International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 4, pp 160-165, April-2026

https://doi.org/10.47001/IRJIET/2026.104023

VOCASIGHT: An AI Assistive Navigation System for the Visually Impaired

[1]Sakshi Asodekar, [2]Madhumita Ghosh, [3]Harshna Patil, [4]Harsh Gaikwad, [5]Tularam Bansode

[1,2,3,4]Student, Department of CSE (AI & ML), Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India

[5]Professor, Department of CSE (AI & ML), Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India

***Abstract* - Vocasight is a hybrid assistive system designed to support visually impaired individuals by integrating software and hardware technologies for real-time navigation and environmental awareness. The system uses computer vision, machine learning, and embedded systems to perform object detection, face recognition, and scene understanding, providing audio feedback through Text-to-Speech. A mobile application developed using Android and Flutter incorporates libraries such as OpenCV, YOLOv8n, and EasyOCR for efficient image processing. Additionally, a portable hardware device based on ESP32-CAM, equipped with ultrasonic sensors and a buzzer, enables real-time obstacle detection and alerts. The system is further extendable with features like navigation assistance and emotion detection. Overall, Vocasight offers a cost-effective and user-friendly solution that enhances safety, independence, and situational awareness for visually impaired users.**

Keywords: Assistive Technology, VOCASIGHT, Visual Impairment, AI-Based Navigation, Smart Navigation System, Computer Vision, Object Detection, Obstacle Avoidance, Wearable Assistive Devices, Real-Time Navigation, Deep Learning, Image Processing, Sensor Fusion, Indoor and Outdoor Navigation, Voice Assistance.

I. INTRODUCTION

In today's technologically advancing world, providing assistive solutions for differently-abled individuals has become essential. Visually impaired individuals often face challenges in navigating their surroundings, recognizing objects, and interacting safely with their environment. Traditional tools such as white canes provide limited information and lack intelligent environmental understanding.

To address these challenges, Vocasight is proposed as an intelligent assistive system that combines mobile-based software and embedded hardware technologies. The system leverages computer vision and machine learning to detect objects, recognize faces, and understand scenes, while providing real-time voice guidance to users. The integration of hardware components further enhances obstacle detection and portability, making the system practical for daily use.

1.1 Project Aims and Objectives

To develop an intelligent and user-friendly assistive system that enhances navigation, safety, and independence for visually impaired individuals.

Objectives and Aims:

1. To implement real-time object detection using deep learning models.
2. To develop a face recognition system for identifying known and unknown individuals.
3. To provide voice-based feedback using Text-to-Speech technology.
4. To design a portable hardware device for obstacle detection using ESP32-CAM.
5. To integrate ultrasonic sensors for distance measurement and alerts.
6. To enable scene understanding for better environmental awareness.
7. To explore advanced features such as navigation assistance and emotion detection.

1.2 Background of Project

Visually impaired individuals face significant challenges in performing daily tasks independently due to lack of real-time environmental awareness. Existing assistive technologies often provide limited functionality and lack integration between intelligent software and portable hardware systems with advancements in artificial intelligence, computer vision, and embedded systems, it is now possible to build smart assistive devices that provide real-time feedback. Vocasight aims to bridge this gap by combining mobile application capabilities with hardware-based sensing to create a comprehensive and reliable assistive solution. Further the integration of advanced technologies such as Artificial Intelligence and Computer Vision has enabled the development of systems that can interpret complex real-world

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 4, pp 160-165, April-2026

https://doi.org/10.47001/IRJIET/2026.104023

environments with high accuracy. By leveraging deep learning models like YOLO (You Only Look Once), assistive applications can perform real-time object detection, scene analysis, and facial recognition. These capabilities significantly enhance the user's ability to understand their surroundings, identify obstacles, and interact more confidently with people and environments.

II. COMPONENTS

2.1 Software components for processing the system

1. YOLOv8n (Object Detection Model)

YOLOv8n (You Only Look Once version 8 – nano) is a lightweight and efficient deep learning model used for real-time object detection. In the VocaSight system, YOLOv8n is used to detect obstacles, people, and objects in the user's surroundings. The model processes input images in a single pass and outputs bounding boxes along with object labels. Its nano variant ensures faster inference while maintaining acceptable accuracy. YOLOv8n performs well in real-time environments and dynamic scenes. The detected information is used to generate navigation instructions for the user. Hence, it plays a crucial role in enabling safe and efficient navigation assistance.

