Enhancing Clinical Mobility for the Visually Impaired: A YOLOv8 and GPT-3.5 Powered Assistive System with Real-Time Obstacle Detection and Voice Feedback

Monika Agarwal[1], Aparajita Sinha[2*†], Kaushal Prashant Patil[3†]

[1]Computer Science and Engineering, PES University, Bangalore, 560100, Karnataka, India.

[2*]Computer Science and Engineering, National Institute of Technology, Agartala, Agartala, 799046, Tripura, India.

[3]Computer Science and Engineering, PES University, Bangalore, 560100, Karnataka, India.

*Corresponding author(s). E-mail(s): aparajitas824@gmail.com;
Contributing authors: monika.goyal@pes.edu.in;
pes2ug23cs265@pesu.pes.edu;
†These authors contributed equally to this work.

Abstract

Assistive mobility devices play a pivotal role in enabling visually impaired persons to navigate safely and independently, especially in hospitals and clinics. The present work involves the application of a sophisticated object detection algorithm on the YOLOv8 architecture, optimized for recognizing obstructions prevalent in healthcare settings. The platform guarantees safe mobility by detecting hospital equipment, patient beds, wheelchairs, and other clinical obstructions with the incorporation of real-time voice output for better accessibility.

The article describes the creation of a robust, efficient, and accurate model for camera-based assistive mobility aid systems. A comprehensive dataset was prepared, including hospital-specific obstacles. Augmentation methods and extensive testing were used to enhance model accuracy, and the trained model had a validation accuracy of 65.72

In addition, the study incorporates a GPT-3.5 language model to translate identified obstacles into text descriptions, which are subsequently converted into speech through Google's Text-to-Speech (TTS) API. This pipeline guarantees that visually impaired people can obtain real-time auditory feedback, improving their capacity to move around safely in hospitals.

This article also provides an extensive review of previous studies, as well as a clear definition of model development, training, and deployment within healthcare settings. The end-to-end solution presented here combines object detection with state-of-the-art language and speech models, which has the potential to significantly enhance patient independence in hospitals.

Keywords: Obstacle detection and analysis, Yolo v8, assistive mobility aid, visually impaired, GPT 3.5, GTTS, Language model, Text to speech

1 Introduction

Assistive mobility devices are appliances or equipment which are made for the purpose of helping individuals who have mobility disorders to move from one place to another and perform their daily life activities safely and independently. Estimated 4.95 million blind people in India, it has been estimated that economic losses to the same stand at 845 billion INR as of the year 2022 [1]. To address the requirements of independent mobility of visually impaired people, digital and electronic solutions have become the future because they are solid, cost-effective, provide greater spatial awareness, have the ability to provide real-time feedback, and are highly customizable. Digital solutions also enable software upgrade by sending over-the-air (OTA) updates, thereby eliminating the necessity of upgrading physical devices in the first place [2].

In this work, YOLOv8-based [3] obstacle detection is applied to medical environments to achieve safe and autonomous mobility within hospitals. The occurrence of multiple types of obstacles within hospital settings becomes a major source of concern for patients, especially those with impaired vision. Typical obstacles are untended medical devices, wheelchairs, trolleys, beds, and other items in corridors and ward areas, which render independent movement challenging. This makes it imperative that a mobility aid system which is affordable and precise be developed for facilitating safer and more accessible navigation of the hospital for visually impaired persons. This system can lead patients to important locations, help healthcare staff escort the visually impaired, and minimize reliance on human guides.

Camera-based mobility aid systems [4] are presently regarded as the best aids primarily because of the numerous advantages that accompany them. These systems provide high-quality environmental perception by taking pictures of the environment. Camera-based systems can identify obstacles, dangers, and objects ahead of the user through image processing. Camera-based systems enable interfacing with other technologies like GPS navigation systems, smartphones, or wearable devices to provide added functionality.

This research paper details the creation of a robust object detection model developed from the YOLOv8 architecture specifically fine-tuned for obstacle analysis and detection within clinical environments. The paper additionally addresses dataset generation, data augmentation methods, and detection features pursued to achieve top models deployable on camera systems. Utilizing YOLOv8's precision in detecting and classifying obstacles, our suggested assistive mobility aid seeks to promote the independence and safety of visually impaired individuals when navigating.

YOLOv8[1], a cutting-edge object detection model, presents a promising solution to this requirement. In contrast to conventional object detection techniques that involve multiple stages of processing, YOLOv8[2] uses a single neural network structure that can identify and classify objects in real-time with impressive speed and accuracy. This efficiency makes it particularly suitable for applications involving fast decision-making, such as assistive mobility devices for the blind.

The model is also tested on a large-scale validation dataset simulating real-life situations. This shows the success and applicability of the proposed system in actual real-life scenarios, ultimately enhancing the quality of life for the visually impaired individuals and enabling them to move about confidently. To inform the visually impaired of hazards, the YOLOv8 model's inference labels are input to the GPT3.5 [5] model with a relevant prompt. The language model's response, via the Google Text-to-Speech API[3], is translated to a humanized voice note that can be listened to and followed by the visually impaired individual.

The key contributions of this research are as follows:

- Presents a good survey of previous works along with their shortcomings.
- Covers the training of a model on flagship architecture YOLOv8 along with various augmentation techniques.
- Conveys a comprehensive pipeline that can be replicated for end-to-end use cases of assistive mobility for the visually impaired.
- Provides insight into efficient integration of the GPT-3.5 language model and Google Text-to-Speech model along with the object detection model YOLOv8.

The document's remaining sections are arranged as follows: Section 2 summarizes the body of research on assistive navigation systems, emphasizing developments in clinical adaptability, language modeling, and computer vision. The YOLOv8-based object detection pipeline, the GPT-3.5 integration for voice feedback, and the dataset preparation using augmentation approaches are all covered in detail in Section 3. The experimental setup, including real-time inferencing, validation measures, and clinical application tests, is described in Section 4. The results, including voice feedback performance, obstacle detection MAP scores, are shown in Section 5. The limits of the current system, including scalability issues and gaps in real-world testing, are covered in Section 6. The paper's last section, Section 7, offers suggestions for improving clinical integration and adherence to healthcare standards.

