International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 14, Issue 1, 2026

Voice Based Assistance For Blind People

B Jyothsna[1], P Shruthi[2], R Sirisha[3], V Sravya Sree[4]
[1]Associate Professor Bhoj Reddy Engineering College for Women Department of Electronics and Communication Engineering, Hyderabad, India.
[2,3,4]B Tech Students Bhoj Reddy Engineering College for Women Department of Electronics and Communication Engineering, Hyderabad, India.
rayabandisirisha@gmail.com, sravyavuligay@gmail.com, shruthipeddolla10@gmail.com

Abstract
Access to digital information and real-world services remains a major challenge for people with visual impairments, particularly in environments that depend heavily on graphical user interfaces. This paper presents the design and development of a low-cost voice-based assistance system intended to support visually impaired users in performing routine daily activities without visual interaction. The proposed system enables users to interact with the device using spoken commands and receive auditory responses through an integrated speech interface. Core functionalities include printed text reading using optical character recognition, voice-driven information retrieval such as time and weather updates, and basic environmental interaction support.
The system integrates speech recognition, text-to-speech synthesis, and optical character recognition modules to provide real-time, hands-free operation. Emphasis is placed on affordability, simplicity of deployment, and usability for non-technical users. Experimental evaluation demonstrates that the proposed solution delivers reliable response times and accurate recognition performance in typical indoor environments. The results indicate that voice-driven assistive technologies can significantly enhance personal independence and digital inclusion for visually impaired individuals.

Keywords: *Assistive technology, voice interface, speech recognition, text-to-speech, optical character recognition, accessibility.*

1 Introduction

Visual impairment considerably restricts an individual's ability to perceive the surrounding environment, access information, and perform everyday activities independently. Although conventional mobility aids such as walking canes and guide dogs provide essential support, they are limited in offering real-time awareness of dynamic objects and obstacles present in complex environments.

Recent advances in artificial intelligence, computer vision, and speech technologies have created new opportunities to design intelligent assistive systems that extend human perception through auditory feedback. Voice-driven interfaces, in particular, enable visually impaired users to interact with digital systems naturally, without relying on visual displays or manual input methods.

This project presents the design and implementation of a voice-based assistance system intended to support visually impaired users by providing real-time object awareness and spoken feedback. The system captures live video through a camera and processes the incoming frames using a deep-learning-based object detection model. Detected objects are converted into meaningful voice descriptions, allowing the user to understand nearby elements in the environment instantly.

Image processing operations are carried out using computer vision techniques, while a lightweight deep learning detector trained on a large-scale object dataset enables recognition of common real-world objects. A text-to-speech module generates clear and continuous audio notifications, ensuring that the user receives timely information without visual interaction.

The proposed solution is designed to be portable and computationally efficient, making it suitable for deployment on low-cost edge devices such as single-board computers and smartphones. The system can be used as a wearable or handheld assistive tool, supporting safe navigation in indoor and outdoor environments.

By transforming visual information into speech, the project demonstrates how artificial intelligence can bridge the gap between machine perception and human auditory understanding. The platform also provides a foundation for future extensions such as location-based navigation, customizable object categories, and multilingual speech support.

The primary aim of this project is to develop a real-time voice-based assistive system that improves the mobility and safety of visually impaired individuals by detecting nearby objects and obstacles and delivering immediate audio feedback.

The system continuously analyzes live camera input using machine learning techniques and informs the user about relevant objects in the surrounding environment through synthesized speech. The proposed solution focuses on portability, low computational overhead, and ease of use, enabling visually impaired users to move more confidently and independently.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

The main objectives of the proposed system are as follows:

- To design a real-time object detection framework capable of identifying commonly encountered objects in everyday environments.
- To convert visual scene information into meaningful and understandable speech feedback.
- To create a hands-free interaction mechanism that allows visually impaired users to receive assistance without relying on visual displays or manual controls.
- To ensure that the system operates efficiently on low-cost and portable hardware platforms.
- To enhance personal independence and situational awareness during daily navigation and exploration.

