International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8s, 2025
https://www.theaspd.com/ijes.php

Real-Time Object Detection For The Visually Impaired Using Yolov8 And NLP On Iot Devices

G.R. Venkatkrishnan[1], R. Jeya[2*], G. Ramyalakshmi[3], S. Sindhu[4], K. Sreenithi[5]

[1]*Associate Professor, Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, Chennai – 603 110, Email: venkatakrishnangr@ssn.edu.in*
[2*]*Associate Professor, SRM Institute of Science and Technology, Kattankalathur, Chennai, Email: jeyar@srmist.edu.in*
[3, 4, 5] *Student*
[3,4,5] *Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, Chennai – 603 110*

**Corresponding Author: R. Jeya*
** Email: jeyar@srmist.edu.in*

Abstract
Object detection is an important development in real-time that integrates and Artificial Intelligence (AI), embedded systems, and Internet of Things (IoT). These innovations aim to address the challenges faced by visually impaired individuals in locating everyday objects independently. Current systems often suffer from limitations like manual tagging, lack of realtime feedback, and poor adaptability in dynamic environments. This paper introduces an IoTbased Voice-Driven Smart Finder that leverages Natural Language Processing (NLP), YOLOv8 object detection, and cloud-based speech recognition for efficient and autonomous object location. The primary objective is to create a voice-interactive, low-cost solution that enhances the independence of visually impaired users by allowing them to locate objects using simple verbal queries. The system's novelty lies in its integration of real-time object detection, speech-based control, and optimized edge-device deployment without the need for predefined object tagging. The proposed model achieved a detection accuracy of 92% on household objects and a fast response time of approximately 1.8 seconds, highlighting its practical effectiveness. By utilizing affordable hardware like Raspberry Pi 4 and integrating cloud APIs for speech processing, this work contributes a scalable and inclusive solution to the assistive technology landscape. Keywords: YOLOv8, Object Detection, Visually Impaired, Speech Recognition

Keywords: *YOLOv8, Object Detection, Visually Impaired, Speech Recognition*

1. Introduction

Object detection has become a cornerstone of assistive technology, especially with the integration of smart embedded systems, IoT, and AI [1]. People with disabilities, particularly those who are blind or visually impaired, now depend on this technology to improve their quality of life [2]. Voice-activated and object-recognition systems are two of the most effective and allow users to engage with their surroundings more freely [3]. With advancements in computer vision and speech processing, real-time systems can now interpret and respond to human inputs, making autonomous navigation and daily object identification more feasible [4]. As society moves towards inclusivity and smart living, the demand for accessible and adaptive assistive tools continues to grow [5]. Visually impaired individuals rely heavily on memory, tactile cues, or external assistance to locate essential objects in their surroundings [6]. These traditional methods become unreliable or impractical when objects are moved or misplaced, resulting in confusion, frustration, and reduced independence. An inability to move or engage with their surroundings effectively is further hampered by congested surroundings, inconsistent illumination, and a lack of real-time spatial awareness [7.8]. Current object identification tools either require manual tagging of items or depend on static models that lack real-time adaptability and hands-free interaction, making them less practical in dynamic, everyday scenarios [9].

In recent research, several assistive systems have been proposed, which frequently face some challenges in smart canes [10], wearable cameras [11], and smartphone [12]-based applications. Many rely on manual tagging, fixed environment setups, or are constrained to detecting predefined objects [10]. Some systems lack effective audio feedback or cannot handle multiple object queries simultaneously [11]. Additionally, smartphone-based models

are often computationally intensive, leading to battery drainage and slower response times [12]. These limitations restrict their usability in real-world scenarios, especially for visually impaired individuals who require continuous, real-time assistance without depending on handheld or visual interfaces [13]. To overcome these limitations, this research introduces an IoT-based voice-driven smart finder designed specifically for real-time object detection and voice-based interaction. Built on a Raspberry Pi 4 platform, the system integrates a camera module, a USB microphone, and a portable speaker to provide a complete hands-free solution. It leverages Deep Learning (DL)-based object detection models (YOLOv8), speech-to-text/text-to-speech conversion, NLP, and s to identify objects through verbal commands. In contrast to earlier approaches, the proposed approach eliminates the need for manual tagging and pre-existing object position knowledge. This paper's primary contributions are listed below:

- A novel IoT-based system was designed using Raspberry Pi, YOLOv8, and cloudintegrated speech processing to allow visually impaired individuals to locate household objects through natural voice commands.
- The system effectively combines the Natural Language Toolkit (NLTK) toolkit and YOLOv8 for accurate extraction of object names from speech and their subsequent realtime detection.
- The YOLOv8 model was optimized for edge deployment on Raspberry Pi, achieve an average response time of 1.8 seconds and 92% detection accuracy, which confirms its suitability for low-power environments.

