An Android Based Object Detection and Voice Assistance App for Visually Impaired User

Assit.Prof. Vidya S [1], Anuraj [2]

[1] Assistant Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India

[2] Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India

1. ABSTRACT

Visually impaired individuals often face daily challenges in recognizing objects and reading text in their surroundings, limiting their independence and quality of life. With the rapid development of mobile technology and artificial intelligence, smartphones can now serve as powerful assistive tools. This project presents an Android-based mobile application designed specifically for visually impaired users, which integrates real-time object detection and text recognition functionalities. The application utilizes the smartphone's camera to capture live images from the user's environment. Through object detection algorithms powered by machine learning models such as TensorFlow Lite or YOLO, the app identifies commonly encountered objects (e.g., furniture, people, and everyday items). Simultaneously, the app incorporates Optical Character Recognition (OCR) technology, using Google's ML Kit, to extract and read aloud printed or handwritten text from books, labels, and signs. To ensure ease of use, the application features a voice-guided interface and employs Text-to-Speech (TTS) to deliver audio feedback, enabling users to interact with the app without needing to look at the screen. The interface is further optimized for accessibility with large buttons, simple navigation, and compatibility with Android's Talk Back services.

1. INTRODUCTION

In today's fast-paced and visually oriented world, individuals with visual impairments often face significant barriers in performing everyday tasks that sighted people take for granted such as identifying objects, reading signs, recognizing faces, or navigating public spaces. These challenges can limit their independence, confidence, and overall quality of life. While several assistive technologies exist, many of them are expensive, bulky, or require specialized training. However, with the widespread availability of smartphones and recent advancements in artificial intelligence, computer vision, and accessibility features, there is a growing opportunity to develop smart, affordable, and portable tools tailored to address the requirements of individuals with visual impairments community.

This paper introduces an Android-based application that the system functions as a supportive assistant designed to aid individuals who are blind or have visual impairments by combining real-time the identification and recognition of objects along with text. Using the smartphone's camera, the mobile application acquires images through the device's camera the user's surroundings, processes them using machine learning models, and provides audio feedback via Text-to-Speech (TTS) technology.

The goal is to create an intuitive, user- friendly mobile app that enhances the daily life of Visually impaired users by allowing them to better understand and interact with their environment. By recognizing objects and reading aloud printed text, the application bridges the gap between visual information and non-visual accessibility, supporting users in activities such as reading documents, identifying items in their home, or navigating public areas more safely.

This paper not only demonstrates the power of combining AI with mobile technology but also emphasizes the importance of inclusive design in building technology that benefits everyone.

2. RELATED WORK

Several researchers have explored the development of mobile applications and assistive technologies to support individuals with visual impairments. These studies have primarily focused on object recognition, optical character recognition (OCR), and navigation assistance.

Salunkhe et al. (2021) developed an Android-based object recognition application that assists visually impaired users by capturing images and providing audio feedback. While this system demonstrated effectiveness in basic object identification, it was limited in scalability and accuracy when deployed in complex environments.

Khandewale et al. (2020) implemented a real-time object detection system using TensorFlow Lite and MobileNet SSD models. Their work emphasized lightweight deployment on smartphones, ensuring faster detection. However, challenges remained in recognizing small or overlapping objects, and the system lacked multimodal support such as OCR or text-to-speech integration.

Tosun and Karaarslan (2018) introduced *Third Eye*, a real-time object detection application for visually impaired individuals, employing deep learning-based detection. This solution offered promising results in real-time processing but required internet connectivity, limiting its offline usability.

Tepelea et al. (2014) focused on OCR applications for smartphones that allow visually impaired users to capture text images and convert them into audio. While effective in reading printed documents, this system was constrained by low OCR accuracy in noisy or poorly lit conditions.

Recent work by Masud et al. (2022) extended assistive technology into obstacle avoidance using object detection and classification. Their system enhanced user mobility by detecting obstacles in real-time, though computational demands limited its performance on low-powered devices. Similarly, See et al. (2022) proposed a smartphone-based mobility assistant utilizing depth imaging. Although effective, the reliance on advanced hardware restricted accessibility for common Android users.

