Real-Time Smart Vision Assistant For Visually Impaired Using Yolov8, OCR And Speech Interaction.

Aastha Kadam[1], Leena Shahare[2], Poonam Fegade[3]
[1,2,3]*K.K. Wagh Institute of Engineering Education and Research, Master of Computer Applications (MCA) Department, Nashik, Maharashtra, India*

***Abstract*—People who are visually impaired might have a hard time being independent. It becomes impossible to see the surrounding things, read the printed text, and learn about the circumstances. Conventional assistive devices, comprising white cane and Braille, provide weak context and lack the ability to interpret visual scenes in action. However, advancements in Artificial Intelligence, Deep Learning and Computer Vision have produced intelligent assistive technologies that convert visual data into meaningful audio output.**

This paper presents the Real-Time Smart Vision Assistant which is composed of YOLOv8, OCR, multilingual language translation, and speech to help the visually impaired. The system uses a camera to capture real-time images and feeds them to deep learning algorithms for object detection and text extraction from printed materials such as books, labels, signaling boards etc. After the extraction, the generated output is converted into speech using TTS technology for English, Hindi and Marathi language. The system uses edge computing architecture to enhance response time and protect data privacy.

According to the experimental evaluation, YOLOv8 detects objects more accurately than any previous detection model. Improving accessibility and independence of the visually impaired by integrating OCR and multilingual speech interaction users. The suggested technology can serve as a low-cost assistive device to understand environment.

***Index Terms*—Assistive Technology, Computer Vision, YOLOv8, Optical Character Recognition, Deep Learning, Text-to-Speech, Accessibility.**

I. INTRODUCTION

Millions of people across the world face the challenge of visual impairment. People with visual impairment or blindness often struggle to identify and understand objects around them. Similarly, they face challenges in reading printed text as well as moving safely in indoor and outdoor locations. Users of classic assistive devices like the white cane can detect nearby objects but are not offered with particular information regarding the environment [5].

Using the most recent techniques in AI and Computer Vision, machines are enabled to interpret visual information like humans. The performance of any image recognition or object detection task is significantly improved using deep learning techniques as these techniques learn features automatically and deliver excellent performance under certain conditions [9].

Object identification is referred to as object detection. The image with object detection gives information about the presented object. YOLO (You Only Look Once) is a widely used object detection algorithm. It predicts bounding boxes and class probabilities directly from images using convolution neural networks [1]. YOLO analyses images in real-time and offers faster detection speed than traditional region-based detection algorithms.

By means of Optical Character Recognition (OCR) technology, information can be extracted from an image. OCR technology enables printed text to be recognized and converted into machine-readable formats. These formats can be generated to speech using text-to-speech engines [3] The combination of OCR and speech technologies enables the visually impaired to read printed materials, be it books, labels, or signboards, independently.

Numerous assistive systems have been created with object detection and speech interaction technology. A voice assistant system utilizes YOLOv3 to detect and announce nearby objects. Nevertheless, pre-deep learning object detection algorithms deliver poorer detection accuracy than contemporary deep learning models. The improved neural network architecture of

YOLOv8 is responsible for its error reduction and improved detection accuracy.
The major contribution of this research is design of intelligent assistive system integrated with real-time object detection, OCR text recognition, multi-lingual speech interaction, and distance estimation in single framework. The system functions offline, utilizing an edge computing architecture, which improves response time and enhances data privacy.

