Audio Sight: A Smart Assistive Device for Visually Impaired People

Athul Deep V.[1]; Yousuf Ali[2]; David Abu Mathews[3]; Meenakshy Bijukumar[4]; Jisha Babu[5]

[1,2,3,4,5]Amal Jyothi College of Engineering

Publication Date: 2025/03/18

Abstract: This paper presents Audio Sight, an innovative assis- tive technology designed to enhance the independence of visually impaired individuals. The smart cap, powered by a Raspberry Pi, integrates real-time facial recognition, OCR-to-speech, object detection, and an emergency alert system. The device helps users recognize family members, detect objects in their surroundings, read text aloud, and ensure safety through an emergency alert mechanism. This paper discusses the design, implementation, and future improvements of Audio Sight, highlighting its significance in smart healthcare.

Keywords: *Assistive Technology, Facial Recognition, Object Detection, OCR-to-Speech, Smart Cap, Raspberry Pi.*

How to Cite: Athul Deep V.; Yousuf Ali; David Abu Mathews; Meenakshy Bijukumar; Jisha Babu. (2025). Audio Sight: A Smart Assistive Device for Visually Impaired People. *International Journal of Innovative Science and Research Technology*, 10(3), 226-230. https://doi.org/10.38124/ijisrt/25mar156.

I. INTRODUCTION

Visual impairment is a major challenge affecting millions of people worldwide, significantly impacting their ability to perform daily tasks independently. Individuals with visual impairment often rely on assistive technologies such as canes, guide dogs, and smartphones to navigate their surroundings. However, these traditional methods have limitations in terms of accessibility, ease of use, and functionality. Recent advances in artificial intelligence (AI), computer vision, and embedded systems have paved the way for the development of more sophisticated assistive devices.

The *Audio Sight* smart cap is designed to address the challenges faced by visually impaired individuals by integrating multiple cutting-edge technologies. This device leverages real-time facial recognition to help users identify family members, friends, and acquaintances, enhancing their social interactions. Additionally, the smart cap features an object detection system that assists users in identifying everyday objects in their environment, significantly improving mobility and safety. The OCR-to-speech functionality enables users to read printed text through audio feedback, providing access to written information such as street signs, books, and labels. Furthermore, the emergency alert system ensures that users can quickly notify caregivers or emergency contacts in case of distress.

By incorporating these advanced features into a lightweight and portable design, *Audio Sight* aims to enhance the quality of life for visually impaired individuals. The following sec- tions detail the system's hardware and software components, methodology, experimental results, and future enhancements, demonstrating its potential as an innovative and practical assistive device.

II. SYSTEM OVERVIEW

A. Hardware Components

The Audio Sight device is equipped with an integrated set of hardware components that enable real-time object and facial recognition, OCR-based text reading, and emergency alert features. These include:

- **Raspberry Pi**: The primary computational unit, responsible for processing image data, running AI algorithms, and controlling the various functions of the device.vice.
- **Camera Module**: Captures images for facial recognition, object detection, and OCR, ensuring accurate real-time analysis.
- **Microphone & Speaker**: Provides real-time auditory feedback, allowing users to hear identified objects, text, and recognized individuals.
- **Power Supply**: The rechargeable battery ensures uninterrupted usage, making it practical for long-term use.

- **Emergency Alert System**: An emergency button is integrated into the device to notify emergency contacts in the event of an emergency.

As shown in Figure 1, the Audio Sight device integrates various hardware components to enhance accessibility and user safety.

B. Software Framework

The software framework integrates various models powered by AI and open source tools to facilitate different functional- ities. These include:

- **OpenCV**: Utilized for real-time image processing, facial recognition, and object detection, enhancing the effi- ciency of the system.
- **Tesseract OCR**: Responsible for converting captured text into digital format and subsequently into speech.

