AI Enhanced Wearable Device for the Visually Impaired

Shobana T S
Information Science and Engineering
B.M.S. College of Engineering
Bangalore, India
0000-0003-3147-4789

Monisha H M
Information Science and Engineering
B.M.S. College of Engineering
Bangalore, India

Rakshitha M
Information Science and Engineering
B.M.S. College of Engineering
Bangalore, India

Himani M
Information Science and Engineering
B.M.S. College of Engineering Bangalore, India

Jyotiprakash Panda
Information Science and Engineering
B.M.S. College of Engineering
Bangalore, India

***Abstract*—Our wearable device is a one stop solution for the daily challenges faced by individuals with both visual and auditory impairments. It is a versatile wearable assistive device designed to enhance their accessibility and readability of braille. The core innovation lies in the integration of three key functionalities: Braille detection, real-time object detection, and intuitive multi-sensory feedback. The device captures real-time images to accurately identify Braille text, which is then promptly converted into an audio output delivered in the user's preferred regional language. Furthermore, it's advanced object detection capabilities continuously monitor the user's surroundings, providing timely alerts upon the detection of potential obstacles, thereby significantly improving mobility. The system's provision of dual feedback mechanisms—both audio signals and vibration alerts—ensures comprehensive accessibility for users with diverse sensory profiles. By effectively combining these features, this wearable aid aims to foster greater independence, bolster user confidence, and offer a practical and versatile solution to the everyday challenges encountered by the visually and hearing-impaired community.**

Keywords—Accessibility, Assistive Technology, Braille Detection, Object Recognition, Text-to-Speech, Raspberry Pi, Multi-Sensory Feedback, Wearable Device.

I. INTRODUCTION

Visual impairment is one of the biggest challenges that millions of people face globally, affecting their independence and quality of daily life. The World Health Organization reports that approximately 1.3 billion individuals experience some form of visual impairment, with 36 million being completely blind. [1] For centuries, individuals with visual impairments have relied on traditional solutions such as human assistance, guide canes, and Braille to navigate their surroundings. The lack of efficient assistive technologies has led to difficulties in education, employment, and other activities for the visually impaired community.

Assistive technology has evolved as early innovations focused on tactile systems like Braille, requiring basic literacy for the visually impaired. The introduction of audio books and text-to-speech systems [2] in the late 20th century expanded access to literature and information, yet these solutions lacked real-time responsiveness and required manual input. Early attempts also included sensor-based devices [3] that provided auditory feedback for obstacle detection, but these were often inaccurate and unreliable. The dependency on static, pre-recorded information meant that users still faced limitations in navigating unfamiliar environments.

Recent advancements have utilized Artificial Intelligence and computer vision to develop more dynamic assistive devices. The integration of AI-powered object recognition and text-to-audio conversion brought a shift, allowing devices to interpret surroundings more effectively. Projects such as Raspberry Pi-based text-to-audio converters and smartphone applications using optical character recognition (OCR) demonstrated the potential of these innovations in improving accessibility [4].

Today, modern assistive devices use deep learning, natural language processing (NLP), and real-time object detection to provide visually impaired individuals with an interactive experience. Systems like YOLO-based object recognition, computer vision-powered autonomous navigation, and mobile-based AI assistants have transformed assistive technology. [5] These solutions provide users with real-time audio feedback about their surroundings, enabling them to identify obstacles, recognize faces, and even interpret digital text on screens. [6] Furthermore, wearable AI-based solutions such as smart glasses and AI-powered navigation aids are pushing the boundaries of accessibility [13].

Despite these advancements, challenges remain in making assistive technology more affordable and widely accessible. Many solutions rely on cloud services, which may not always be available due to internet dependency. Future research should focus on developing cost-effective, offline-capable solutions that support local AI processing. By continuously innovating and integrating AI-driven accessibility tools, society can work toward creating an inclusive environment where visually impaired individuals can navigate the world with confidence and independence.

II. RELATED WORKS

S. Zafar et al., 2022 [1] conducted a comprehensive review of assistive devices for visually impaired persons, categorizing them based on functionality and performance. They highlighted essential characteristics, conducted a feature-based quantitative analysis, and identified limitations for future improvements in device design.

D. T. V. Pawluk et al, 2015 [2] reviewed haptic technology design for visually impaired individuals, emphasizing the importance of behavioural research, user characteristics, and training in developing effective assistive devices.

