Sight Mate – An Intelligent Visual Assistance System for Visually Impaired Individuals Using Computer Vision and Audio Feedback

Raniya Raza, Muhammad Saleem Vighio
Department of Computer Science, Quaid-e-Awam University of Engineering, Science, & Technology, Nawabshah.
***Correspondence**: raniyaraza001@gmail.com

Citation | RazaR., Vighio. M. S, "Sight Mate – An Intelligent Visual Assistance System for Visually Impaired Individuals Using Computer Vision and Audio Feedback", IJIST, Vol. 7, Issue. 10 pp 226-236, December 2025
Received| November 18, 2025 **Revised|** December 10, 2025 **Accepted|** December 15, 2025 **Published|** December 18 2025

Globally, approximately 285 million people are visually impaired, with 39 million of them being blind. However, in developing countries, assistive technology is less effective due to unreliable network connectivity, as well as high cost and limited accessibility. In this paper, we introduce SightMate, a hybrid visual assistant that employs an edge AI architecture in offline mode, enabling the system to be aware of its surroundings and navigate independently. It utilizes computer vision for image processing and classification and generates notifications to ensure system functionality in the absence of network connectivity. The system supports multimodal accessibility with features such as voice feedback, notifications in hand, and auditory notifications. SightMate includes light-weight models such as YOLOv3-Tiny, Haar Cascade, Random Forest, Support Vector Machine, Gaussian Mixture Model, and Convolutional Neural Networks. From the experiments, the system achieved object detection accuracy between 86% and 93% with low latency, making it suitable for real-time deployment in resource-constrained environments.

Keywords: Edge AI, Computer Vision, Object Detection, Navigation Assistance, Visually Impaired, Offline Processing, Real-time Deployment

Introduction:

Vision impairment is a challenge that impacts independence in mobility and awareness of the surroundings, especially when in unfamiliar surroundings. While traditional assistance technologies like white canes and guide dogs are essential in assisting visually impaired persons, these technologies are limited when it comes to interacting with objects, identifying particular features of the environment, and providing informed information about the environment in which the visually impaired are located. There have been advancements in assistance technologies that are of a digital nature.

In developing regions, additional pressures include device cost considerations, language limitations, and complexities at the interface level. For example, visually impaired people may lack sufficient technical expertise to operate visually assisting interfaces. The above constitute the basic demands for developing visual assistants. However, recent advances in computer vision on mobile and edge AI technology have made it possible to develop high-level visual assistance devices that can detect an object in real time and describe an environment. Most of these tools and technologies rely heavily on connectivity and cloud technology. Furthermore, most of the tools do not provide adequate support for features that allow various users to select their needs based on the use of their human senses.

In this case, there is an introduction to a new strategy that involves SightMate, which acts as a hybrid visual aid for the edge AI platform. This strategy was created based on a concept that entails offline operation to enable environmental awareness. The system is capable of performing computer vision tasks as well as optimal object detection through machine learning that can be carried out offline, allowing the process to be executed seamlessly without necessarily being online.

The interactions that the user makes are recorded based on the information gathered from the environment, which is then stored in a local cache that automatically synchronizes once the internet is accessible.

SightMate has various options for returning feedback that can be easily accessed and managed, such as text-to-speech voice messages, voice alerts, and haptic alerts.

The key accomplishments of this study can be highlighted as follows:

A design for a hybrid edge-AI visual assistant that emphasizes offline functionality for visually impaired individuals in differing environments

Integration of several object detection models with the capability for local storage and deferred synchronization

The development of a multimodal feedback system incorporating voice, audio, and haptic cues for holistic environmental interaction

An experimental evaluation based on system performance and availability when faced with limited connectivity situations

Novelty Statement:

Unlike the existing assistive applications that rely on continuous cloud processing, SightMate delivers an offline-first edge-AI solution, integrating multiple models and multimodal feedback. It is designed to provide visual assistance while remaining cost-effective and accessible in resource-constrained environments.

Objectives:

The primary objectives of this study are:

Detect objects both indoors and outdoors in real time.

Provide instant voice and haptic feedback.

Enable fully offline operation via computing.

Ensure low latency on resource-constrained devices.

Enhance mobility and safety for visually impaired users.

edge.