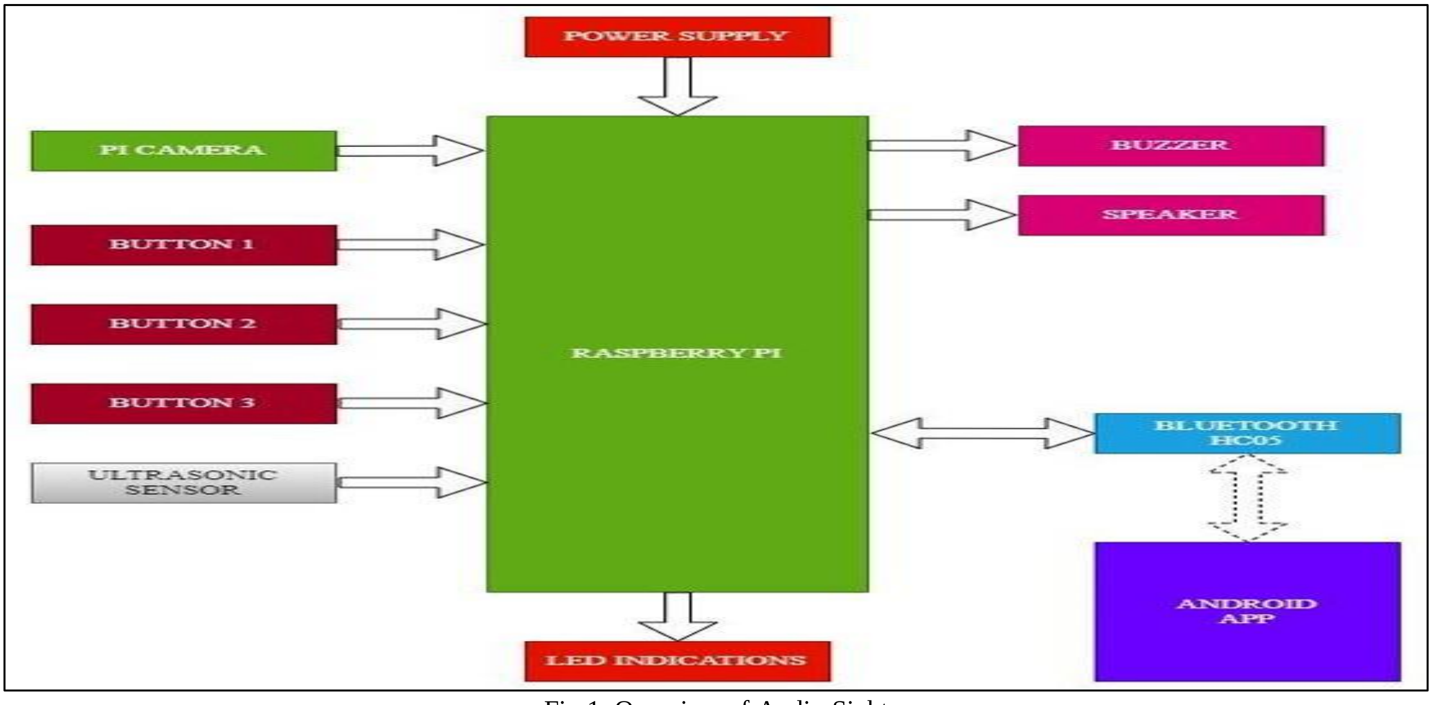

Fig 1: Overview of Audio Sight

- **Google Text-to-Speech (gTTS)**: Converts the extracted text into an audible format for the user.
- **YOLO Object Detection**: Helps recognize and classify objects in the user's environment, improving situational awareness.

III. METHODOLOGY

A. Facial Recognition

Facial recognition is implemented using OpenCV's FaceNet algorithm [3]. The system captures an image, encodes the facial features, and compares them with a stored database. If a match is found, the name of the recognized individual is announced through the speaker. Steps in the facial recognition process:

- Capturing facial images through the camera module.
- Extracting facial landmarks using OpenCV.
- Encoding the extracted features using FaceNet.
- Comparing the encoded face with stored profiles.
- Announcing the identified name through the speaker.