The paper's remaining portions are arranged as follows. In Section 2, the literature review is covered in detail. Section 3 describes the methodology. The experiment's results and observations are shown in Section 4. Future work and conclusion and are included in Section 5.

2. Literature review

In 2021, Rahman et al., [14] developed an IoT-based system that automatically recognizes things for visually impaired persons. The system recognized objects and provides auditory feedback in real time using computer vision and DL. Their study improved accessibility and independence, it facilitated everyday navigation and environmental engagement.

In 2023, Ahammed et al., [15] introduced an IoT-based smart guide stick capable of assisting the visually impaired in navigating a space. The guide stick incorporates functionalities, such as obstacle detection and water detection capabilities, gave direction guidance, voice alerts, and sharing of location through GPS and GSM modules. Therefore, the guide stick fulfils all aspects of safety and confidence in free movement through known and unknown spaces.

An ultrasonic technology-enabled computer vision-based navigation system by camera with obstacle detection has been proposed by Ghatkamble et al. (2023) [16]. The system gives both auditory and tactile feedback, supporting more accessible navigation in a diverse environment. The dual-modality feedback system comes with alerts through tactile, which addresses the differences in the hearing ability of various users, and as another layer of interaction for safe guidance.

In 2023, Guravaiah et al., [17] created the third eye technology, which allowed visually challenged people to recognize objects and generate speech. The system recognizes things and provides real-time auditory feedback using computer vision and artificial intelligence. By enhancing environmental awareness and independence, the third eye improved accessibility and daily navigation for visually impaired users.

In 2023, Leong et al., [18] developed a system that can recognize barriers and calculate distance for individuals with visual impairments. The technology uses computer vision and ultrasonic sensors to detect obstacles and estimate distance in real time. Their study offered auditory feedback and notifications, it improved mobility and safety, empowering users to comfortably traverse their surroundings.

In 2023, Senarathna et al., [19] developed a Machine Learning (ML)-based assistive system that uses ML to identify and detect items for people with vision impairments. The device provided real-time auditory feedback and object identification using AI and computer vision. It improved everyday navigation and interactions with the environment by increasing accessibility and independence.

In 2024, Kumar et al., [20] created Echo Guidance, a voice-activated assistance app for those with visual impairments that is connected to a smart stick. The technology improves autonomous navigation, detects obstructions, and sends out real-time speech notifications using ML and the IoT. It improves mobility and safety, making daily movement more accessible for blind users.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8s, 2025
https://www.theaspd.com/ijes.php

3. Proposed Methodology

The proposed system is composed of interconnected modules designed for accurate and efficient object detection. A USB or PiCamera captures live video, which is pre-processed through resizing, normalization, and color space conversion. A DL model then performs object detection. Detections are filtered based on confidence scores and annotated accordingly. To minimize latency, the system uses multi-threading for simultaneous image capture, processing, and feedback. The final output is delivered to the user through audio and optionally via a display. Figure 1 represents an IoT-based voice-driven smart finder system, where a Raspberry Pi handles voice commands using speech-to-text and NLP to locate and report object positions audibly.

