IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

AI Powered Visual Assistance for Visually Challenged using YOLOv3 Algorithm

Kavana N[1], Lakshmitha R[2], Ganavi H R[3], Nisarga T K[4]
Studets, B.E, Computer Science and Engineering[1-4]
Kalpataru Institute of Technology, Tiptur, India

Abstract: *Visual impairment significantly limits independent access to information and daily activities for millions of individuals worldwide. Traditional assistive tools such as Braille are gradually becoming less practical, while commercial screen-reading software remains costly and inaccessible to many users. To address these challenges, this research presents an AI-based vision and voice assistant designed specifically to support visually impaired individuals through real-time, voice-driven interaction.*

The proposed system integrates machine learning, computer vision, and natural language processing to deliver intelligent assistance using a camera and microphone for environmental perception. Core functionalities include object detection, scene understanding, obstacle awareness, text-to-speech conversion, and personalized voice guidance. Visual data captured by the camera is processed through a structured pipeline consisting of image acquisition, pre-processing, object detection, and classification, followed by natural language generation for meaningful auditory feedback. Pre-trained deep learning models such as YOLO, SSD, or Faster R-CNN are employed and fine-tuned to enhance accuracy and performance in diverse real-world environments.

Speech synthesis enables clear and natural feedback, while optional speech recognition allows users to issue voice commands for seamless interaction. The system is designed with an emphasis on accuracy, speed, affordability, and ease of use, making it suitable for everyday navigation and smartphone-based applications. By improving situational awareness and accessibility, the proposed AI vision assistant aims to enhance independence, safety, and overall quality of life for visually impaired individuals.

Keywords*:* Visual Impairment, AI Vision Assistant, Assistive Technology, Computer Vision, Object Detection, Voice Assistant, Speech-to-Text, Text-to-Speech, Deep Learning, YOLO, Scene Understanding, Obstacle Detection, Natural Language Processing, Accessibility Technology

I. INTRODUCTION

Visual impairment significantly restricts an individual's ability to perceive surroundings, identify objects, and navigate safely in daily life. Traditional assistive tools such as white canes and Braille provide limited environmental awareness and lack real-time contextual information. Recent advancements in artificial intelligence, computer vision, and deep learning have enabled the development of intelligent assistive systems capable of interpreting visual data and converting it into meaningful auditory feedback. This paper presents an AI-powered vision assistance system for visually challenged individuals using the YOLOv3 object detection algorithm. The system captures real-time video through a camera, detects and classifies surrounding objects using a deep learning model trained on the COCO dataset, and delivers audio feedback via text-to-speech synthesis. Designed to be accurate, fast, and cost-effective, the proposed system enhances situational awareness, mobility, and independence for visually impaired users in real-world environments.

II. PROJECT OBJECTIVE

- To develop an AI-powered assistive system for visually impaired individuals using YOLOv3.
- To detect and recognize surrounding objects in real time from live video input.
- To provide clear audio feedback through text-to-speech conversion.

DOI: 10.48175/IJARSCT-30585

ISSN 2581-9429 IJARSCT

- To ensure fast, accurate, and offline performance on low-cost hardware.
- To improve environmental awareness and independent navigation for visually challenged users.

III. SCOPE OF THE PROJECT

The scope of this study is limited to real-time object detection and audio feedback for visually impaired users. The system focuses on detecting commonly encountered objects in indoor and outdoor environments using a YOLOv3 model trained on the COCO dataset. The proposed framework operates offline and is suitable for deployment on portable and low-cost hardware platforms. While the current implementation emphasizes object recognition and speech output, the system architecture is scalable and can be extended to include advanced features such as distance estimation, object tracking, navigation assistance, and integration with newer deep learning models in future work.

IV. OUT-OF-SCOPE

- Autonomous navigation and path planning for visually impaired users.
- Distance measurement, depth estimation, or 3D environment mapping using LiDAR or stereo vision.
- Advanced scene understanding, semantic segmentation, or full environmental description.
- Facial recognition, emotion recognition, or person identification features.
- Integration with GPS, maps, or real-time navigation services.