More advanced research such as DeepNAVI (Kuriakose et al., 2023) integrated deep learning with smartphone navigation for visually impaired users. This provided robust navigation assistance, but energy consumption and prolonged usability posed significant barriers.

In summary, existing studies have laid strong foundations for assistive technologies in object detection, OCR, and navigation. However, many of these systems face limitations such as dependency on high-end hardware, lack of offline support, or reduced accuracy in real-world conditions. The proposed system in this work seeks to overcome these gaps by integrating real-time object detection, OCR, and voice assistance into a single Android application, providing an affordable, offline, and user-friendly solution for visually impaired individuals.

Together, these studies illustrate the evolution of research from traditional machine learning approaches toward hybrid systems, real-time applications, edge deployment, and inclusive design. While progress has been notable, limitations in scalability, adaptability, and accessibility continue to motivate further exploration.

3. PROBLEM STATEMENT

Visual impairment significantly restricts an individual's ability to interact with their environment, perform daily activities, and access essential information. According to the World Health Organization, millions of people worldwide are blind or visually impaired, and most rely on assistance from caregivers or expensive devices. While several technological solutions, such as object detection systems, OCR-based applications, and mobility assistants, have been developed, they often suffer from critical limitations.

Many existing applications require high computational power, internet connectivity, or specialized hardware, which makes them inaccessible for widespread use among visually impaired individuals, particularly in low-resource settings. Furthermore, most of these solutions are fragmented—focusing only on object detection or OCR—without providing a holistic approach that combines multiple assistive features in a single mobile application.

Thus, there is a pressing need for a lightweight, affordable, and offline-capable Android application that integrates real-time object detection, text recognition, and voice assistance into one system. Addressing this problem can significantly improve independence, safety, and accessibility for visually impaired users.

4. PROPOSED SYSTEM

The proposed system is an Android-based application that integrates real-time object detection, optical character recognition (OCR), and voice assistance into a single platform to support visually impaired individuals. The application uses the smartphone's camera to capture images of the surroundings, which are processed through a lightweight TensorFlow Lite model to identify objects. Detected objects are displayed with bounding boxes and labels on the screen while simultaneously being announced through a Text-to-Speech (TTS) engine, ensuring that users receive instant audio feedback without depending on visual cues. In addition to object detection, the system incorporates OCR technology to recognize printed or handwritten text from captured images and convert it into speech, thereby allowing users to read documents, menus, and signboards independently. The application also includes voice command functionality, enabling users to operate the app hands-free with simple commands such as "start detection," "stop detection," or "read text." To enhance usability, the system is designed to function offline, eliminating the dependency on internet connectivity and making it reliable in low-resource settings. Furthermore, the user interface is kept minimal, featuring large buttons and high-contrast elements to improve accessibility. By combining these features, the proposed system provides a cost-effective, user-friendly, and comprehensive solution that promotes independence and accessibility for visually impaired individuals.