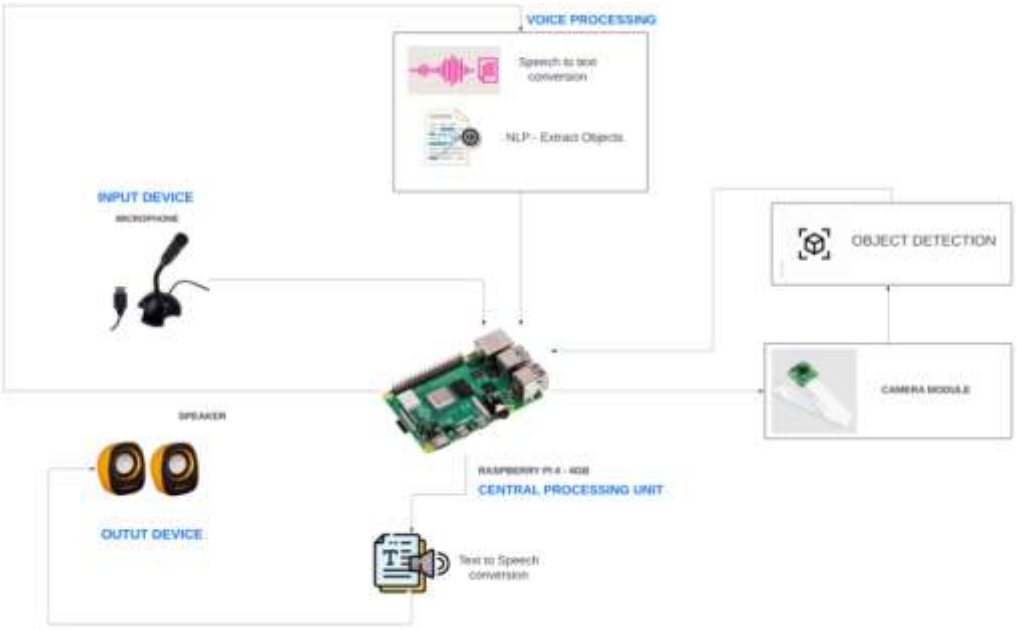

Figure 1: Overall framework of proposed methodology

3.1 Data collection

The microphone is a key input device, enabling hands-free voice control for visually impaired users. It captures real-time audio, which speech recognition processes to detect commands and provide feedback. Voice input enhances accessibility, making interactions more intuitive. An external microphone offers better audio quality, noise reduction, and voice recognition accuracy than built-in options. Its sensitivity and frequency response ensure clear speech detection, even in noisy environments, improving usability in outdoor or public spaces.

3.2. Data preprocessing

Preparing voice commands for item recognition through the use of various NLP approaches and the NLTK requires data preprocessing. The first step, Tokenization, breaks down the spoken query into individual components like words or subwords, allowing for easier analysis of each term. Next, Part-of-Speech (POS) Tagging assigns grammatical roles to these tokens (e.g., noun, verb, pronoun), which helps the system understand the context and structure of the sentence. Following this, Stopword Removal eliminates common words such as “is,” “the,” and “my,” which do not contribute meaningful information for object identification, thus reducing processing load. Finally, Named Entity Recognition (NER) identifies and classifies important entities like brand names, locations, or specific object names (e.g., "Apple" as a brand), ensuring accurate interpretation of the user's intent. Together, these four steps enhance the system’s ability to extract relevant information from voice commands efficiently and accurately.

3.3. Object detection using YOLOv8

YOLOv8 does not require predetermined anchor boxes because it uses a single-stage, anchor-free detection method. Instead, it employs key-point and center-based detection strategies, enhancing object detection

efficiency. Unlike traditional models, YOLOv8 does not rely on edge-detection kernels, making it lightweight and suitable for real-time processing. The Backbone, which extracts key features; the Neck, which combines data from deep and shallow layers to enhance feature resolution; and the Head, which handles object detection, are the three primary parts of Figure 3. Non-Maximum Suppression (NMS), a post-processing technique that eliminates superfluous bounding boxes, further refines the results. YOLOv8's efficiency and adaptability make it an excellent choice for applications requiring accurate and rapid object detection. Figure 3 shows the distinctions between object detection models with one step and those with two stages.