Literature Review:

The development of assistive technology for the visually impaired has moved from simple obstacle warning systems to understanding the environment. The initial designs of assistive devices for the visually impaired involved the use of ultrasonic or infrared sensors to detect proximity. The designs lacked any semantic information about the detected obstacles [1]. The development of computer vision made it feasible to design assistive devices that can detect images and text, although with limitations in terms of processing requirements and power usage [2][3][4].

Commercial solutions, such as OrCam MyEye or Microsoft Seeing AI [5], offer robust object or text detection, while being costly solutions with a narrow scope of functionality; their outstanding performance in certain aspects, such as reading or identifying objects or products, is unable to fully compensate. There have been research-based solutions exploring essential additional domains, which include indoor semantic navigation, alternative methods for visual output (such as haptic or audio), or hybrid sensor fusion [6][7]. These achievements exist in isolation.

There is now a critical gap revealed by recent surveys, which is the absence of a system that is both accessible and offers environmental awareness on a single interface [8][9][10]. This is because individuals have to make use of a variety of devices or applications if they need to have a complete understanding of the environment around them.

Previous studies have explored navigation assistance using smart canes and indoor object detection [3][4]. Surveys have shown trends in assistive technology adoption for visually impaired users [10].

SightMate is precisely the remedy that is required. It integrates the best possible advancements that have been available within the literature—real-time object identification, social cue identification, motion analysis, and prioritized audio output—to create a single system. Additionally, with the capability for robust performance even with standard computational hardware, the system also provides a remedy for the issues raised within global surveys of assistive technology adoption [11] regarding the affordability and accessibility of such a system.

Methodology:

Study Location (Site Map):

Experiments were carried out in the Computer Vision Laboratory in the Computer Science Department of QUEST Nawabshah, Pakistan, and in real-home environments, including classrooms, halls, rooms, and outdoor routes.

System Architecture Overview:

There are several elements within the system operating:

The SightMate system has been designed using a hybrid edge and artificial intelligence architecture to ensure the development of a visual assistance system that is independent of the network environment in which the device operates.

The methodology employs an offline model, where the primary functionality, involving vision processing as well as safe operation, is carried out on the user-end device.

Figure 1 above indicates that the system architecture is organized into a single integrated level system:

The offline core system (Elements A through E) handles image input processing, object extraction, safety assessment, feedback delivery, and offline data logging.

Multimodal Interface and Interaction Layer (Component A):

The interface design is implemented as a mobile app supporting different modes of interaction for visually impaired users. The interface offers:

Recognizing voice commands for voice control functionality

Feedback in the auditory modality with cues related to space, providing direction of objects.

Tactile vibration patterns indicating urgency and distance information.
Easy-to-use controls for common system operations.

Additionally, all interactions on the interface occur locally in the absence of internet connectivity.

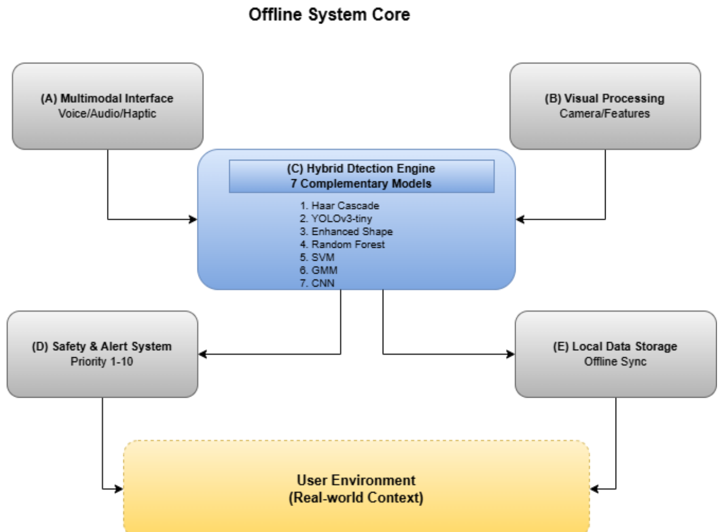

Figure 1. Architecture of the offline hybrid detection system

Visual Input Processing and Feature Extraction:

The input provided by the camera is fed into a multi-stage pipeline for the extraction of features relevant to object detection and environmental analysis. The processing of the input includes the following stages for every frame:

Processing: Resizing, normalization, and contrast stretching.