2 Literature Survey

Recent progress in artificial intelligence and embedded vision has enabled the development of intelligent assistive systems aimed at improving environmental awareness and independence for people with visual impairments. Research efforts increasingly focus on combining real-time computer vision, speech processing, and portable hardware to deliver context-aware auditory feedback.

One of the most well-known commercial initiatives in this area is the accessibility application developed by **Microsoft**, which employs a smartphone camera to recognize objects, read printed text, and describe surrounding scenes through speech output. Similarly, wearable solutions introduced by **OrCam Technologies** provide camera-based visual perception and real-time audio feedback by mounting a miniature vision unit on eyeglass frames. Although these products demonstrate high practical value, their high cost and dependence on continuous connectivity restrict their adoption in economically constrained environments.

From an academic perspective, numerous studies have examined the use of lightweight deep learning models for object detection on embedded and edge platforms. In particular, single-stage detectors and compact convolutional neural network architectures have been widely explored because of their reduced computational requirements and faster inference speed. Among these approaches, mobile-oriented neural models combined with single-shot detection strategies offer a suitable compromise between detection accuracy and processing latency, making them well suited for assistive applications that demand real-time response.

Open-source development frameworks have further accelerated research in this domain. Computer vision libraries such as OpenCV, along with mobile-optimized inference engines, enable researchers to implement vision pipelines on resource-limited devices. Several studies integrate object recognition modules with offline and lightweight speech synthesis engines to provide spoken descriptions of detected objects and surroundings.

Despite these advances, existing systems often suffer from at least one of the following limitations: high hardware cost, restricted portability, insufficient real-time performance, or dependence on cloud services. The literature highlights a continuing need for compact, offline, and affordable solutions that can be deployed on low-power edge platforms. The proposed work directly addresses this gap by adopting efficient neural models and offline text-to-speech processing on portable devices such as the **Raspberry Pi Foundation** platform.

3. Software Requirements

This section outlines the minimum software resources required to design, implement, and execute the proposed voice-based assistance system. All selected tools are open-source or freely available, which supports the objective of building a low-cost and easily reproducible assistive solution.

The essential software components are summarized below:

- **Programming Language:** Python
- **Dataset Format:** CSV (for configuration, logging, and experimental records)
- **Development Environment:** Anaconda distribution with Python IDE support
- **Operating System:** Windows 10

The above configuration is sufficient to support computer vision processing, machine learning inference, speech synthesis, and system integration required by the proposed application.

3.2.1 Python Programming Language

Python is a high-level, object-oriented programming language widely adopted in scientific computing, artificial intelligence, and application development. It is particularly suitable for rapid prototyping and experimental system development due to its simple syntax, extensive standard library, and strong community support.

In the proposed voice-based assistance system, Python serves as the primary implementation language for image processing, object detection, audio processing, and system control. The availability of mature libraries for computer vision, deep learning, and speech technologies enables efficient integration of multiple functional modules within a single application framework.

Key characteristics that make Python suitable for this project include:

- Clear and readable syntax that simplifies development and debugging.
- Cross-platform compatibility, allowing the same code base to be deployed on different operating systems and hardware platforms.

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

- Strong support for scientific and artificial intelligence libraries required for object detection and speech processing.
- High flexibility for integrating external libraries and hardware interfaces.

Python also provides robust support for modular programming, enabling the system to be structured into independent components such as video acquisition, object detection, speech generation, and user interaction.

Anaconda Installation on Windows

To prepare the development environment, the following steps are performed:

1. Download the Anaconda distribution for Windows from the official source.
2. Install the distribution using default configuration settings.
3. Create a dedicated virtual environment for the project.
4. Install the required Python libraries for computer vision, machine learning, and speech synthesis within the created environment.
5. Configure the development environment to access the system camera and audio devices.