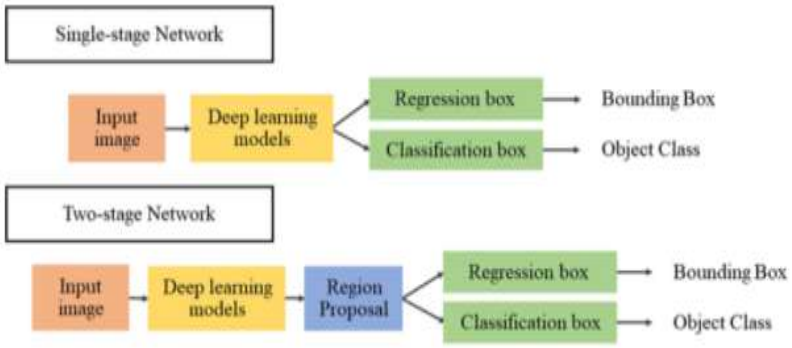

Figure 3: Single-Stage and Two-Stage Object Detection Networks

3.3.1. Text-To-Speech (TTS) Conversion

One crucial technological tool that connects written and spoken communication is textto-speech (TTS) conversion. It plays a crucial role in assistive technologies, AI-driven virtual assistants, automated customer service systems, and multimedia applications. This section explores Text-to-Speech conversion in-depth, focusing on its implementation using pyttsx3, the text processing techniques involved, the speech synthesis process, execution methods, and a comparative analysis with cloud-based TTS services

3.4. Model Deployment on Raspberry Pi

Following refinement, the Raspberry Pi 4B has the improved YOLO model installed to enable real-time object detection in a resource-constrained environment. To ensure smooth performance, the model is first converted into a lightweight ONNX format suitable for edge deployment. For file transfer, FileZilla a secure FTP client is used to migrate the model and related files from the development system to the Raspberry Pi over an SSH connection. This approach ensures efficient and reliable deployment, allowing the embedded system to execute the model locally with minimal latency. Figure 4 shows the FileZilla interface used during model transfer.

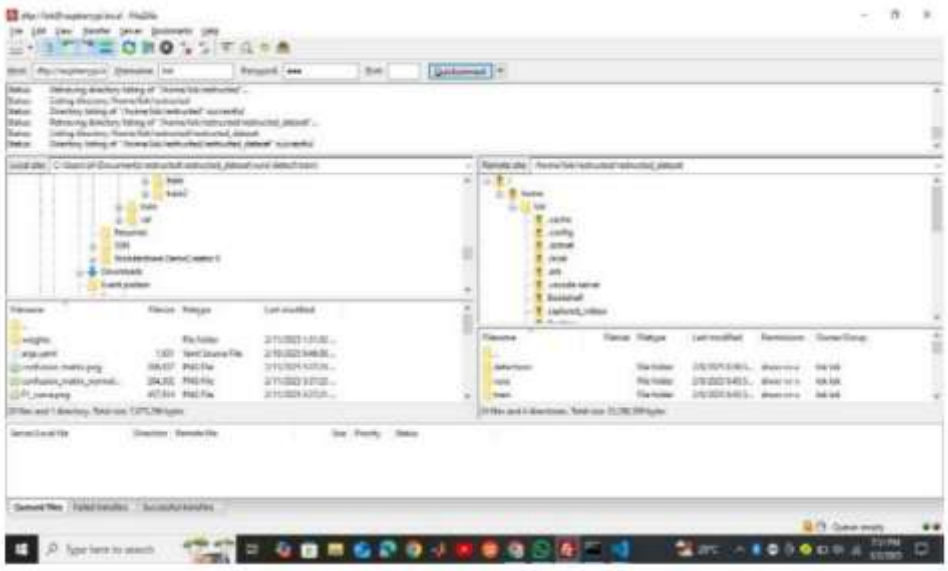

Figure 4: FileZilla interface while handling the file transfer

4. Results and Discussion

The results demonstrate the effectiveness of the proposed approach in accurately and instantly identifying items with voice instructions. The integration of DL and speech processing techniques has led to reliable performance in identifying everyday objects. The discussion highlights how the system responds efficiently to user input, processes images accurately, and delivers audio feedback with minimal delay.

4.1. Dataset description

This study uses a curated dataset of 10,000 images derived from the COCO (Common Objects in Context) dataset, specifically formatted for YOLO-based object detection. The dataset includes labeled bounding boxes converted from COCO's JSON format to YOLO format for better model compatibility. To ensure balance and diversity, 25 relevant object classes were selected, with each class having approximately 400 samples. The dataset was split into 20% testing (2,000 images) and 80% training (8,000 images) using stratified sampling. This configuration makes it possible to train the object detection model effectively and assess its effectiveness in real-time situations.

4.2. Hardware-Software Integration

The effectiveness of the system relies on a carefully structured interaction between various hardware components, including the processing unit, input devices, and output peripherals, all of which work together to capture, process, and deliver information efficient which is displayed in Figure 5.