Sahoo and Choudhury et al, 2024 [3] investigated computer vision for use in assistive technology, highlighting object and face recognition and gestures as interfaces. They highlighted the promise of increased mobility and independence for users with multiple disabilities, along with ethical issues, user-centric design, and the necessity of more research to enhance accessibility and practical effectiveness.

Myo Min Aung et al, 2024 [4] created a YOLOv4 object recognition system for the visually impaired using a Raspberry Pi and the COCO dataset. With 100% real-time object detection efficiency, the system focuses on user-centered design. User feedback from a hospital exhibition validated its usability and ability to improve safety and independence for users.

Parenreng et al, 2024 [5] created a visual aid system for the blind using Convolutional Neural Networks for object detection and ultrasonic sensors for measuring distance. The prototype glasses had 84.3% accuracy in detecting objects and a 2.1 cm error in distance measurement, improving navigation and mobility for users.

K. Thopate et al, 2024 [6] created "Vision Voice," a Raspberry Pi-based application that reads printed text aloud through OCR and gTTS for visually impaired individuals. Though effective, it has problems with small print and distance precision. Future developments will improve noise reduction, multilingual functionality, and ease of use.

Md. Zahidul Hasan et al, 2022 [7] employed the YOLO V3 model for real-time object detection and audio feedback distance measurement for navigation. Although effective, the system is challenged by dense or low-light environments. The future will focus on enhancing adaptability and sensor fusion.

Naga Prabhas Katakam et al, 2024 [8] surveyed image-to-audio methods for the blind using CNNs, LSTMs, and TTS systems. They pointed out problems with background management, character separation, and dataset dependency. Future research will focus on enhancing adaptability, decreasing processing time, and increasing datasets for enhanced user experience.

Hasventhran Baskaran et al, 2019 [9] created Smart Vision, a Raspberry Pi device based on Microsoft's Computer Vision API to provide real-time audio descriptions. Though efficient, it is dependent on stable internet and poses privacy issues. Improvements in future include higher accuracy, facial expression recognition, and multi-language capabilities.

Abhishek Brajvasi and Bhupendra Singh Kirar, 2024 [10] created an Android-based real-time picture-to-audio converter using TensorFlow Lite for object identification and TTS. It is effective, but it struggles in low-visibility conditions and may be improved with increased processing speed.

O. Ceoca and E.-H. Dulf, 2024 [11] developed an assistive helmet for visually impaired individuals featuring obstacle avoidance, fall detection, and track monitoring, enhancing mobility and independence while maintaining affordability.

K. Swathi et al, 2024 [12] designed smart glasses for the visually impaired, utilizing Raspberry Pi 4, a Pi Camera to detect objects, a GPS module to track locations, and ultrasonic sensors to measure distances. The system provides audio feedback and remote monitoring through caregivers, boosting safety and independent navigation. Usability and adaptability challenges exist despite remarkable enhancements in mobility.

Abidi et al, 2024 [13] surveyed navigation devices for the visually impaired, classified as Visual Imagery, Non-Visual Data, Map-Based, and 3D Sound Systems. Advances in sensors, AI, and smartphones improving autonomy were noted while emphasizing the displacement of traditional solutions by electronic systems and advocating a need for continual innovation.

Ruiqi Cheng et al, 2018 [14] developed a real-time visual localization approach for the visually impaired, utilizing multi-modal images and GNSS signals. The method significantly enhances positioning accuracy at key locations, improving navigation effectiveness for users in real-world scenarios.

S. R. Sankaranarayanan et al, 2024 [15] developed an innovative sensor-free obstacle detection system for visually impaired individuals using computer vision and audio feedback. The prototype enhances navigation safety and autonomy, receiving positive results during testing with real-world scenarios.

Fig. 1: Basic Prototype of the proposed system

III. Proposed System for the Wearable Device

The wearable device in Fig. 2 captures real-time images to analyze its surroundings. An image processing unit determines whether the captured image contains Braille text or requires object detection. If Braille is detected, it is translated into readable text and converted into audio, which is then played for the user. If no Braille is found, the system performs object detection to identify obstacles, providing tactile feedback through vibrations. Together, audio and vibration outputs offer real-time assistance for individuals with visual, hearing, or dual impairments.

A. Training the Braille-to-English Translation Model

The Braille translation model was developed using the open-source tool Roboflow, leveraging a labeled dataset of 10,382 images. Of these, 8,305 were used for training, 1,038 for validation, and 1,039 for testing.