1. Click on the link below to open the download page

https://www.anaconda.com/download/#windows

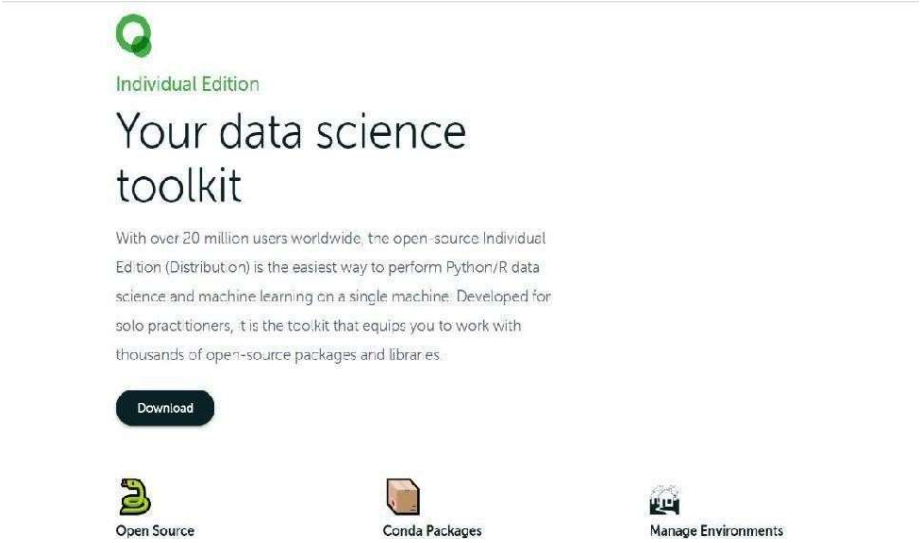


Fig 1 Anaconda Individual Edition

Click on the download button and check for the compatibility of your system. Then, it will start downloading.

Fig 2 Anaconda Installers for Windows, macOS, and Linux

2. Double click the installer to launch

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

3. Click on Next.

Fig 3 Installation setup window

4. Read the License Agreement and then click on I Agree

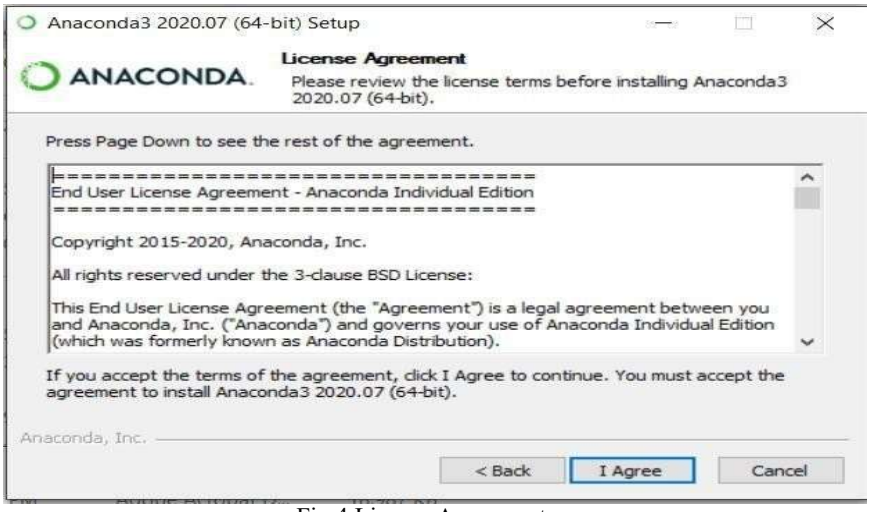


Fig 4 License Agreement

5. Select installation type “Just Me” unless you’re installing it for all users (which require Windows Administrator privileges) and click on Next.