Figure 5: Hardware setup of Smart Finder

The Raspberry Pi 4 Model B's quad-core Cortex-A72 SOC, the Broadcom BCM2711, performs better than previous versions. In addition to having two USB 3.0 and two USB 2.0 connections for quick data transfer, it supports a variety of RAM configurations. Gigabit Ethernet, Bluetooth 5.0, and dual-band Wi-Fi guarantee smooth connectivity. It is portable and can be powered by a 5V/3A USB-C source or a power bank. Table 1 shows the specifications.

Table 1: Specifications of Raspberry Pi 4

Specification	Details
USB Ports	2 × USB 2.0, 2 × USB 3.0
Storage	microSD card slot (up to 512GB)
RAM Options	8GB LPDDR4-3200, 4GB, or 2GB
Processor	Quad-core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5GHz, Broadcom BCM2711
GPIO Pins	40-pin header with various functionalities

4.3. Performance Evaluation

An object detection model is evaluated in this work using common assessment measures, such as Mean Average Precision (MAP), Intersection over Union (IoU), Precision, Recall, F1 Score, and Confusion Matrix Analysis.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8s, 2025
https://www.theaspd.com/ijes.php

Mean Average Precision

MAP is determined using each object class's average precision (AP). The capacity of the model to identify items with Intersection Over Union (IoU) ≥ 0.5 is measured by MAP @O.5. The model can recognize and localize objects with a good degree of accuracy if the MAP @0.5 is greater. With the MAP @0.95 metric, AP is averaged in steps of 0.05 across IoU thresholds from 0.5 to 0.95. Table 2 displays the MAP scores for the learned object detection model.

Table 2: MAP scores for the trained object detection model

Metric	Value (%)
MAP@0.5	87.5
MAP@0.5:0.95	64.2

The higher MAP @0.5 score suggests that the model performs well when considering detections with at least 50% overlap. The moderate MAP @0.5:0.95 score indicates that performance decreases as stricter localization accuracy is enforced.

Precision, Recall, and F1 Score

Recall, F1, and Precision score are three crucial metrics for assessing object detection models. These metrics demonstrate the model's ability to fairly balance false positives and false negatives while accurately detecting and classifying items. The following formulas are used to evaluate the metrics.

Precision = True Positives (TP) / (True Positives (TP) + False Positives (FP)) (1)

Recall = True Positives (TP) / (True Positives (TP) + False Negatives (FN)) (2)

F1 Score = 2 × (Precision × Recall) / (Precision + Recall) (3)

4.4. Performance Analysis

In this section, the main objective is to assess how well the proposed system performs in real-time object detection utilizing the YOLOv8 model combined with Natural Language Processing (NLP) for audio feedback. The experiments were conducted on IoT platforms such as Raspberry Pi to assess system efficiency, model accuracy, processing speed, and usability for visually impaired users. Several performance indicators are examined to confirm the efficacy of the applied solution, such as inference time, precision, and recall.

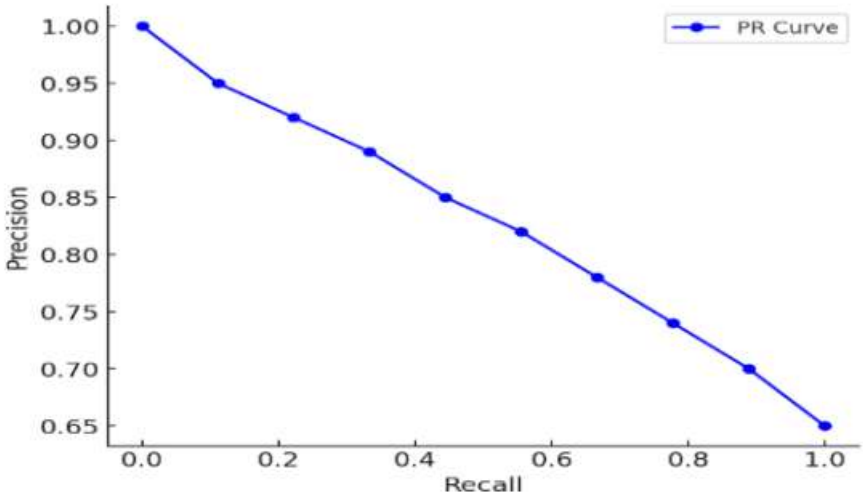

Figure 6: Precision-Recall graph of trained model

A classification model's precision-recall trade-off is depicted by the Precision-Recall (PR) curve in Figure 6. Precision steadily declines with increasing recall, suggesting that the model is increasingly inclined to detect more true positives at the expense of more false positives. This curve is particularly valuable in evaluating performance

on imbalanced datasets, such as object detection tasks in YOLOv8, where high precision and recall are both critical.