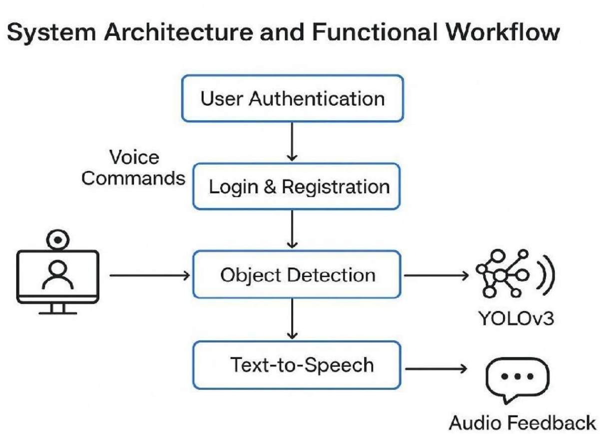
V. PROJECT CONTEXT AND STRATEGIC IMPERATIVE

Assistive technology for visually impaired individuals is a strategic priority in building inclusive and accessible digital societies. Increasing urban complexity, dense environments, and reliance on visual information create significant challenges for visually challenged individuals in daily navigation and object recognition. Conventional assistive tools provide limited situational awareness and do not scale effectively to dynamic real-world conditions.
AI-based vision assistance systems offer a strategic advantage by enabling:

- Real-Time Environmental Awareness: Detect and identify surrounding objects instantly through computer vision.
- Enhanced Safety: Reduce risks from obstacles and moving objects in indoor and outdoor environments.
- Cost-Effective Accessibility: Provide affordable alternatives to expensive commercial assistive devices.
- Offline and Portable Operation: Ensure usability without internet dependency across diverse settings.
- Digital Inclusion: Support equitable access to smart technologies for visually impaired users.

This project aligns with global initiatives in AI-driven accessibility, inclusive design, and smart assistive systems, demonstrating the role of deep learning and computer vision in enhancing independence, safety, and quality of life for visually impaired individuals.

VI. METHODOLOGY



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

Step 1: User Authentication
The system begins with user authentication to ensure secure and personalized access. Users can interact with the system using both manual input and voice commands. Authentication helps maintain user-specific settings such as preferred language, speech rate, and detection mode. Voice-guided prompts assist visually challenged users throughout the authentication process.

Step 2: Login and Registration Using Voice Commands
New users can register by providing personal details such as name and credentials through voice input or keyboard entry. Existing users can log in using spoken commands or typed credentials. Speech recognition technology converts voice commands into text, enabling hands-free interaction. Upon successful login, the system grants access to the object detection module.

Step 3: Image Capture and Object Detection
After authentication, the system activates the camera to capture live video frames from the user's surroundings. Each frame is pre-processed and passed to the YOLOv3 deep learning model for real- time object detection. YOLOv3 detects multiple objects simultaneously by predicting bounding boxes, class labels, and confidence scores in a single forward pass. Non-Maximum Suppression (NMS) is applied to remove duplicate detections and improve accuracy.

Step 4: Object Classification Using YOLOv3
The YOLOv3 model, pre-trained on the COCO dataset, identifies common objects such as people, vehicles, furniture, doors, and obstacles. The system analyses the spatial position of detected objects to determine their relative direction (front, left, or right). Priority is given to critical objects to ensure user safety.

Step 5: Text-to-Speech Conversion
Once objects are detected and classified, the corresponding object names and directional information are converted into text messages. A Text-to-Speech (TTS) engine processes this information and generates natural-sounding speech output. The speech parameters are optimized for clarity and accessibility.

Step 6: Audio Feedback to the User
The generated audio output is delivered to the user through speakers or earphones. Feedback such as "Person detected in front of you" or "Chair on your left" helps the user understand their surroundings in real time. This continuous audio feedback enables safe navigation and enhances situational awareness.

Step 7: Real-Time Operation and System Continuity
The entire workflow operates continuously in real time. The system updates detections frame by frame and avoids repetitive announcements using cooldown mechanisms. This ensures smooth interaction, minimal latency, and a user-friendly experience suitable for visually challenged individuals.

Outcome
By integrating authentication, YOLOv3-based object detection, and text-to-speech feedback into a single workflow, the proposed methodology effectively converts visual information into audible guidance, improving independence, safety, and mobility for visually impaired users.

