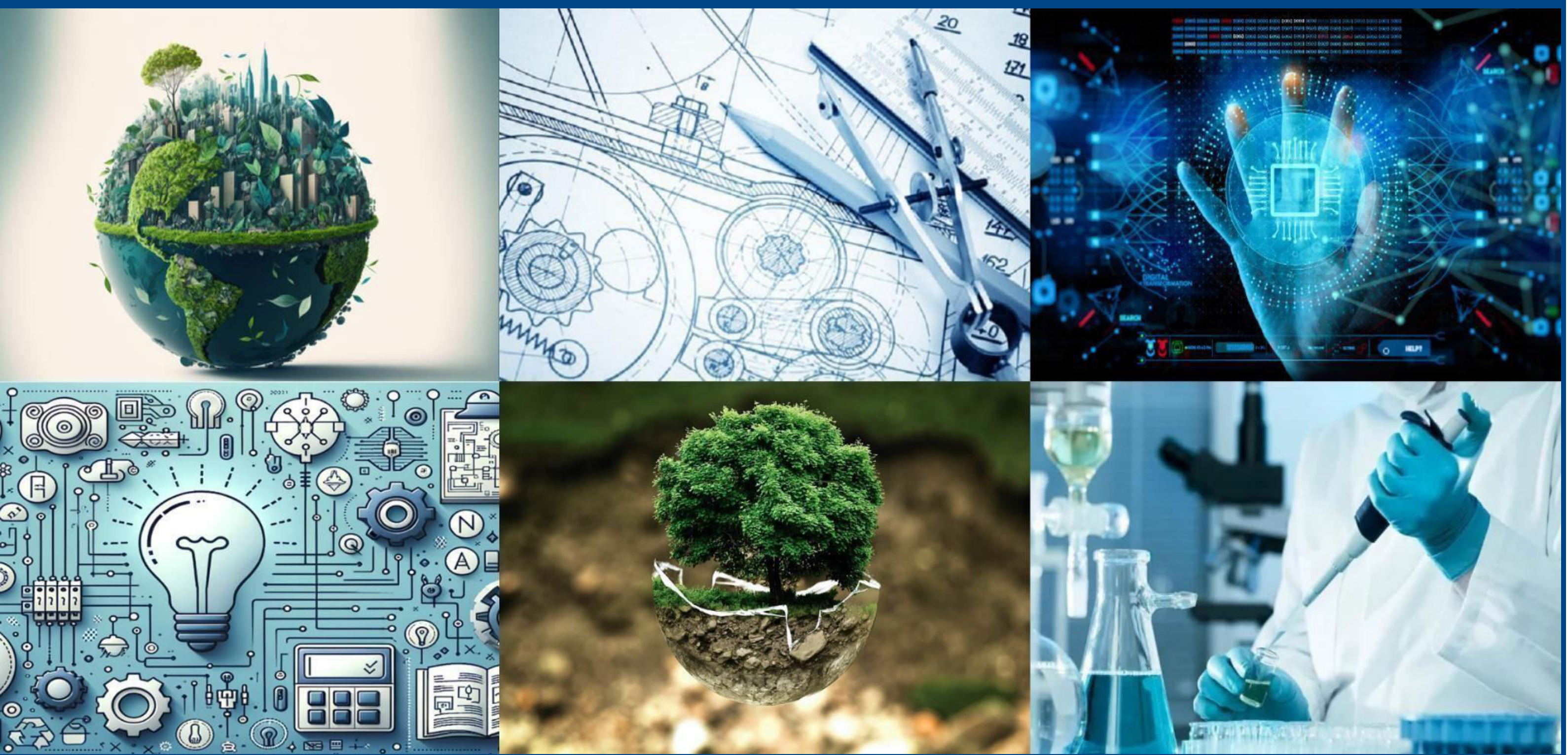
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808037

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

LumiMate AI Assistant for Visually Impaired

Annapurneshwari, Kavana N, Vanashree R, Sushmita E H, Rudresh N C

U.G. Student, Department of Information Science and Engineering, PESITM, Shivamogga, Karnataka, India

U.G. Student, Department of Information Science and Engineering, PESITM, Shivamogga, Karnataka, India

U.G. Student, Department of Information Science and Engineering, PESITM, Shivamogga, Karnataka, India

U.G. Student, Department of Information Science and Engineering, PESITM, Shivamogga, Karnataka, India Assistant Professor, Department of Information Science and Engineering, PESITM, Shivamogga, Karnataka, India

ABSTRACT: This work presents a real-time intelligent system that integrates object detection, face analysis (with age and gender prediction), and speech synthesis into a unified web-based framework. The proposed system employs the YOLOv8 deep learning model for robust and accurate object detection in live video streams. For facial analysis, a specialized YOLO- based face detector is combined with pre-trained deep neural networks to perform age and gender classification. The outputs are further enhanced through pyttsx3-based speech synthesis, enabling interactive and accessible communication of results. The complete system is deployed using the Flask web framework, providing browser-based interaction and real-time streaming capabilities. Unlike conventional static detection methods, the proposed approach delivers multi-modal, real-time, speech-assisted AI perception, with strong potential for applications in assistive technology, surveillance, and intelligent automation.