Confusion Matrix Analysis

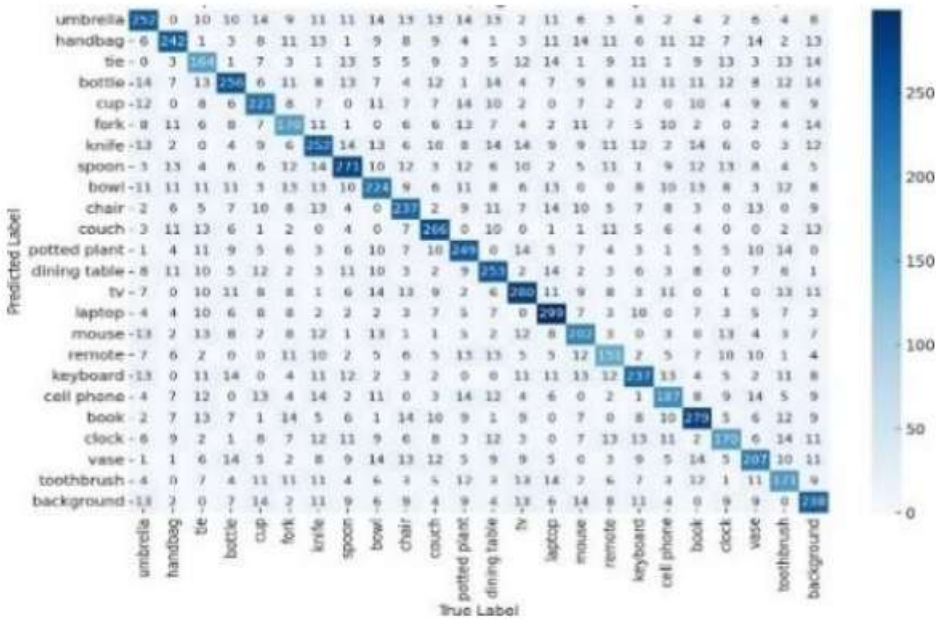

Figure 7: Confusion matrix analysis graph of the trained model

The confusion matrix in Figure 7 illustrates the classification performance of the YOLOv8 object detection model across various item categories. Each column in the matrix represents the real (true) class, whereas each row represents the anticipated class. Higher values are represented by darker hues in the diagonal cells, which show cases that were successfully classified. Off-diagonal entries represent misclassifications between object classes. This visual evaluation helps in identifying which categories the model predicts accurately and where it tends to confuse objects, thereby offering insights into areas for further model optimization.

5. Conclusion

The IoT-Based Voice-Driven Smart Finder showed a meaningful step forward in inclusive technologies, enhanced the life of those who are blind or visually impaired. By fusing AI-driven object detection with IoT and natural language voice interfaces, the system enables users to locate personal items using simple spoken commands. The deployment of the YOLOv8 model on resource-efficient hardware like Raspberry Pi ensures real-time detection capabilities with a high accuracy of 92%, providing a reliable and accessible solution. Additionally, the system's quick response time of 1.8 seconds supports smooth user interaction, thereby enhancing overall usability and satisfaction. This study also emphasized accessibility and affordability by using cost-effective components and cloud-based speech services, making the solution practical for a wide user base. However, limitations such as reduced performance in low-light or noisy environments highlight the need for further refinements. Future work will focus on integrating infrared sensors, adaptive camera settings, and noise-cancellation techniques to enhance performance. Expanding the system's connectivity with other smart home devices and developing a mobile companion app could further streamline user control and customization.

References

[1] Joshi, R.C., Yadav, S., Dutta, M.K. and Travieso-Gonzalez, C.M., 2020. Efficient multiobject detection and smart navigation using artificial intelligence for visually impaired people. Entropy, 22(9), p.941.

[2] Messaoudi, M.D., Menelas, B.A.J. and Mcheick, H., 2022. Review of navigation assistive tools and technologies for the visually impaired. Sensors, 22(20), p.7888.