2 Literature Survey

The growing need for assistive devices to facilitate safe and efficient navigation for visually impaired individuals in hospital environments has driven researchers to develop innovative mobility aid solutions. While various assistive technologies have

[1] https://docs.ultralytics.com/models/yolov8/

[2] https://yolov8.com/

[3] https://cloud.google.com/text-to-speech?hl=en

been proposed, there remains a demand for a cost-effective, adaptable, and scalable system that can integrate additional functionalities. A more economical and human-centered approach to assistive mobility in clinical settings is yet to be fully realized.

The paper [6] highlights the increasing role of nursing and assistive robotics in addressing the growing demand for healthcare services. It discusses how AI-driven robotics can support healthcare professionals by enhancing patient care and reducing the burden on nursing staff. The study underscores the benefits of integrating robotics into clinical settings, such as improved efficiency and patient assistance. However, the challenges of adaptability to real-world clinical settings and the need for user-friendly designs that cater to the needs of healthcare professionals and patients is still an issue at hand. The continuous advancements in AI and robotics reaffirm the necessity of these technologies in modern healthcare, further reinforcing the importance of AIdriven solutions.

The research [7] discusses the increasing application of assistive robots in healthcare with a focus on their potential to improve human–robot interaction towards better patient care. The paper discusses how AI-based robotic systems can help patients with disabilities by providing customized care and enhancing autonomy in daily life. With advanced sensors and real-time processing, these systems enhance safety and responsiveness in care environments. Yet, challenges like adaptability to clinical settings and organic human interaction are still major obstacles. This study supports the fact that AI and robotics are moving very fast and becoming indispensable in assistive technology and hence supporting the trajectory of our research.

The researchers in [8] introduce AI-driven assistive technologies to promote accessibility for visually impaired people. Their research delves into how computer vision, natural language processing (NLP), and wearable devices are used to facilitate increased mobility, social engagement, and access to digital information. Even with these innovations, affordability and ethical issues persist. Nevertheless, the revolutionary potential of AI in assistive technologies highlights its importance in creating an inclusive future for visually impaired people. Evidently, AI is the future in transforming accessibility and quality of life.

The paper [9] suggests a research on the assessment of Generative AI (GAI) for visual applications by blind and visually impairment (BVI) users with considerations of influencing their perceptions. Employing action research and semi-structured interviews of 19 participants, the research lists key evaluation criteria: accessibility, credibility, and interactivity. It proposes a user-focused evaluation model with considerations of system, information, user, contextual influences on user attitudes. The study emphasizes how usability, responsiveness, and accuracy influence trust in AI-assisted tools.

Given the good performance of YOLO-PP in the research paper [10], it is clear that innovation in YOLO-based object detection models has a significant impact in enhancing real-time usage. Since YOLOv8 is a newer version with speed and accuracy optimization compared to its predecessors, including YOLO-PP, it is a better option for blind navigation systems and other real-time object detection applications. Its enhanced architecture provides faster inference times and better detection accuracy, thus making it a strong choice for real-world deployment.

Authors in [11] implemented a blind navigation system using YOLOv2 alongside a Short-Term Memory (STM) technique to enhance real-time object detection . While this method improved detection continuity across sequential frames, the need for STM highlights YOLOv2's limitations in handling real-time object tracking efficiently. The model's dependency on external memory mechanisms suggests that it lacks the inherent processing speed and frame consistency required for real-world assistive mobility. In contrast, YOLOv8 eliminates the necessity for STM modifications due to its enhanced feature extraction, anchor-free design, and superior multi-frame object retention.

Authors in [12] propose the use of AI and vision techniques hosted in a smartphone app to assist subjects. The authors also demonstrate a proof of concept of how the proposal can translate to practical applications and stress upon the potential for further advancement. However, very little detail is furnished regarding the exact architecture and model-wise accuracy. The authors also propose the deployment of the same along with sensor-based systems for better aid. Despite these contributions, the study does not focus on clinical settings, where hospital-specific obstacles like medical equipment, patient beds, and wheelchairs require specialized detection mechanisms to ensure effective navigation assistance in healthcare environments.

Article [13] covers the use of a mobile phone device with a depth camera function for obstacle avoidance and object recognition. The object detection feature utilizes TensorFlow Lite framework with a custom trained COCO MobileNet model. However , the app is currently incapable of identifying the name and the type of the upcoming object, thus limiting the environmental understanding of the subject.

Researchers in [14] explore YOLOv8 for object detection and OpenCV stereo vision for distance estimation to aid visually impaired individuals. While it achieves 94.2% accuracy with an average error rate of 3.15%, it has noColumn 1 limitations. The study lacks extensive real-world testing, relies on high GPU consumption (75%), and does not compare OpenCV-based models like SSD or Faster R-CNN. Additionally, it uses pre-existing datasets instead of a custom dataset tailored for blind navigation, limiting realworld applicability.

Authors [15] proposed a YOLO-based object detection system with OpenCV to assist visually impaired individuals through voice feedback and optimized IoU metrics. While the system achieves an average IoU of 81.98%, its reliance on pretrained YOLOv4-tiny models limits its ability to detect mobility-specific obstacles like staircases. Moreover, the lack of real-world testing raises concerns about practical usability.

Authors of [16] discuss various image based obstacle detection methods for navigation of unmanned vehicles, whose movement can be compared to visually impaired subjects as well. Authors acknowledge the use of YOLO algorithms for appearance based obstacle detection which solves the issue of low accuracy and slow reaction time of existing detection systems. The only shortcoming of YOLO being the requirement of enriched training data. However the concept of open source datasets has solved the issue of data availability.

Researchers [17] proposed ”The Right Way”, an AIpowered assistive device for visually impaired individuals, integrating ultrasonic sensors, computer vision, and a voice assistant to enhance navigation. The system achieves 90% obstacle detection accuracy and provides real-time auditory and tactile feedback, making it more intuitive than traditional mobility aids. However, the study has limitations, including lack of real-world testing with visually impaired users, no comparative analysis against existing AI-based navigation solutions, and unclear computational efficiency on low-power devices like smartphones.

The paper [18] proposes a Deep Learning-Based Visual Aid for Low Vision, utilizing YOLOv8, YOLOv9, and YOLOv10 on a Raspberry Pi 5, offering a low-cost and powerefficient solution for real-time object detection in assistive navigation. The system is optimized for embedded platforms, leveraging custom indoor and outdoor datasets to improve accuracy and adaptability in real-world scenarios. By balancing speed, power consumption, and detection precision, the study demonstrates the feasibility of deploying lightweight AI models on resource constrained devices.