I. INTRODUCTION

Artificial intelligence (AI) and computer vision have become revolutionary technologies that are changing the way machines understand and respond to the world around them. These innovations are especially useful in the field of assistive technology, where they can support individuals with visual impairments. Traditional aids such as walking canes or guide dogs provide only limited guidance, mainly focused on obstacle avoidance or navigation. They often cannot deliver information about dynamic objects, people, or other environmental details that are essential for safe and independent living.

Recent advancements in computer vision, particularly real-time object detection models like YOLOv8, have opened new possibilities. These models are capable of identifying and classifying multiple objects within a live scene quickly and with high accuracy. In addition, face analysis techniques contribute another layer of contextual awareness by estimating features such as age and gender. This allows assistive systems to provide users with richer information about the people in their surroundings, making interactions more meaningful and informative.Despite these improvements, many existing assistive solutions remain task-specific and provide only one type of feedback, such as object recognition without meaningful narration or text-to-speech without visual analysis. Such limitations reduce their effectiveness in real-world conditions, where users require comprehensive support that combines multiple modes of recognition and communication.

To address these challenges, a real-time intelligent assistive system has been developed that integrates object detection, facial analysis, and audio narration within a single framework. The proposed system uses YOLOv8 to detect and classify objects in real time, ensuring robust and accurate recognition. For face analysis, deep neural networks are employed to predict age and gender, thereby enriching the information available to the user. To communicate these insights, the system incorporates the pyttsx3 library, which converts the detected outputs into natural-sounding speech. This narration allows users to receive immediate auditory feedback about their environment

A Flask-based web interface provides the platform for real-time interaction. Through this interface, live video from the camera is streamed, processed, and displayed in the browser. Users can interact with the system directly without requiring specialized hardware, making the solution more practical and accessible. The combination of visual recognition with spoken guidance transforms the system into an interactive tool that significantly enhances the independence and safety of visually impaired individuals.

II. LITERATURE REVIEW

[1] Voice Based Assistance About Surroundings of Blind and Visually Impaired People – Nishikanth Annamaneni, Vuppu Shankar, Hanumanthu Bhukya, and Bojja Vani (2023), published in the International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), presents a voice-commanded assistance system that combines YOLOv5 for object detection with ClipCap and GPT-2 for image captioning. The system provides descriptive feedback of the surroundings, enabling hands-free interaction for visually impaired individuals, though it suffers from high computational cost and sometimes irrelevant outputs.

[2] Text and Speech Recognition for Visually Impaired People using Google Vision – Navya Edupuganti, Sanjana Manda, and Tulasi Sree Surapaneni (2021), published in the International Conference on Smart Electronics and Communication (ICOSEC), describes an Android application that uses the Google Vision API to extract text from images and converts it into speech via the Android TTS API. It is useful for reading labels and signboards but depends on internet connectivity, limiting offline use.

[3] Visual Recognition of Text Messages to Communicate with Visually Impaired Subjects – Shruti Rathi, Swati Chandel, and Devender Sharma (2024), presented at the International Conference on Advanced Computing and Communication Systems (ICACCS), develops a system using Residual CNNs with BiLSTM and CTC loss to recognize handwritten text from the IAM dataset, which is then converted to speech using gTTS. The system performs well on clean handwriting but struggles with messy inputs and requires preprocessing, slowing real-time application.

[4] Mobile App for Enhancing Accessibility Among the Visually Impaired – Vinish Pujari, Karthik Madnal, and Divya Premchandran (2024), published in the DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAEI), proposes a mobile application integrating YOLOv10 for real-time object detection, Tesseract OCR for text and currency recognition, and audio feedback. It empowers visually impaired users with independence but faces latency and camera quality issues.

[5] Helping Visually Impaired People Take Better Quality Pictures – Maniratnam Mandal, Deepti Ghadiyaram, Danna Gurari, and Alan C. Bovik (2023), published in IEEE Transactions on Image Processing, introduces the LIVE-Meta VI-UGC dataset with 40K images annotated for perceptual quality and distortions. It proposes the P2P++ model to provide real-time quality predictions and feedback, helping visually impaired users capture clearer photos. The system is accurate but requires continuous camera use and large memory.