DOI: 10.48175/IJARSCT-30585

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

VII. RESULTS AND DISCUSSION

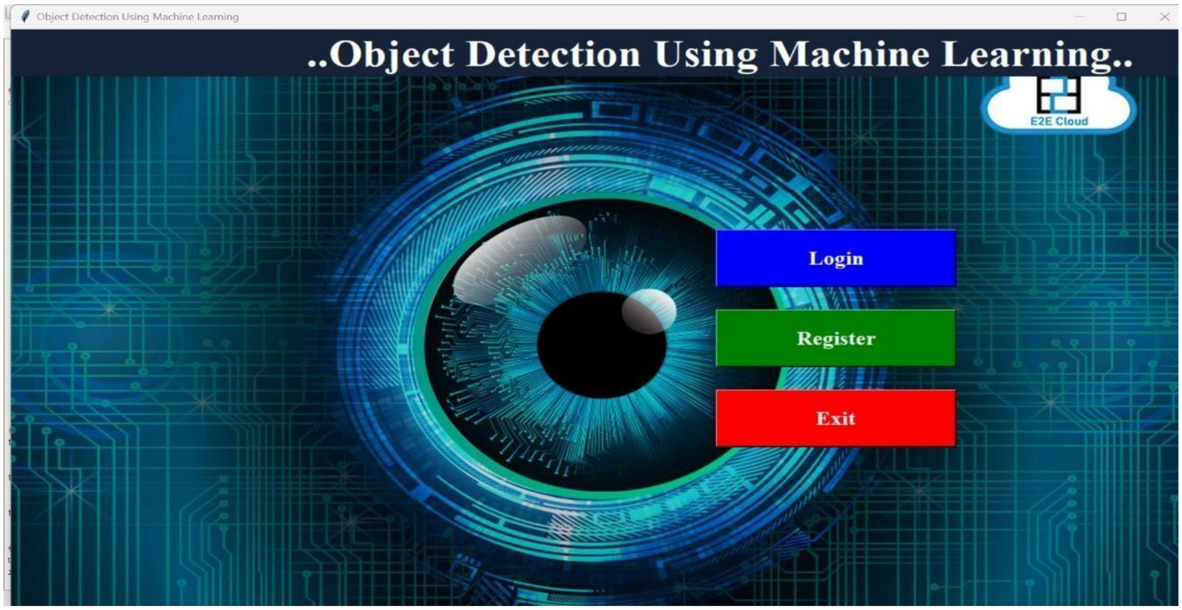


Fig 1: User Dashboard

System Dashboard Overview

• When the program runs:

• This main window appears first.

The user selects an option:

• Login/Register → proceeds to authentication.

• Exit → closes the app.

After login, the system moves to the object detection interface, where: This main window appears first.

The user selects an option:

• Login/Register → proceeds to authentication.

• Exit → closes the app.

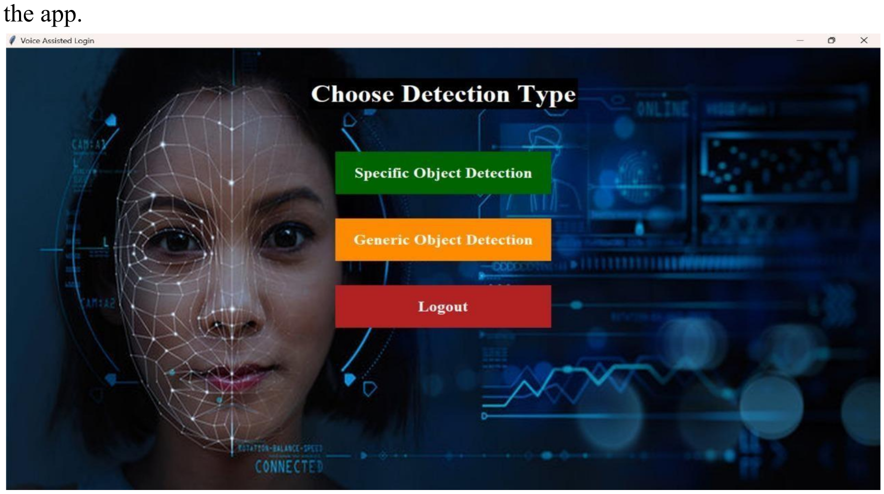


Fig 2: Detected Objects Display Interface

DOI: 10.48175/IJARSCT-30585

Interface Overview : This interface allows the authenticated user to choose between different object detection modes, enabling flexible interaction based on their needs. The title “Choose Detection Type” guides the user to select the preferred detection function.

Options Provided:
Specific Object Detection: Activates the mode to detect a particular object (e.g., “Find chair”).
1. Generic Object Detection: Enables the system to identify and announce all visible objects in the camera’s view.
2. Logout: Safely ends the session and returns the user to the login interface.