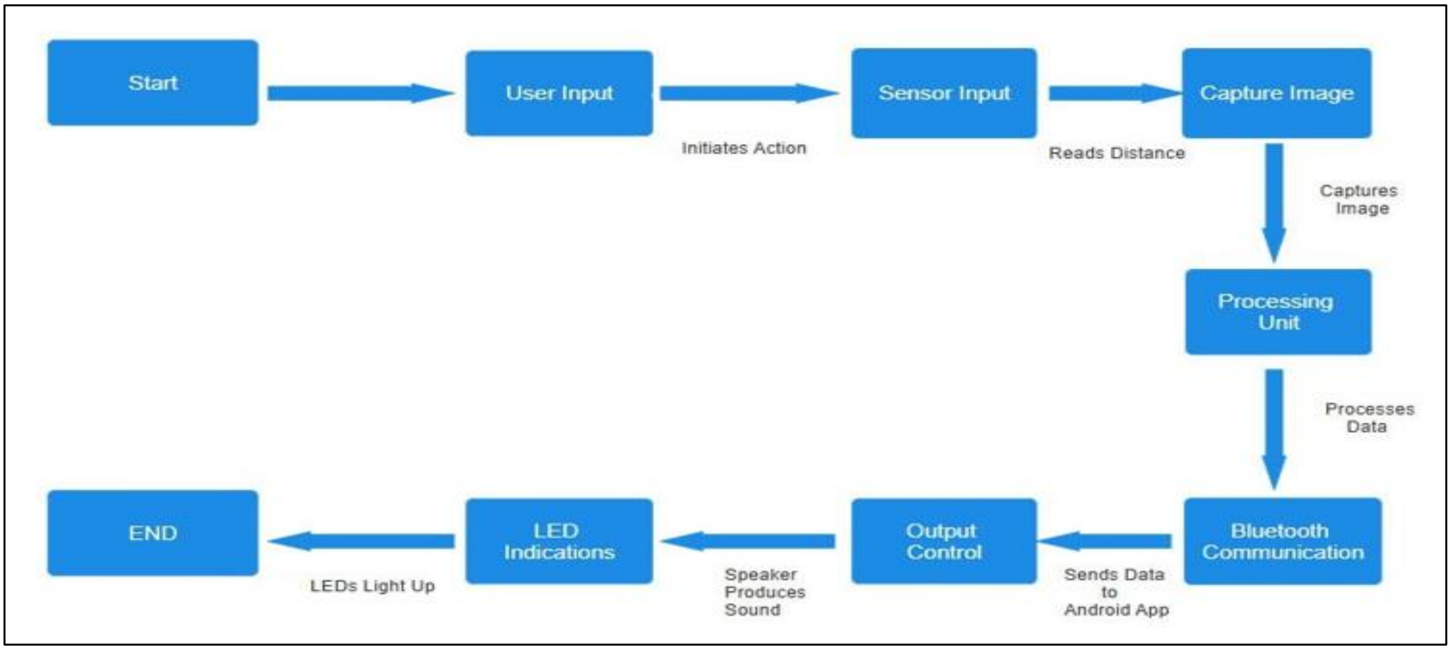

Fig 2: Block Diagram of Audio Sight

As shown in Figure 2, the facial recognition module follows a structured process to ensure accurate identification.

B. Object Detection

Object detection is performed using the YOLO deep learning algorithm [4]. The system processes the environment in real time, identifying objects and providing auditory feedback about their presence and location. Steps in the object detection process:

- Capturing an image from the camera.
- Processing the image with the YOLO model.
- Detecting and classifying objects.
- Announcing the detected objects through the speaker.

C. OCR-to-Speech

OCR-to-speech functionality enables users to read printed text. Using Tesseract OCR [5], the system extracts text from captured images and converts it into speech through gTTS.

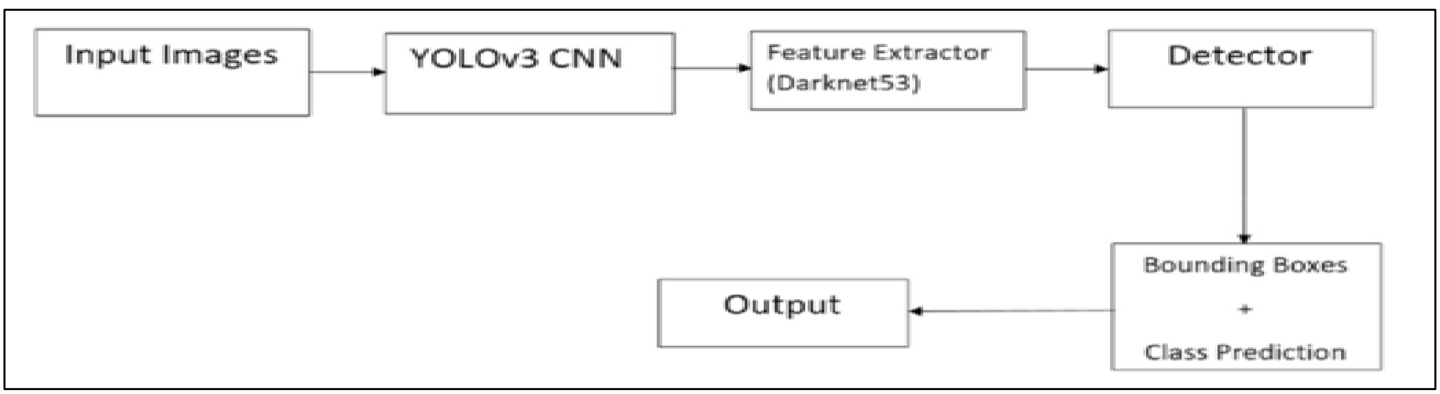

Fig 3: DFD of OCR-to-Speech Module

As shown in Figure 3, the OCR-to-Speech module processes text extraction and conversion efficiently.