[6] Android Based Object Detection System for Visually Impaired – Ajinkya Badave, Rathin Jagtap, Rizina Kaovasia, Shivani Rahatwad, and Saroja Kulkarni (2020), published in the International Conference on Industry 4.0 Technology (I4Tech), presents an Android app that employs TensorFlow Object Detection API to identify objects through the mobile camera and provides audio feedback about their name, direction, and distance. It is low-cost and accessible, but less advanced than YOLO-based systems.

[7] A Review Paper on AIoT as Vision Aid for Visually Impaired People – Praveen Barapatre, Nikita Bhosage, Mukta Gaikwad, Rajani Pawar, and Kshitija Kamble (2024), presented at the IEEE International Conference for Convergence in Technology (I2CT), reviews AIoT-based assistive technologies including smart glasses, smart shoes, and mobile apps. It highlights how AI and IoT enhance navigation, text recognition, and object detection, though it lacks a detailed comparative evaluation of the approaches.

[8] An Approach to the Use of Stereo Vision System and AI for the Accessibility of the Visually Impaired – José Ricardo B. da Silva, Joao Victor B. Soares, Luana R. Gomes, and José R. Sicchar (2024), published in the International Conference on Control, Automation and Diagnosis (ICCAD), proposes a stereo vision-based approach integrated with YOLO object detection to assist navigation. Using disparity maps from binocular cameras, the system estimates depth and detects obstacles, providing audio and vibration feedback. It performs in real time on Jetson Nano but is limited by hardware cost and restricted field of view.

[9] Artificial Intelligence for Face Recognition and Assistance for the Visually Impaired – Thomas (2023), presented at the IEEE ICEPE, introduces a smart walking stick that integrates ultrasonic sensors for obstacle detection,

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808037

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

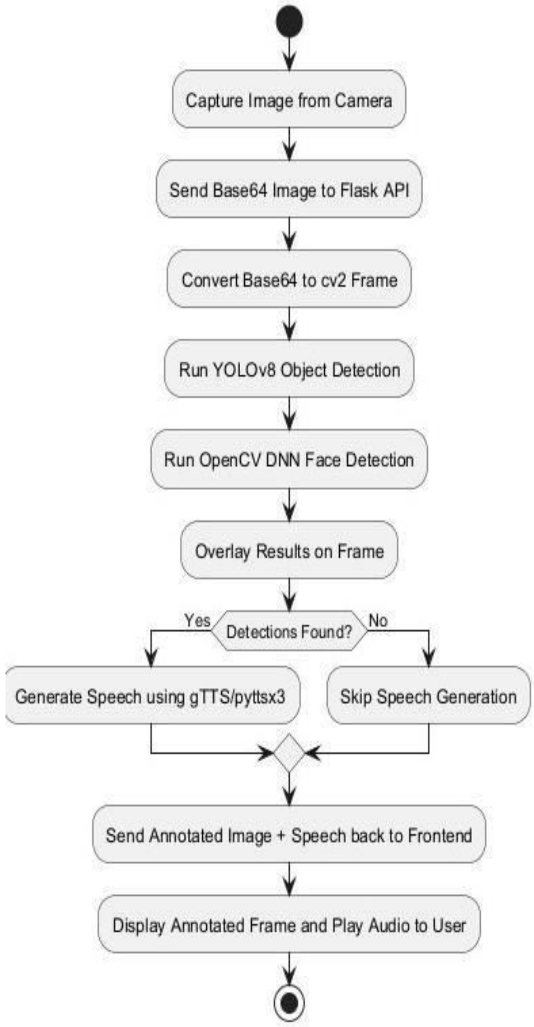
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

SSD for object recognition, and SSIM with Dijkstra's algorithm for indoor navigation. Outdoor guidance is supported with GPS and PubNub API, while emergency alerts are sent via Twilio. Although effective in providing safety and assistance.

[10]

[11] Nyx – An Educational Assistant for the Visually Impaired – Mathew (2022), presented at the IEEE ICOEI, introduces an Android-based voice-user interface called Nyx, designed to support educational accessibility. It integrates OCR for reading printed materials, gTTS for speech output, voice-based note-taking, and Selenium for web-based query resolution. While the system improves access to learning materials, it has limited offline capabilities and reduced performance in noisy environments.