[3] Hariharan, S., Abinaya, A., Anjuga, V. and Bhuvaneshwari, V., 2025. Voice Controlled Wheelchair for Physically Disabled People and Blind People. Asian Journal of Applied Science and Technology (AJAST), 9(1), pp.21-28.
[4] Mahendran, J.K., Barry, D.T., Nivedha, A.K. and Bhandarkar, S.M., 2021. Computer vision-based assistance system for the visually impaired using mobile edge artificial intelligence. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2418-2427).
[5] Almufareh, M.F., Kausar, S., Humayun, M. and Tehsin, S., 2024. A conceptual model for inclusive technology: advancing disability inclusion through artificial intelligence. Journal of Disability Research, 3(1), p.20230060.
[6] Hussan, M.I., Saidulu, D., Anitha, P.T., Manikandan, A. and Naresh, P., 2022. Object detection and recognition in real time using deep learning for visually impaired people. International Journal of Electrical and Electronics Research, 10(2), pp.80-86.
[7] Nair, V., Olmschenk, G., Seiple, W.H. and Zhu, Z., 2022. ASSIST: Evaluating the usability and performance of an indoor navigation assistant for blind and visually impaired people. Assistive Technology, 34(3), pp.289-299.
[8] AL-Najjar, M., Suliman, I. and Al-Hanini, G., 2018. Real Time Object Detection and Recognition for Blind People.
[9] Kappers, A.M., Holt, R.J., Junggeburth, T.J., Oen, M.F.S., van de Wetering, B.J. and Plaisier, M.A., 2024. Hands-free haptic navigation devices for actual walking. IEEE Transactions on Haptics, 17(4), pp.528-545.
[10] Baldonado, J., 2024. An Enhanced audio-based smart cane for visually impaired people. waves beneath sunrise, p.203.
[11] Li, G., Xu, J., Li, Z., Chen, C. and Kan, Z., 2022. Sensing and navigation of wearable assistance cognitive systems for the visually impaired. IEEE Transactions on Cognitive and Developmental Systems, 15(1), pp.122-133.
[12] Khan, A. and Khusro, S., 2021. An insight into smartphone-based assistive solutions for visually impaired and blind people: issues, challenges and opportunities. Universal Access in the Information Society, 20(2), pp.265-298.
[13] Jafri, R., Ali, S.A., Arabnia, H.R. and Fatima, S., 2014. Computer vision-based object recognition for the visually impaired in an indoors environment: a survey. The Visual Computer, 30, pp.1197-1222.
[14] Rahman, M.A. and Sadi, M.S., 2021. IoT enabled automated object recognition for the visually impaired. Computer Methods and Programs in Biomedicine Update, 1, p.100015.
[15] Ahammed, F., Adnan, M.A., Rahman, M.F., Alam, N., Paul, H. and Jibon, Z.A., 2023, October. Development of an IoT-Based Intelligent Guide Stick to Provide Improved Navigation Skills to Blind People. In 2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 1106-1111). IEEE
[16] Ghatkamble, R., Kumar, K.R., Hrithik, S.J., Kumar, J.H. and Sujan, P.S., 2023, April. Computer Vision and IoT-Based Smart System for Visually Impaired People. In 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1) (pp. 1-5). IEEE.
[17] Guravaiah, K., Bhavadeesh, Y.S., Shwejan, P., Vardhan, A.H. and Lavanya, S., 2023. Third eye: object recognition and speech generation for visually impaired. Procedia Computer Science, 218, pp.1144-1155.
[18] Leong, X. and Ramasamy, R.K., 2023. Obstacle detection and distance estimation for visually impaired people. IEEE Access, 11, pp.136609-136629.
[19] Senarathna, P., Pigera, I., Dodanduwa, S., De Silva, H., Amarakoon, T. and Thelijjagoda, S., 2023, December. Machine Learning Based Assistive Technology for Object Detection and Recognition for Visually Impaired Individuals. In 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) (pp. 1-7). IEEE.
[20] Kumar, A., Surya, G. and Sathyadurga, V., 2024, April. Echo Guidance: VoiceActivated Application for Blind with Smart Assistive Stick Using Machine Learning and IoT. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) (pp. 01-06). IEEE