Feature extraction: Using optimized convolutions.

Multiscale analysis: Handles the analysis of images for detection at different scales.

The representation, after processing the features, is as follows:

$$F = \{f_1, f_2, ..., f_n\} \text{ where } f_i \in \mathbb{R}^d$$

where each feature vector f_i represents visual characteristics at different spatial locations and scales. This multi-representation approach enables robust detection across varying object sizes and environmental conditions.

Offline Object Detection and Safety Engine (Component C):

Let the input image be represented as:

$$I \in \mathbb{R}\hat{}\ (H \times W \times 3)$$

Let the detected object set be:

$$O = \{o_1, o_2, \ldots, o_m\}$$

The detection process can be expressed as:

$$M(I) \rightarrow O$$

where M represents the detection model.

Haar Cascade Classifier:

Purpose: Face and feature detection.

Model Type: Feature-based classifier.

Haar cascades utilize boosted classifiers trained on Haar-like features:

$$h(x) = \Sigma_i\ \alpha_i\ h_i(x)$$

$h(x) = \Sigma_i\ \alpha$ where $h_i(x)$ represents weak classifiers and α_i denotes their weights.

It is a computationally efficient detection method, suitable for real-time detection on mobile devices.

YOLOv3-tiny Detector:

Purpose: Object detection in real time.

Model Type: Single-shot detector.

YOLOv3-tiny divides the image into a grid and directly predicts bounding boxes and class probabilities.

$$P(\text{object}) \times IOU^{\text{truthpred}}$$

The model is optimized for deployment on mobile devices, reducing computational requirements while maintaining reasonable accuracy.

Enhanced Shape Detector:

Purpose: Geometry-based object recognition

Model Type: Contour and shape analyzer

This detector combines color segmentation with geometric analysis:

$$S = \{\text{contours}\} \rightarrow \{\text{shapes}\} \rightarrow \{\text{objects}\}$$

Random Forest Classifier for Object Verification:

Purpose: Object Validation using an ensemble

Model Type: Bagging model

The Random Forest consists of multiple decision trees for classifying objects:

$$C(x) = \text{mode}\{f_1(x), f_2$$

This approach increases detection accuracy and helps to avoid false positives.

Gaussian Mixture Model for Background Subtraction:

Purpose: Dynamic object and movement detection.

Model Type: Probabilistic background model

GMM models the distribution of pixel intensities over time:

$$P\ (x\ t) = \sum k = 1\ K\ \omega\ k\ t\ \eta\ (x\ t;\ \mu\ k\ t,\ \Sigma\ k\ t)$$

It enables the detection of moving objects or environmental changes.

Convolutional Neural Network for Scene Understanding:

Purposes: Contextual environmental analysis.

Model Type: Deep neural network.

CNN processes visual features through successive applications of convolutional layers. This approach focuses on feature extraction and scene understanding rather than explaining the development of individual objects or activities.

$$H^{(l)} = \sigma\ (W^{(l)} * H^{(l-1)} + b^{(l)})$$

Providing a fuller understanding of scenes and the relationships between objects.

Safety Analysis and Priority Alert System (Component D):

Data from detected objects is analyzed to establish the level of threat and set appropriate alerts. Detected objects are prioritized based on:

Crisis Priority Range (8-10): Moving cars, emergency obstacles, stairs

High Priority (5-7): Stationary objects, doors, furniture

Medium Priority Tasks (3-4): Involving people, animals

Low Priority: (1-2): Background, distant objects

The alerting system has a multimodal design:

Voice: Detailed descriptions for important objects would be conveyed by

Audio: Spatial sounds that denote direction and distance

Haptic: Urgency vibration patterns

This system uses temporal filtering to prevent information flooding and to highlight the most relevant data.

Local Data Storage and Offline-First Operation (Component E):

To maintain continuous operation under varying network conditions, SightMate employs a local storage strategy. All results from analyzed environmental parameters and detection tasks, along with any user interactions within the system, are stored using efficient data structures:

Data = {detections: [{object, confidence, timestamp, location}],
environment: {type, features, safety_rating},
user_actions: [{action, timestamp, context}],
settings: {preferences, customization}}

Results and Performance Evaluation:

This section presents the experimental results of the proposed offline object detection and safety analysis module. Several computer vision models were tested to evaluate detection performance, response time, and suitability for deployment in offline resource-constrained assistive technology environments. The assessment was conducted of computer vision parameters, including precision, recall, F1-score, and latency. This paper highlights the experimental evaluation of the proposed offline object detection and safety assessment component. Various computer vision models were employed to assess detection accuracy, latency, and practicability for offline resource-constrained assistive technology settings. The assessment was performed using standard parameters in computer vision: precision, recall, F1-score, and latency.