D. Emergency Alert System

The emergency alert system is designed for quick distress signaling. When the emergency button is pressed, the system sends a predefined message with the user's GPS location to emergency contacts via a GSM module. Steps in the emergency alert process:

- User activates the emergency button.
- The system retrieves the user's current location.
- A distress message is sent to the emergency contacts.
- The system provides audible feedback confirming the alert has been sent.

As shown in Figures 4 and 5, the emergency alert system ensures timely notifications for enhanced user safety.

IV. RESULTS AND DISCUSSION

A. Performance Evaluation

The system was evaluated with 20 visually impaired participants in different environments. The evaluation included recognition accuracy, response time, and usability. The results showed:

- Facial Recognition Accuracy: 92% in well-lit condi- tions, with minor performance drops in low-light envi- ronments.
- Object Detection Accuracy: 87%, with challenges identifying small objects.
- OCR Text Recognition Rate: 85%, particularly efficient on printed materials.

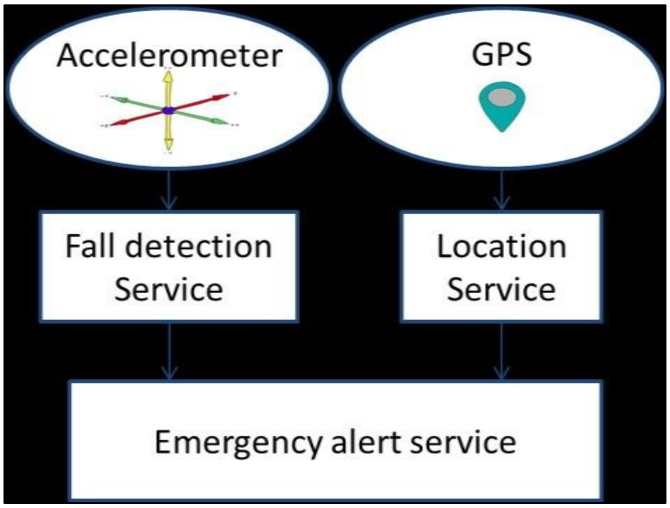

Fig. 4. DFD of Emergency Alert Module

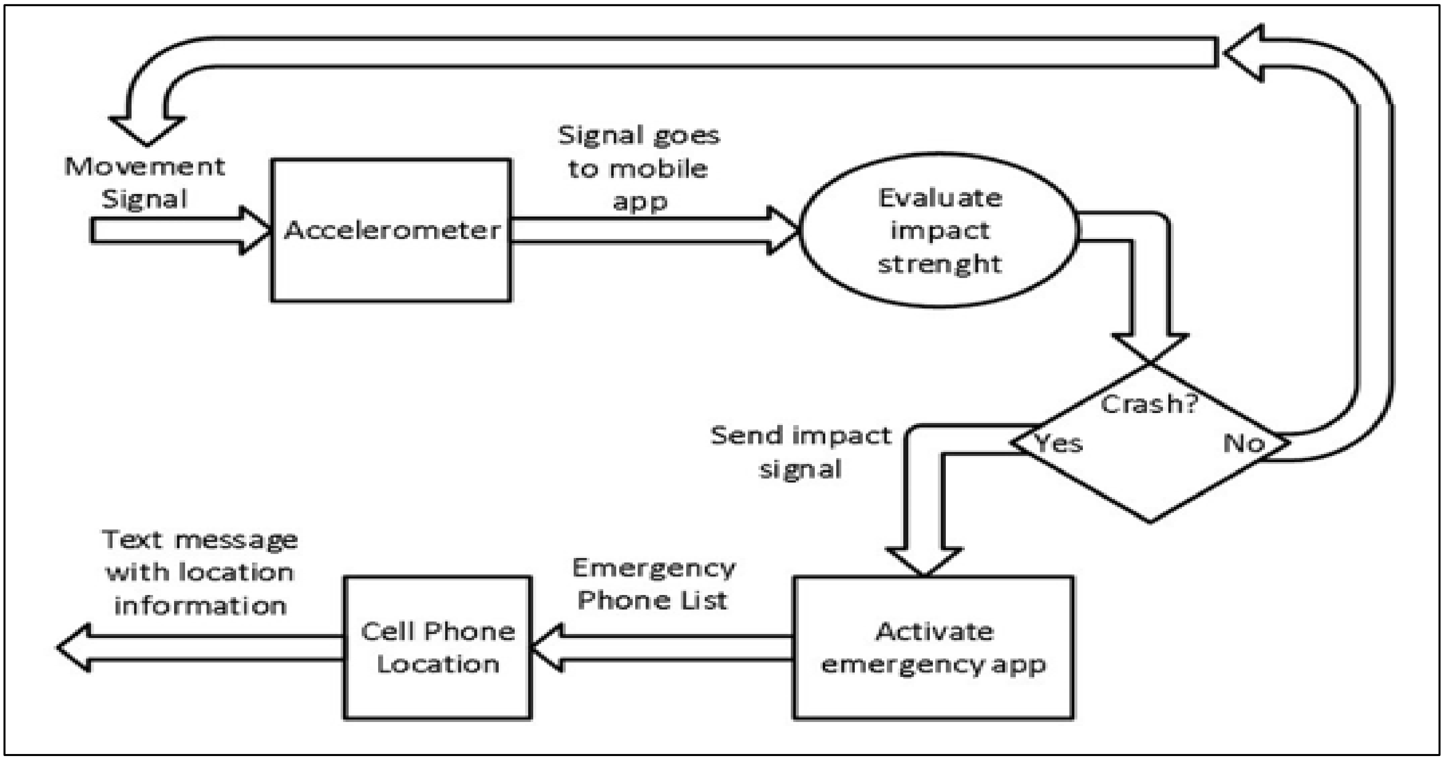

Fig 5: Emergency Alert Flowchart

- Emergency Alert Response Time: The system success- fully sent alerts within 5 seconds on average.

B. Challenges and Improvements

➢ *Several limitations were observed during testing:*

- **Low-Light Performance**: Infrared sensors are being explored for night-time use.
- **Handwritten OCR Accuracy**: Handwriting recognition remains a challenge.
- **Small Object Detection**: Further training of the YOLO model is required for better accuracy.

V. CONCLUSION

The Audio Sight smart cap provides an affordable and effective assistive solution for visually impaired individuals. By integrating facial recognition, object detection, OCR-to-speech, and emergency alert features, it enhances user inde- pendence and safety. Future work will focus on adding GPS

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar156

Table 1: Experimental Results

Metric	Generated Example Value
Accuracy	89.1%
Precision	91.2%
Recall (Sensitivity)	88.0%
F1-Score	89.6%