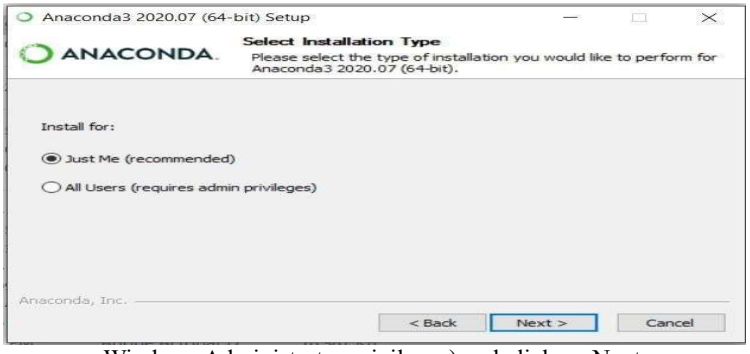


Fig 5 Installation Type Selection Window in Anaconda3 2020.07 Setup

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

6. Select a destination folder to install Anaconda and click the Next button.

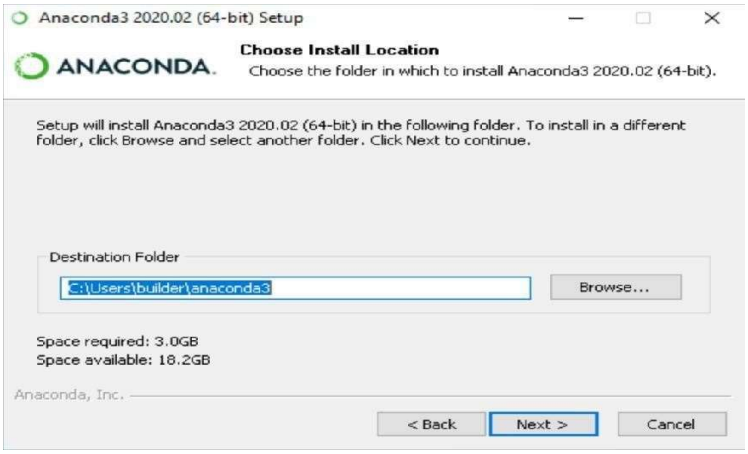
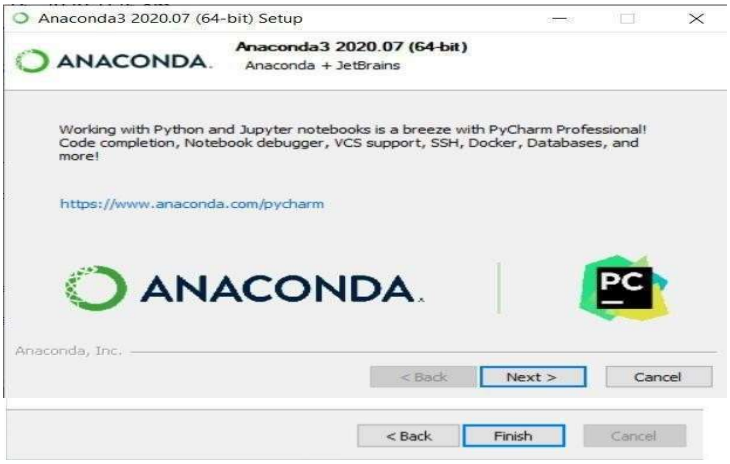


Fig 6 Choose Install Location Window in Anaconda3 2020.02 Setup

Fig 7 Choose Install Location Window in Anaconda3 2020.02 Setup



8. And then click the Finish button.

Fig 3.8 Completing Anaconda3 2020.02 (64-bit) Setup Window

After a successful installation you will see the "Thanks for installing Anaconda" dialog box.

TensorFlow API

The TensorFlow API is an open-source machine learning framework developed by Google that allows developers and researchers to build, train, and deploy machine learning models easily. It supports a wide range of applications, including image classification, natural language processing, and object detection. One of its most powerful components is the TensorFlow Object Detection API, which provides a streamlined way to train object detection models with minimal code. This API comes with a variety of pre-trained models, configuration files, and tools for evaluating and exporting trained models. It supports both custom datasets and popular ones like COCO, and integrates well with other TensorFlow tools for preprocessing, model optimization, and deployment.