B. Data Acquisition

Live images of the Braille text and the nearby surroundings are captured using Raspberry Pi Camera. This camera is connected to the Raspberry Pi processor via the CSI Port.

C. Braille to Text Model

Prior to prediction, the image captured by the Raspberry Pi Camera undergoes preprocessing with OpenCV to enhance the visibility of Braille dots using grayscale conversion and thresholding techniques. The trained model then detects Braille patterns from the live stream and translates them into the corresponding English alphabet.

D. Regional Text Translation Module

The translated text from the Braille translation model is fed into the regional language translation module through a pipeline. This module utilizes IndicTrans2, a powerful model by AI4Bharat, designed for seamless translation of English text into multiple Indian regional languages.

E. Text-to-Speech Module

The regional text is subsequently processed by the text-to-speech module, developed using the IndicParler TTS package. This model supports 69 distinct voices, offering a rich range of emotional expressions. The final audio output is generated and saved as a .wav file.

F. Playing the final Audio translation

The output audio file is played back, with the sound transmitted through earphones connected to the Raspberry Pi.

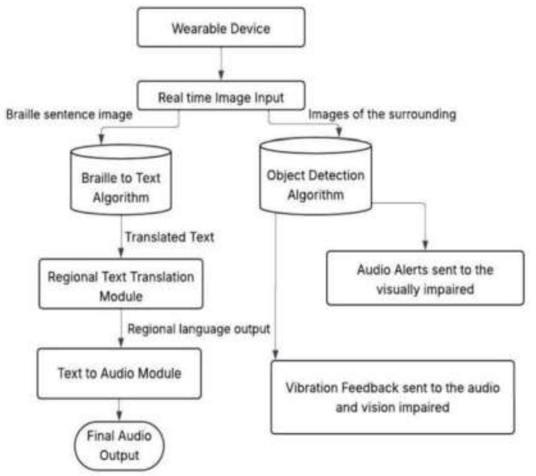

Fig. 2: Architecture of the proposed system

G. Object Detection Using YOLOv8

YOLOv8 You Only Look Once performs object detection in a single neural network pass resulting in fast predictions. The input image is divided into an S×S grid, where each grid cell predicts bounding boxes along with their associated class probabilities and confidence scores. Each predicted bounding box includes coordinates, width, height, and an objectness score indicating the likelihood of containing an object. During training, predictions are matched with ground truth bounding boxes using the Intersection over Union (IoU) metric. A prediction is considered accurate if its IoU with the ground truth exceeds a certain threshold. Non-Maximum Suppression (NMS) is applied to eliminate overlapping predictions by retaining only the highest-confidence boxes. The loss function in YOLOv8 combines localization loss, confidence loss, and classification loss to train the network efficiently. The model is ideal for deployment in applications requiring low latency and high throughput.

H. Model Optimization Using OpenVINO

Direct deployment of the model can result in increased inference latency and higher memory consumption, which is not appropriate for real-time environments. Therefore, the model is converted into Open Visual Inference and Neural Network Optimization (Open VINO's) Intermediate Representation (IR) format (XML and BIN). Open VINO is designed to accelerate deep learning models for edge devices and different computing environments). Then we optimise the model by using OpenVINO's Neural Network Compression Framework (NNCF), specifically Post-Training Quantization (PTQ) method. This process converts the model from high-precision (FP32) to low-precision (INT8), reducing memory usage and increasing inference speed while maintaining accuracy. A representative subset of the COCO dataset is used as a calibration dataset to keep a check on the activation ranges and select appropriate scaling factors.

I. Direction Estimation Based on Bounding Box Position

To prompt the direction of the obstacles in the user's path the object direction is inferred using the horizontal position of the detected object's bounding box. For each object detected by the YOLOv8 model, the center x-coordinate of its bounding box is computed. Based on its relative position within the frame, the object is classified into one of three spatial zones: Left, Center, or Right. This classification is done by dividing the frame width into three equal vertical regions. If the bounding box center falls in the leftmost third, the direction is Left, if in the middle third, it is labeled Center and otherwise, it is categorized as Right.

