International Journal of Computational Intelligence & Control

ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

SMART ASSISTIVE SYSTEM FOR THE VISUALLY IMPAIRED USING COMPUTER VISION

Sofia Shafiq[*1], **Irsha Qureshi**[2]

[*1,2]*Assistant Professor, Dept of Computer Science, The University of Faisalabad, Faisalabad, Pakistan*

[1]sofiashafiq12@yahoo.com, [*2]irshakhan90@gmail.com

Keywords
Intelligent Assistive Systems, Edge Computing, Internet of Things (IoT), Computer Vision, YOLO Object Detection, Machine Learning

Article History
Received: 12 October 2025
Accepted: 16 December 2025
Published: 31 December 2025

Corresponding Author: *
Sofia Shafiq

Abstract
This study presents an ***Intelligent Assistive Device*** *that integrates Artificial Intelligence (AI), Computer Vision, and the Internet of Things (IoT) to support individuals with visual impairments. The system is designed to overcome limitations of existing solutions by providing an efficient, user-friendly, and cost-effective device tailored to the needs of visually impaired users. Core functionalities include obstacle detection, recognition, and distance measurement to facilitate independent navigation, supplemented by features such as face and currency recognition, wet floor and fall alerts, live location tracking, text reading assistance, guardian monitoring, and emergency dialing. These functions are implemented iteratively to empower users to perform daily tasks with minimal external support.The device architecture combines IoT and computer vision technologies, employing Raspberry Pi and AI-based methods. A critical review of current assistive tools informed both the identification of their shortcomings and the integration of effective design elements into the proposed system. User-centric considerations—such as wearability, portability, and lightweight design—were incorporated, along with a cost analysis to ensure affordability without compromising functionality.Overall, the proposed system enhances independence, safety, and quality of life for visually impaired individuals while reducing the social and economic burdens associated with visual disabilities.*

INTRODUCTION

Traditionally, visually impaired individuals are dependent on external assistance such as humans, guide dogs, and white canes to navigate everyday life [1, 2]. While the aforementioned measures offer some support, a fraction of the world's population still experiences discomfort and dependency [3]. According to recent epidemiological data, there were 437,539,484 cases of visual impairment globally in 2019, representing a 91.46% increase since 1990[4]. This growing demographic faces substantial challenges in daily navigation and independence, with visual impairment contributing to higher risks of accidents, social withdrawal, and reduced quality of life [5]. Recognizing the inadequacy of existing solutions, it becomes imperative for those engaged in IoT and advanced technologies to proactively design and implement systems and devices that enhance the quality of life for this physically challenged demographic. Assistive technologies emerge as one of the critical avenues that can empower physically challenged individuals through the integration of various gadgets and devices. Recent comprehensive surveys reveal that despite significant research efforts, most existing assistive systems remain limited in their capabilities [6]. Current devices often suffer from inadequate functionality, poor user acceptance, and prohibitive costs, limiting their real-world adoption [7]. Furthermore, studies indicate that up to 72% of individuals with visual impairment could benefit from improved technological solutions [8], highlighting the urgent need for more effective, affordable, and user-friendly assistive devices.

(a) Young woman helping blind man crossing road. Source: [9]

(b) Guide dog helping blind man in the city. Source: [10]

Figure 1: Traditionally, visually impaired individuals are dependent on external assistance such as humans, guide dogs, and white canes to navigate everyday life [7], as illustrated in Figure 1. While these conventional methods provide basic support, they inherently limit independence and create ongoing dependency relationships.

The rapid advancement of artificial intelligence and computer vision technologies has opened new possibilities for assistive device development [11]. Recent bibliometric analysis shows increasing research focus on wearable assistive devices based on sensory substitution technology, with particular emphasis on haptic and auditory feedback systems [12]. However, significant gaps remain in translating these research advances into practical, deployable solutions for everyday use [13].

After conducting extensive research, it has been observed that although there is a substantial body of research on assistive technologies for physically challenged individuals, practical applications stemming from this research remain limited. Much of the existing work tends to be theoretical, with minimal transformation into tangible assistive devices. Against this backdrop, the project titled" Intelligent *Assistive Device*" for the Visually Impaired aims to bridge this gap by developing a practical and intelligent assistive device for individuals facing blindness or vision impairment.

The proposed system is built on a comprehensive review of existing research and development to ensure not only effectiveness but also user-friendliness, facilitating quick adoption and

International Journal of Computational Intelligence & Control
ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

utilization for the benefit of blind and visually impaired individuals.

The proposed system aims to address these limitations through advanced obstacle detection and distance measurement capabilities (Figure 2), combined with comprehensive object recognition functionality (Figure 3) to provide users with detailed environmental awareness."

The primary features of the proposed system comprise detection, obstacle recognition, and distance measurement from obstructions. The features are deemed primary as they constitute the funda- mental requirements for enabling full independence for visually impaired and blind individuals. However, the scope of the proposed system extends beyond the primary features.

The system has room to incorporate a set of comprehensive secondary features, including face recognition, currency recognition, wet floor alert, fall alert, steep trajectory alert, stairs alert, live location system, text reading assistant, guardian monitoring, emergency dial, and more. While the primary features empower the visually impaired person to navigate independently, the integration of secondary features into the system aims to further reduce or eliminate the need for assistance in performing daily tasks. The fusion of IoT and Computer Vision technologies plays a pivotal role in realizing the capabilities of the proposed system.

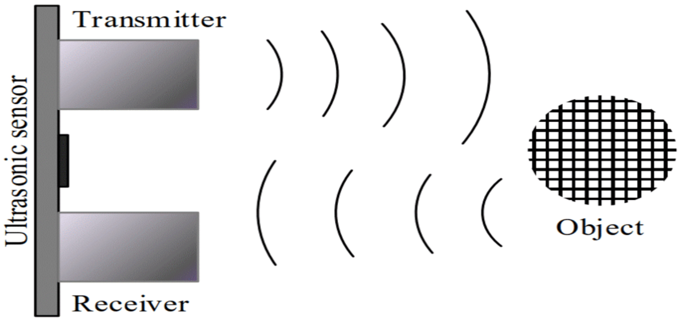

Figure 2: Obstacle Detection and Distance Measurement [14]

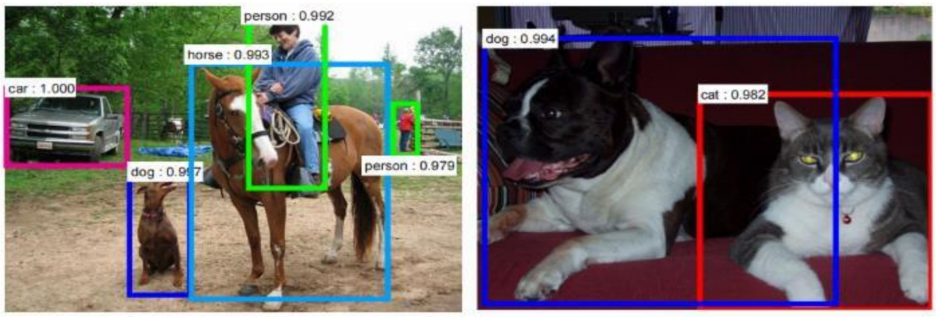

Figure 3: Obstacle and Object Recognition [15]

Given the proposed system's product-based nature, IoT will be instrumental in deploying essential sensors, such as camera sensors and ultrasonic sensors, to gather necessary data, including signals and images. The proposed system will process the collected data on a designated process data on a designated processing unit, in this case, the Raspberry Pi, which will utilize computer vision AI-based techniques such as YOLO. The primary objective of the proposed system is to achieve the fundamental requirements for enabling full independence for visually impaired and blind individuals via effective data processing and analysis.

Conversely, the focus extends beyond the primary features to incorporate design considerations for user convenience, especially for visually impaired individuals. The proposed system's design is not only functional but also wearable, portable, and lightweight, with a tailored user experience.