SSD (Single Shot MultiBox Detector)

The SSD (Single Shot MultiBox Detector) algorithm is a real-time object detection method widely used for its balance between speed and accuracy. Unlike two-stage detectors like Faster R-CNN, SSD performs object localization and classification in a single forward pass of the network, making it highly efficient for applications requiring fast detection. It uses multiple feature maps to detect objects at different scales, allowing it to detect both large and small objects effectively. TensorFlow provides several pre-trained SSD models, such as ssd_mobilenet_v2, which are commonly used in edge devices due to their lightweight nature and fast inference speeds.

COCO (Common Objects in Context)

The COCO (Common Objects in Context) dataset is a large-scale benchmark dataset used for training and evaluating object detection algorithms. It contains over 330,000 images and more than 1.5 million object instances annotated across 80 different categories, such as people,vehicles, animals, and household items. What sets COCO apart is that it includes objects in their natural

context, often with occlusions and complex backgrounds, making it ideal for building robust detection systems. The dataset is widely used in the research community and is supported out of the box by the TensorFlow Object Detection API, making it convenient to use for training and testing models like SSD.

Block Diagram and its Working

Visually impaired individuals face significant challenges in navigating their surroundings and accessing information independently. With the advancement of technology, particularly in the field of artificial intelligence and machine learning, it is now possible to create intelligent
systems that assist the blind in their daily activities. This project, Voice-Based Assistance for Blind People using Machine Learning, aims to develop a smart, user-friendly system that acts as a personal assistant through voice interaction.

The system leverages machine learning algorithms, speech recognition, and text-to-speech
(TTS) technologies to understand user commands and provide verbal feedback. It is designed to help users perform tasks such as reading text from images, recognizing objects, identifying currency, or navigating environments, all through voice-based communication. By integrating computer vision techniques with audio processing, the system can interpret visual information and relay it audibly to the user.

This project not only enhances independence and safety for the visually impaired but also demonstrates how machine learning can be applied to improve accessibility and inclusion. The ultimate goal is to bridge the gap between technology and disability by creating a low- cost, reliable, and effective assistive solution.

Based on the recognized command, the system performs different tasks. For example, if the user asks to read text from an image, the system activates the camera module, captures the image, and applies Optical Character Recognition (OCR) to extract any readable text. The extracted text is then converted into audio using a Text-to-Speech (TTS) engine and played back to the user. If the user requests object detection, the camera captures the scene, and a pre-trained deep learning object detection model (such as SSD or YOLO) is used to identify and label objects in the image.

The names and positions of the detected objects are then described to the user via voice output

This project not only enhances independence and safety for the visually impaired but also demonstrates how machine learning can be applied to improve accessibility and inclusion.