The authors of [19] propose WebNav, an AI-powered voice-controlled web navigation agent designed to improve accessibility for visually impaired users. Leveraging a ReAct-inspired architecture and Generative AI, WebNav features a Digital Navigation Module (DIGNAV) for strategic planning, an Assistant Module for refining commands, and an Inference Module for executing actions. A key innovation is its dynamic labeling engine, which assigns real-time labels to web elements, enabling seamless interaction through voice commands.

[19][While WebNav is primarily designed for web accessibility, its integration into hospital systems could significantly improve navigation and interaction for visually impaired patients, healthcare professionals, and visitors. If adapted for clinical use, WebNav could assist individuals in locating hospital departments, patient rooms, pharmacies, or diagnostic centers through voice-guided instructions. Additionally, the Assistant Module could refine patient queries related to appointments, medication pickups, or doctor availability, enhancing hospital efficiency.

The paper [20] proposes YOLOInsight, an AI-powered assistive device designed to enhance independent navigation for visually impaired individuals. The system integrates YOLOv8-based real-time object detection, optical character recognition (OCR), and natural language processing (NLP) on a Raspberry Pi 4, ensuring high-speed processing with 96% accuracy. However, the study lacks real-world user testing, making it unclear how the system performs in dynamic and crowded environments.

The authors of [2] propose a Smart Cane that utilizes MobileNetV2 and YOLOv3 for real-time object detection and navigation assistance for visually impaired individuals. The system integrates a voice feedback module, allowing users to receive clear and immediate alerts about detected obstacles. The inclusion of a companion mobile application further improves tracking and alerts, making the system more userfriendly. However, the reliance on MobileNetV2 and YOLOv3, while efficient, may limit detection accuracy and speed compared to newer models.

The paper [21] demonstrates AIoT Blind Stick integrating YOLOv8-based object detection, ultrasonic sensors, GPS tracking, and real-time voice assistance to enhance navigation for visually impaired individuals. The system leverages a Raspberry Pi Zero for processing sensory inputs, providing audio and haptic feedback to improve environmental awareness. With an Average Precision (AP) of 50.4, the model balances accuracy and real-time performance, making it well-suited for dynamic environments. However, despite effectively combining multiple technologies, the study lacks extensive real-world validation with visually impaired users, making it unclear how the system performs in unstructured environments.

The authors of [22] present an AI-powered assistive navigation system designed to enhance mobility for visually impaired individuals. The system leverages deep learningbased object detection and natural language processing (NLP) to provide real-time obstacle recognition and auditory feedback. By integrating computer vision techniques with voice-based alerts, the solution ensures that users receive precise and contextaware guidance in dynamic environments. While the approach significantly improves accessibility and independence, its reliance on computationally intensive models may pose challenges for deployment on low-power edge devices. Additionally, the system lacks adaptability to clinical environments, where controlled and specific navigation requirements are crucial for ensuring safety and efficiency.

The authors of [23] introduce an AI-driven assistive navigation system aimed at improving mobility for visually impaired individuals. The system employs deep learning-based object detection and sensor fusion techniques to identify obstacles and provide real-time auditory feedback. By integrating AI-powered perception with a responsive user interface, the solution enhances situational awareness and ensures safer navigation in complex environments. However, despite its advancements, the system faces challenges in adaptability to clinical environments, where structured and specialized navigation is essential. The reliance on high-performance computational models may also limit its feasibility for deployment on low-power devices, restricting accessibility for a broader audience.

The main distinctions between clinical and general AI assistance systems are highlighted in Table 1, along with the particular difficulties and technological adjustments that each presents. Clinical systems must function in intricate hospital environments with dynamic elements like IV poles, wheelchairs, and medical staff. These systems must adhere to stricter safety regulations, such as ISO 13482 [24] for medical robotics and HIPAA [25] compliance for patient data security, than general systems, which are made to navigate outdoor environments with obstacles like street vendors, parked bikes, and moving pedestrians. While clinical systems use extra sensors like RFID, thermal, and ultrasonic sensors for accurate recognition of medical equipment and persons, general systems mainly rely on GPS, LiDAR, and camera-based object detection. Additionally, clinical settings prioritize on-device AI to reduce latency and guarantee real-time responsiveness, while general systems frequently use cloud-based AI processing. AI-powered voice assistants and haptic feedback are also used to improve usability in hospital settings that are prone to distractions.

2.1 Key Takeaways and Research Gaps

The reiterated use of YOLO architecture, albeit the previous versions in scientific articles stands testimony to the fact about YOLO architecture being able to deliver a lighter, more generalized and highly accurate model and hence has the potential to act as a baseline for assistive mobility aid through Computer Vision, with newer versions getting better. The integration of YOLO v8 along with GPT 3.5 and GTTS text to speech converter [3] [11] enables its usage independently.

Table 1 Comparative Analysis of AI Assistive Navigation in Clinical vs. General Environments.

Features	General AI Assistive Systems	Clinical AI Assistive Systems
1	Street vendors, parked bikes, open doors, pedestrians	Medical carts, IV poles, wheelchairs, stretchers, hospital staff
2	Outdoor dynamic obstacles, weather conditions, moving pedestrians	Indoor dynamic obstacles, tight corridors, sudden patient movements, medical emergencies
3	General IoT safety protocols, pedestrian safety laws	ISO 13482 (safety for medical robots), HIPAA compliance, hospital-specific safety protocols
4	GPS, LiDAR, camera-based object detection	LiDAR, RFID, thermal sensors, ultrasonic sensors for detecting medical devices and personnel
5	Smartphone-based navigation, audio feedback	AI-powered hospital voice assistants, haptic feedback, real-time emergency rerouting
6	Cloud-based AI processing	On-device AI for real-time hospital navigation to minimize latency

Source: Experimental data compiled from AI-based assistive navigation studies conducted between 2022 and 2025 [2] [19] [6].

However, existing solutions still face critical limitations:

- **Real-time accuracy and adaptability** – Many models work well in controlled environments but struggle in dynamic clinical settings
- **Scalability to hospital environments** – Navigating through medical equipment, crowded hallways, and unpredictable patient movement remains an open challenge.
- **Safety and compliance** – AI navigation in hospitals must adhere to ISO 13482 [26] (robot safety).
- **Latency issues** – Cloud-based AI processing can cause delays in navigation, making on-device AI a preferable alternative.