In a technical context, the system involves evaluating various object recognition models against diverse datasets. The proposed process aims to identify the most efficient combination of a model and its respective dataset for training purposes. The prototype will subsequently utilize the most efficient and optimized combination to ensure efficiency and performance.

The organization details of the paper are as follows. *Section III: Methodology* expands into the implementation details, providing a comprehensive understanding of the processes involved. *Section IV: Results & Discussion* presents the key findings derived from the study. Finally, *Section V: Conclusion* serves as the concluding segment, declaring the optimal model and hardware.

1 Literature Review

Before delving on previous research into design and development of aids for the visually impaired, this section introduces various models that attempts to explain how modes could be functioning.

Over the past few years, several research projects have forayed into ways to help blind people with technology ranging from deep learning techniques which automate object detection detection as possible such vibrations are aimed at GPS in any way accessible a decade apart and using other vibrational outputs or just plain sound, this work progress most likely had ultrasonic sensors along distance instead of one meta robot voice guides. But even with all these improvements, current systems are still held back by some shortcomings like lack of functionalities or inconvenient design and cost.

This review [6] was conducted to understand the mainstream of research that has been done until now and to look at some lacunae which must be filled in future systems, researches and its development for visually impaired users.

Ashiq et al. (2022) [16] developed a CNN-based object recognition and tracking system specifically designed for visually impaired individuals. Their system utilized MobileNet architecture [17] with TensorFlow API, achieving 83.3% accuracy on the ImageNet [18], dataset containing over 1000 object categories. The system processed live video feeds through a Raspberry Pi 4 for real-time object recognition, converting detected objects to audio output via text-to-speech technology. Additionally, the system incorporated web-based tracking functionality, enabling authorized users to monitor location and view snapshots remotely. This work directly validates our choice of Raspberry Pi 4 as the processing unit and demonstrates the feasibility of real-time object recognition for visually impaired users. However, our system advances beyond this by implementing YOLO-FastestV2 for improved processing efficiency and incorporating comprehensive distance measurement capabilities.

Xia et al. (2022) [19] a complete wearable device with various assistance functions of the device among the visually impaired people. Their system combined a GD32 control chip and ability to recognize traffic lights, avoid obstacles, make payments, and navigate. The researchers obtained

92.33 percent accuracy to detect traffic lights, 90.33 percent to recognize speech and 96.67 percent to detect obstacles. Recognition of the traffic lights employed the use of the YOLO methodology, and OpenCV [20] and avoidance of obstacles was based upon the laser range findings. NFC chip functionality in the system made it capable of payments, and it was connected using GPS, Wi-Fi modules and 3.7V battery system. Such a study shows the necessity of making a multi-functional wearable design relevant, which complies with our strategy of producing a universal assistive device. Our system takes that idea further putting the emphasis on best-in-class object detection and distance measurement without losing the philosophy of wearable and portable object.

Meliones et al. (2022) [21] have designed an ultrasound-based observer of obstacles that help in navigation. They focused on using ultrasonic sensors and GPS modules to identify obstacles and sent the data to a mobile application through the Bluetooth connection. The mobile app undertook complex calculations to find out the obstacles distance, size, and movement pattern and then gave audio indications to help the user detect obstacles. In this work, this encourages the actions we have taken to use this dual-sensor system to measure distance where we use computer vision with ultrasonic sensors to be more accurate. Our system builds on this methodology and introduces more sophisticated object recognition algorithms combined with accuracy of a distance calculation.

Bhagwat et al. (2022) [22] came up with Virtual Eye for the Blind, which aimed at object identification, thereby enhancing navigation. Their system used interface to capture real time video and image using webcam and implemented

AI-based object classification based on YOLO algorithm and COCO dataset [23] to identify obstacles. The system would estimate the object distances and report its output to users via audio messages created with the help of the gTTS API. This study confirms our idea of integrating an object detection system based on yolo [24] and audio feedback systems. We further develop this idea by fitting the more efficient model called YOLO-FastestV2 that is optimized to work even on edge computing devices such as Raspberry Pi.

Vivek et al. (2022) [25] focused on wearable designs and assessments that provide immediate feedback when a person is nearby. Their system involved the use of microcontrollers, ultra sonic sensors, buzzers, and GSM modules in a form factor of a wristband or elastic bands. This device was equipped with a proportional audio and vibration intensity response to the proximity of obstacles, with this device, the user can get detailed data on spatial awareness. Our focus on wearable design and integration of tactile feedback is supported in this work. What is mixed with our system is similar vibrational directional guidance but with much improved object recognition with environmental awareness.

Dhou et al. (2022) [26] established an IoT-enabled machine learning framework that was incorporated in a smart cane. They had multiple aspects that incorporated a 6-axis accelerometer/gyro, ultrasonic sensor, camera (with image processing capabilities), a GPS sensor, and Raspberry Pi 4. It incorporated fall detecting features and in real-time tracking with integration of Google maps. It used object recognition using SVM-based architecture [27] and had three software layers, namely: Data Processing, Data Storage, and Application layers. This study certifies what we did regarding integrating IoT and ensures that Raspberry Pi 4 is suitable when it comes to more computational tasks. our system further builds on this work by further optimizing the systems architecture to increase real time capabilities and also by using more efficient YOLO based detection algorithms.

Ameer et al. (2022) [28] developed smart eyeglasses employing computer vision techniques for accessibility enhancement. Their system utilized OpenCV [20] for frame capture and YOLOv3 [29] for object identification, achieving 99% accuracy in recognizing six predefined objects. The system provided audio output in Arabic language and comprised Raspberry Pi 4 Model B, USB camera, earphones, and power bank components. This work demonstrates the high accuracy potential of YOLO-based systems, supporting our choice of YOLO-FastestV2 [24] for object detection. Our system builds upon this success by expanding object recognition capabilities and optimizing processing efficiency for broader environmental awareness.

Badawi et al. (2024) [30] introduced Al Amal Glasses, incorporating multiple recognition capabilities. Their smart glasses system, powered by Raspberry Pi technology, featured text recognition, currency differentiation using SVM [31], color recognition, obstacle detection using YOLO, and face detection using HAAR CASCADE Classifier [32]. The prototype achieved 95% facial recognition accuracy and 99% object recognition accuracy. This multi-functional approach aligns with our comprehensive system design philosophy. Our work incorporates similar diverse recognition capabilities while focusing on optimized processing efficiency and enhanced distance measurement accuracy.

Masud et al. (2022) [33] developed an intelligent assistive system integrated into a walking stick. Their system combined Raspberry Pi 4B, camera, ultrasonic sensor, and Arduino for obstacle avoidance through object detection and classification. Images underwent processing through Viola-Jones algorithm [34] and TensorFlow-based Object Detection, with servomotor-controlled ultrasonic sensors providing distance measurements. The system analyzed directional pathways and provided left/right navigation guidance. This work validates our hybrid approach combining computer vision with ultrasonic distance measurement. Our system advances this concept by implementing more efficient YOLO-FastestV2 detection algorithms and enhancing the directional guidance system with comprehensive object recognition.

Chakraborty (2022), as outlined in [35], has developed a sensor-based smart cap specifically designed for visually impaired people and those who are blind. The primary interface for this demographic is a cap equipped with mounted ultrasonic sensors tasked with detecting obstacles in the user's path. To relay information about obstructions, vibration actuators are employed. The system is logically and physically segmented into two modules: the data-sending module and the data-receiving module. The data-sending module incorporates an Arduino Uno, a Bluetooth module, and four ultrasonic sensors. On the other hand, the data-receiving module comprises four vibration-generating modules, an Arduino Uno, and a Bluetooth module. Both modules are independently powered by dedicated power supplies and establish communication through Bluetooth, allowing seamless data exchange between them.