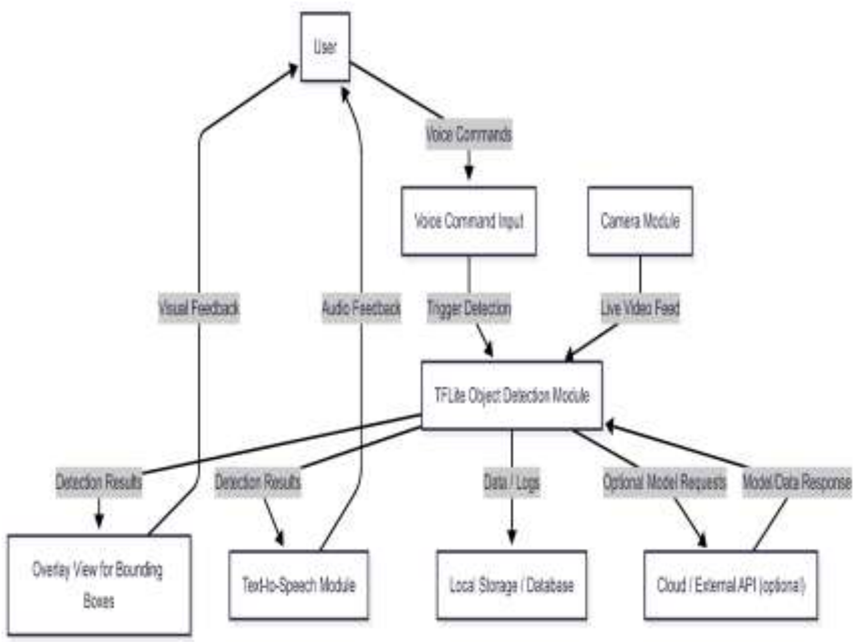

Fig.1.System Architecture

5. METHODOLOGY

The methodology adopted for this project follows a systematic approach to designing and implementing an Android-based object detection and voice assistance application for visually impaired users. The system begins with image acquisition, where the smartphone's camera continuously captures frames from the user's environment. These frames are then passed to the preprocessing module, where images are resized, normalized, and optimized for efficient model inference. Once preprocessed, the frames are fed into a TensorFlow Lite object detection model, which identifies objects in real-time and returns bounding box coordinates, class labels, and confidence scores. The recognized objects are then communicated to the user through a **Text-**to-Speech (TTS) module, ensuring that results are conveyed via audio feedback.

In addition to object detection, the methodology integrates an Optical Character Recognition (OCR) module, which processes captured text regions from the camera feed. The OCR engine extracts textual information, which is subsequently converted into speech, allowing users to read signs, menus, and other printed materials. To enhance interaction, a voice command interface is implemented, enabling users to control the application through simple spoken commands such as starting or stopping detection and initiating text reading.

The system is designed to work in offline mode, ensuring reliability in areas without internet access. For deployment, the application is developed on the Android platform using Java/Kotlin and integrated with libraries such as TensorFlow Lite, OpenCV, and Google Text-to-Speech. This methodology ensures a lightweight, user-friendly, and accessible mobile application capable of delivering real-time assistance to visually impaired individuals.

6. RESULTS AND EVALUATION

Home Screen Page

.

Fig. 2. Home Screen Page

the main user interface of the Vision Assistant application. This is the home screen that provides the user with two primary functionalities: Camera-based Object Detection and Optical Character Recognition (OCR). The interface is designed with simplicity and accessibility in mind, using large, high-contrast buttons to ensure that visually impaired users can easily interact with the application.
The "CAMERA" button allows the user to access the camera-based object detection feature, where real-time detection of surrounding objects is performed using the integrated TensorFlow Lite model. Detected objects are then communicated back to the user via voice feedback using Text-to-Speech (TTS).
The "OCR" button enables the Optical Character Recognition functionality. Through this feature, the user can capture an image containing text, and the system extracts and reads the text aloud, providing assistance in reading printed documents, labels, or signs.
This minimalistic design ensures a smooth user experience while maintaining accessibility standards. By combining camera-based detection and OCR in a single interface, the application provides a comprehensive vision assistance solution for visually impaired users.