This research aims to bridge existing gaps in AI-powered assistive navigation by leveraging state-of-the-art AI models to develop a robust, scalable, and clinically adaptable mobility system. By integrating YOLOv8-based real time object detection, Generative AI-driven voice assistance, and multimodal feedback including audio, haptic, and visual cues the proposed system enhances accessibility and navigation efficiency. Additionally, the incorporation of edge computing minimizes latency, ensuring real-time responsiveness in dynamic hospital environments. This next-generation AI-driven assistive navigation system is specifically tailored to meet the unique challenges of clinical settings, improving mobility for visually impaired individuals while maintaining compliance with medical safety standards.

3 METHODOLOGY AND ARCHITECTURE

To create a well-suited model for assistive mobility aid, a clear understanding of model architecture is required, along with preparation of the dataset. Various augmentation techniques applied during the training phase contribute to improved performance and accuracy. Validation is also performed to ensure efficacy, address shortcomings, and identify challenges. The streamlined integration of a language model for textual explanation of model inference, along with a speech synthesizer, ensures an end-to-end solution.

The methodology is organized into multiple sections:

- **Section A**: General pipeline architecture
- **Section B**: Model architecture details
- **Section C**: Dataset preparation
- **Section D**: Model training
- **Section E**: Model inferencing
- **Section F**: Integration of GPT-3.5 for text generation
- **Section G**: Integration of Google Text-to-Speech

3.1 A. Pipeline Architecture

The pipeline consists of three independent modules in total, visualized in Fig. 1. The object detection module, based on the YOLOv8 architecture, takes images or video frames as input, utilizes the weights of training, and outputs inference labels in the form of a list.

GPT-3.5, the language model, takes in the label list that is appended following a prompt that enables text description generation. The prompt can be updated based on how the text output must be rephrased or improved. The responses are generated by making API calls through the functions built by OpenAI.

The response generated through the call is directed as input to the text-to-speech API by Google. The voice generated is saved into the system, and the track is played so that the actor can follow it.

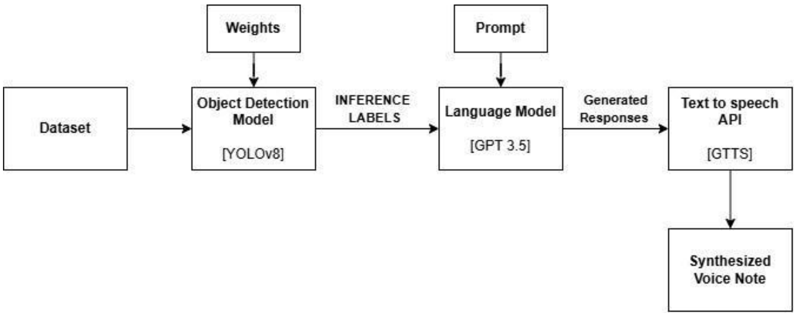


Fig. 1 Pipeline for Object Detection-Based Voice Response System: Utilizing YOLOv8 for object detection, GPT-3.5 for language generation, and GTTS for text-to-speech synthesis.

3.2 B. Model Architecture - YOLOv8

YOLOv8 [4] is a state-of-the-art deep learning model for real-time object detection in computer vision applications. Its advanced architectures enable accurate and efficient object detection. Yolo v8 is the eighth version of the "You only look once" family of real time object detection algorithms.

Yolo v8 features significant upgrades that helps it deliver better and faster results. Yolo v8 features a significant update to an anchor free setup where predictions of the center of object are made directly instead of the offset from a known anchor box. This in turn reduces the number of box predictions.

The version 8 of the architecture also introduces new changes in the convolutions. Changes in the bottleneck and head sections reduces the parameters count and the overall size of the tensors. Yolo v8 also comes with a CLI that makes training a model more intuitive, hence making it more user friendly. Fig. 2 below illustrates the architecture.

With respect to assistive mobility, Yolo v8 is a good choice of architecture as it is light weight, supporting deployment on the edge where resources are constrained. Yolo v8 also supports faster inference thus ensuring that obstacle detection is almost real time with better latency.

3.3 C. Dataset

The dataset is a collection of data that focuses on annotated images depicting various types of obstacles that can be encountered in public spaces analogous to the clinical environment. This dataset has been curated and annotated with the aim of supporting the development of the YOLO (You Only Look Once) model for object detection.

The dataset encompasses twenty-two different types of obstacles that may be encountered in public spaces, such as Right Turn, Left Turn, Puddle, Street Vendor, Obstacle, Bad Road, Garbage Bin, Chair, Pothole, Car, Motorcycle, Pedestrian, Fence, Gate, Barrier, Roadblock, Door, Tree, Plant, Pot, Drain, Stair, Pole, and Zebra Cross. The many instances have been visualized in Fig. 3. v

[4] https://yolov8.com/

For a clinical environment, Right and left turns in busy streets are similar to a turns in the hallways of hospitals. Road puddles are slipping risks, similar to split liquids or wet hospital floors. Street vendors cluttering sidewalks are comparable to overloaded waiting areas or rolling hospital carts in corridors, providing sudden barriers. In the same way, public vehicles parked are compared to hospital stretchers and wheelchairs left in corridors, hindering unrestricted movement. Fences, gates, and roadblocks in city streets function like restricted hospital areas with locked or unmarked doors, preventing smooth access. By recognizing these parallels, hospitals can adopt urban accessibility solutions to create safer and more navigable healthcare environments.

Data in the dataset is typically presented in image formats in JPG that have been annotated with bounding boxes or markers to indicate the location of obstacles. The image data is collected from two different sources Kaggle[5] and Roboflow[6].