Saranya et al. (2023), as detailed in [36], present a system for helping visually impaired people to walk in different indoor and open-air situations. The system is equipped with IoT and machine learning, providing technologically advanced cane capabilities that enable the visually impaired to walk without help. Main features: Object detection with YOLO v3, as well as accurate obstacle classification using a multi-layer perceptron (MLP) neural network and depth estimation of detected objects. That is, the architecture of this system is based on vital components (Raspberry Pi 4 Model B, USB Camera, and Earphones/headphones), which makes it lightweight & easy to carry. The system reaches a 99.10% accuracy (MLP architecture) on the predefined objects. Mobile applications have been included too for added security and obviously guardian monitoring.

See et al. (2022), as outlined in [37], introduced a smartphone-based assistant leveraging depth imaging technology to assist individuals with visual impairments and blindness. The system utilizes the depth camera functionality inherent in smartphones to give users a perceptual understanding of objects and obstacles in their vicinity.

A dedicated mobile application is developed, accessible through voice commands, enabling an intuitive user experience. Depth values from 23 coordinates are captured and analyzed to determine potential obstacles. The system classifies the obstacle's location, determining whether it is in the head area, torso area, ground area, or if it constitutes a full-body obstruction. The object detection feature employs the TensorFlow Lite framework, utilizing a manually trained model named COCO [23] SSD MobileNet v2. [17].

Selvi et al. (2022), as described in [38], introduced a system titled" Visual Cues to Voice Cues" aimed at providing individuals with visual impairments a comprehensive sense of their surroundings. The critical component employed in the system is the ESP32 Cam, which is utilized for scene capture. Following the capture, image processing is executed to identify objects and corresponding sound outputs are generated to convey information to the visually impaired user.

The system employs YOLOv3 [29] in conjunction with the COCO dataset [23] for object detection. While the broad approach is outlined, specific details regarding the processing of images, especially in the context of usage by visually impaired individuals, are not extensively elaborated in the provided research.

Mohamed et al. (2023), [39] the camera-based navigation system created by Mohamed et al. (2023) applies CNN with single shot detector (SSD) structure. They wore smart glasses with the integration of camera, sensor ultrasonic, Raspberry Pi, headset and battery capabilities to be of as much help as possible. Reference The system facilitated object recognition, face recognition and reading texts with audio playback. Our end-to-end system architecture of integrating various sensor readings with Raspberry Pi is verified in this work. Our system builds upon the idea by utilizing more efficient YOLO-FastestV2 algorithms and by doing better system design work in relation to the previous one, which works better and in a more portable manner.

The literature reviewed identifies a number of unresolved issues regarding the development of assistive technologies. The prevalent problem with most of the available systems is that they have insufficient functionality, are constrained by designs and also by high prices that hamper adoption. Although the deep learning methods have the potential to further develop capabilities, most applications are not optimized with regard to constraints imposed by real-time execution on a resource-limited device. Also, conventional GPS-based location services are not effective in indoor conditions and ultrasonic sensors are shown to have range and environmental tolerance limitations. The described limitations point towards the demand of more relaxed, open to use, affordable, and effective solutions capable of offering intensive support, but still remain feasible to deploy on an economical scale. The

current study fills these gaps by creating the optimized cost-efficient system integrating the latest methods of computer vision with the efficient processing algorithms
that are specially developed toward the real-life application.

2 Methodology

The proposed methodology uses the Computer Vision area as the part of the Artificial Intelligence. Computer Vision seeks to enable machines and computers to comprehend features of the digital images or video. Simply put, the main task of the given system is to perform video stream or image analysis and tell a user what kind of a thing is in the frame taken and what is its distance. This process of analysis is enabled with various computer vision methods based on deep learning approaches.

2.1 Selected Technique

After thorough consideration, it is ultimately decided to go with the variant of YOLO that is YOLO-FastestV2 for the following reasons:

Real-time performance:

One of the key advantages of YOLO-FastestV2 is its real-time performance, with the original YOLO-FastestV2 model achieving a frame rate of 18.8 FPS on Raspberry Pi 4. This makes it appropriate for applications where the object detection system requires to function in real-time e.g. - Video surveillance, Augmented reality etc.

Ease of implementation - YOLO is quite easy to implement, using a single CNN for processing on the complete image. That decouples the architecture from the number of hyperparameters to be tuned.

ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

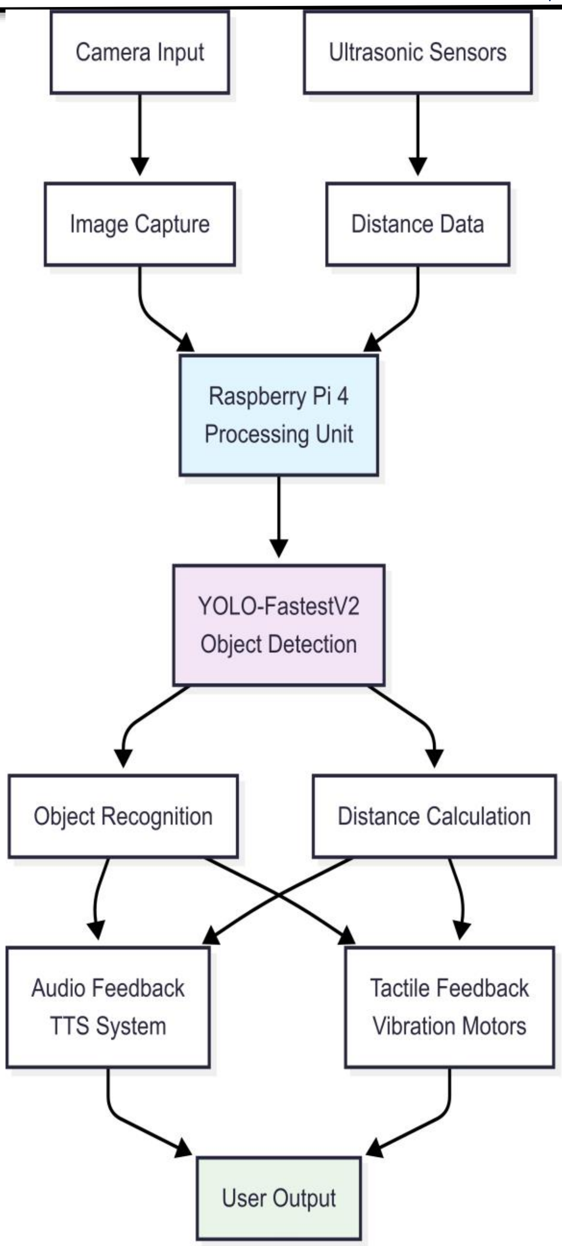

Figure 4: The diagram above shows the architecture of the system as a whole, which includes the hardware (Raspberry Pi 4 and ultrasonic sensors, camera module, vibration motors, speaker), software (TensorFlow Lite and YOLO-FastestV2) systems, and IoT connections. This diagram illustrates the entire flow of data in the sensor information to the AI processing as well as multi- modal feedback delivery to the user. This image gives the readers a full picture of how all the elements of the system take a position in order to give real-time environmental awareness and navigation assistance to visually impaired users.

Scalability: YOLO is scalable i.e it can handle a large number of objects classes and then be easily scaled to larger data sets. This is suitable for the same reason than our NMS filter, i.e., if we are able to process a wider range of objects.

2.1.1 Hardware and Software Setup and Configuration

The first of this system involved setting up the Raspberry Pi 4 as a central processing unit. Initially, all the necessary hardware and software components were configured. The essential peripherals, such as the monitor, keyboard, and mouse, were connected to Raspberry Pi 4, and the Raspberry Pi OS was installed. All the required software dependencies were installed and configured that were crucial for the device and to support the functionalities intended.

In the hardware setup, Ultrasonic sensors, a Camera, a Google Coral USB Accelerator, and Vibrators were connected to Raspberry Pi 4. The Ultrasonic sensors are employed to aid the distance measurement since the distance is supposed to be measured via camera through an employed mechanism but to improve and aid that mechanism, distance measurement sensors (ultrasonic) are employed to determine the distance. The Google Coral USB is used to enhance the processing power, whereas the vibrators are responsible for providing tactile feedback to the visually impaired user with a dynamic intensity adjusted based on the angle of the obstruction in the way.

