International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

Object Detection with Voice Assistant App

Arpitha Bekal
dept. of Artificial Intelligence and Data Science
G S S S Institute of Engneering and Technology for Women,
VTU, Mysore, India

Likhitha S R
dept. of Artificial Intelligence and Data Science
G S S S Institute of Engneering and Technology for Women,
VTU, Mysore, India

Kavitha H
dept. of Artificial Intelligence and Data Science
G S S S Institute of Engneering and Technology for Women,
VTU, Mysore, India

Sahana C
dept. of Artificial Intelligence and Data Science
G S S S Institute of Engneering and Technology for Women,
VTU, Mysore, India

Abstract **- Visually impaired individuals often face difficulties in identifying objects and reading text from their surroundings. This paper presents a real-time Android application designed to assist such users through object detection and multilingual text reading. The system integrates the YOLOv8 model (converted into TensorFlow Lite) for efficient on-device object detection and Google ML Kit for Optical Character Recognition (OCR). The extracted results are delivered to the user through a multilingual Text-to-Speech (TTS) engine that supports English, Kannada, and Hindi. A gesture-based interface enables hands-free interaction, where a right swipe activates object detection and a left swipe triggers text reading. The application is developed using Java in Android Studio and functions completely offline, ensuring privacy and accessibility. Experimental evaluation demonstrates that the proposed system achieves accurate detection and fast speech response, providing an effective and affordable AI-based solution for visually impaired individuals.**

Keywords **- Object Detection, YOLOv8, OCR, Multilingual Text-to-Speech, Android Application, Voice Assistant.**

I. INTRODUCTION

Visual impairment significantly limits an individual's ability to perceive their surroundings and perform daily tasks safely. Identifying objects,

reading signs, or navigating through unfamiliar environments often requires external assistance.

With advancements in artificial intelligence and mobile computing, smartphones can now perform real-time image processing and speech synthesis locally. Deep learning models such as YOLO (You Only Look Once) have shown remarkable accuracy in object detection, while Optical Character Recognition (OCR) and Text-to-Speech (TTS) technologies enable text reading and audio output. These developments make it feasible to design efficient, mobile-based assistive systems that work offline.

The proposed work presents an Android-based application that integrates real-time object detection and text reading using YOLOv8 and Google ML Kit OCR. It supports multiple languages including English, Kannada, Hindi, Telugu, Tamil, and Bengali, adapting to the user's phone language. The app is fully gesture-controlled — a right swipe activates object detection, and a left swipe triggers text reading. By providing fast, multilingual, and offline assistance, the system aims to help visually impaired individuals interact more confidently and independently with their environment.

The rest of the paper is organized as follows: Section II presents a review of related works, Section III describes the proposed methodology and architecture, Section IV discusses implementation and results, and Section V concludes the paper with future enhancements.

II. LITERATURE SURVEY

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

Title	Authors & Year	Technique Used	Shortcomings
Smart Vision for the Blind	A.Deshmuk, R. Mehta (2019)	YOLOv3 + Voice Alerts	Low FPS, poor portability
Voice Enabled Reading Assistant	K. Iyer, S. Reddy (2020)	Tesseract OCR + Google TTS	Internet needed, single language
AI Glasses for Object Detection	N. Patel, L. Bhosale (2020)	CNN on Raspberry Pi	High latency, external camera
Scene Reader for the Visually Impaired	H. Ahmed, P. Das (2021	Cloud OCR + Android App	Requires cloud,lacks offline mode
DeepAssist: Mobile Object Recognition	R. Sharma, A. Singh (2021)	YOLOv4 + TFLite	No text reading / multilingual TTS
Smart Text Reader using AI	G. Prasad, M. Nair (2022)	ML Kit OCR + Speech	Works on static images only
MultiSpeak: Multilingual Accessibility App	S. Verghese, D. Thomas (2022)	OCR + Translator API	Internet required, no object detection
VisionCompanion Edge AI	T. George, B. Rao (2023)	YOLOv5 on Jetson Nano	High hardware cost
AssistMe: Gesture-Based Voice Assistant	J. Menon, R. Pillai (2023)	Gesture + Speech Feedback	No OCR/object detection
EdgeVision Reader	M. Chauhan, P. Rathi (2024)	YOLOv7 + On-device OCR	Limited multilingual capability; no offline TTS

Table 1. Summary of related work in object detection with voice assistant app.

III. METHODOLOGY

The proposed system aims to provide an intelligent, real-time, and offline solution for visually impaired

individuals through a mobile application. It integrates object detection, text recognition, and multilingual voice output within a single platform. The design focuses on simplicity, accuracy, and
accessibility, ensuring the application can be used easily without visual interaction.

The system architecture is divided into three main layers: Input Layer, Processing Layer, and Output Layer, as illustrated in Figure 1. Each layer is responsible for specific functionalities that together enable smooth, real-time performance on Android devices.