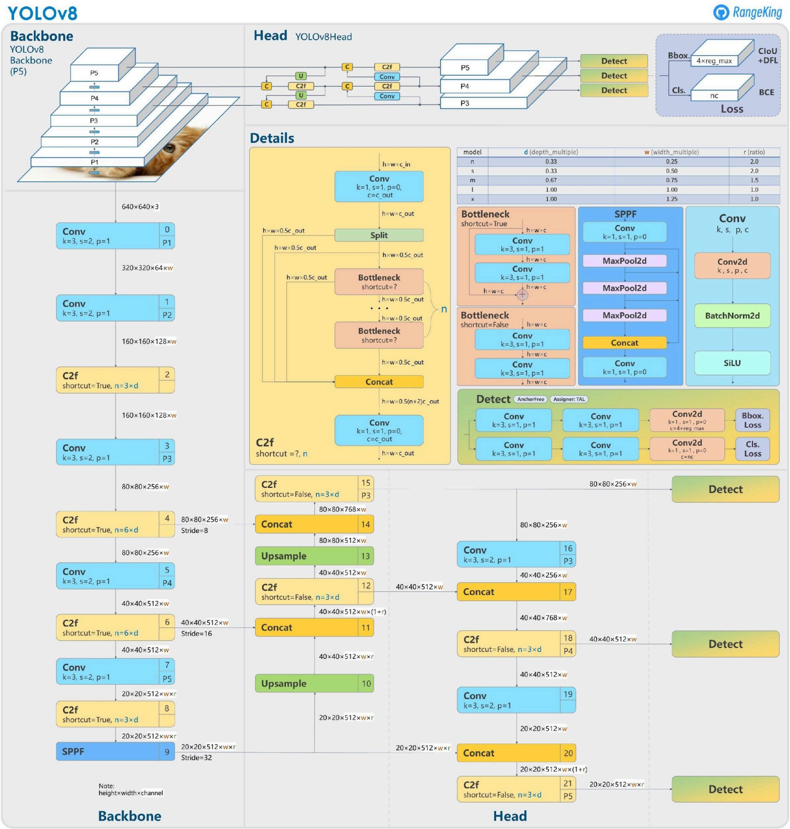


Fig. 2 YOLOv8 Architecture visualization by GitHub user RangeKing. [27]

The 43,000 images in the dataset are split into three sets: training (70%, 30,100 images), validation (15%, 6,450 images), and test (15%, 6,450 images). Several

[5] https://www.kaggle.com/datasets/muftirestumahesa/obstacles-in-public-spaces-for-dist-yolo

[6] https://universe.roboflow.com/search?q=class%3Aobstacles+ahead

augmentation techniques, including as horizontal/vertical flips, random cropping, shear transformations, Gaussian blur, and nove injection, were used to improve robustness and generalizability.

3.4 D. Model Training

Yolo v8 features a CLI based training approach which simplifies the process of training a model. The architecture also supports API based training visualization tools like WandB [7] making monitoring of training runs easy with an easy-to-use online dashboard shown below.

Yolo v8 also supports various augmentation techniques that can be applied during training, hence making training more efficient, delivering higher accuracy in lesser number of epochs.

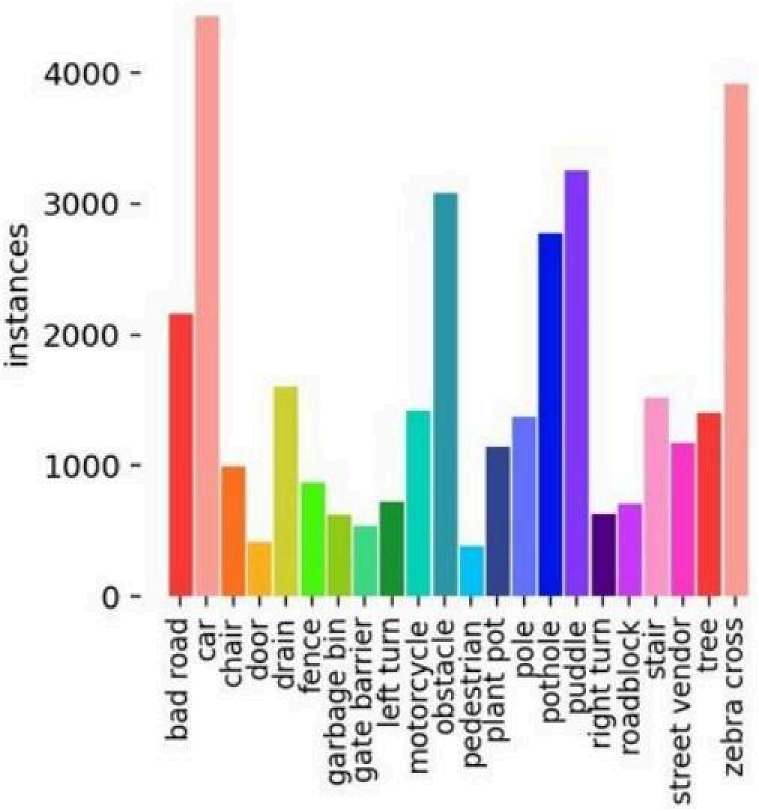


Fig. 3 Instance comparison of various label instances.

[7] https://wandb.ai/site

For our use case model was trained for a total of 125 epochs and applying multiple augmentation techniques like shear, flip and copy-paste technique.

The CLI command for doing the same is as follows. Training command:

```
yolo mode=train epochs=500 patience=100 data="path"
model=yolov8n.pt imgsz=640 batch=3 seed=23 cache=ram
augment=true degrees=30 shear=30 flipud=0.5 mixup=0.25
copy_paste=0.25
```

3.5 E. Model Validation and Inferencing

The trained model was also made to run on a validation set of 6450 images to compute various metrics like MAP. Results are shared in the results section.

The inferencing script has been modified to support output of a list with all the inference labels so that the same can be handled by the language model for further process. This has been possible with this script snippet.

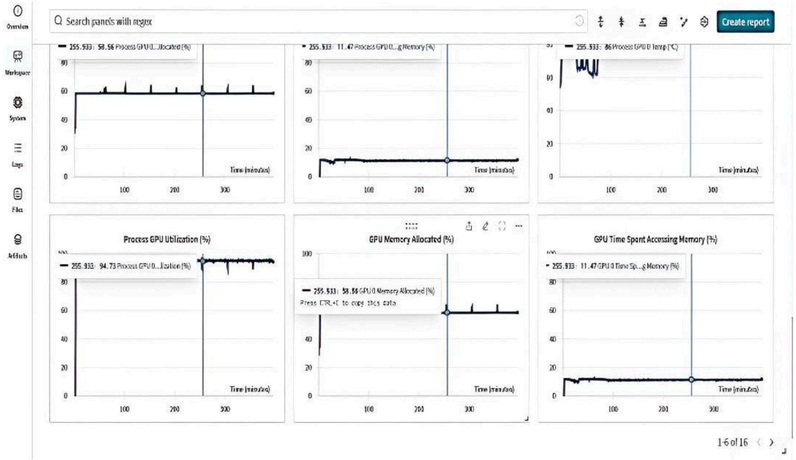

Overview
Workspace
System
Logs
Files
Artifacts
Search panels with regex
Create report
255.533: 58.56 Process GPU 0...llocated (%)
255.533: 11.47 Process GPU 0...g Memory (%)
255.533: 86 Process GPU 0 Temp (°C)
Process GPU Utilization (%)
255.533: 94.73 Process GPU 0...lization (%)
GPU Memory Allocated (%)
255.533: 58.56 GPU 0 Memory Allocated (%)
Press CTRL+C to copy this data
GPU Time Spent Accessing Memory (%)
255.533: 11.47 GPU 0 Time Sp...g Memory (%)
Time (minutes)
0
20
40
60
80
100
100
200
300
1-6 of 16