Design:
The background features a futuristic AI interface with facial mapping graphics, highlighting the system’s advanced computer vision and deep learning capabilities.

Purpose:
This screen acts as the central navigation hub for users, offering an intuitive and accessible way to start specific or generic object detection operations after successful authentication.
The title “Choose Detection Type” clearly guides the user to select the preferred detection function.

Options Provided:
3. Specific Object Detection: Activates the mode to detect a particular object (e.g., “Find chair”).
4. Generic Object Detection: Enables the system to identify and announce all visible objects in the camera’s view.
5. Logout: Safely ends the session and returns the user to the login interface.

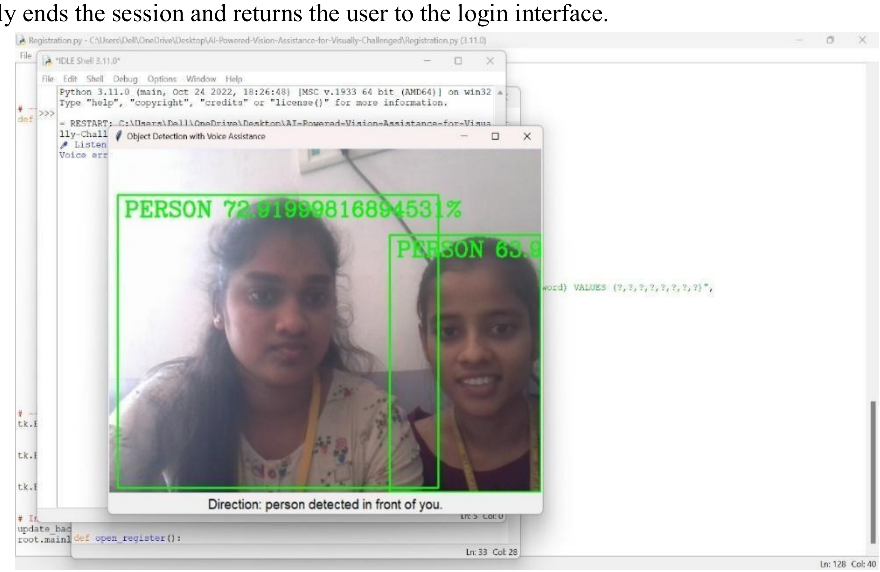


Figure 3: Specific Object Detection

The figure above shows the real-time output of the AI-Powered Visual Assistance System during object detection using the YOLOv3 algorithm.
The system captures live video input through the connected camera and processes each frame to identify visible objects. In this instance, the system successfully detects two persons in the camera frame. Each detected object is enclosed within a green bounding box, and the label “PERSON” along with the confidence score (e.g., 72.91% and 63.9%) is displayed at the top of each bounding box, indicating the model’s certainty of its prediction. At the bottom of the window, a directional message — “Direction: person detected in front of you.is generated and displayed, while the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, December 2025
Impact Factor: 7.67

same message is audibly announced through the text-to-speech module. This confirms that: The YOLOv3 model is correctly identifying human objects in real time.

The voice assistance module is functioning as intended, providing audio feedback to the user. The system is capable of handling multiple object detections simultaneously with high accuracy.

Overall, this output validates the system's functionality in providing real-time visual interpretation and voice-based feedback, demonstrating its potential as an assistive tool for visually impaired users.