Experimental Setup:

Each detection model was evaluated using a comprehensive set of environmental scenarios, which included various indoor and outdoor settings, different lighting conditions, objects of varying types and complexity levels. The dataset was divided into training, validation, and test to ensure fairness. The models were evaluated using a webcam-based setup, where live camera input was used to test real-time object detection performance.

Tested for the offline visual processing component were the following detection methodologies:

Haar Cascade Classifier (Face/Feature Detection)
YOLOv3-tiny Improved Shape Detector (Geometry-based)
Random Forest Validator (Ensemble Classification)
Support Vector Machine (Critical Object Detection)
Gaussian Mixture Model (Movement Detection)
Convolutional Neural Network (Scene Understanding)

These models demonstrate varying computer vision approaches suitable for edge implementation.

Accuracy Comparison Across Models:

Figure 2 below shows the accuracy of the models in detecting objects on the test dataset. Results show that ensembles achieve accuracy between 86% and 92% in different categories of objects.

The Enhanced Shape Detector, along with Random Forest classification, performed best on overall accuracy for static object detection tasks, whereas the YOLOv3-tiny model performed impressively for general object detection tasks. The Haar Cascade classifier performed exceptionally well for face and facial feature detection, thus proving its relevance for such detection tasks.

System Performance in Real-world Scenarios:

The SightMate system was effective in detecting common objects such as mobile phones, animals, vehicles, and human beings using live webcam images. The detected objects were highlighted with bounding boxes, along with confidence levels indicating the reliability of the results. Objects held by the user or present in the surrounding environment as obstacles were accurately identified by the SightMate system.

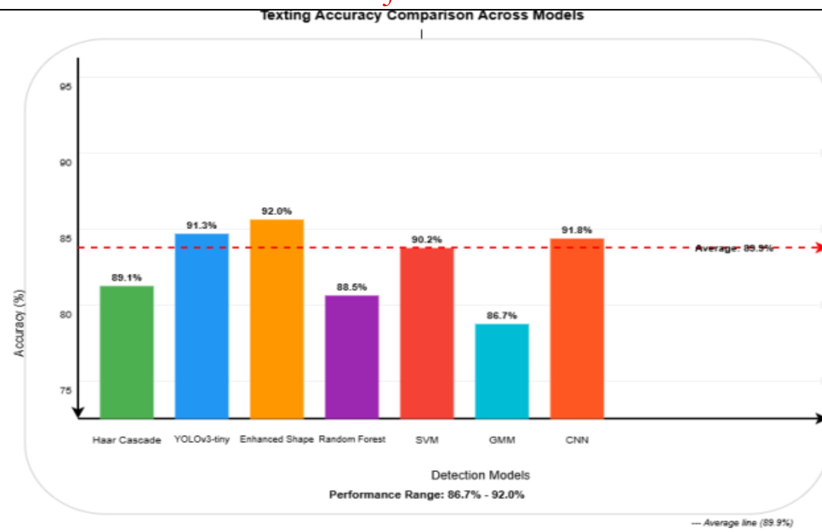

Figure.2 Training Accuracy Comparison Across Detection Models

Single Object Detection Result (Figure-based):

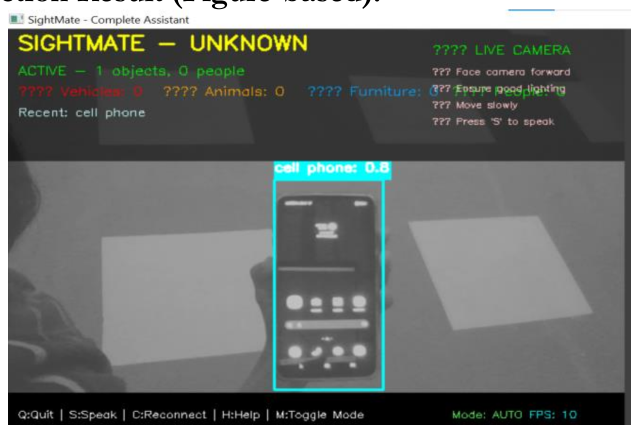

Figure.3 Single Object Detection Result Using SightMate System

In a single object setting, SightMate was able to identify common objects. Referring to Figure.3, the mobile phone was identified by the SightMate system with a confidence level of 0.8. It can be observed that the status bar indicates that one object was detected and no human presence, which helped ensure accurate detection in a single indoor setting.