True Positive (TP)	280
True Negative (TN)	185
False Positive (FP)	35
False Negative (FN)	30

Navigation and AI-based improvements for real-time object classification and enhanced text recognition.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to all individuals and organizations that contributed to the successful completion of this research. The authors extend theirxtend their deepest appreciation to our mentors and academic advisors for their invaluable guidance, feedback, and encouragement throughout the project.

Special thanks to the participants who volunteered for testing and provided crucial insights into improving the system. We acknowledge the support of the institutions and research facilities that provided access to essential resources and tools for experimentation and development.

We are also grateful to our families and friends for their unwavering support and motivation. Lastly, we recognize the contributions of the open-source community for making available tools such as OpenCV, YOLO, and Tesseract OCR, which were instrumental in the development of *Audio Sight*.

REFERENCES

[1]. K. K. Reddy et al., "IoT-driven Accessibility: A Refreshable OCR- Braille Solution for Visually Impaired and Deaf-blind Users through WSN," *Journal of Economy and Technology*, 2024.

[2]. A. Sobowale et al., "Enabling Social Interaction: A Face Recognition System for Visually Impaired People using OpenCV," *FUOYE Journal of Engineering and Technology*, vol. 9, no. 2, pp. 203–212, 2024.

[3]. J. A. Mensah et al., "FaceNet Recognition Algorithm Subject to Multiple Constraints: Assessment of the Performance," *Scientific African*, vol. 23, e02007, 2023.

[4]. M. Sneha Merin et al., "Raspberry Pi-based Object Detection and Reading for Blind Person," *International Conference on Artificial Intelligence in Information and Communication*, 2019.

[5]. T. Khete and A. Bakshi, "Autonomous Assistance System for Visually Impaired using Tesseract OCR and gTTS," *Journal of Physics Confer- ence Series*, vol. 2327, no. 1, pp. 012065, 2022.

[6]. Uppin and G. Gorge, "VisioFace: An Android App Utilizing Deep Learning for Facial Recognition to Aid the Visually Impaired," 2023.

[7]. P. Bhargavi et al., "Smart Cap for Visually Impaired People," 2023.