III. METHODOLOGY



3.1 Video Acquisition and Pre-processing

The system begins by capturing live video streams through a webcam or an external video source using the OpenCV library. OpenCV provides efficient access to camera frames in real-time and ensures smooth frame-byframe processing at high frame rates. Each frame is acquired as a raw image matrix (BGR format) which can then be pre-processed for further analysis.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808037

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Pre-processing involves resizing the input frames to match the required input dimensions of the detection models, such as YOLOv8 and the deep neural networks for age and gender classification. Normalization is also performed to scale pixel values between 0 and 1, which accelerates computation and improves model accuracy.

3.2 Object and Face Detection using YOLOv8

The core detection module of the system is powered by YOLOv8, a state-of-the-art deep learning framework for object detection. YOLOv8 was selected due to its balance of high accuracy and low inference time, making it particularly suitable for real-time applications. For object detection, each frame is passed through the YOLOv8 network, which returns bounding boxes, confidence scores, and class labels for the detected objects. These detections are drawn directly on the video feed, allowing users to visually interpret what the system is detecting at any moment. For face detection, YOLOv8 is further fine- tuned or combined with a specialized face-detection model to ensure accurate localization of human faces within frames. Once faces are detected, they are cropped and forwarded to additional classifiers for age and gender prediction. This two- step pipeline (face detection followed by classification) improves accuracy and reduces computational redundancy.

3.3. Age and Gender Classification

After a face is detected, the system uses a Deep Neural Network (DNN) for age and gender classification. Pretrained convolutional neural networks are employed, which have been trained on large datasets such as IMDB-WIKI or Adience, to categorize gender as either "Male" or "Female" and estimate age into predefined ranges (e.g., 0–2, 4–6, 8–12, 15–20, 25– 32, 38–43, 48–53, 60+).The input face images are resized and standardized before being fed into the classifier. Once predictions are obtained, the results are displayed as labels on the video stream. This enables users not only to identify the presence of a face but also to receive immediate demographic insights.

3.4 Speech Synthesis using pyttsx3 and gtts

To complement the visual outputs, the system incorporates pyttsx3, a Python-based text-to-speech (TTS) library, to generate spoken feedback. Detected object names, recognized text, and demographic information from face analysis are converted into speech in realtime.The pyttsx3 engine is platform-independent and runs offline, making it suitable for continuous use without requiring an internet connection. The speech rate, pitch, and voice can be customized to improve accessibility, particularly for visually impaired users. This multimodal feedback loop ensures that users receive both visual and auditory information, enhancing usability and accessibility.

3.5 Flask-Based Web Deployment

All components are integrated into a Flask web application to provide a user-friendly interface. Flask enables the system to host a lightweight web server that streams the live video feed, overlays detection results,The Flask backend manages data flow between the detection models, OCR engine, and TTS module, ensuring synchronized and low-latency performance. By deploying the solution on a web server, the system becomes accessible across devices connected to the same network, extending its usability beyond the host machine

IV. RESULTS

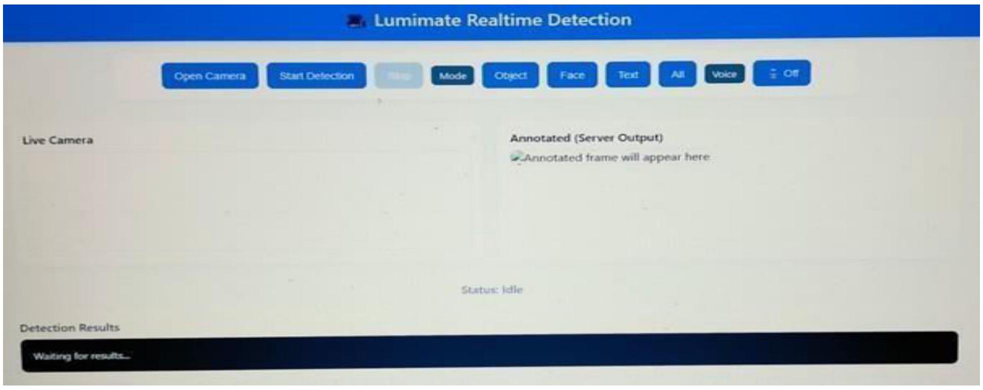


Figure 1: Interface for real time analysis

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The LumiMate Realtime Detection system provides a structured and user-friendly interface that makes real-time AI analysis simple and accessible. The top section contains interactive buttons for opening the camera, starting or stopping detection, switching between different modes such as object, face, text, or all, and enabling or disabling voice interaction. This design ensures that users can easily control the system with minimal effort. On the left side, the live camera panel displays the real- time feed from the environment, while on the right side, the annotated output panel presents processed frames with AI-based detection results, allowing users to compare input and output simultaneously.