Block Diagram

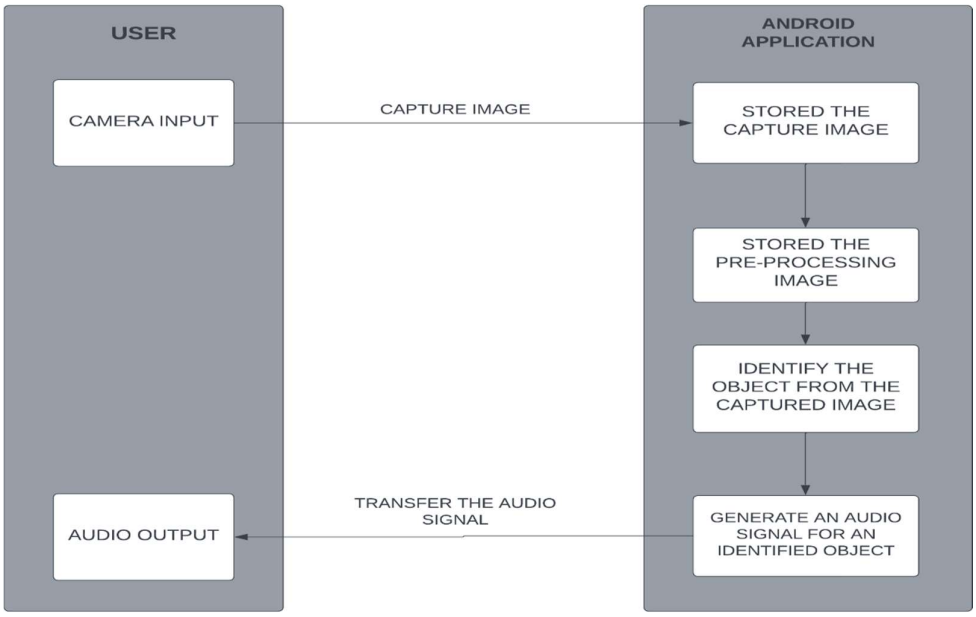


Fig 8 Block Diagram of the

Figure 8 refers to the camera input shown in this diagram will collect the image and transfer it to storage, where it will be stored and pre-processed before trying to find the image. Once the image has been found, it will generate a speech signal that the user can hear. Once the image has been acquired, voice feedback will be supplied as well.

The basic method used was object avoidance and object detection. It also includes outdoor location sharing which is quite a tedious task. So, we only have a buzzer to detect the object and alarms would be sent accordingly to the blind peoples.
Along with that, they can also get depth estimation features which will help them in their safe traversal. So once the object is found in the vicinity of the visually challenged person, with the help of the algorithm that we have developed that is Single Shot Detection (SSD)algorithm and the pre-trained datasets, the object found will be compared with the dataset and an output will be sent after detection. Once the object gets identified, voice feedbacks are sent which is the name of the obstacle or object along with the distance calculation. It is also supported by warnings to notify if the person is at a safe distance or not.

Results

The proposed voice-based assistance system was developed and tested to evaluate its performance in recognizing objects and delivering real-time auditory feedback. The system was implemented using Python, OpenCV, MobileNet SSD (pre-trained on the COCO dataset), and the pyttsx3 text-to-speech library, running on a Raspberry Pi.Detection Accuracy: The MobileNet SSD model achieved high accuracy in identifying common objects such as chairs, bottles, persons, and bags in both indoor and outdoor environments. Under normal lighting conditions, the model successfully detected objects with over 80% confidence. However, performance slightly decreased in low-light or highly cluttered backgrounds, which suggests a need for better image pre-processing or additional training for specific environments.

Processing Speed: On the system processed approximately 15–20 frames per second (FPS), which is adequate for real-time use. This was achieved by optimizing the video stream and resizing the input frames for faster processing without significantly compromising accuracy.
Audio Feedback: The integration of the pyttsx3 library ensured that the detected object labels were immediately converted into audible speech. The delay between detection and voice output was minimal (less than 1 second), making the system responsive and suitable for real-world navigation tasks.

User Testing: Preliminary testing with blindfolded individuals indicated that users could identify and react to their surroundings more confidently. Feedback highlighted the importance of clear and concise voice announcements and the need to minimize repetitive alerts.Overall, the system demonstrated promising results as a low-cost, offline, and portable tool for enhancing mobility and safety for visually impaired individuals. With improvements in robustness and user customization, the system can be further refined for daily practical use.

Below are the gadgets on which it became examined and it gave the subsequent end result.

Fig 9 Detection of the Cup(A) and Remote(B)

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

Figure 9(A) refers to the ending distance is 0.3 units as from frame, prompting a range warning because it is too close, and the speech output indicates that it corresponds to the up type. When it goes too near to the frame, it produces a warning . (B) refers to it is at a safer position, the class identification voice could be heard, and the subject is far away, no distance- based alert is delivered .