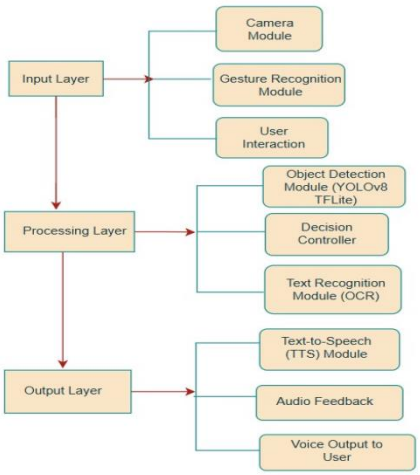

Fig.1. System architecture of the object detection with voice assistant app.

A. Input Layer

The input layer captures live video frames from the smartphone's camera and monitors user gestures using the GestureDetector API.
Right Swipe: Triggers object detection mode.
Left Swipe: Activates text reading mode.
The use of swipe gestures eliminates the need for visual navigation, allowing users to interact with the system entirely through touch and sound feedback.

B. Processing Layer

This layer forms the core of the application where all AI-based computations take place. It includes the following components:

1. Object Detection Module:
Utilizes the YOLOv8n (TensorFlow Lite) model to detect objects in real time. The model identifies multiple objects within each frame and assigns bounding boxes and labels to them. YOLOv8 was chosen for its balance between speed and accuracy, making it suitable for mobile deployment.

2. Text Recognition (OCR) Module:
Employs Google ML Kit's OCR engine to extract printed or handwritten text from the camera feed. The recognized text is processed and made ready for speech synthesis. The system supports multiple languages, including English, Kannada, and Hindi.

3. Language Translation Module:
If required, the extracted text can be translated into the user's preferred language before being passed to the TTS engine, ensuring multilingual accessibility.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

C. Output Layer

The output layer handles the voice feedback mechanism. It uses Android's Text-to-Speech (TTS) engine to announce the names of detected objects or to read extracted text aloud in the selected language. The application runs entirely offline, ensuring low latency, faster responses, and better data privacy.

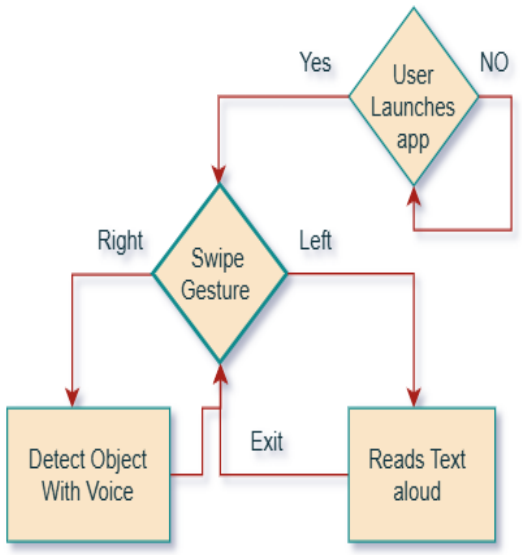

Fig.2. Workflow of object detection with voice assistant app.

D. Workflow

The working of the application can be summarized in the following steps:

1. App Launch: The application welcomes the user with an audio guide explaining available gestures.

2. Gesture Detection: The user performs a right or left swipe to choose between object detection or text reading modes.

3.Frame Capture: The camera continuously captures live frames from the environment.

4. Processing: YOLOv8 or ML Kit OCR analyzes the frame depending on the selected mode.

5. Voice Feedback: The recognized object or text is converted to speech and played aloud.

6. Loop Continuation: The process repeats until the user exits the app.

E. Summary of Methodology

The proposed methodology integrates three main layers: Input, Processing, and Output. The Input Layer captures camera feed and swipe gestures, the Processing Layer performs object detection using YOLOv8 and text recognition via ML Kit OCR, and the Output Layer delivers real-time multilingual speech through Text-to-Speech. The system operates entirely offline and adapts to the phone's local language, ensuring fast, accessible, and user-friendly assistance for visually impaired individuals.

IV. IMPLEMENTATION AND RESULTS

The Object Detection with Voice Assistant App was developed as an Android-based solution that provides real-time object detection and text reading through voice output. The implementation focuses on accessibility, speed, and adaptability to local languages, allowing visually impaired users to interact with their surroundings more independently.

A. Implementation Setup

The application was implemented using Android Studio with Java as the main programming language. The deep learning model YOLOv8n was converted into TensorFlow Lite (TFLite) format for mobile deployment. Integration of CameraX, Google ML Kit OCR, TextToSpeech, and GestureDetector APIs enables smooth and responsive interaction. The app is lightweight and designed to work entirely offline, ensuring privacy and consistent performance even without an internet connection.