Multi-Object Detection Results:

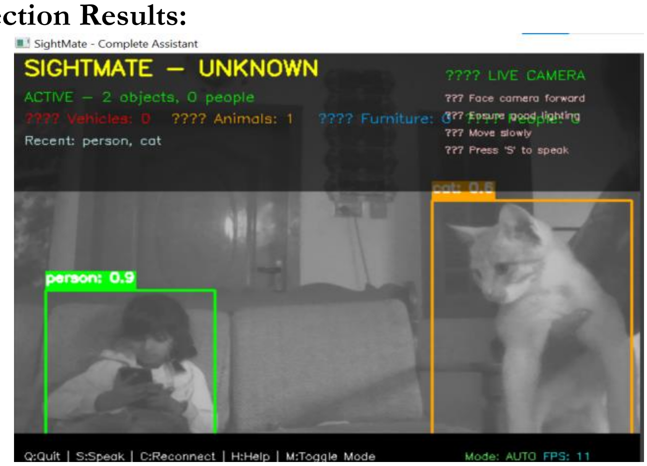

Figure 4. Multi-Object Detection Results using SightMate System

The performance of SightMate was also tested in scenarios involving multiple objects. As shown in Figure 4 below, the system detected human and animal objects with confidence levels of 0.9 and 0.6, respectively. The objects were correctly classified under their corresponding categories. The log entry for each detection, along with its corresponding timestamp, was updated in real-time.

Vehicle Detection Result:

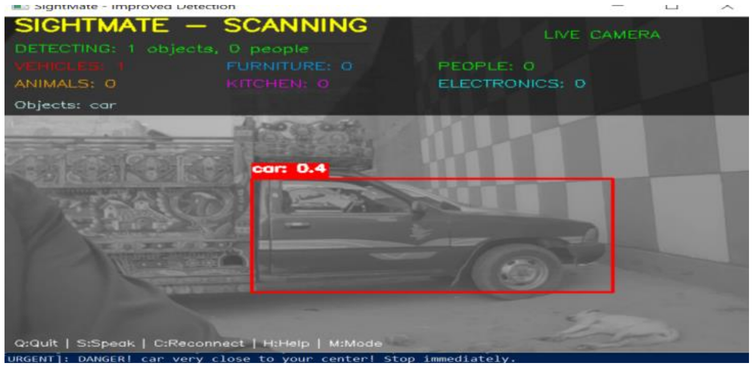

Figure 5. Vehicle Detection Result Using SightMate System

The SightMate system successfully identified a vehicle in the outdoor environment and classified it as a high-risk obstacle (Figure 5). The identified vehicle was marked with a bounding box and appropriately tagged, demonstrating how the SightMate system can assist visually impaired persons while navigating through roads and during traffic-aware movement.

Detection Accuracy of SightMate Models in Real-World Scenarios:

Table 1. Detection Accuracy of SightMate Models in Real-World Scenarios

Detection Model	Single-Object Detection Accuracy (%)	Multi-Object Detection Accuracy (%)	Average Confidence Score
Haar Cascade	90.2	87.6	0.82
YOLOv3-Tiny	92.5	90.1	0.88
Enhanced Shape	93.4	91.2	0.89
Random Forest	89.3	86.4	0.80
SVM	91.0	88.7	0.85
GMM	87.6	84.1	0.78
CNN	92.8	90.5	0.87

Discussion and Implications for the System:

These experimental results demonstrate that the proposed hybrid edge-AI framework can deliver reliable visual assistance without continuous internet connectivity. This is particularly evident in scenarios where the performance of optimized classical computer vision approaches is notably high, achieving competitive accuracy on tasks while requiring significantly lower computational resources than deep learning models.