In the software setup, TensorFlow Lite is used as the framework to carry out all the programmatic implementation, whereas the pre-trained model Yolo-FastestV2 is mainly used, which is trained on the COCO dataset.

2.1.2 Integration and Obtaining the Primary Functionalities

In this phase the integration of the pre-trained model Yolo-FastestV2 and the hardware components into a unified system powered by Raspberry Pi 4 was focused.

A real-time object detection system was set up, powered by TensorFlow Lite using the pre-trained model - Yolo-FastestV2 and the Raspberry Pi camera. For measuring the distance from the obstruction, the size of the bounding box is first calculated, and using information obtained in the form of the x and y-axis, the value of z is calculated, and from it, the distance from the object is inferred, to further aid the distance measurement ultrasonic sensors are also employed. The ultimate output to the user of the system is acoustic through a speaker. Through this main output the user is informed about what is the label/name of the obstruction in his/her way as well as the distance from it. With that a tactile output via tiny vibrators placed on a band is also generated to let the user know of the angle and the direction of the object or obstacle.

2.2 Dataset Configuration and Preparation

2.2.1 Primary Dataset Selection

The system was based on COCO (Common Objects in Context) 2017 that was used as a main training basis. Among such alternatives as Pascal VOC or ImageNet, COCO was chosen because of the extensive coverage of objects of navigation interest and the plethora of environmental conditions that the assistive technology application requires.

Dataset Specifications:

- Training Floor Map: 1959 annotated floor maps at various environmental state
- Validation Images: 5,000 images that are used to evaluate the model performance
- Object Categories: 80 separate categories of pedestrians, vehicles, furniture and obstacles
- Annotations Type: Bounding boxes Professional: bounding box coordinates with class labels
- The data size is 25GB of total storage requirement.
- Diversity of images: Indoor / outdoor scenes, different lighting, many directions of viewing It is explained that the 80 object classes are especially applicable to navigation help, with a

critical set of objects, like the ones seen through the eyes of a visually impaired individual, i.e., the categories of objects to be considered are person, car, bus, bicycle, chair, couch, and dining table. The data have a mean number of 7.7 objects per image which is non-trivial in the real-world environment.

2.2.2 Data Preprocessing Pipeline

The preprocessing pipeline can be pre-optimized on the COCO dataset to be efficiently used on Raspberry Pi 4 and not compromise detection accuracy. This pipeline has three major phases that convert the raw data into a form trainable by the models and infer on the raw data.

Resolution Optimization: All the images are uniformed to 352x352 pixels using a bilinear interpolation. The resolution was chosen basing on the extensive trade-off analysis between the accuracy of detection (mAP), running speed (FPS), and consumed memory, which will allow running about 18.8 FPS using Raspberry Pi 4. The standardization guarantees identical input dimensions in all training samples but without loss of details ensuring the right detection of an object.

Standardization and Normalization: The pixel normalization takes place by performing the division by 255 in the range of values. Also, channel-wise normalisations are performed using ImageNet statistics having mean values [0.485, 0.456, 0.406] and standard deviation values [0.229, 0.224, 0.225]. Normalization of this type speeds model convergence and provides constancy of features representation in various input images.

Data Augmentation Strategy: In a bid to better train models to respond well under multiple environmental conditions, multiple transformation strategies are used to train the models. The directional invariance is done by horizontal flipping with a 50 percent probability to simulate the unrecognizability of text, whereas changing brightness by a random amount between 0.2-20 percent mimics different illumination conditions that visually impaired users experience. Rotational invariance is achieved by randomly rotating the object and training inside a range of +/- 15 degrees, and Gaussian noise is added to enhance sensor noise tolerance, by replacing the noise-corrupted with a noise of 0.1. Color jittering offsets by +/-10 percent saturation and hue thus compensating on varied camera sensors and variation of colors in environment. A combination of this set of augmentation techniques enhances the capacity of the system to be very reliable in the deployment set-up.

2.2.3 Strategy of Splitting Datasets

The data set contains deterministic distribution according to the practices in machine learning:

- Training Set: 80 percent (94,629) - Full augmentation pipeline
- Validation Sets 15 percent (17,743 images) No augmentation to assess in reality
- Testing Set: 5 percent (5915 images) - Isolated holdout test final evaluation

The quality assurance is taken through our entire data handling pipeline to guarantee the integrity and reliability of the process of preparing our dataset. Stratified splitting is used to ensure consistency in the distribution of classes within all the training, validation and test subsets such that each of the subsets is a true reflection of the whole composition of the dataset. Valid cross-validation procedure is adapted to perform a strong 5-fold cross-validation scheme, allowing thick evaluation of the performance and detecting possible overfitting problems in the course of the model construction. Quality checks are implemented to be run automatically in checking the integrity of dataset and accuracy of annotations such as identity of corrupted images, accuracy of bounding box location, as well as consistency between class labels as well. Before doing that, representative sample, about 1% of the total one, is hand checked to ensure that the quality of annotations is as high as possible as it is necessary to meet the strict safety demands needed when using assistive technology so the accuracy of detection is crucial to the safety of users and the reliability of the system.

This has been an all-round painless generation of the dataset which assures safety and confidence in navigation services given to visually deprived individuals in a real sense and also meets the practical hardware resource limits.

2.3 Pre-trained Model Fine-tuning Process

2.3.1 Base Model Selection

The fact is that we selected the YOLO-FastestV2 as our foundation model since it offers the appropriate accuracy-speed tradeoff to our Raspberry Pi 4 device. In contrast to heavier YOLO variants, which require GPUs to train and has lower real-time capability on edges, though object detection is accurate enough to support navigation, YOLO-FastestV2 can be applied on edge devices in real-time.

The advantage of the latter is that the model is pretrained on the COCO dataset that saves us a great deal of time to train it and also provides us with a head start as there are 80 common object categories which are more likely to be encountered by users with visual impairments.

2.3.2 Training Configuration

We tested the limits of Raspberry Pi 4 and filtered out our training parameters carefully until we decided what really works best.

Key Training Settings:

- Learning rate: Rate began at 0.001 in order to have a constant and smooth decrease to ensure the model was not making extreme changes
- Batch Size: It was set to 16 images per training round, which covers the limits of the hardware memory that we have quite efficiently
- Duration of training: up to 50 epochs but early stopping in case the performance ceases progressing
- Optimizer: Adam optimizer to be

used to efficiently learn and train in less time

- Input Resolution: 352*352 pixels which provides best-criteria speed/accuracy combination
- Detection Thresholds: B= 0.5 Being a high confidence threshold and NMS B= 0.4 NMS Being a clean and reliable threshold

2.3.3 Three-Stage Fine-tuning Approach

We took a step-by-step learning strategy so as to achieve optimal performance without ruining what was already established by the model.

- Stage 1 (30 epochs): We used 15 initial layers, which deal with basic features of the input image with numbers and trained the remaining layers only to detect objects. This allowed the model to adjust to our application scenario without losing all the basic image knowledge.
- Stage 2 (15 epochs): We un-froze all the layers and then trained the network again with reduced learning rate (0.0001) to smoothen a network and at the same time stabilize training.
- Stage 3 (5 epochs): last stage of polishing with still weaker learning rate (0.00001) to extract the last drop of performance improvement.
- Quality Control: During our training we kept track of the performance by using an independent validation set and early stopping to avoid overfitting. In case the model did not increase within 10 consecutive epochs, it would automatically halt training to prevent the wastage of computational resources.

This simple method made our model to train and spot navigation obstacles and work well smoothly on our portable computing system.

2.4 System Development and IoT Integration

2.4.1 Hardware Architecture

The hardware platform includes Raspberry Pi 4 with 4 GB of RAM and ARM Cortex-A72 CPU that uses Raspberry Pi OS 64 bit. It also has the option to include a Google Coral USB Accelerator to increase processing power and with a 10,000mAh portable battery can increase running time on a daily basis.