II. LITERATURE REVIEW

Artificial intelligence-based assistive technologies have given a new dimension to accessibility for the visually impaired. Deep learning methods enhance the performance of object recognition systems through automatic extraction of hierarchical features [9] [10].
The YOLO object detection algorithm, proposed by Joseph Redmon treat the problem of object detection as a regression problem. It predicts different bounding boxes and class probabilities directly from images. YOLO has a very high detection speed that can detect multiple objects at once.
Syahrudin and his colleagues developed an assistive object detection system using YOLOv8 integrated with OpenCV distance estimation technique for visually impaired assistance [2]. The study shows better precision than previous YOLO models.
Assistive systems that are built on OCR let visually impaired read printed text. Saju and colleagues introduced a reading system that utilizes OCR technology for image text extraction. The project made use of the Tesseract OCR engine and speech synthesis. This system of smart reader works on raspberry pi and OCR engine to convert print text to Speech output [5].
The voice assistant system using Yolov3 gives audio feedback on the detected objects and assists the visually impaired in identifying the surrounding environment [6] [11].
Smart mobility assistant utilizes algorithm and smart sensors to detect obstacles and support navigation [7].
Despite many assistive systems have been developed and proposed, much existing system only focus on a single functionality such as object detection, text recognition, etc. According to studies, ultrasonic sensors, GPS modules and similar hardware components are required in certain systems. This increase system complexity and cost [7]. Systems based on OCR can read text, but cannot detect any objects in their environment. The accuracy of YOLOv3 and other early object detection models is lower than YOLOv8 model [2]. There is limited research on the integration of YOLOv8 object detection, OCR text recognition, multilingual speech interaction and distance estimation into a single assistive framework. That's why integrated smart vision assistant can be made for better accessibility and independence of visually impaired people.

III. METHODOLOGY

The Smart Vision Assistant comprising computer vision and speech processing aims to enable visually impaired users to perceive their surrounding environment in real-time. The architecture of the framework consists of an image acquisition module, a preprocessing module, an object detection module, etc.
A camera is used to capture real-time images of the surrounding environment. The OpenCV library applies image preprocessing techniques to the image for better image quality and accurate detection such as Grayscale conversion, noise removal, image resizing, etc.
The object detection task relies on the YOLOv8 deep learning model. The input image is divided into grid cells, and then it predicts bounding boxes and confidence score of detected objects [1]. YOLOv8 features an enhanced neural network architecture for improved detection performance [2].
The OCR module retrieves text information from images taken from printed texts like books, labels, and signboards. Tesseract OCR engine utilizes shape identification techniques to convert text from image to machine readable format [3].
The extracted text is translated in English, Hindi, Marathi, etc. A multilingual speech interface will improve usability for all users. processed speech is produced by the text-to-speech engine.
The architecture of edge computing allows the system to handle the data on-site without the need to connect to the internet, thus allowing quicker response time as well as privacy [7].
3.1 Experimental Setup
The programming for the system was done using Python with OpenCV and deep learning. YOLOv8 was employed for the object detection assignment. The Tesseract OCR engine was used to recognize text. I

used the pyttsx3 Speech Synthesis library to implement Text to Speech conversion.
The evaluation used a dataset containing real-time webcam images from an indoor setting. The dataset contains common objects like chair, bottle, person, mobile phone, laptop, etc. Printed books, labels and signboards, as captured under normal lighting conditions. Accuracy metric was used to assess performance evaluation.
Accuracy = Correct Predictions / Total Predictions

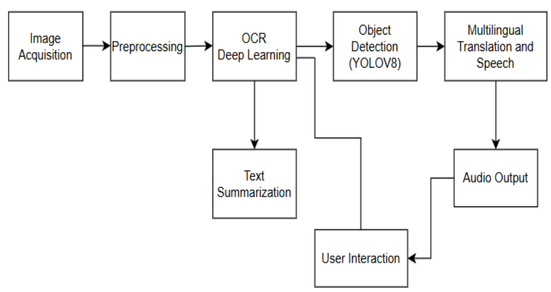


Fig 1: System Architecture

3.2 System Overview
3.2.1 The model that we propose consist of two layers:

- Hardware Layer: responsible for capturing images.
- Software Layer: software component processes the captured image via artificial intelligence codes and generates speech.