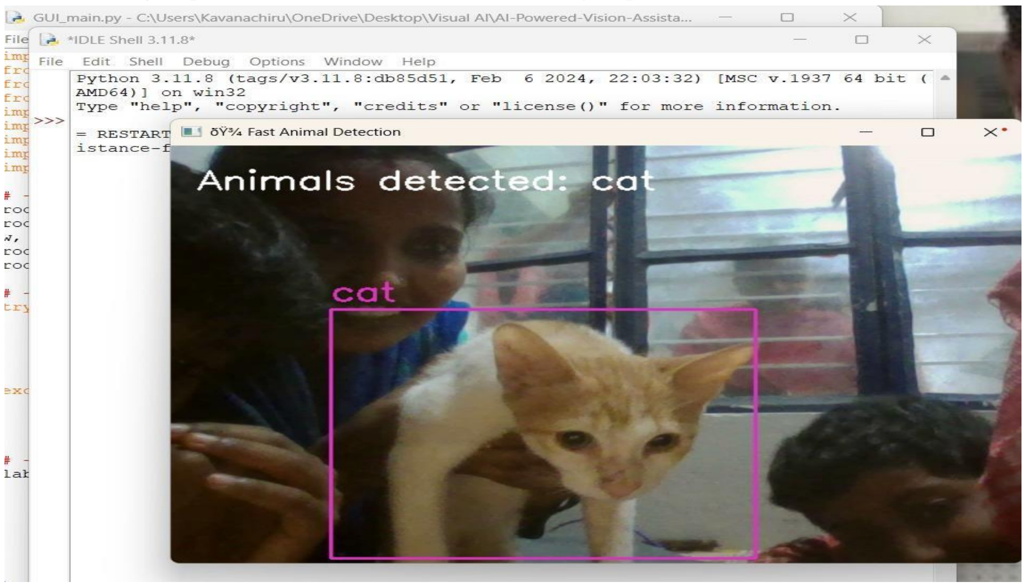


Figure 4: Generic Object Detection

The image shows a computer application window titled "Fast Animal Detection." Inside the window, a live camera feed displays two young women sitting next to each other, looking toward the camera. Both appear to be indoors, possibly in a classroom or lab environment. Text in large red letters at the top-left of the camera feed says "No animals detected." This indicates that the animal detection system is running, analysing the live video feed, and currently not identifying any animals within the frame. The rest of the desktop background displays the Python editor and parts of the project script in the background.

VIII. CONCLUSION

AI and Machine Learning enhance railway safety by enabling automated detection of track defects and obstacles with high accuracy. These intelligent monitoring systems reduce inspection costs, improve reliability, and provide early warnings to prevent accidents. By integrating computer vision, deep learning, and sensor technologies, the system supports continuous surveillance and predictive maintenance, contributing to smarter, safer, and more efficient railway infrastructure.

REFERENCES

[1] A. M. George, A. Ramachandran, M. C. M., M. A. T., and B. A. R., "YOLO-based object recognition system for visually impaired," International Journal of Science and Engineering Applications, vol. 14, no. 1, pp. 34–42, Jan. 2025, doi: 10.7753/IJSEA1401.1009.

[2]A. A. Salokhe, D. Kumawat, S. Potdar, V. Patil, K. Gavali, and S. Chindarkar, "AI-powered assistive system for the visually impaired," International Journal of Creative Research Thoughts (IJCRT), vol. 13, no. 4, Apr. 2025, ISSN: 2320-2882.

[3] S. Surendarkumar, M. Ranjithkumar, B. Deepanraj, M. Sivashankar, and M. Umapathy, "Blind people assistance for object detection using AI," International Journal of Innovative Research in Engineering, vol. 4, no. 3, pp. 222–226, May–Jun. 2023, doi: 10.59256/ijire.2023040378.

[4] J. Wang, S. Wang, and Y. Zhang, "Artificial intelligence for visually impaired," Displays, vol. 77, Art. no. 102391, Feb. 2023, doi: 10.1016/j.displa.2023.102391.

[5] S. A. Hande and P. B. Bilawar, "Digital voice assistant for visually impaired users," International Journal of Advance and Applied Research, vol. 10, no. 1, pp. 595–600, Sep.–Oct. 2022, doi: 10.5281/zenodo.7219852.

[6] M. Pranav, M. E. Anand, and A. S. Kumar, "An intelligent assistant for visually impaired people," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 6, pp. 1042–1046, Jun. 2020.

[7]. A. K. Nair, R. Pradhan, and M. V. P. Siva Kumar, "A Survey on Artificial Intelligence-Based Solutions for Visually Impaired People," 2021 International Conference on Artificial Intelligence and Applications (ICAIA), Chennai, India, 2021.

[8]. M. A. Ahmed, M. A. Almansob, and S. J. Hussain, "An Artificial Intelligence-Based System for Visually Impaired People," 2021 International Conference on Advanced Science and Engineering (ICOASE), Sana'a, Yemen, 2021.

[9]. S. K. Gupta, A. H. J. J. Prasad, and P. K. Sahu, "Smart Phone Based System for the Visually Impaired People using Artificial Intelligence," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (UP section), Noida, India, 2020.

[10]. R. Bharti, A. K. Singh, and S. K. Singh, "An Artificial Intelligence-Based System for Visually Impaired People," 2019 IEEE 2nd International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 2019.

[11]. A. Almomani, R. Almomani, and K. Y. Quwaider, "A Review on Computer Vision-Based Assistive Technology for Visually Impaired People," International Journal of Advanced Computer Science and Applications, 2019.