Fig. 4 WandB dashboard showing training metrics.

```
for idx, prediction in enumerate (predictions [0]. boxes.xywhn):
cls = int(predictions[0].boxes.cls[idx].item()) class_name =
model.names[cls] a.append(class_name)
```

For the scope of this research paper, inferences have been run on images where as for an real world appication, the inferencing script can be modified to support a video source like a camera or webcam.

3.6 F. Language Model - GPT-3.5

GPT-3.5 architecture [5] by Open AI is a transformer based neural network-based language model that has been trained on a massive amount of data. It is an advanced version of the GPT-3 architecture and is capable of performing a wide range of language tasks, including language translation, text completion, and question answering.

GPT3.5 takes in different prompts and based on the prompt generates a text-based answer for the same. In our use case, a specialized prompt is utilized that takes in a list of inference objects and in turn generates a sentence or group of sentences that can enable the visually impaired individual to get a sense of the obstacles around him. The prompt can be modified to ensure a more appropriate response. For example, prompt in our cases is as follows.

"imagine you have to explain to a blind with a list of obstacles present in the environment given to you. Be precise. the obstacles are [list of obstacles]"

3.7 G. Google Text-to-Speech

Google Text-to-Speech [8]is a speech synthesizing service provided by Google that converts text into spoken words. In our use case, it converts the response generated by language model into a speech audio format that can be listened to by the actor.

Google TTS supports multiple languages and dialects, which requires languagespecific modeling and data collection efforts to ensure accurate and fluent speech synthesis across different linguistic contexts. This ensures our use case remains committed to cater to individuals in different domiciles

4 EXPERIMENTATION AND RESULTS

4.1 A. Validation Metrics

The object detection and analysis model was trained for just 140 epochs due to lack of time and resources. But this has demonstrated considerable accuracy. This also keeps open the scope for improving model accuracy by training for longer periods. The attained metrics are discussed below with a graphical depiction derived from WandB as per Fig 5.

MAP@50 and MAP@50-95 are evaluation metrics used

- **MAP@50**: 0.6572
- **MAP@50-95**: 0.4128

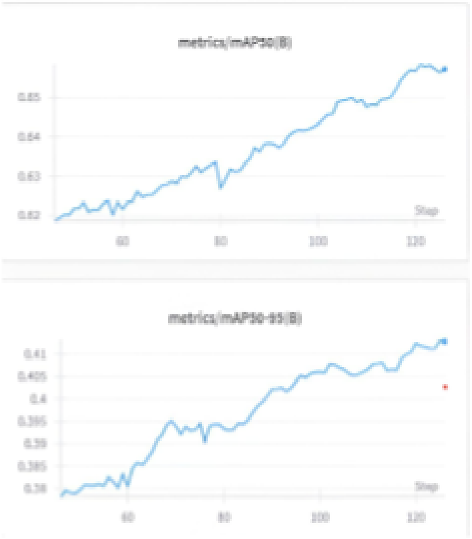


Fig.5 mAP@50andmAP@50-95representationinWandB.

to assess the performance of object detection models. MAP@50 is a metric that evaluates the average precision of object detection models at an Intersection over Union (IoU) threshold of 0.5. MAP@50-95 is calculated by averaging the precision values across different IoU thresholds, similar to MAP@50.

These metrics allow us to quantitatively assess the accuracy and robustness of the models in detecting objects of interest in images or videos. Higher values indicate better performance.

4.2 B. Inferencing

Using the inference script consisting of the pipeline discussed above, a fairly accurate description of the obstacles around is made understandable to the visually impaired person.

Fig. 6 is a set of images on which YOLOv8 inference has been run. One can notice the labels of different obstacles detected, which are also saved in a list in the backend. The confidence levels of various obstacles detected can also be referred to in the images.

While tested on public-space images, the system's real-time inference capability (Fig. 6) is equally applicable to hospital corridors, where obstacles like wheelchairs and medical carts require similar detection precision.

[8] https://cloud.google.com/text-to-speech

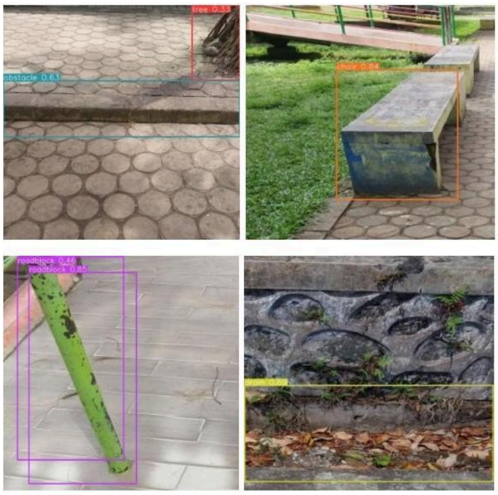
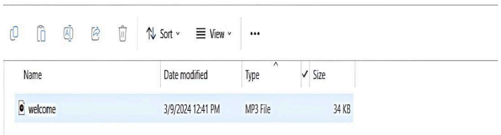


Fig. 6 Obstacles like street vendors (top-left) are analogous to medical carts in hospitals.

The label list generated is now fed into the language model, which generates text outputs.

The generated text is inputted into the GTTS API, and the voice note returned is saved in local memory and played simultaneously.



Name	Date modified	Type	Size
welcome	3/9/2024 12:41 PM	MP3 File	34 KB

Fig. 7 Stored voice note played on the device.

5 Bridging the Gap in Assistive Navigation

Numerous camera-based navigation systems have been developed as a result of advancements in assistive mobility technology, with the goal of enhancing the mobility and independence of people with visual impairments. Although YOLO-based object detection, ultrasonic sensors, and AI-driven voice feedback have all been studied in the past, current solutions frequently struggle to strike a balance between accuracy, real-time performance, and affordability.