3.2.2 The system components consist of:

- Camera Module: Captures real-time images or text.
- Processing Unit: refers to the main system where the AI algorithms interpret the images taken.
- Object Detection Module (YOLOv8 Identifies objects like chairs, tables, people, and bottles in captured images.
- OCR Module (Tesseract OCR): Extracts printed text from images, such as books, labels, signboards et
- Text-to-Speech Module: Recognized object names and extracted text will be spoken aloud.
- Audio Output: Provides voice output to the user.

3.3 Processing Steps: -
3.3.1 Image Acquisition: The camera takes images of the surrounding environment instantly in real-time which may consist of objects and printed text.

3.3.2 Image Preprocessing: It enhances picture quality and improves detection accuracy. The OpenCV library helps in preprocessing the image by conversion into grayscale, resizing and noise removal to enhance the performance of OCR.

3.3.3 Object Detection (YOLOv8): It detects surrounding objects, such as human, vehicles, furniture.

3.3.4 Text Recognition using OCR: The OCR part pulls out text from images having printed text such as books, labels, and signboards. The Tesseract OCR engine transforms image texts into machine-readable format.

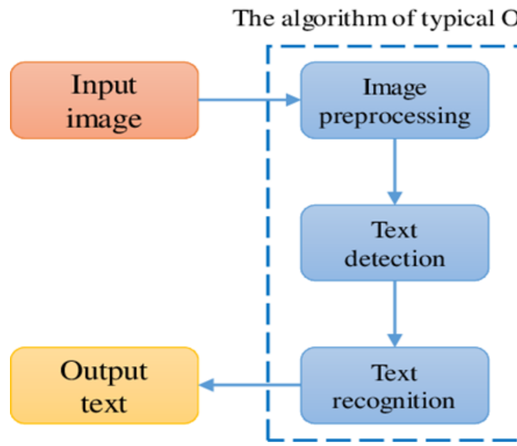


Fig 2: Algorithm of Typical OCR

3.2.5 Text Processing: The pyttsx3 library is used in the text-to-speech module to convert the object name and the extracted text into speech output.

3.2.6 Multilingual Translation and Speech: Switches from text or summarized content to speech on the given content using pyttsx3 or gTTS. The translation of Marathi, Hindi, and English id done using Marian MT or Google Translate API.

3.2.7 Speech Output: The module uses pyttsx3 library to convert the detected object names and text extracted into speech.

Table 1: Tools and Technologies

Sr. No	Component / Module	Tool / Technology Used
1.	Programming Language	Python
2.	Object Detection	YOLOv8 (Deep Learning Model)
3.	OCR (Text	Tesseract OCR

	Recognition)	
4.	Image Processing	OpenCV
5.	Speech Conversion	pyttsx3
6.	Libraries	NumPy, TensorFlow / PyTorch
7.	Camera	Webcam / Mobile Camera
8.	Audio Output	Speaker / Headphones

IV. RESULTS AND DISCUSSIONS

4.1 Result

The following results were obtained from a vision camera system after the implementation.

The efficiency of the proposed Smart Vision Assistant was tested using real-time image input taken from a cam in the indoor environment. The authors tested the YOLOv8 object detection model against widely available object classes which include person, chair, bottle, phone, laptop. The system has shown consistent detection capability and high confidence scores for most objects in the scene. The model efficiently managed to detect both single and multiple objects at the same time very successfully.