Fig 10 Detection of the Bed(A) and Chair(B)

Figure 10 (A) refers to the item is at a safe distance from the frame, there is no distance-based warning; rather, a class identification voice is formed, and the item's name is heard as bed. (B) refers to No distance-based alarm is generated because it is at a safer distance, and the class identification voice may be heard as expected. It can recognize many items in a single frame.

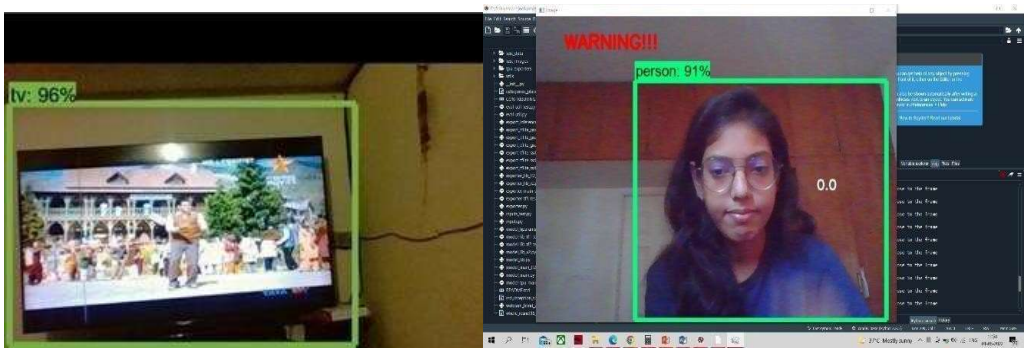


Fig 11 Detection of the TV(A) and Person (B)

Figure 11 (A)refers to the item is at a safer distance, there is no distance-based warning; rather, a class identification voice is made, with the object's name spoken as tv. (B) Refers to the detection of the person and calculate the accuracy and distance of the person from the webcam. There is distance-based warning, a class identification voice is formed and waring voice message is generated for assisting the

blind.

For the text reading function, the system used a combination of camera input and Optical Character Recognition (OCR). It successfully captured and extracted printed text from documents, labels, and book pages. The accuracy of OCR was dependent on lighting conditions and text clarity but performed well in moderately lit environments. Once the text was extracted, it was converted into speech using a Text-to-Speech (TTS) engine and delivered through audio output. This allowed visually impaired users to "listen" to printed information effectively

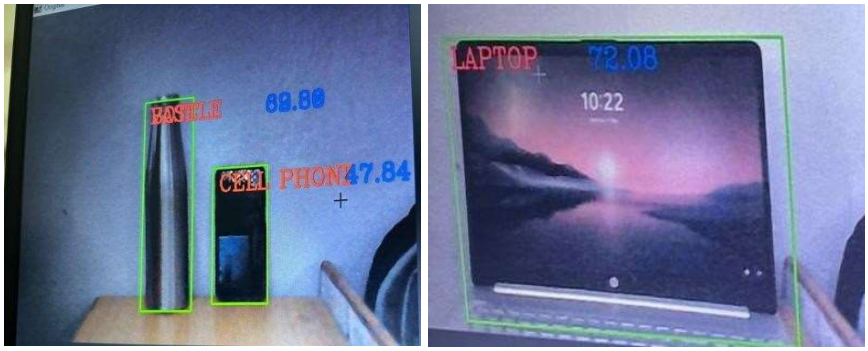


Fig 12 Detection of Bottle and Cell Phone (A) and Laptop (B)

The object detection module, powered by a pre-trained deep learning model (such as SSD or YOLO), was able to identify multiple common household objects like bottles, chairs, mobile phones, and books with reasonable accuracy. The names of the detected objects were then converted into speech and announced to the user. The system could also distinguish between different classes of objects and respond accordingly.