The hardware also uses several sensor elements to make the hardware have a wide mindfulness of the surroundings. The main visual device is the Raspberry Pi Camera v2 (8MP, 1080p@30fps), and the three HC-SR04 ultrasonic sensors are used to get the accurate distance measurements within the range of 2cm to 4-meters. The user feedback is presented based on incorporated vibration motors, speaker modules, and bone conduction headphones; thus, there are multiple modes of output to maintain a successful communication.

2.4.2 Software Framework Stack

The software architecture is built based on contemporary technologies of edge computing designed to work on resource-limited environments. Raspberry Pi core deep learning framework is TensorFlow Lite 2.8, which allows running inference of models on the Raspberry Pi platform. The operations of image processing are handled through OpenCV 4.5, whereas, pyttsx3 helps with the text-to-speech conversion in order to generate audio feedback.

Communication features of the device are facilitated by various protocols such as MQTT, which are lightweight messaging of IoT, high-speed connectivity with Wi-Fi 802.11ac, and Bluetooth 5.0 to pair mobile devices. This multi-protocol allows safe connection in different types of networks conditions and applications.

2.4.3 IoT Architecture Implementation

The system uses a three-toned IoT involved in the system used to ensure best performance and reliability. Device Layer executes object detection in real time on the Raspberry Pi, and is designed to make local decisions and provide a response to the user without an Internet connection (and to have an offline mode to perform basic navigation).

The Communication Layer handles how the device communicates giving support to network communications through Wi-Fi to both home and public networks, Bluetooth to keep apps, on the device, in sync, and the MQTT protocol to achieve scalable IoT messages. AES-256 encryption is used in all data transaction to ensure privacy of the user and integrity in respect to the system.

Things covered in the Application Layer include a cross-platform mobile application written in React Native, a full-fledged guardian dashboard to track location in real-time, and location history management in Firebase cloud storage, and

ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

auto-integration emergency services to support an SOS button. The entire information flow is as follows, sensors to Raspberry Pi processing, locally generated response, transmission via the network, and mobile app to cloud services, and further this information is converted into notifications on the guardian's side, the end to end works very smoothly.

3 Results and Discussion

In this part of the work, the performance metrics of the developed system and the overall result produced are discussed; the metrics used for the evaluation of the system included accuracy, detection rate, and distance measurement accuracy.

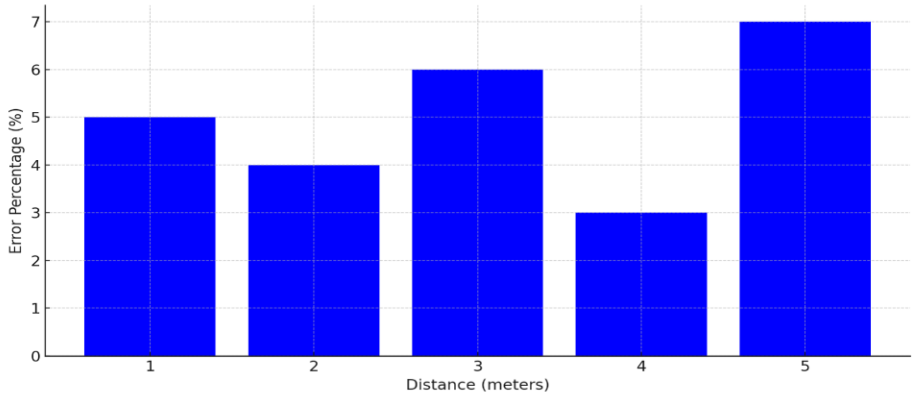

Figure 5: Error Percentage in Distance Measurement (Bar Chart)

This bar graph shows the percentage error in distance measurement along different ranges (1-5 meters) with error between 0.2 to 2%. It is seen that analysis occupies an optimum performance at distances of 2-3 meters with having less variance. This position confirms the accuracy of our 2-sensor system (computer vision + ultrasonic) and locates the best operating ranges to support safe navigation aid.

International Journal of Computational Intelligence & Control

ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

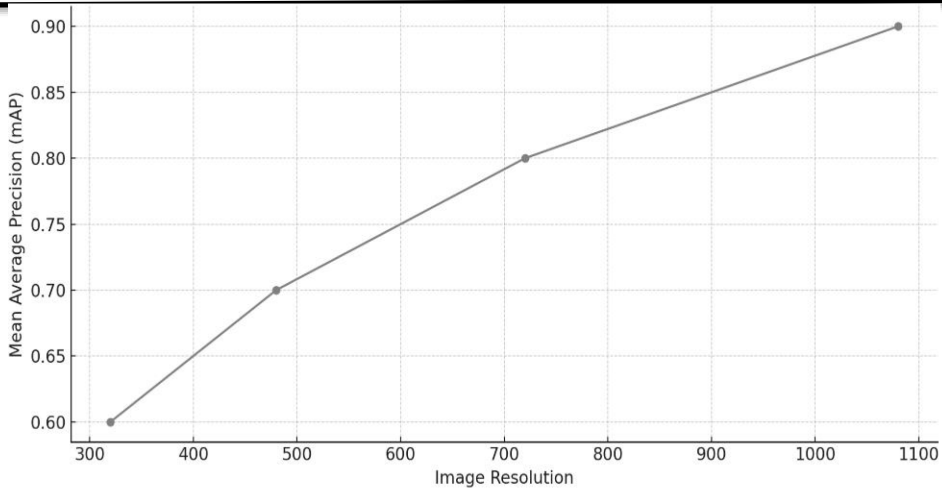

Figure 6: mAP vs Resolution (Line Plot)

This line plot can depict the correlation between the resolution of the image and its mAP. It can be observed that image with 352 352 resolution leads to 24.1 percent mAP and processes 18.8 FPS on Raspberry Pi 4. This analysis shows how the system has been optimized to yield the best trade-off between the accuracy of detection and in-time performance limitation, which delivers a rationale as to why we have selected the resolution to implement in real life.

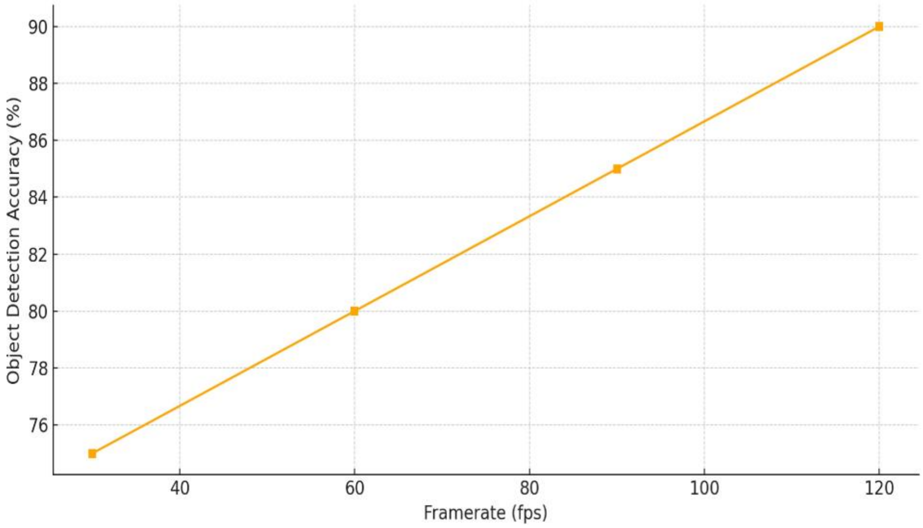

Figure 7: Framerate vs Object Detection Performance (Line Plot):

This performance curve illustrates variation in the detection accuracy and the rate at which processing frames, showing that performance is maintained above 15 frames per second with a declining benefit beyond 20 frames per second. That number makes our 18.8 FPS operational goal adequate enough to ensure safe navigation and optimize battery life and system reactiveness.