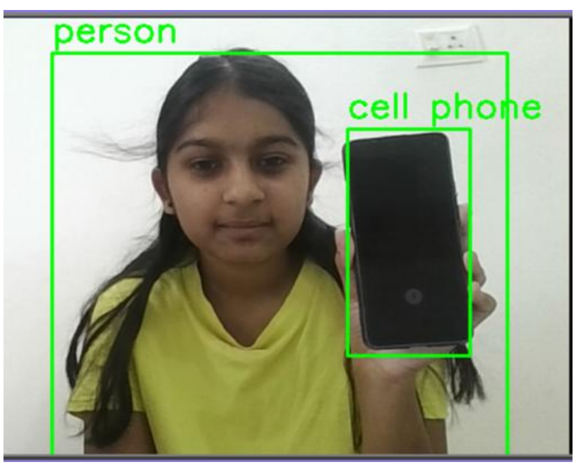


Fig 3: Object detection output using YOLOv8

Table 2: Object Detection Accuracy

Sr. No	Object	Accuracy (%)
1.	Person	95
2.	Chair	90
3.	Bottle	95
4.	Mobile	95

Table 2 showcases the results of object detection demonstrate that YOLOv8 gives consistent results when detecting frequently found indoor items. Detection accuracy is high when detected objects can be easily seen, and adequate light levels exist in the area.

The YOLOv8 framework offers a notable level of enhancement in detection performance as a result of the optimization of neural network architecture for performance and the improved feature extraction capability. When compared with previous object detection models, YOLOv8 can better cope with objects of different sizes and orientations [2]. The system could handle frames in real time with very little delay, which is very useful for applications that require constant awareness of the environment.

We have tested the OCR module on printed text images collected from books, product labels, and signboards. The Tesseract OCR engine has managed to decode text from captured images and converted it into machine readable text. By changing the image to grayscale, thresholding the images, removing noise, and so on, the text would become clearer and enhance recognition accuracy [3]. OCR results demonstrate that the system can accurately recognize printed text with clear font structure.

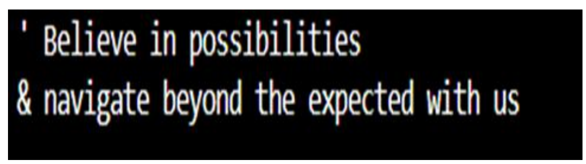


Fig 4: OCR text recognition output using Tesseract OCR

Table 3: OCR Recognition Accuracy Results

Sr. No	Text Type	Total Samples	Correct Recognition	Accuracy (%)
1.	Book Text	10	9	95
2.	Printed Label	10	9	95
3.	Signboard	10	7	85
4.	Document	10	7	85

Table 3 provide results from OCR demonstrate that the system extracts printed text well from images of printed material taken under normal lighting conditions. The accuracy of recognition is affected minimally by blurriness in an image as well as font type and background noise.

Multilingual translation module worked well. It changed the extracted text into languages. These languages include English, Hindi and Marathi. This helps people who use these languages. They can use the system easily.

The speech synthesis module made a voice. It used a text-to-speech engine. The voice output was easy to understand. The system gave live feedback. It described objects and text.

The distance estimation function gave an idea. It told the distance, between the object and the camera. This helps people who're blind. They can know how close objects are. They can avoid hitting things.

The test results are good. They show that object detection, text recognition and speech interaction work together. They help people who're blind. They make it easier for these users to get around.

4.2 Discussion

The results show that learning based object detection models like YOLOv8 enable people who are visually impaired to understand their environment better. YOLOv8 provides improved detection accuracy compared to previous versions due to enhanced feature extraction capability.

YOLOv8 improves environmental understanding by detecting multiple objects simultaneously with higher accuracy. The OCR module works well for extracting text from images. To get results from the OCR module the images need to be cleaned up first. This means making the text clearer and removing anything that's not text. The OCR module can work differently depending on the light the size of the text and the type of font used. The results are good enough for people who need help.

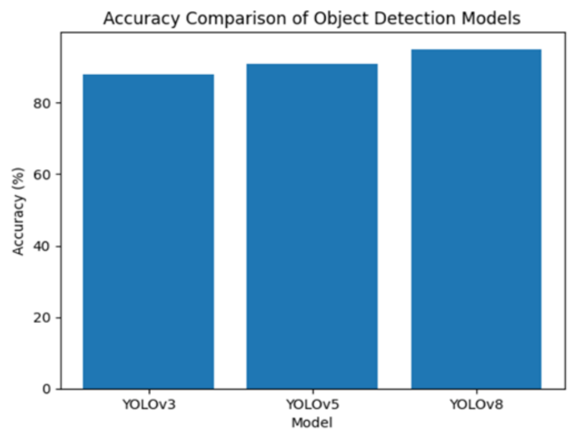
People who use this system can get information in the language they prefer because the system can talk to them in languages. The system supports English, Hindi and Marathi which means many people can use it. The system can also generate audio in time so people who are blind can understand what is being displayed.