2. OpenCV(computer vision)

Figure 1: OpenCV

OpenCV is an open-source computer vision library widely used for image processing and real-time visual analysis. In the VocaSight system, OpenCV is responsible for capturing frames from the camera and performing preprocessing operations such as resizing, filtering, and color conversion. These preprocessing steps are essential to improve the performance of detection and recognition algorithms. OpenCV also facilitates efficient handling of image data, enabling smooth and continuous processing. It acts as a bridge between the hardware camera input and the AI models used in the system.

3. EasyOCR

Figure 2: EasyOCR

In the VocaSight system, EasyOCR is used to detect and read text present in the surroundings, such as signboards, labels, and documents. It works efficiently even in complex backgrounds and varying lighting conditions. The extracted text is then converted into speech using Text-to-Speech (TTS) modules to assist visually impaired users. EasyOCR is lightweight, easy to integrate, and suitable for real-time applications. It enhances the system's ability to provide reading assistance and improves overall accessibility.

2.2 Hardware Components

1. ESP32-CAM Module

Figure 3: ESP32-CAM

The module captures real-time images of the user's surroundings, which are then processed by the system. It supports Wi-Fi connectivity, enabling communication with the mobile application or server. The compact size of the ESP32-CAM makes it suitable for portable assistive devices. It is capable of performing basic image processing tasks and transmitting data efficiently.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 4, pp 160-165, April-2026

https://doi.org/10.47001/IRJIET/2026.104023

2. Ultrasonic Sensor

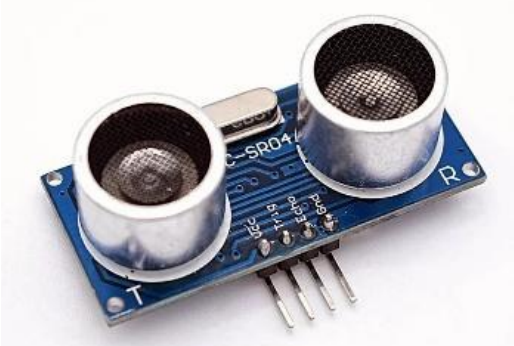

Figure: 4: Ultrasonic Sensor

The ultrasonic sensor is used to measure the distance between the device and nearby objects. It works by emitting ultrasonic waves and calculating the time taken for the waves to reflect back from an object. In the VocaSight system, it is used to detect obstacles that may not be clearly visible through the camera. The sensor provides real-time distance information, which helps in avoiding collisions. It enhances the reliability of the system by complementing the vision-based detection module.

3. Buzzer

Figure 5: Buzzer

The buzzer is an output device used to provide immediate audio alerts to the user. In the VocaSight system, it is activated when obstacles are detected within a certain distance. The buzzer produces a sound signal that alerts the user about potential hazards. This is particularly useful in situations where quick response is required. The buzzer operates with low power and can be easily integrated with the ESP32-CAM module.

III. METHODOLOGY

The development of Vocasight is carried out in the following phases:

1. Requirement Analysis:

Identify the needs of visually impaired users and define system functionalities.

2. System Design:

Design architecture for both mobile application and hardware integration.

3. Software Development:

Implement object detection, face recognition, OCR, and TTS functionalities using AI libraries.

4. Hardware Development:

Develop a portable device using ESP32-CAM and ultrasonic sensors for obstacle detection.

5. Integration:

Connect software and hardware modules to enable real-time communication.

6. Testing:

Perform functional and performance testing to ensure system reliability.

7. Deployment:

Deploy the system as a mobile application with a portable hardware device.

IV. RESULT

The Vocasight system successfully detects objects, recognizes faces, and provides scene descriptions through voice feedback. The hardware module effectively detects nearby obstacles using ultrasonic sensors and alerts users via a buzzer.

The system demonstrates low latency and reliable performance, ensuring timely alerts for user safety. Initial testing indicates improved navigation capability and situational awareness for visually impaired users.

4.1 Object Detection and Navigation

The system successfully detects obstacles in real time using the YOLO model and segments the scene into regions for navigation. Based on obstacle position, directional guidance such as “Move Left” is generated to assist the user in safe navigation.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 4, pp 160-165, April-2026

https://doi.org/10.47001/IRJIET/2026.104023

Figure 6: Object Detection

4.2 Face Recognition

The face recognition module accurately identifies known individuals and labels unknown faces using pre-trained models. The system provides real-time identification, enabling users to recognize people in their surroundings.