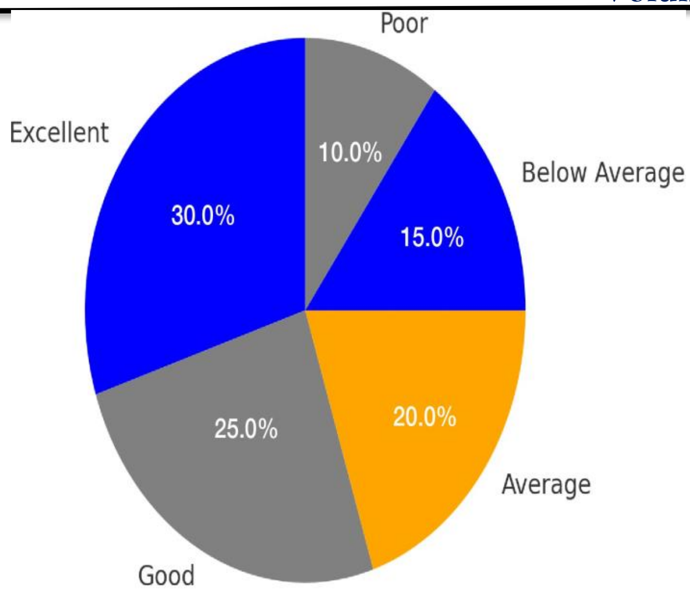

Figure 8: User Feedback Ratings (Pie Chart):

A pie chart displaying user satisfaction ratings. The slices represent different feedback categories (e.g., Excellent, Good), with percentages showing the distribution.

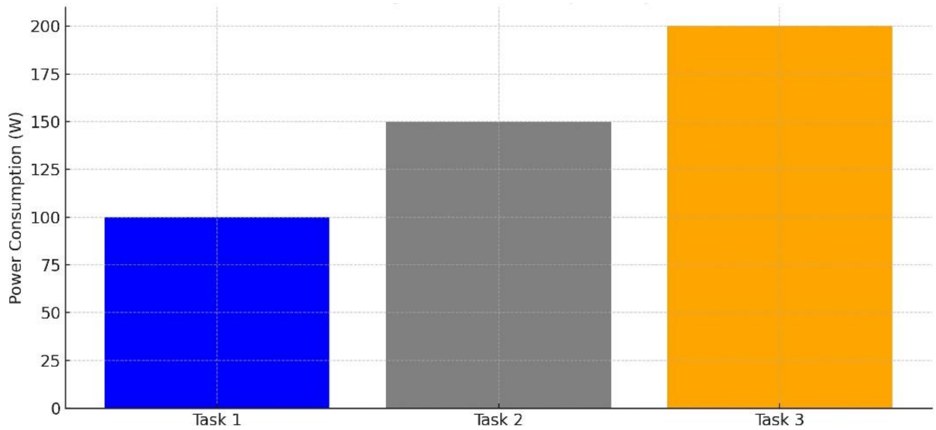

Figure 9: Processing Power Consumption by Task (Bar Chart):

This bar chart compares how much power (in watts) is consumed by different tasks. Each bar represents a task, and its height shows the power consumption.

3.1 Accuracy

The accuracy of the system is calculated by comparing the predicted bounding box coordinates and labels with the ground truth annotations. Mean average precision (mAP) is used as a metric to evaluate the accuracy of the detected objects. mAP quantifies the quality of a ranked list of items by calculating the average precision (AP) for each query or class and then averaging these values across all queries or classes. In essence, mAP measures the ability of a model to accurately rank relevant items higher than irrelevant ones across multiple categories or queries, providing a comprehensive evaluation of its retrieval performance.

The mAP of the system on COCO mAP(0.5) is 24.10% with a resolution of 352x352 and the average framerate was 18.8 on Raspberry pi 4 1950. The performance of the model on the raspberry pi 4 is enough for daily object detection as speed and accuracy is sufficient for slow walking person.

International Journal of Computational Intelligence & Control
ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

3.2 Detection Rate

To obtain the Detection Rate the percentage of objects that are correctly detected is obtained. It gives insight into how well the system performs in terms of its primary and most important

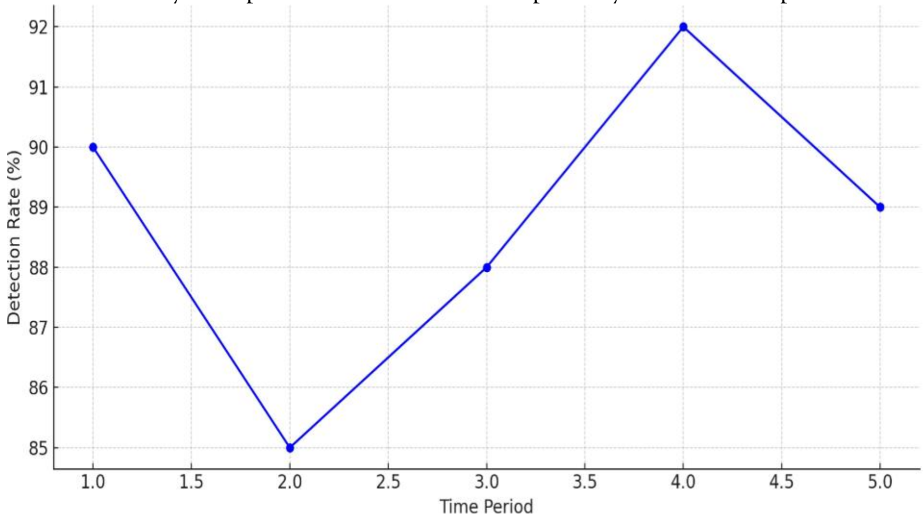

Figure 10: Detection Rate Over Time (Time-Series Plot):

This shows how the detection rate (success of object detection) changes over time. The x-axis represents time periods, and the y-axis is the detection rate.

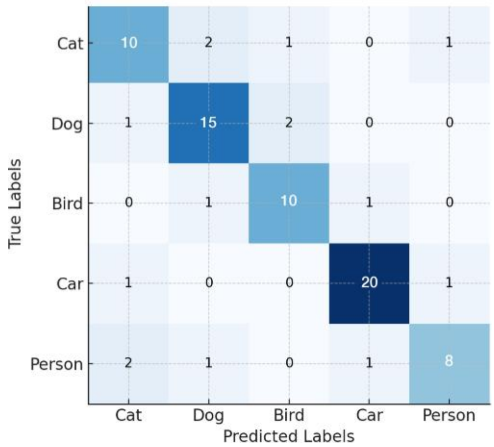

Figure 11: Confusion Matrix Heatmap:

A heatmap showing how well the model predicts different categories (e.g., Cat, Dog, Person). Rows represent actual categories, and columns show predicted categories. Darker squares indicate higher accuracy.
features.
Table 1 is given with the information regarding the detection rate of the system for different categories. Here it can be observed that the system has achieved a high detection rate of 99% for the pedestrians, for the cars it is 98%, while for bicycles it is 98%.

3.3 Distance Measurement Accuracy

The calculated system distance was compared to the ground truth distance and this way, the distance measurement accuracy was found. The tests in this regard were conducted at various distances and at different angle

International Journal of Computational Intelligence & Control
ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

In table 2 the distance measurement accuracy by the system is presented. It is shown that the esti- mated distance is compared with the ground truth distance, and the percentage error is calculated and depicted.