The system works well without internet because it uses edge computing. This means that the system can process information locally which makes it faster and more private.

The proposed system integrates multiple assistive features in a single framework. Other systems can only detect objects. Recognize text. This system can do many things, which makes it more useful for people who need help.

The results show that using YOLOv8 for object detection OCR, for text recognition and multilingual voice communication really helps people who're visually impaired. It makes them more independent and able to do things on their own.

Future improvements may include integration of GPS navigation, object tracking, and mobile application interface to enhance usability.



V. CONCLUSION

In this paper, an intelligent Smart Vision Assistant capable of assisting visually impaired individuals through the integration of YOLOv8 object detection, optical character recognition (OCR), multilingual translation, and speech interaction is presented. The deployment of the system has made it practical to comprehend the environment by detecting surrounding objects and reading the text from the images in real-time.

The YOLOv8 deep learning model has improved accuracy over earlier object detection algorithms. With the help of the OCR module, printed text can be recognized from books, labels and signboards allowing visually impaired to access text independently. Users can get info in English and Hindi and Marathi through multilingual speech interaction.

The experimental evaluation shows that the proposed system performs reliably for an object detection and text recognition task. The system has an edge computing architecture that minimizes response time while preserving the privacy of captured images.

The Smart Vision Assistant set to be produced would be a low-cost solution the disabled can use to enhance independence. The combination of computer vision and speech technologies shows how AI can create inclusive assistive systems.

In future work, the mobile application interface, GPS-based navigation assistance, face recognition, and real-time object tracking may be integrated to improve the system.

The development of intelligent assistive technologies is catered through the proposed system thus improving accessibility of visually impaired.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016

[2] E. Syahrudin, A. S. Prabuwono, and M. R. Yaakob, "YOLOv8-based object detection for assistive systems," *International Journal of Advanced Computer Science and Applications*, 2024

[3] S. Saju, R. S. Kumar, and A. R. Nair, "Optical character recognition based assistive system for visually impaired," *International Journal of Engineering Research and Technology (IJERT)*, 2019

[4] D. Raj, S. Karthik, and M. Kumar, "Assistive reading system for visually impaired using OCR and text-to-speech," in *Proc. IJERT Conf.*, 2018

[5] S. Sharma, P. Gupta, and A. Verma, "Drushti: A smart reader for visually impaired people," *International Journal of Engineering Research and Technology (IJERT)*, 2018S. Madhar, A. Khan, and R. Patel, "Voice assistant for visually impaired using YOLO algorithm," *International Journal of Innovative Research in Technology*, 2022

[6] S. Mallikarjun and S. R. Kawale, "Smart mobility assistant for the visually impaired using machine learning," *Journal of Information Systems Engineering and Management*, 2025

[7] S. Mubeen *et al.*, "Optical character recognition using Tesseract," *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, 2022

[8] R. Goel and P. Sharma, "Deep learning-based object recognition techniques," *International Journal of Computer Science*, 2021

[9] Rohini and A. Avinash, "Smart vision-based navigation system for visually impaired using deep learning," *Journal of Computer Technology and Applications*, pp. 417–425, 2021

[10] J. Dharanidharan, R. Puviarasi, and S. R. Boselin Prabhu, "Object detection system for blind people," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 1675–1676, 2019

[11] H. K. Rathod *et al.*, "Smart assistive system for visually impaired using computer vision," *International Journal of Innovative Research in Computer Science and Technology*, 2022