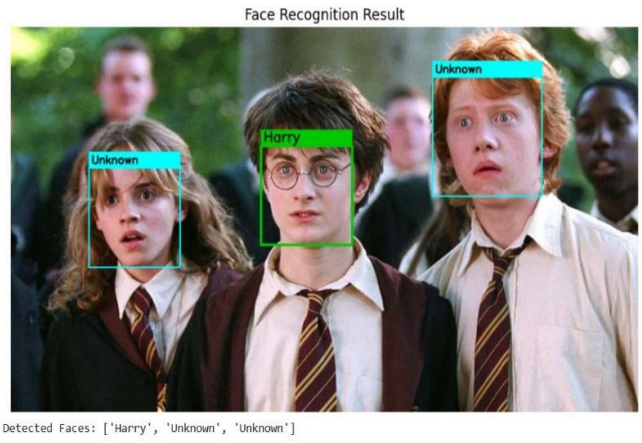

Figure 7: Face Recognition

4.3 Emergency Detection

The system analyzes the scene and detects hazardous situations such as fire using image captioning and classification techniques. It generates safety alerts and provides immediate instructions to ensure user safety.

Figure 8: Emergency Detection

4.4 Application Interface

The VocaSight mobile application interface is designed to be simple, intuitive, and accessible for visually impaired users. The interface includes a central display area showing real-time detection results, such as recognized faces or objects. A microphone button is provided to enable voice commands, allowing users to control the system hands-free using instructions like “scan” or “stop.” The application also displays the current system status (e.g., idle or active) and identified outputs, such as recognized names. The use of clear visuals and voice interaction enhances usability and ensures ease of operation. Overall, the interface focuses on providing a seamless and user-friendly experience for assistive navigation.

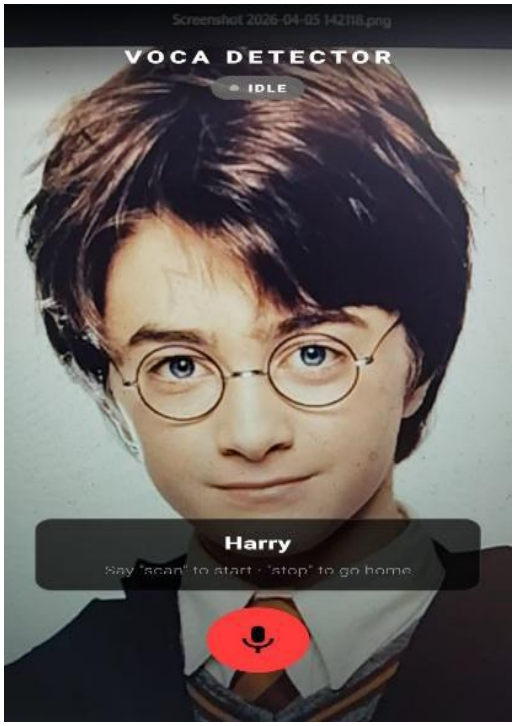

Figure 9: Application Interface

ISSN (online): 2581-3048
Volume 10, Issue 4, pp 160-165, April-2026
https://doi.org/10.47001/IRJIET/2026.104023

4.5 Hardware Implementation of the Vocasight system

The hardware prototype of the VokaSight system integrates the ESP32-CAM module, ultrasonic sensor, and buzzer on a breadboard to enable real-time obstacle detection and alert generation. The ultrasonic sensor measures the distance of nearby objects, while the ESP32-CAM processes the input and controls the system operations. Upon detecting obstacles within a threshold distance, the buzzer is activated to provide immediate audio alerts to the user. The setup is compact, cost-effective, and suitable for portable assistive applications.

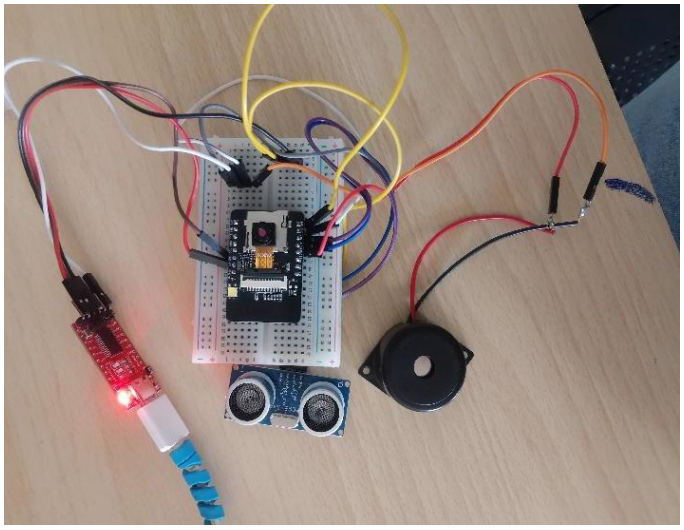

Figure 10: Hardware Implementation in real life

V. CONCLUSION

VocaSight provides an effective and innovative solution for assisting visually impaired individuals by integrating artificial intelligence and embedded systems. The system enhances user independence by offering real-time navigation support, object detection, and voice-based interaction.