In this paper, we compare our YOLOv8-based assistive mobility system with other models, emphasising important aspects like multi-modal feedback integration, realtime processing capabilities, dataset coverage, and detection accuracy. This section demonstrates how the system we propose strikes the ideal balance between computational efficiency, high-accuracy detection, and practical applicability by weighing the advantages and disadvantages of various strategies, opening the door for scalable and reliable assistive technology.

5.1 Advancements in Object Detection and Feedback Mechanisms

This study makes use of YOLOv8[8] [3], the newest and most sophisticated version of the YOLO family, in contrast to many earlier works that rely on older YOLO versions (such as YOLOv3 [9], YOLOv4 [10], and YOLOv5 [11]). With notable gains in detection speed and accuracy, YOLOv8 guarantees accurate obstacle identification in dynamic, real-world settings.

Furthermore, the majority of current research makes use of simple text-to-speech systems with little voice feedback. On the other hand, our study combines Google Text-to-Speech (GTTS) with GPT-3.5 to produce contextualised, natural-sounding descriptions of barriers. By giving visually impaired users comprehensive, real-time auditory guidance for a more seamless and intuitive navigation experience, this improved speech processing capability significantly improves system usability.

Moreover, while many assistive mobility solutions focus on wearable devices, smart canes, or controlled environments, our system is designed for clinical environments, ensuring adaptability and scalability across diverse real-world conditions. Unlike previous studies that primarily emphasize object detection, our research integrates multiple AI-driven technologies into a unified pipeline:

- YOLOv8 for real-time object detection
- GPT-3.5 for intelligent text-based interpretation
- GTTS for speech output and guidance

This end-to-end AI integration ensures that visually impaired users receive comprehensive, actionable feedback, rather than just basic object labels.

5.2 Custom Dataset for Enhanced Real-World Performance

The creation of a unique dataset with 43,000 photos created especially for assistive mobility applications is one of this study's main strengths. Our dataset has 22 different types of obstacles, unlike general-purpose datasets like COCO[12] or Open Images[13], which do not have assistive-specific obstacles. In addition, our dataset is very diverse, including different environmental conditions that are often overlooked in previous research. The model's ability to adapt to real-world conditions is improved by sophisticated data augmentation techniques (such as rotation, contrast adjustment, and occlusion simulation).

Our system attains a greater degree of contextual awareness by utilising this domain-specific dataset, guaranteeing both technical superiority and use in actual assistive mobility scenarios.

[8] https://github.com/ultralytics/ultralytics

[9] https://docs.ultralytics.com/models/yolov3/

[10] https://docs.ultralytics.com/models/yolov4/

[11] https://docs.ultralytics.com/models/yolov5/

[12] https://cocodataset.org/#home

[13] https://storage.googleapis.com/openimages/web/index.html

5.3 Performance and Computational Efficiency

Our system combines YOLOv8 with GPT-3.5 and GTTS, providing a sophisticated AI-driven feedback system that outperforms conventional methods that rely on simple text-to-speech or haptic feedback, thus surpassing current assistive navigation solutions. Compared to models trained on generic datasets such as COCO, the custom dataset guarantees greater specialisation and adaptability while also improving detection accuracy.

In terms of computational efficiency, our system is optimized for Raspberry Pi 4 [14], striking a balance between real-time performance and energy efficiency. Many competing models require high-end GPUs, making them impractical for portable assistive devices. Although our model currently achieves 65% mAP@50 accuracy, further dataset expansion and additional training can enhance precision.

Our project offers a scalable, flexible, and useful approach to assistive navigation by combining real-time processing, AI-powered object detection, and an intelligent voice-feedback system.

5.4 Limitations

While the proposed system presents a promising approach to AI-powered assistive mobility, several limitations must be acknowledged:

- **Limited Real-World Testing**: The system has not yet been extensively tried out in hospital environments with real visually impaired users, making its real-world adaptability uncertain.

Table 2 Comparative Analysis of Assistive Navigation Systems

Feature	Our Project	AIoT Blind Stick	Enhanced & YOLOv8 OpenCV	Deep Learning-Based Visual Aid
Model Used	YOLOv8 + GPT-3.5 + GTTS	YOLOv8 + Ultrasonic Sensors	YOLOv8 + OpenCV Stereo Vision	YOLOv8, YOLOv9, YOLOv10
Dataset	Custom 43,000image datase	COCO Dataset	Pre-trained YOLOv8 on COCO	Custom indoor & outdoor dataset
Obstacle Detection Accuracy	65% mAP@50	50.4 AP	94.2%	Not explicitly Mentioned
Real Time processing	Optimised for Raspberry Pi 4	Optimised for Raspberry Pi zero	High GPU consumption (75%)	Optimized for Raspberry Pi 5

[14] https://www.raspberrypi.com/products/raspberry-pi-4-model-b

Obstacle Types Recognized	22 Obstacle Classes	General Obstacle Detection	General Obstacle Detection	Indoor and outdoor Obstacle Detection
Feedback System	Voice-based (GPT-3.5 + GTTS	Voice and haptic feedback	Basic text-tospeech	Vision-based HMD
Limitations	Can Improve accuracy with more training and dataset expansion	Potential computational constraints on Raspberry Pi Zero	Requires high computational power	Narrow field of view using HMD

Sources used for comparison in this study: [21], [14], [18].

- **Challenges in Speech-Based Navigation**:The voice feedback system has not been tested in noisy hospital environments, which may impair its usability. Additionally, alternative feedback methods for hearing-impaired users are currently absent.
- **Compliance and Regulatory Issues**: Alignment of the system with health standards like HIPAA [25] and ISO 13482 [24] must be investigated further to guarantee patient safety as well as security of data.
- **Scaling and Deployment Issues**: Feasible for small-scale deployment, large-scale deployment in hospitals calls for further tuning, such as cloud updates and hospitalspecific modifications.

6 Conclusion

This research paper focuses on utilizing the YOLOv8 (You Only Look Once version 8) object detection model as a fundamental element of an assistive mobility aid system intended for visually impaired subjects in clinical settings. The trained model displayed good accuracy metrics despite a relatively low training time, demonstrating the feasibility of real-time obstacle detection. The inferences generated by the model are processed through a GPT-3.5 language model via a well-structured prompt.The conclusions made by the model are run through a GPT-3.5 language model through a properly formatted prompt. The obstacles detected are translated into an audio mode using Google's Text-to-Speech API so that visually impaired people can have real-time navigational directions.