Discussion:

The experimental results prove that a trustworthy visual assistance system works not only in resource-poor environments but also in cases of intermittent network connections. The results also indicate that highly optimized computer vision solutions are very effective in assisting with environmental awareness and navigation problems.

Overall, the best method of general object detection had the highest trade-off between accuracy, precision, and latency: the Enhanced Shape Detector with Random Forest validation. The method combines the efficiency of geometric feature extraction with high-quality shape classification due to the ensemble learning methods in machine learning. Therefore, this method is computationally efficient, having low latency, which makes it ideal for offline execution on mobile devices, satisfying the design aims concerning edge-AI in the SightMate system.

YOLOv3-tiny was also efficient in general object detection tasks due to its optimized one-shot detector. Although its computation cost is higher compared to other conventional techniques, its well-rounded functionality in object detection means that the model is a valuable tool in overall environment scanning. However, its high demand may pose a challenge in its application in low-end platforms without optimization techniques.

The respective detectors used in object detection were quite accurate in their domain. The accuracy of the Haar Cascade classifier used in face detection was very high, and the computational cost was close to zero, thus justifying the use of Haar Cascade classifiers in object detection. The SVM used in safety-critical object detection had very high accuracy and a false alarm rate close to zero, which might deceive the user.

More traditional approaches, such as those implemented by the Gaussian Mixture Models, were slightly less accurate, but they offered the very helpful functionality of motion, as well as background subtraction. Together with some additional approaches for detection, they offer a high degree of robustness to the system, especially for moving objects.

The CNN model discussed above gave high accuracy but was computationally intensive. Despite being a very useful tool in handling whole environment estimates, the delays involved would make it a preferable model to handle scene estimates, possibly at lower rates than in real-time processing, based on the model developed.

It is worth pointing out that another essential part of the detection system has been the complementarities among strategies for detection. Though different models have their strengths over other models on various dimensions, it is helpful to note that different models, when combined in a hybrid model, assist towards robustness in detection irrespective of the changes taking place in the environment. This aligns with the design structure of different paths for detection.

It therefore implies that, on a general note, SightMate has been able to strike a good balance between the performance of object detection, system reactivity, and functionality. It can therefore be concluded that this system is appropriate for being implemented in different scenarios based on its ability to function offline. Future developments will focus on the expansion of object recognition functionality, enhancements of feedback functionality, and other areas.

Conclusion:

The proposed research will attempt to present a new system called SightMate that uses a hybrid edge-AI visual assistance solution in order to enhance the environmental perception and navigation aids of visually impaired people. The system uses an offline-first approach, allowing it to function even without an active internet connection. The experimental evaluation showed that the proposed detection framework can achieve object detection accuracy between 86% and 93%, proving the system's reliability in object detection.

The experiment showed that the proposed system, using both deep learning and classical computer vision, can achieve an effective balance between detection accuracy and efficiency, making it an effective solution for real-time object detection and visual assistance.

Based on the experiment, it can be concluded that the proposed system can be an effective solution in assisting visually impaired people in their mobility and environment perception. However, the system might face certain challenges, such as device battery consumption, hardware compatibility, and training the users on how to effectively use the system.

Limitations:

Despite the promising results, there are several limitations in the proposed system. First, performance in environmental factors for this proposed system can be determined based on structured environment datasets, which may not represent actual environments well. For a more detailed investigation, performance in extreme environments could be of interest.

The system design uses a purely visual analysis of the environment, and it could benefit from the incorporation of other techniques such as ultrasonic ranging or thermographic analysis.

Thirdly, even though the multimodal feedback system presents plural forms of communication, there might be quite substantial variation regarding what constitutes the optimal feedback approach for different users. There is not much given to the system for individual preferences.

Finally, the current system depends on the good quality of the device camera and processing power, which, in most cases, is lacking in mobile devices.

Future Work:

Going forward, there will be emphasis on integrating smart glasses capabilities, allowing users to use their hands freely. Algorithms will also be improved to work effectively in low-light environments and harsh weather. All these will be achieved through multi-sensor fusion, using LiDAR, ultrasonic, and IMU sensors.

Hardware optimization techniques such as model pruning, quantization, and edge computing would allow real-time processing capabilities in wearable technology. Personal feedback systems would be adaptable based on individual users' choices and mobility patterns. Offline conversational AI with an optimized local language model would allow improved interaction irrespective of connectivity.