The combination of mobile application and portable hardware ensures usability, reliability, and efficiency. This project demonstrates the potential of assistive technologies in improving quality of life and safety for visually impaired individuals.

VI. FUTURE SCOPE

1. Integration of GPS-based navigation for outdoor assistance.
2. Implementation of emotion detection using deep learning models.
3. Development of wearable devices such as smart glasses.
4. Multi-language voice support for wider accessibility.
5. Cloud-based processing for improved accuracy.
6. Enhanced battery efficiency and compact hardware design.

ACKNOWLEDGEMENT

As every project is ever complete with the guidance of experts. So, we would like to take this opportunity to thank all those individuals who have contributed in visualizing this project. We express our deepest gratitude to our project guide and coordinator Prof. Tularam Bansode (CSE (AIML) Department, Smt. Indira Gandhi College of Engineering, Ghansoli) for his valuable guidance, moral support and devotion bestowed on us throughout our work. We extend our sincere appreciation to our HOD ma'am and entire professors from Smt. Indira Gandhi College of Engineering for their valuable inside and tip during the designing the project. Their contributions have been valuable in many ways that we find it difficult to acknowledge them individually.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
[2] Open-Source Computer Vision Library, "OpenCV Documentation," [Online]. Available: https://opencv.org/
[3] A.Geitgey, "Face Recognition Library," [Online]. Available: https://github.com/ageitgey/face_recognition
[4] Google, "Tesseract OCR Engine," [Online]. Available: https://github.com/tesseract-ocr/tesseract
[5] Espressif Systems, "ESP32-CAM Technical Reference Manual," [Online]. Available: https://www.espressif.com/
[6] D. Jurafsky and J. H. Martin, Speech and Language Processing, *3rd ed., Pearson*, 2020.
[7] Flutter Documentation, "Flutter SDK," [Online]. Available: https://flutter.dev/
[8] Android Developers, "Android OS Documentation," [Online]. Available: https://developer.android.com/
[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE CVPR*, 2016.
[10] World Health Organization (WHO), "World Report on Vision," 2019.
[11] A.Howard et al., "Searching for MobileNetV3," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019.
[12] M. Sandler et al., "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE*

ISSN (online): 2581-3048
Volume 10, Issue 4, pp 160-165, April-2026
https://doi.org/10.47001/IRJIET/2026.104023

Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[13] W. Liu et al., "SSD: Single Shot MultiBox Detector," *European Conference on Computer Vision (ECCV)*, 2016.

[14] S. Zhang et al., "Deep Learning-Based Object Detection for Assistive Navigation of Visually Impaired Individuals," *IEEE Access*, 2021.

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

[16] C. Szegedy et al., "Going Deeper with Convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[17] A.Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[18] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.

[19] S. Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[20] H. Bay et al., "SURF: Speeded-Up Robust Features," *European Conference on Computer Vision (ECCV)*, 2006.

[21] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[22] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning," *USENIX Symposium on Operating Systems Design and Implementation*, 2016.

[23] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, 1997.

[24] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv*:1412.6980, 2014.

AUTHORS BIOGRAPHY

Sakshi Asodekar, Pursuing Final year in B.E. CSE (AI&ML) at Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Madhumita Ghosh, Pursuing Final year in B.E. CSE (AI&ML) at Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Harshna Patil, Pursuing Final year in B.E. CSE (AI&ML) at Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Harsh Gaikwad, Pursuing Final year in B.E. CSE (AI&ML) at Smt. Indira Gandhi College of Engineering, Ghansoli, New Mumbai, Maharashtra, India.

Citation of this Article:

Sakshi Asodekar, Madhumita Ghosh, Harshna Patil, Harsh Gaikwad, & Tularam Bansode. (2026). VOCASIGHT: An AI Assistive Navigation System for the Visually Impaired. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(4), 160-165. Article DOI https://doi.org/10.47001/IRJIET/2026.104023

****

