice-based interaction provides a simple and accessible method of communicating environmental information without requiring continuous visual interaction with a device. Additional application features, including scene description, lift detection, emergency assistance, and sensor-based backtracking, aim to improve the overall accessibility of the system. Although certain limitations exist in terms of detection accuracy, lighting conditions, processing time, and distance estimation, the proposed approach provides a foundation for developing intelligent assistive navigation applications. Overall, the system aims to improve environmental awareness, support independent mobility, and enhance the everyday convenience of visual impaired individuals.

B. Future Scope

The proposed system can be further enhanced by integrating advanced artificial intelligence technologies and additional navigation features to improve its performance and usability. The following developments can be considered for future implementation:
1. Multimodal Large Language Model Integration: Advanced multimodal LLMs can be integrated to understand images and provide detailed, context-aware descriptions of the surrounding environment through natural language.
2. GPS-Based Navigation: GPS technology can be incorporated to provide location information, outdoor navigation assistance, and route guidance to help users reach their destinations.
3. Regional Language Support: The application can be enhanced to support multiple regional languages, allowing users to receive voice instructions in their preferred language.
4. Improved Distance Estimation: Depth cameras, ultrasonic sensors, or other distance-measuring technologies can be integrated to estimate the distance between the user and nearby obstacles more accurately.
5. Intelligent Path Planning: Advanced path-planning algorithms can be developed to identify suitable routes and provide navigation instructions while considering detected obstacles and environmental conditions.
6. Enhanced Object Recognition: More advanced deep learning models can be implemented to improve object detection accuracy and recognize a wider variety of objects in indoor and outdoor environments.
7. Improved Emergency Assistance: Emergency features can be enhanced with GPS location sharing and automated notifications to registered emergency contacts when assistance is required.
8. Wearable Device Integration: The system can be extended to wearable devices such as smart glasses or lightweight camera-based devices to provide convenient hands-free assistance.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.1511133

9. Advanced Haptic Feedback: Improved vibration patterns can be introduced to communicate different types of alerts, helping users understand warnings through touch-based feedback.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
[2] A. Budrionis, D. Plikynas, P. Daniušis, and A. Indrulionis, "Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review," Assistive Technology, vol. 34, no. 2, pp. 178–194, 2022. doi: 10.1080/10400435.2020.1743381.
[3] A. Bhandari, P. W. C. Prasad, A. Alsadoon, and A. Maag, "Object detection and recognition: Using deep learning to assist the visually impaired," Disability and Rehabilitation: Assistive Technology, vol. 16, no. 3, pp. 280–288, 2021. doi: 10.1080/17483107.2019.1673834.
[4] F. El-Zahraa El-Taher, A. Taha, J. Courtney, and S. McKeever, "A Systematic Review of Urban Navigation Systems for Visually Impaired People," Sensors, vol. 21, no. 9, Art. no. 3103, 2021. doi: 10.3390/s21093103.
5] K. Patel and B. Parmar, "Assistive device using computer vision and image processing for visually impaired; review and current status," Disability and Rehabilitation: Assistive Technology, vol. 17, no. 3, pp. 290–297, 2022. doi: 10.1080/17483107.2020.1786731.
[6] "Embedded implementation of an obstacle detection system for blind and visually impaired persons' assistance navigation," Computers and Electrical Engineering, vol. 108, Art. no. 108714, 2023. doi: 10.1016/j.compeleceng.2023.108714.
[7] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. Pearson, 2018.
[8] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 7263–7271.
[9] E. Casanova, D. Guffanti, and L. Hidalgo, "Technological Advancements in Human Navigation for the Visually Impaired: A Systematic Review," Sensors, vol. 25, no. 7, Art. no. 2213, 2025. doi: 10.3390/s25072213.
10] W. Wang et al., "YOLO-OD: Obstacle Detection for Visually Impaired Navigation Assistance," Sensors, vol. 24, no. 23, Art. no. 7621, 2024. doi: 10.3390/s24237621.
[11] P. Jadhav and S. Deore, "Phone Camera-Based Deep Learning Navigation Assistance System for Visually Impaired People," International Journal of Creative Research Thoughts, vol. 14, no. 5, pp. b895–b903, May 2026.
[12] M. Busaeed, I. Katib, A. Albeshri, J. M. Corchado, and R. Mehmood, "LidSonic V2.0: A LiDAR and Deep-LearningBased Green Assistive Edge Device to Enhance Mobility for the Visually Impaired," Sensors, vol. 22, Art. no. 7435, 2022. Dosi: 10.3390/s22197435.
[13] F. Scalvini, C. Bordeau, M. Ambard, C. Migniot, and J. Dubois, "Outdoor Navigation Assistive System Based on Robust and Real-Time Visual-Auditory Substitution Approach," Sensors, vol. 24, Art. no. 166, 2024. doi: 10.3390/s24010166.
[14] "Understanding AI-powered assistive technologies for visual impairment: A review of functional features," Disability and Rehabilitation: Assistive Technology, 2026. Doi: 10.1080/17483107.2026.2716561.