Below the main panels, the interface features a status indicator that continuously updates the current activity, such as idle or detecting, along with a detection results section that lists recognized objects, faces, or text in real time. By combining visual, textual, and voice-based outputs, the system provides inclusive feedback suitable for both general users and visually impaired individuals. Its clear layout, easy navigation, and multimodal design make it an effective platform for assistive applications while also being adaptable for broader uses such as smart surveillance, intelligent monitoring, and real-time security systems.

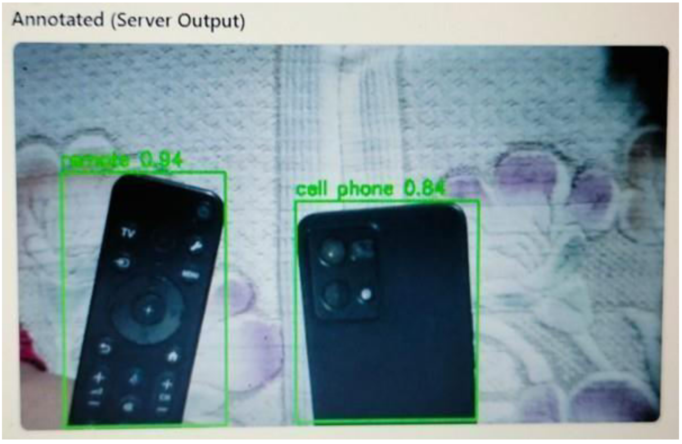


Figure 2: Object Detection Output with Bounding Boxes

This frame illustrates LumiMate's real-time object detection, identifying a remote and a phone with green bounding boxes. Confidence scores show prediction reliability, while spoken feedback helps visually impaired users recognize essential items. The system's ability to detect multiple objects simultaneously highlights its robustness and practical value as an intelligent visionbased assistant.

Figure 3: Face recognition output with gender and age

The LumiMate Realtime Detection system demonstrates its advanced capability for real-time face detection by not only identifying the presence of faces but also providing demographic insights such as gender classification and estimated age range. Each detected face is marked with a bounding box, while demographic details are displayed instantly on the live feed, allowing users to interpret the information without delay. The processing is handled efficiently on the server side, ensuring that results are generated with minimal latency and maintaining a smooth, uninterrupted experience. This demographic analysis adds a predictive layer to the system's outputs, enabling situational awareness by offering users a deeper understanding of the people in their environment. Such functionality is particularly valuable for visually impaired individuals, as it complements the system's object detection, text recognition, and voice interaction modules to deliver a holistic form of AI-powered assistance. By combining face recognition with auditory feedback, the system ensures that users not only know when someone is present but also receive descriptive details about them, thereby enriching interaction and communication. Beyond accessibility, this feature highlights the system's robustness and potential for applications in areas such as smart surveillance, personalized assistance, and human-computer interaction, making LumiMate a versatile and practical intelligent vision-based solution.

V. CONCLUSION AND FUTURE WORK

The Blind Assistance Object Detection and Navigation system offers a practical and reliable approach to supporting visually impaired individuals by integrating YOLO-based object detection and facial recognition within a Flask-powered web framework. By delivering real-time feedback about objects and faces in the environment, the system strengthens accessibility, independence, and situational awareness beyond the limits of conventional aids. Despite its effectiveness, certain challenges such as reduced accuracy in low-light conditions, cluttered surroundings, and processing delays remain. Future improvements can focus on designing lightweight and optimized models suitable for mobile platforms, making the system more portable and efficient. The inclusion of conversational AI could further enhance interaction and usability. Expanding the solution with features like currency recognition, scene description, and intelligent navigation assistance may also broaden its scope. Furthermore, leveraging edge or cloud computing could minimize latency while enabling ongoing updates. With these advancements, the system holds significant potential to evolve into a scalable and comprehensive assistive technology that greatly enhances the quality of life for visually impaired individuals.

REFERENCES

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection.
2. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788.
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
4. OpenCV Documentation. (2023). Open Source Computer Vision Library. Available at: https://docs.opencv.org
5. World Health Organization (WHO). (2022). World Report on Vision. Geneva: World Health Organization.
6. Kumaran, M., & Kumar, A. (2021). Assistive Technologies for the Visually Impaired: A Review. Journal of Ambient Intelligence and Humanized Computing, 12(6), 5673–5689.
7. Szegedy, C. (2015). Going Deeper with Convolutions. Proceedings of the IEEE CVPR.
8. Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444.