Avenues of further functionality in the realm of mobility would come from the navigation system's integration with map services and transportation APIs. Moreover, it would adjust to regional languages and cultures, i.e., Urdu and Sindhi. It would also seek clinical validation and approval for "Medical Device" classification following controlled trials on users with vision impairments.

Functions such as text reading, currency recognition, and face recognition (with privacy provisions), if more complex, would be introduced later. The long-term aim would be to provide an assistive space that would seamlessly interface with one's perception itself and function as an extension of it, to enable the user to live, work, and engage with the world through learning from experiences.

Acknowledgment:

The authors are grateful for the help and supervision rendered by the Department of Computer Science, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Sindh, Pakistan, during the entire process of this research work. The authors are thankful to the entire faculty and the supervisor for providing feedback and technical inputs for the development and assessment of the proposed concept named "SightMate."

References:

[1] Ruxandra Tapu, Bogdan Mocanu, Titus Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognit. Lett.*, vol. 137, pp. 37–52, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167865518308602

[2] Faiyaz Ahmad, Mohd Zeeshan Ansari, "Virtual Assistance System for Visually Impaired people using Deep Learning," *Procedia Comput. Sci.*, vol. 258, pp. 67–75, 2025, doi: https://doi.org/10.1016/j.procs.2025.04.196.

[3] Cheng Su, Litao Zhu, "Nav-YOLO: A Lightweight and Efficient Object Detection Model for Real-Time Indoor Navigation on Mobile Platforms," *ISPRS Int. J. Geo-Inf*, vol. 14, no. 9, p. 364, 2025, doi: https://doi.org/10.3390/ijgi14090364.

[4] Omar Kanaan Taha Alsultan, Mohammad Tarik Mohammad, "A Deep Learning-Based Assistive System for the Visually Impaired Using YOLO-V7," *IIETA J.*, 2023,

[Online]. Available: https://www.iieta.org/journals/ria/paper/10.18280/ria.370409
[5] "Seeing AI: New Technology Research to Support the Blind and Visually Impaired Community - Microsoft Accessibility Blog." Accessed: Feb. 07, 2026. [Online]. Available: https://blogs.microsoft.com/accessibility/seeing-ai/
[6] Alexandru Lavric, Cătălin Beguni, "A Comprehensive Survey on Emerging Assistive Technologies for Visually Impaired Persons: Lighting the Path with Visible Light Communications and Artificial Intelligence Innovations," *Sensors*, vol. 24, no. 15, p. 4834, 2024, [Online]. Available: https://www.mdpi.com/1424-8220/24/15/4834
[7] Wei Wang, Bin Jing, "YOLO-OD: Obstacle Detection for Visually Impaired Navigation Assistance," *Sensors*, vol. 24, no. 23, p. 7621, 2024, doi: 10.3390/s24237621.
[8] Lalitha Manirajee, Siti Qatrunnada Hanis Shariff, "Assistive Technology for Visually Impaired Individuals: A Systematic Literature Review (SLR)," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 14, no. 2, 2024, [Online]. Available: https://www.researchgate.net/publication/378254054_Assistive_Technology_for_Visually_Impaired_Individuals_A_Systematic_Literature_Review_SLR
[9] Xiaochen Zhang, Xiaoyu Huang, "Advancements in Smart Wearable Mobility Aids for Visual Impairments: A Bibliometric Narrative Review," *Sensors*, vol. 24, no. 24, p. 7986, 2024, doi: https://doi.org/10.3390/s24247986.
[10] Mohammad Moeen Valipoor, Angélica de Antonio, "Recent trends in computer vision-driven scene understanding for VI/blind users: a systematic mapping," *Univers. Access Inf. Soc.*, vol. 22, pp. 983–1005, 2023, [Online]. Available: https://link.springer.com/article/10.1007/s10209-022-00868-w
[11] Marwa Obayya, Fahd N. Al-Wesabi, Wafi Bedewi & Menwa Alshammeri, "An intelligent framework for visually impaired people through indoor object Detection-Based assistive system using YOLO with recurrent neural networks," *Sci. Rep.*, vol. 15, 2025, [Online]. Available: http://nature.com/articles/s41598-025-27603-8