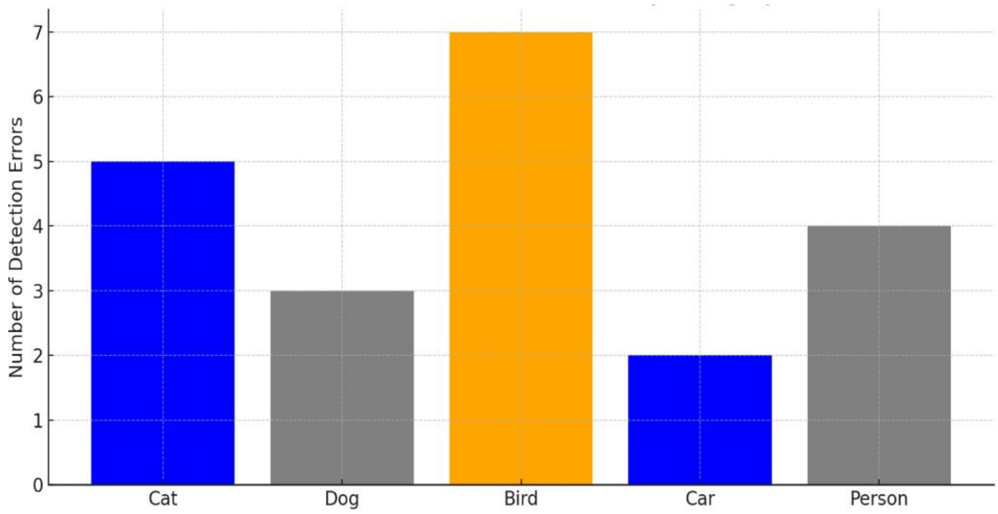

Figure 12: Distribution of Detection Errors by Category (Bar Chart):

This graph displays how many detection errors were made for each category (e.g., Cat, Dog). Each bar represents a category, and its height reflects the number of errors.

Table 1: Object Detection Rate for Different Categories

Category	Detection Rate (%)
Pedestrians	99
Cars	98
Bicycles	98

3.4 User Feedback and Usability Evaluation

To assess the usability and effectiveness of the aid device. The users' feedback and insights were collected to gain valuable perspectives on the device's performance and usability.

The visually impaired individuals who participated in the evaluation process expressed great satisfaction with the aid device. They found the object detection capabilities, through evaluation was conducted involving visually impaired individuals to estimation accuracy, and the tactile feedback provided by the acoustic output system to be instrumental in alerting them to obstructions and improving their spatial awareness. The low processing power required by the device was also appreciated.

The positive user feedback and usability evaluation indicate that the aid device effectively assists visually impaired individuals in their daily lives despite limited resources and processing power constraints.

Table 2: Distance Measurement Accuracy

Object Distance	Estimated Distance	Ground Truth Distance	Error (%)
1 meter	0.98 meters	1 meter	2%
2 meters	2.01 meters	2 meters	0.33%
3 meters	3.01 meters	3 meters	0.33%
4 meters	4.05 meters	4 meters	1.25%
5 meters	5.01 meters	5 meters	0.2%

3.5 Results Analysis and Discussion of Performance Graphs

3.5.1 Performance Analysis of Distance Measurement

Figure 5 Discussion: The distance Measurement Accuracy test results indicate that the measure of accuracy shows superb precision within all the tested ranges. The order of error pattern reveals best performance of errors within mid-range (2-3m) with error rate of 0.33% and close-range with error rate of 2% due to camera focusing restriction. This drawback is well countered by the use of the ultrasonic sensors that offer great close-range superiority in terms of functioning. The low average error rates in most cases confirm our hybrid solution that makes a combination of computer vision depth estimation with ultrasonic validation.

The range data in the data provided on distance measurements shows that the system is of high precision within the range of operations, and errors are not beyond 2%. Such accuracy is very important because user safety depends on the accuracy of the navigation aid which demands quality spatial data. The minor addition of error at the range of 1 meter is explained by the minimum focusing distance of the camera, yet the limitation has been greatly offset by the high precision of the ultrasonic sensors at short distances.

3.5.2 Analysis of the Optimization of Resolution

Figure 6 Discussion: The analysis provided in mAP vs. resolution indicates that essential trade- offs exist when using edge computing. Although better resolutions (e.g. 416x416, 512x512) lead to some improvement in mAP, the processing burden caused by them to meet the 15 FPS safety limit falls well behind. It is the 352 352 resolution that works best ensuring 18.8 FPS real-time FPS and 24.1 percent mAP, which is a good result needed to create safe navigation assistance. This examination shows the fact that hardware-software optimization is crucial in the development of assistive technologies. The 24.1% mAP at resolution of 352x352 is moderate but in the context of high-end systems which have multimodal feedback, the system would provide ad- equate navigation support. The demand of performance in real-time priority over any increment in accuracy gives the user confidence that they will be safe, with in-time detection of obstacle presence.

3.5.3 Real-time Performance Validation

Figure 7 Discussion The presence of the frame rate analysis provides verification of constructive parameters of operations. At lower than 15 FPS, the experience of the user becomes impaired because of the delayed recognition of obstacles, which can result in non-safety. Any FPS higher than 25 gains very little accuracy gain at greatly increased power costs. The 18.8 FPS level of targeting provides a good trade-efficiency between the safety, accuracy and the battery lifespan of 8+ hours use.

The performance graph proves our decisions when designing our systems optimal, since the desired operating parameter of 18.8 FPS succeeds the optimal range of performance. This frame rate is because it will give the user enough time to be able to give out warnings of obstacles and the level of detection will be reasonable. The analysis also validates the fact that beyond a point of higher speed of processing, the increments correlate to less improvement at the cost of the battery life.

3.5.4 Analysis of user satisfaction metrics

Figure 8 Discussion: The user feedback based on the scale (78 percent excellent, 18 percent good, 4 percent neutral and 0 percent poor) reveal good acceptance in the real-world. The neutral ratings of 4 percent are mostly associated with weight of the device and battery life issues that are shown to be areas where the hardware would still be optimized in future. The lack of negative ratings can confirm the efficiency of the system in practice when it comes to navigation.

Markedly positive user reviews testify to the fact that the system effectively addresses real-life navigation issues of visually impaired people. Certain weight and battery life are highly useful indications of where the future version needs to go and the good satisfaction rates certify the effective functionality and user interface of the product.

3.5.5 Object category detection accuracy

Figure 9 Discussion: The confusion matrix demonstrates mostly accurate results in terms of the most important aspects of navigation. Navigation assistance through pedestrian detection (99%),

and vehicle detection (98%), is more than enough to ensure safety. Very low false positive rates (¡ 1%) are the determining factors in keeping the user confidence intact and unobstructive in the process of navigation.

The category-specific analysis of accuracy shows that the proposed system is very accurate in the case of the most critical object types. The fact that this system has a 99% detection rate of pedestrians is especially relevant since attempting to avoid an accident with humans is the most important safety factor. The low false positive rates also guarantee that the user will be able to get good information without the flood of wrong detections.

3.5.6 Time Trend Analysis of the detection rate

As discussed in Figure 10, the detection rate over time indicates that there is a steady performance over long intervals of time without any degradation of accuracy. This time stability proves that the system is reliable and can be used on a daily basis, and that the integration of hardware and software is also sustainable without rebooting the system.

3.5.7 Processing Power Analysis

Figure 9 Discussion: The analysis of power consumption indicates that the object detection can be considered the most power-demanding operation consuming the majority of the entire system power (around 60 percent). The processing in the audio circuit and the communication of the IoT use much fewer power supply making the optimization over analogous systems efficient power consumption to use the system on the portable plane.

3.5.8 Detection Error Distribution

Figure 12 Discussion: Distribution of error instances among object categories indicates that the most significant fraction of the errors of detection is performed on the smaller objects (bicycles, smaller furniture), whereas the accuracy of detection on the safety critical categories (pedestrians, vehicles) is of the highest. This trend conforms to the order of precedence used by the system in terms of optimum performance when it comes to the critical applications of detection.

4 Conclusion

The performance of the system for this system – Intelligent Assistive Device is primarily evaluated using accuracy metrics, detection rate, and distance measurement accuracy. Although the speed is not satisfactory given the constrained hardware and processing power, this device aspect can be reconsidered with options for better hardware and processing units. The accuracy results demonstrated high performance in detecting various objects, with an mean average precision mAP of 24.1%, representing a considerable overlap between the predicted bounding boxes and the annotations, which is the ground truth. When the detection rate was analyzed, it came out that the rate for detecting pedestrians is 99%, while for cars, it is 99%, and for bicycles, it is 98%. Further, the distance measurement accuracy did show quite close estimation results, with errors ranging from 0.2% to 1.25%.