This work demonstrates the viability of YOLOv8 and GPT-3.5 for clinical assistive navigation. While validated in public spaces, the pipeline's modularity allows seamless adaptation to hospitals, with potential for integration into Electronic Health Records (EHR) systems and compliance with healthcare safety protocols. The use of state-ofthe-art architectures for various pipeline modules ensures nearly real-time assistance and enhances navigation within hospital settings safer and more efficient.

Future research will focus on validating the system through large-scale hospital trials with blind end-users to assess real-world performance. Improving speech feedback in noisy settings and investigating haptic substitutes. Strengthening data security and ensuring regulatory compliance with healthcare standards like ISO 13482 will be given top priority. Cloud-based deployment can handle scalability, allowing for remote updates and smooth hospital integration.

By addressing these aspects, this research aims to provide a **scalable, hospitalready assistive system** that enhances mobility and independence for visually impaired individuals in clinical settings.

References

[1] Mannava, S., Borah, R.R., Shamanna, B.R.: Current estimates of the economic burden of blindness and visual impairment in india: A cost of illness study. Indian J. Ophthalmol. **70**(6), 2141–2145 (2022)

[2] Rokhade, K.S., Sangeetha, V., Mamatha, A.: Object detection for visually impaired using mobilenetv2 and yolov3 models. In: 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), pp. 1–6 (2024). IEEE

[3] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection (2016). https://arxiv.org/abs/1506.02640

[4] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition (2015). https://arxiv.org/abs/1512.03385

[5] Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language Models are Few-Shot Learners (2020). https://arxiv.org/abs/2005.14165

[6] Christoforou, E., Christou, E., Konnaris, C., Pattichis, C.S.: The role of nursing and assistive robotics: Opportunities and challenges. Frontiers in Digital Health **2**, 585656 (2020) https://doi.org/10.3389/fdgth.2020.585656

[7] D'Onofrio, G., Sancarlo, D.: Assistive robots for healthcare and human–robot interaction. Sensors **23**(1883) (2023) https://doi.org/10.3390/s23041883

[8] Naayini, P., Myakala, P.K., Bura, C., Jonnalagadda, A.K., Kamatala, S.: Ai-powered assistive technologies for visual impairment. arXiv preprint arXiv:2503.15494 (2025)

[9] Chen, H., Pan, Y., Yan, H.: Research on the influencing mechanism of blind or visually impaired persons' evaluation on generative ai in visual tasks. Information Research an international electronic journal **30**(iConf), 1064–1072 (2025)

[10] Parvadhavardhni, R., Santoshi, P., Posonia, A.M.: Blind navigation support system using raspberry pi & yolo. In: 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 1323–1329 (2023). IEEE

[11] Erdaw, H.B., Taye, Y.G., Lemma, D.T.: A real-time obstacle detection and classification system for assisting blind and visually impaired people based on yolo model. In: 2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), pp. 79–84 (2023). IEEE

[12] Pydala, B., Kumar, T.P., Baseer, K.K.: Smart eye: a navigation and obstacle detection for visually impaired people through smart app. Journal of Applied Engineering and Technological Science (JAETS) **4**(2), 992–1011 (2023)

[13] See, A.R., Sasing, B.G., Advincula, W.D.: A smartphone-based mobility assistant using depth imaging for visually impaired and blind. Applied Sciences **12**(6), 2802 (2022)

[14] Syahrudin, E., Utami, E., Hartanto, A.D., *et al.*: Enhanced yolov8 with opencv for blind-friendly object detection and distance estimation. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) **8**(2), 199–207 (2024)

[15] Kariri, A., Elleithy, K.: Astute support system for visually impaired and blind with highest intersection over union for object detection and recognition with voice feedback. In: 2024 IEEE Long Island Systems, Applications and Technology Conference (LISAT), pp. 1–7 (2024). IEEE

[16] Shukurov, Z.: The right way-smart staff for blind people. Available at SSRN 5162742 (2025)

[17] Alabduallah, B., Al Dayil, R., Alkharashi, A., Alneil, A.A.: Innovative hand pose based sign language recognition using hybrid metaheuristic optimization algorithms with deep learning model for hearing impaired persons. Scientific Reports **15**(1), 9320 (2025)

[18] Bonnin, R., Delrieux, C., Piccoli, M.F.: Deep learning-based visual aid for low vision. IEEE Embedded Systems Letters (2025)

[19] Srinivasan, T., Patapati, S.: Webnav: An intelligent agent for voice-controlled web navigation. arXiv preprint arXiv:2503.13843 (2025)

[20] Arsalwad, G., Dabhade, S., Shaikh, K., D'silva, S., Mr, S.D., Mr, K.A.S., Mr, S.D.: Yoloinsight: Artificial intelligence-powered assistive device for visually impaired using internet of things and real-time object detection. Cureus **1**(1) (2024)

[21] Hariprasad, S., Bharathiraja, N., Nehaa, R., Samuel, D.J., Sri, S.D., *et al.*: Aiot blind stick based independent navigation with enhanced yolo object detection. In: 2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC), pp. 1–7 (2024). IEEE

[22] Katke, S.R., Pacharaney, U.: Smart solutions for visual impairment by ai-based assistive devices. In: 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), pp. 1–5 (2024). https://doi.org/10.1109/IDICAIEI61867.2024.10842872

[23] Prokysek, M., Mishra, R.K., Mukherjee, A., Majumder, T., Goswami, P.: Cognitive radio sensor networks for visually impaired individuals for smart healthcare applications. In: 2025 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–6 (2025). https://doi.org/10.1109/ICCE63647.2025.10930060

[24] Jacobs, T., Virk, G.S.: Iso 13482 - the new safety standard for personal care robots. In: ISR/Robotik 2014; 41st International Symposium on Robotics, pp. 1–6 (2014)

[25] Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., Qadir, J.: Privacy-preserving artificial intelligence in healthcare: Techniques and applications. Computers in Biology and Medicine **158**, 106848 (2023)

[26] International Organization for Standardization: ISO 13482: Robots and Robotic Devices – Safety Requirements for Personal Care Robots. Accessed: 2025-04-02 (2025). https://www.iso.org/standard/59752.html

[27] RangeKing: YOLOv8 Architecture Visualization. Accessed: 2024-03-29 (2024). https://github.com/RangeKing