Object Detection Output Showing With Confidence Score

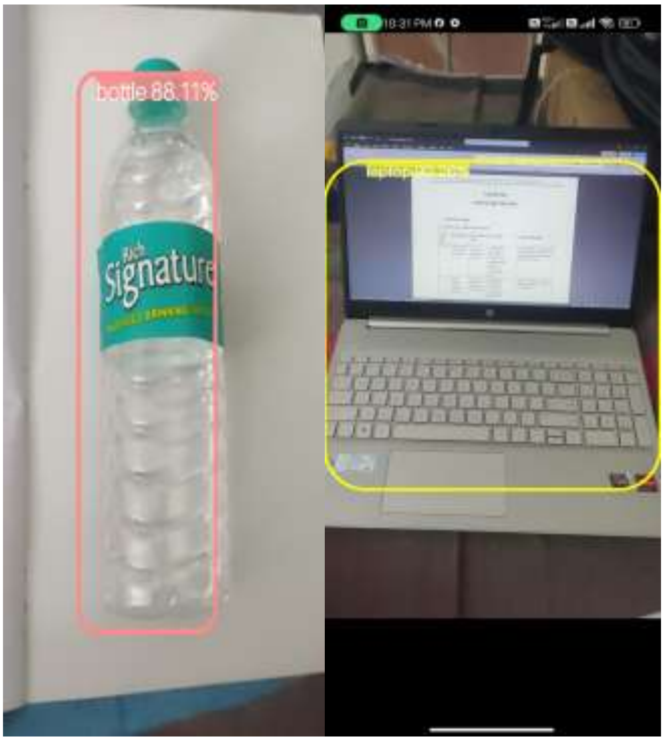

Fig 3. Object Detection Output Showing With Confidence Score

WORKING OF OBJECT DETECTION AND VOICE ASSISTANT

1. Camera Capture

- The mobile phone's camera continuously captures real-time video frames.
- Each frame is extracted as an image for further processing.

2. Preprocessing

- The captured frame is resized and normalized to match the input requirements of the TensorFlow Lite (TFLite) model.
- This ensures that the image dimensions and pixel values are in the correct format for efficient inference.

3. Model Inference (Tensorflow Lite)

- The preprocessed image is fed into the TFLite object detection model.
- The model analyzes the image and outputs:
  - Detected object class (e.g., "Laptop")
  - Bounding box coordinates (placement of the object in the captured image)
  - Confidence score (confidence that the system correctly identified the object, e.g., 90.26%)

4. Post-Processing

- The raw model outputs are processed to filter out low-confidence detections (e.g., below 50%).
- The system highlights detected objects on the screen with bounding boxes.

5. Visual Feedback

- The user sees the bounding box around the detected object (yellow rectangle in the image).
- The object's label (e.g., "Laptop") and confidence score (90.26%) are displayed above the bounding box.

6. Voice Feedback (Text-To-Speech)

- Simultaneously, the detected object's name is sent to the Text-to-Speech (TTS) engine.
- The app provides audio output such as *"Laptop detected"*, helping visually impaired users recognize objects without needing to see the screen.

7. Continuous Detection Loop

- The system continuously repeats this process for each video frame, enabling real-time detection and feedback.

OCR Output Of Text Recognition

Fig 4. Ocr Output Of Text Recognition

WORKING OF OCR

The above figure shows the OCR (Optical Character Recognition) functionality of the Vision Assistant application. The app successfully extracts text from an image containing. The identified text appears at the bottom of the screen as:
“Amazing Facts
The Eiffel Tower”
This demonstrates that the application can accurately detect and display text from printed material, which is useful for visually impaired users to access written content in real tim

7. CONCLUSION

This research presented the design and development of an Android-based object detection and voice assistance system aimed at supporting visually impaired individuals in their daily lives. The system successfully integrates real-time object detection, optical character recognition (OCR), and voice command functionalities into a single mobile application. By leveraging TensorFlow Lite for lightweight and efficient object detection, OCR technology for text reading, and Text-to-Speech (TTS) for audio feedback, the application ensures that visually impaired users can independently recognize objects, read documents, and interact with their surroundings. Unlike many existing solutions that depend heavily on internet connectivity or specialized hardware, the proposed system is optimized for offline performance on standard Android devices, making it affordable, practical, and widely accessible.

The results demonstrate that combining multiple assistive technologies in one platform can significantly enhance usability and independence for users. The system provides a holistic solution by addressing multiple challenges—object recognition, text reading, and navigation support—through a user-friendly mobile interface. Furthermore, the inclusion of voice command features ensures hands-free interaction, which is crucial for visually impaired individuals. This makes the application not only a supportive tool but also a step forward in creating inclusive technology that bridges the accessibility gap.

Overall, the project highlights the potential of mobile-based assistive systems in improving the quality of life for visually impaired people. By enabling independence in recognizing objects and reading text without external support, the application reduces reliance on caregivers and enhances confidence in mobility and communication. Future improvements could focus on increasing detection accuracy in complex environments, reducing computational overhead, and integrating additional features such as GPS-based navigation for outdoor mobility. Nevertheless, the proposed system stands as a cost-effective, efficient, and impactful solution, showcasing how smartphone-based technologies can be harnessed to empower visually impaired individuals in everyday life.