The formula in Eq. (1) is used to determine the spatial direction (Left, Center, or Right) of a detected object within a video frame by calculating the center x-coordinate of its bounding box relative to the frame's width.

$$center_x = (X_{min} + X_{max}) / 2 \qquad (1)$$

Left: $center_x < (1/3) \times frame\_width$

Center: $(1/3) \times frame\_width \leqslant center_x < (2/3) \times frame\_width$

Right: $center_x \geqslant (2/3) \times frame\_width$

Here,
X_{min} = The X coordinate of the left boundary of the object's bounding box
X_{min} = The x-coordinate of the right boundary of the object's bounding box
$center_X$ = The horizontal center point of the bounding box, calculated as $(x_{min} + x_{max}) / 2$
frame_width = The total width (in pixels) of the video frame or image in which the object is detected.

J. *Distance Estimation Using HC-SR04 Ultrasonic Sensors*

Multiple HC-SR04 ultrasonic sensors are arranged in a semi-circular configuration. Each sensor is oriented to cover a specific zone (left, center and right) corresponding to the directional zones used in visual detection. The HC-SR04 sensor operates by emitting ultrasonic pulses and measuring the time interval between transmission and reception of the echo. The distance is calculated using the standard formula in Eq. (2):

$$\text{distance} = (\text{Echo Time} \times \text{Speed of Sound}) / 2 \quad (2)$$

where the distance is in cm.

When a YOLOv8 detection is labeled as being in a certain direction (e.g., "Left"), the corresponding sensor's reading is taken to estimate the distance of the object from the user.

K. *Audio and tactile feedback*

Audio feedback is provided through earphones, conveying detected object names, directions, and distances. Feedback via vibrations is also provided by the ERM (Eccentric Rotating Mass) vibration motors. When an object is detected within a predefined safety threshold, for example set at 1.5 meters, the motor is triggered to vibrate. The vibration intensity is set proportionally to the proximity of the obstacle, that is stronger vibrations are emitted as the object approaches the user.

Table I lists all the hardware components necessary for assembling the wearable device.

Component	Specification
Processor	Raspberry Pi 4
Camera Module	Raspberry Pi Camera Module V1
Vibration Motors	10mm Coin-Type ERM Vibration Motors
Ultrasonic Sensor	HC-SR04 Ultrasonic Sensor

Table I: Hardware specifications of the device

IV. System Evaluation

A. *Evaluation of Braille to Text Model*

To evaluate the performance of our Braille text recognition system, we experimented with four different models. Among various evaluation metrics, we prioritized precision because accurate identification of Braille characters is critical—false positives can lead to incorrect translations, significantly impacting the usability of the device for visually impaired users. The precision scores of the four models are presented in Table II, providing a clear comparison of their performance

Model	Precision
Roboflow 3.0 Model	97.7%
Azure AI Vision Model	92%
OpenCV Computer Vision Model	83 %
Convolutional Neural Network Model with Optical Character Recognition	78.60%

Table II. Precision Scores of the various trained models

Among the models tested, the one trained and deployed using Roboflow achieved the highest precision, ensuring the most reliable character detection. Additionally, the Roboflow model demonstrated the lowest inference latency, which is essential for real-time applications. Its seamless integration pipeline further simplified deployment on the Raspberry Pi, making it the most practical and efficient choice for our wearable assistive system.

B. *Evaluation of the Object Detection Model*

The optimization of the YOLO object detection model using OpenVINO's INT8 quantization significantly enhanced performance. The optimized model (Model 2) demonstrated a 35.46% reduction in latency, dropping from 51.59ms to 33.30ms, while also improving throughput by 54.95%, achieving 30.03 FPS. Despite a slightly higher load time, the model size was nearly halved, and memory usage decreased, making it more efficient for real-time applications. These improvements highlight how model optimization not only speeds up inference but also reduces resource consumption, which is crucial for edge deployments like Raspberry Pi. Fig. 3 illustrates the substantial reduction in latency.

We evaluated multiple versions of the YOLO object detection models—YOLOv6, YOLOv7, and YOLOv8—for our Braille translation system. While YOLOv6 and YOLOv7 offered moderate performance with frame rates around 4–8 FPS and latency ranging between 120–250 ms, YOLOv8 showed improved efficiency, delivering up to 10 FPS and lower latency. The optimized version of YOLOv8 significantly outperformed all others, achieving a high frame rate (~38 FPS), minimal latency (~33 ms), and a compact model size (~25 MB), making it highly suitable for real-time Braille detection on resource-constrained devices. These values are depicted in Table III. Additionally, the optimized

YOLOv8 model exhibited reduced memory consumption and faster inference speeds, contributing to smoother and more responsive performance. It's ease of deployment and compatibility with OpenVINO further reinforced its selection for integration into the wearable system.