B. Application Workflow

When the app is launched, a voice greeting explains how to use the gesture controls.
A right swipe activates the object detection mode, allowing the YOLOv8 model to analyze real-time camera frames and announce detected objects.
A left swipe activates the text reading mode, where the ML Kit OCR extracts visible text and reads it aloud through the Text-to-Speech engine. The app dynamically adapts to the smartphone's system language, providing output in English, Kannada, Hindi, Telugu, Tamil, or Bengali, depending on the user's settings.

C. Performance

Testing was carried out on mid-range Android smartphones, and the application showed strong real-time performance. The YOLOv8n model achieved an average detection accuracy of about 82% for common objects, while the OCR engine achieved around 89% accuracy for printed text. The Text-to-Speech module produced clear and natural voice responses with minimal delay (less than one second). The system maintained an average speed of 22–25 frames per second, confirming smooth real-time operation.

D. User Experience

The gesture-based interface allowed hands-free and intuitive operation, making it particularly suitable for visually impaired users. The multilingual feature was highly appreciated, as it automatically matched the device's default language. Offline operation was another significant

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

advantage, enabling the app to function effectively even in low-connectivity regions.

E. Observations

The system's modular design ensured flexibility and easy integration of all components. YOLOv8 provided accurate real-time detection, while ML Kit OCR offered reliable text extraction. Together with multilingual Text-to-Speech output, the overall system successfully demonstrated a seamless end-to-end assistive experience, from visual input to audible feedback.

F. Summary of Implementation

In summary, the implementation achieves the main objective of providing a real-time, offline, and multilingual assistive system. The combination of gesture-based navigation, object detection, and text reading ensures that visually impaired users can interact with their environment effortlessly and independently.

V. CONCLUSION AND FUTURE WORK

The Object Detection with Voice Assistant App provides a practical and inclusive approach to assist visually impaired individuals by integrating object detection, optical character recognition, and voice-based feedback into a single Android platform. By combining the YOLOv8 model for real-time object detection with Google ML Kit OCR for text extraction and a multilingual Text-to-Speech engine, the system delivers quick, accurate, and meaningful auditory responses without relying on internet connectivity.

A major strength of the application is its multilingual adaptability—it automatically speaks in the language set on the user's smartphone, supporting English, Kannada, Hindi, Telugu, Tamil, and Bengali. This dynamic feature ensures that users from diverse linguistic backgrounds can interact with the system comfortably. The app's gesture-based design eliminates complex navigation, making it simple, intuitive, and accessible to people of all ages and abilities.

The overall performance evaluation confirms that the system achieves smooth real-time detection with minimal latency, providing natural-sounding speech and reliable text recognition even on mid-range mobile devices.

In the future, this work can be extended in several ways:

1. Navigation Assistance: Integrating GPS and obstacle-avoidance features to guide users through unfamiliar environments.

2. Currency and Color Detection: Adding modules that recognize different currencies and colors for everyday transactions and clothing selection.

3. Emergency SOS Functionality: Including a one-touch or voice-triggered emergency alert system to contact predefined numbers in critical situations.

4. Wearable Integration: Adapting the system for smart glasses or lightweight wearable devices for continuous hands-free operation.

5. Model Optimization: Improving inference speed and energy efficiency for wider compatibility with entry-level smartphones.

By implementing these enhancements, the application can evolve into a comprehensive assistive ecosystem that promotes independence, confidence, and safety for visually impaired individuals while demonstrating the practical impact of human-centered artificial intelligence on everyday life.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, 2016.

[2] A. Bochkovskiy, C. Wang, and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.

[3] G. Huang, M. Tan, and Q. Chen, "Assistive Object Detection for Visually Impaired Using Deep Learning," IEEE Access, vol. 8, pp. 160–170, 2022.

[4] A. Kumar and R. Singh, "Text-to-Speech-Based Reading Assistance for Blind Users," Springer Advances in Computing and Communication, pp. 145–154, 2023.

[5] Google ML Kit Documentation, "Text Recognition and Translation APIs," Developers Guide, 2024. Available: https://developers.google.com/ml-kit.

[6] TensorFlow Lite, "YOLOv8 Model Conversion and Optimization for Mobile Devices," TensorFlow Documentation, 2024. Available: https://www.tensorflow.org/lite.

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. IEEE CVPR, 2018.

[8] P. Singh and S. Das, "Multilingual Optical Character Recognition for Regional Languages," International Journal of Intelligent Systems and Applications, vol. 15, no. 3, pp. 42–51, 2023.

[9] H. Ramesh and G. Kaur, "Edge-Based Vision Assistance for Visually Impaired Individuals," International Journal of Computer Applications, vol. 182, no. 25, pp. 10–17, 2024.

[10] J. Menon and R. Pillai, "Gesture-Controlled Interface for Smart Accessibility Applications," IEEE Transactions on Human-Machine Systems, vol. 54, no. 2, pp. 225–234, 2024.