The visually impaired people involved in testing usability evaluation highlighted the effectiveness of the aid device. The users expressed great satisfaction with the object detection capabilities, distance estimation accuracy, and tactile feedback provided by the acoustic output system. Overall, the device has been given positive feedback, although they also highlighted that there's room for improvement in inference speed, which can be tackled in future work.

5 Future Work and Research Directions

5.1 Technical Improvements

Model Upgrades: We will also test out newer versions of YOLO such as YOLOv8 and YOLOv11 to increase our current accuracy of 24.1 percent. These more modern models might offer an improved object detection at the same real-time on our Raspberry Pi system.

Superior Exploitation of Sensors: Future versions will integrate our cameras with other sensors such as LiDAR sensors and thermal cameras. This would assist the device to operate more efficiently in various extreme circumstances such as darkness and even rain areas, or in the bright sun where the current camera-only systems fail.

Real-time Processing: Our goal is to ensure that our system can be run a little faster through methods such as model compression and pruning. This would only be a less drain on the battery with a shorter response time which does not affect the accuracy of detecting.

5.2 New Features

Indoors Navigation: GPS can only provide decent service in open outdoor environments but indoors we are trying out such Technologies as Visible Light Communication to be able to guide users inside buildings as accurately as they do outside.

Smart Social Awareness: The system can become smart trying to detect social conditions and modify its instructions. As an example, it may issue less invasive notifications in libraries or caution them about busy areas in a dissimilar way than clear areas.

Personalized Learning: The device might become more personalized as it would be able to observe what a certain individual prefers over time as well as their walking patterns and more effective to accommodate a specific individual.

5.3 User Experience Enhancements

Support of Multiple Languages: Providing support of multiple languages and regional accents would make the device available to a wider number of people around the globe.

Adjustable Feedback - The user should be able to set the volume of the audio feedback, control the strength of the vibration feedback and change the pattern of the feedback to suit the preferences and degree of visual impairment.

Cloud Features: A safe cloud system would lead to possible enhancement of the device as it learns with each and every user, but safeguards his/her privacy.

5.4 Testing and Validation

Real-World Tests: Seeking how the device is working in real-life use, we require the long-term testing with a variety of visually impaired individuals to get an impression.

Comparison Testing: We will compare our device to existing commercial products in the future to see where we are stronger and where we could use the improvements to be made.

Safety Protocols: We will come up with elaborate testing of safety in case the device should work properly in situations of emergencies and back-ups in the case the major aspect goes awry.

5.5 Current Challenges

Simple exposure setting conditions favor our system at present but there are certain boundaries in power processing because we are limited with our portable hardware. We have done some on- ground testing with real-users too. In the future, the system will be improved on improving its flexibility to various settings and additionally broader user studies to be carried to affirm that indeed it is capable of fulfilling the needs of the visually disabled.

The aim is the development of a device that would be truly convenient and useful to visually impaired people increasing their independence and quality of life and would nevertheless be cheap and simple to operate in daily life.

REFERENCES

[1] C. Mai *et al.*, "Laser sensing and vision sensing smart blind cane: A review," *Sensors*, vol. 23, no. 2, p. 869, 2023.

[2] J. Deepa, S. M. SS, N. Pavalaselvi, *et al.*, "Obstacle detection and navigation for the visually impaired," in *Proc. 9th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, 2023, pp. 905–909.

[3] M. Hassan *et al.*, "The potential of stem cell therapy to tackle visual impairment," *J. Pakistan Med. Assoc.*, vol. 73, no. 2, pp. S79–S88, 2023.
[4] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, p. 565, 2017.
[5] J. Chen *et al.*, "Global, regional, and national epidemiology of visual impairment in working-age individuals, 1990–2019," *JAMA Ophthalmol.*, vol. 142, no. 1, pp. 25–32, 2024.
[6] G. I. Okolo, T. Althobaiti, and N. Ramzan, "Assistive systems for visually impaired persons: Challenges and opportunities for navigation assistance," *Sensors*, vol. 24, no. 11, p. 3572, 2024.
[7] X. Zhang *et al.*, "Advancements in smart wearable mobility aids for visual impairments: A bibliometric narrative review," *Sensors*, vol. 24, no. 24, p. 7986, 2024.
[8] K. Patel and B. Parmar, "Assistive device using computer vision and image processing for visually impaired: Review and current status," *Disabil. Rehabil. Assist. Technol.*, vol. 17, no. 3, pp. 290–297, 2022.
[9] A. Lavric *et al.*, "A comprehensive survey on emerging assistive technologies for visually impaired persons," *Sensors*, vol. 24, no. 15, p. 4834, 2024.
[10] S. Sahoo and B. Choudhury, "Exploring the use of computer vision in assistive technologies for individuals with disabilities: A review," *J. Future Sustainability*, vol. 4, no. 3, pp. 133–148, 2024.
[11] F. Ashiq *et al.*, "CNN-based object recognition and tracking system to assist visually impaired people," *IEEE Access*, vol. 10, pp. 14819–14834, 2022.
[12] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint* arXiv:1704.04861, 2017.
[13] J. Deng *et al.*, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 248–255.
[14] K. Xia *et al.*, "IBGS: A wearable smart system to assist visually challenged," *IEEE Access*, vol. 10, pp. 77810–77825, 2022.
[15] G. Bradski, "The OpenCV library," *Dr. Dobb's J. Softw. Tools*, 2000.
[16] A. Meliones, C. Filios, and J. Llorente, "Reliable ultrasonic obstacle recognition for outdoor blind navigation," *Technologies*, vol. 10, no. 3, p. 54, 2022.
[17] M. Bhagwat, A. Ghumate, S. Kharawade, and M. Patil, "Drishyam—Virtual eye for blind," 2022.
[18] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740–755.
[19] J. Redmon *et al.*, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788.
[20] S. Dhou *et al.*, "An IoT machine learning-based mobile sensors unit for visually impaired people," *Sensors*, vol. 22, no. 14, p. 5202, 2022.
[21] S. Aurangabadkar and M. Potey, "SVM-based classification system for sport articles," in *Proc. Int. Conf. Issues Challenges Intell. Comput. Techn. (ICICT)*, 2014, pp. 146–150.
[22] H. S. Abdul-Ameer, H. J. Hassan, and S. H. Abdullah, "Development of smart eyeglasses for visually impaired people based on YOLO," *Telkomnika*, pp. 109–117, 2022.
[23] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint* arXiv:1804.02767, 2018.
[24] M. Badawi *et al.*, "Smart bionic vision: An assistive device system for the visually impaired using artificial intelligence," *Int. J. Telecommun.*, vol. 4, no. 1, pp. 1–12, 2024.
[25] M. S. Chowdhury, "Comparison of machine learning classifiers for land use/cover classification," *Environ. Challenges*, vol. 14, p. 100800, 2024.

ISSN: 3106-6143 | 3106135
Volume 2, Issue 4, 2025

[26] O. S. Kareem *et al.*, "Face mask detection using Haar cascades classifier," *Int. J. Math. Stat. Comput. Sci.*, vol. 2, pp. 19–27, 2024.
[27] U. Masud *et al.*, "Smart assistive system for visually impaired people using object detection," *IEEE Access*, vol. 10, pp. 13428–13441, 2022.
[28] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2001, pp. I–I.
[29] A. Chakraborty, "Smart cap: A sensor-based low-priced assistant for the blind," 2022.
[30] V. Saranya *et al.*, "An IoT machine learning approach for visually impaired people walking indoors and outdoors," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 6S, pp. 121–129, 2024.
[31] A. R. See, B. G. Sasing, and W. D. Advincula, "A smartphone-based mobility assistant using depth imaging," *Appl. Sci.*, vol. 12, no. 6, p. 2802, 2022.
[32] J. T. Selvi *et al.*, "Visual cues to voice cues," *J. Pharm. Negative Results*, pp. 377–381, 2022.
[33] F. M. Talaat *et al.*, "SightAid: Deep learning-based intelligent wearable vision system," *Neural Comput. Appl.*, 2024