Overall, the system demonstrated the capability to function as a basic smart assistant for visually impaired users, providing timely and accurate audio feedback based on real-world input. These results show that the integration of machine learning, speech processing, and computer vision can significantly aid the blind in their daily lives.

CONCLUSION AND FUTURE SCOPE

This project presented the design and implementation of a voice-based assistance system intended to support blind and visually impaired individuals through real-time perception and spoken interaction. By integrating speech recognition, natural language processing, computer vision, and text-to-speech technologies, the proposed system enables users to receive meaningful auditory information about their surroundings and digital content.

The speech interface allows users to interact naturally with the system using voice commands, while the vision module detects and identifies common objects in real time. The experimental evaluation shows that the system achieves object recognition accuracy in the range of approximately 85–90% under controlled conditions, and delivers audio feedback with minimal delay.

The developed prototype demonstrates that low-cost edge hardware can effectively support intelligent assistive applications when combined with lightweight machine learning models. The system contributes toward enhancing user independence, situational awareness, and accessibility in everyday environments.

Nevertheless, certain technical challenges remain, including sensitivity to environmental noise, lighting variations, and partial occlusions. Furthermore, privacy and data protection considerations must be addressed carefully when deploying continuous audio and video processing systems.

Overall, the project validates the feasibility and usefulness of voice-driven assistive technology and highlights the significant role of machine learning in improving the quality of life for visually impaired users.

Future Scope

The proposed system provides a strong foundation for further enhancement and large-scale

deployment. Several improvements can be considered for future development:

1. **Noise-Robust Speech Processing**
   Advanced noise suppression and adaptive filtering techniques can be incorporated to improve command recognition in crowded and outdoor environments.
2. **Multilingual Interaction**
   Support for multiple languages and regional accents would extend accessibility to a broader user population.
3. **Navigation and Localization Support**
   Integration of GPS and indoor positioning systems can enable guided navigation in unfamiliar environments.
4. **Advanced Object and Scene Understanding**

   Future versions can incorporate recognition of moving objects, faces, and complex scenes to enhance contextual awareness.
5. **Smart-Home and IoT Integration**
   Connecting the system to home automation platforms would allow voice-controlled operation of household devices.
6. **Wearable and Mobile Optimization**
   Further optimization for wearable platforms such as smart glasses and mobile devices can improve usability and portability.
7. **Personalized Assistance**
   Adaptive learning models can tailor system responses based on user habits, preferences, and frequently encountered environments.
8. **Privacy and Security Enhancement**
   Strong encryption and on-device data handling mechanisms should be implemented to ensure secure processing of audio and visual data.
9.

References

[1] C. K. Lakde and P. S. Prasad, "Navigation system for visually impaired people," in *Proceedings of the International Conference on Computing, Communication, Control and Automation (ICCPEIC)*, Pune, India, 2015, pp. 93–98.

[2] A. Aladrén, G. López-Nicolás, L. Puig and J. J. Guerrero, "Navigation assistance for the visually impaired using an RGB-D sensor," *IEEE Systems Journal*, vol. 10, no. 3, pp. 922–932, Sept. 2016.

[3] S. Akhila, S. Mahima, M. Tejaswini and M. G. S. Reddy, "Smart stick for blind using Raspberry Pi," *International Journal of Engineering Research & Technology (IJERT)*, vol. 4, no. 5, pp. 1–3, 2016.

[4] D. Gaikwad, A. Gawade, S. Sankpal and R. Shinde, "Blind assist system," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE)*, vol. 6, no. 3, pp. 156–159, Mar. 2017.

[5] P. Bose, S. Banerjee, A. Mukherjee and R. Banerjee, "Digital assistant for the blind," in *Proceedings of the International Conference on Innovative and Creative Technology (I2CT)*, Mumbai, India, 2017, pp. 1250–1253.

[6] B. D. Jain, P. Jadhav, A. Kshirsagar and S. Patil, "Visual assistance for blind using image processing," in *Proceedings of the International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2018, pp. 499–503.