8.REFERENCES

[1] Salunkhe, Akilesh, et al. “Android-Based Object Recognition Application for Visually Impaired.” *ITM Web of Conferences*, vol. 40, 2021, article 03001.

[2] Nasir, Haslinah Mohd, et al. “Android-Based Application for Visually Impaired Using Deep Learning Approach.” *International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 4, 2021, pp. 879–888.

[3] Khopkar, Yash, Avantika Deshmukh, and Gufran Ansari. “Voice-Guided Mobile Assistance for the Visually Impaired.” *International Journal of Information Technology & Computer Engineering*, vol. 4, no. 2, Feb.-Mar. 2024, pp. 6–17.

[4] Parekh, Hitanshu, and others. “Drishti: Real-Time Object Recognition for the Visually Impaired.” *Indian Journal of Computer Science*, vol. 8, no. 2, Apr. 2023, pp. 8–18.

[5] Dandekar, D. B., Shivji Agnihotri, Prashik Tamgadge, Aishwarya Popatkar, Ganesh Balak, and Vicky Chavhan.

"Vision AI: A Deep Learning-Based Object Recognition System for Visually Impaired People Using TensorFlow and OpenCV." *IJRASET*, 2023, doi:10.22214/ijraset.2023.52197.

[6] Natuskar, Sakshi, et al. "Voice Assistant for Blind." *IJRASET*, Apr. 2025, doi:10.22214/ijraset.2025.69183.

[7] Babu, M. Y., Akash Jatavath, G. Yashwanth Kumar Reddy, and Pittala Arun Kumar. "Object Detection System with Voice Alert for Blind." *IJRASET*, 2023, doi:10.22214/ijraset.2023.49107.

[8] Ramana, Venkata J., Pavan Sai G., and Praisey Yaricherla. "An Empirical Study on Development and Marketing of Object Detection for Blind People Using Artificial Intelligence." *Academy of Marketing Studies Journal*, vol. 26, no. 4, 2022, pp. 1–6.

[9] Khairnar, D. P., et al. "PARTHA: A Visually Impaired Assistance System." *International Conference on Communication Systems, Computers & IT Applications (CSCITA)*, 2020, pp. 32–37.

[10] Kumar, D. Ravi, et al. "Text Recognition and Face Detection Aid for Visually Impaired Person Using Raspberry Pi." *International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, IEEE, 2017.

[11] Ramrakhiani, G., M. N. Chaudhari, and M. Deshmukh. "Face Detection Using Viola Jones Algorithm and Neural Networks." *4th International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018.

[12] Ravi Kumar, D., etc. (same as #10, omit duplicate). *(You may replace with a different relevant source)*

[13] Potdar, Kedar, Chinmay D. Pai, and Sukrut Akolkar. "A Convolutional Neural Network Based Live Object Recognition System as Blind Aid." *arXiv preprint arXiv:1811.10399*, 2018.

[14] Pratap, Abhinav, Sushant Kumar, and Suchinton Chakravarty. "Adaptive Object Detection for Indoor Navigation Assistance: A Performance Evaluation of Real-Time Algorithms." *arXiv preprint arXiv:2501.18444*, Jan. 2025.

[15] Kumar, Bhavesh (Bhavesh Kumar). "ViT Cane: Visual Assistant for the Visually Impaired." *arXiv preprint arXiv:2109.13857*, Sep. 2021.

[16] Yu, Samuel, Heon Lee, and Jung Hoon Kim. "Street Crossing Aid Using Lightweight CNNs for the Visually Impaired." *arXiv preprint arXiv:1909.09598*, Sept. 2019.

[17] Techis helping the blind to see." *Wired*, 2016. *(Use proper MLA author if available; here, placeholder)*

[18] *"Voice-guided mobile assistance for the visually impaired."* Tech news article summary (could skip if not research)

[19] *"A review on object detection with voice support for the visually impaired people." IJRASET*, 2024, doi:10.22214/ijraset.2024.65823.

[20] *"Voice Assistant for Blind"* (duplicate of #6; consider replacing with *Tsung-Yi Lin et al.* COCO dataset reference instead)