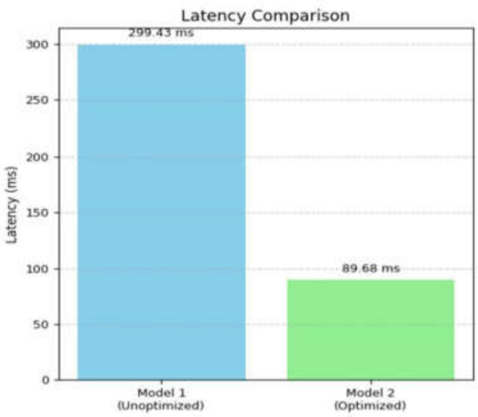

Fig. 3: YOLO Optimization Performance Comparison

YOLO Version	**FPS (ms)**	**Latency (ms)**	**Model Size (MB)**
YOLOv6	~5-8	~120-200	40-100
YOLOv7	~4-7	~140-250	60-120
YOLOv8	~6-10	~100-170	20-70
YOLOv8 (optimised)	~38	~33	~25

Table III. Performance Measure of YOLO Models on CPU

V. Results

Fig. 4 displays the output generated by our Braille translation model trained using Roboflow. The model successfully detects and labels 21 Braille characters, each outlined with distinct colored bounding boxes and class labels. This demonstrates the model's strong capability in accurately identifying and localizing Braille symbols, a crucial step in translating tactile Braille inputs into readable text. The consistent detection performance highlights the effectiveness of the Roboflow training pipeline for fine-grained visual recognition tasks like Braille decoding.

Fig. 4: Output of Braille Character Detection Using Roboflow-Trained Model

Fig. 5 shows the real-time output of an object detection system, identifying and localizing a person and a bottle with associated confidence scores and estimated distances. The model runs efficiently at around 37 FPS with 27 ms latency, making it suitable for real-time applications. The bounding boxes display object type, confidence (e.g., person: 0.96), position (e.g., center or left), and proximity (e.g., 1.5m or 0.3m). A summary at the bottom highlights detected close objects, which can be useful for applications like assistive navigation or obstacle avoidance.

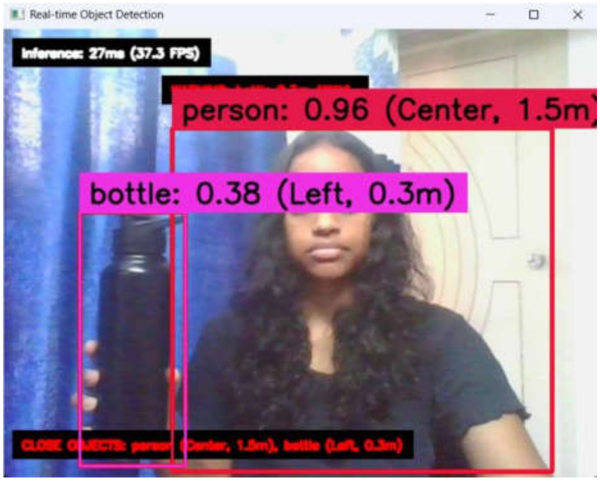

Fig. 5: Real-time Object Detection with Distance and Position Estimation

Fig. 6: Prototype of the Wearable Device

Fig. 6 illustrates the wearable device developed by integrating various hardware components along with the deployed models, designed to provide a smooth and seamless user experience.

Conclusion

This study delves into the breakthroughs in visually impaired aid technologies, facilitated by the collaboration of artificial intelligence, sensor fusion, and multimodal data processing. The research emphasizes the ways in which AI-based orientation systems, in combination with real-time obstacle detection and adaptive feedback processes, are significantly enhancing mobility and spatial perception among users. The importance of haptic feedback, auditory notifications, and computer vision-based identification in empowering intuitive environmental interaction is also analyzed. Additionally, the potential of AI-based Braille text identification and translation to different Indian regional languages as an important value addition to accessibility for easy understanding of written text is introduced.

The study also takes into consideration the significance of vibration feedback during object detection in enhancing accessibility for hearing-impaired individuals, allowing for the delivery of vital spatial awareness information. Riding on these developments, the study expects subsequent breakthroughs in the fields of depth perception, advanced AI-based obstacle detection, and advanced feedback mechanisms to further optimize these assistive technologies. The final aim is to develop assistive technologies that are more efficient, accessible everywhere, and very reliable, thus enabling visually impaired people to move around the world in a more confident and independent manner.

VI. REFERENCES

[1] [1] S. Zafar, M. Asif, M. B. Ahmad, T. M. Ghazal, T. Faiz, M. Ahmad, and M. A. Khan, "Assistive Devices Analysis for Visually Impaired Persons: A Review on Taxonomy," IEEE Access, vol. 10, pp. 13354-13366, 2022, doi: 10.1109/ACCESS.2022.3146728.

[2] D. T. V. Pawluk, R. J. Adams, and R. Kitada, "Designing Haptic Assistive Technology for Individuals Who Are Blind or Visually Impaired," IEEE Transactions on Haptics, vol. 8, no. 3, pp. 258-273, July-Sept. 2015, doi: 10.1109/TOH.2015.2471300.

[3] S. K. Sahoo and B. B. Choudhury, "Exploring the use of computer vision in assistive technologies for individuals with disabilities: A review," Journal of Future Sustainability, vol. 4, no. 3, pp. 133-148, Jul. 2024, doi: 10.5267/j.jfs.2024.7.002.

[4] M. M. Aung, D. Maneetham, P. N. Crisnapati, and Y. Thwe, "Enhancing Object Recognition for Visually Impaired Individuals using Computer Vision," International Journal of Engineering Trends and Technology, vol. 72, no. 4, pp. 297-305, Apr. 2024, doi: 10.14445/22315381/IJETT-V72I4P130.

[5] J. M. Parenreng, A. B. Kaswar, and I. F. Syahputra, "Visual impaired assistance for object and distance detection using convolutional neural networks," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 8, no. 1, pp. 26-32, Jan. 2024, doi: 10.29207/resti.v8i1.5417.

[6] K. Thopate, K. Amrutkar, T. Kasliwal, S. Karvir and O. Kumbhar, "Vision Voice: A Raspberry Pi-Based Text-to-Audio Converter for the Visually Impaired," 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP), Sonipat, India, 2024, pp. 231-234, doi: 10.1109/INNOCOMP63224.2024.00045

[7] M. Z. Hasan, S. Sikder and M. A. Rahaman, "Real-Time Computer Vision Based Autonomous Navigation System for Assisting Visually Impaired People using Machine Learning," 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2022, pp. 1-6, doi: 10.1109/STI56238.2022.10103268.

[8] N. P. Katakam, H. Mynedi, S. Imambi, and C. Nikhileswar, "Deep learning techniques for real-time image to voice for the visually impaired: A review," Degres Journal, vol. 9, no. 3, pp. 53-54, 2024.

[9] H. Baskaran, R. L. M. Leng, F. A. Rahim and M. E. Rusli, "Smart Vision: Assistive Device for the Visually Impaired Community Using Online Computer Vision Service," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 730-734, doi: 10.1109/CCOMS.2019.8821635

[10] A. Brajvasi and B. S. Kirar, "Android Based Real-Time Picture-to-Audio Converter for the Visually Impaired," 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2024, pp. 1-6, doi:10.1109/SCEECS61402.2024.10482202..

[11] O. Ceoca and E.-H. Dulf, "Assistive Helmet for Visually Impaired Human Beings," *2024 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, Cluj-Napoca, Romania, 2024, pp. 1-10, doi: 10.1109/AQTR61889.2024.10554135.

[12] K. Swathi, P. Gayathri, D. Manasa, G. Pavani, and B. Likhitha, "Smart glasses for visually impaired people using IoT," International Research Journal on Advanced Engineering and Management, vol. 2,no. 5, pp. 1390-1394, May 2024, doi: 10.47392/IRJAEM.20

[13] M. H. Abidi, A. N. Siddiquee, H. Alkhalefah, and V. Srivastava, "A comprehensive review of navigation systems for visually impaired individuals," Heliyon, vol. 10, no. e31825, pp. 1 -19, May 2024, doi: 10.1016/j.heliyon.2024.e31825

[14] R. Cheng, K. Wang, L. Lin, and K. Yang, "Visual Localization of Key Positions for Visually Impaired People," 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 2018, pp.2893-2898, doi: 10.1109/ICPR.2018.8546264.

[15] S. R. Sankaranarayanan, M. R. Manjushree, H. P. Harshini, and G. V. Jeeshitha, "Obstacle Detection for Visually Impaired: Real-Time Assistive System," Proceedings of the IEEE, Apr. 2024.