

Advancements in AI-Based Object Detection for Visually Impaired Individuals

Dr. Annapoorna B R¹, Arjun Amit Kharade², Ayush Gakre³, Bhoomi Jain⁴ and Nikhitha S⁵

¹Assistant Professor, Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, Bangalore-560078, India

Email: annapoorna-cs@dayanandasagar.edu

²⁻⁵Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, Bangalore-560078, India

Email: arjunkharade1643@gmail.com, ayushgakre2000@gmail.com, jain.bhoomi2512@gmail.com, snikhitha632@gmail.com

Abstract—This study proposes an AI-driven assistive system aimed at enhancing autonomous navigation and situational awareness for visually impaired individuals. The system integrates real-time object detection using YOLOv5, monocular depth estimation via MiDaS, and heuristic-based pathfinding through the A* algorithm to ensure accurate spatial understanding and obstacle avoidance. Contextual environmental descriptions are generated using BLIP and GenAI, while auditory guidance is provided through the Google Text-to-Speech (gTTS) engine. The proposed solution demonstrates robust performance in both indoor and outdoor settings, offering a scalable, portable, and efficient tool that significantly improves independent mobility, safety, and quality of life for visually impaired users.

Index Terms— YOLOv5, BLIP, MiDaS, A* Heuristic and GenAI.

I. INTRODUCTION

Visual impairment affects a vast number of individuals around the world, significantly limiting their ability to move independently and carry out daily activities. Tasks such as navigating busy or unfamiliar spaces, identifying objects, and avoiding physical obstacles remain challenging and often require external assistance, such as white canes or guide dogs. Although these traditional aids offer some support, they come with limitations in terms of range and functionality. To overcome these barriers, there has been a growing interest in leveraging artificial intelligence (AI) and computer vision to design smarter assistive tools. Object detection, in particular, has emerged as a promising solution, enabling systems to recognize and locate objects in real-time and convert this visual information into audio or tactile feedback that users can easily understand. This technological shift is helping to improve the safety and confidence of visually impaired users as they navigate their surroundings. Over the last ten years, advancements in deep learning have transformed computer vision, especially in the area of object detection. Convolutional neural networks (CNNs) and real-time detection algorithms have made it possible for machines to identify numerous objects with high accuracy. However, for assistive technologies to be practical and usable in daily life, these systems must also be lightweight, fast, and efficient enough to run on mobile or wearable devices with limited processing power. This paper provides an in-depth review of various object detection models designed for assistive use, exploring their effectiveness, performance, and limitations in real-world scenarios. By aligning the technical aspects of AI with the unique needs of visually impaired users,

this study highlights how modern computer vision can contribute to greater independence and improved quality of life for this community.

II. BACKGROUND

Developing AI-based object detection systems for visually impaired users involves more than simple object recognition. These systems must identify and locate multiple objects accurately, while operating efficiently on low-power, portable devices. Unlike standard classification, object detection must also determine the position and boundaries of objects, which is especially difficult in real-world settings with varying light, cluttered backgrounds, and occlusions. Misidentifications or delays can pose serious risks, making both accuracy and speed crucial. Additional challenges include handling low-light conditions, overlapping objects, and diverse object appearances. Since most assistive devices have limited computing power and battery life, models must be carefully optimized to balance accuracy, speed, and efficiency for practical use.

TABLE I: COMPARISON OF VARIOUS OBJECT DETECTION DATASETS

Dataset	Classes	Train			Validation			Test
		Images	Objects	Objects/Images	Images	Objects	Objects/Images	
PASCAL VOC 12	20	5,717	13,609	2.38	5,823	13,841	2	10,991
MS_COCO	80	1,18,287	8,60,001	2.38	5,000	36,781	7	40,670
ILSVRC	200	4,56,567	4,78,807	1.05	20,121	55,501	2.76	40,152
OpenImage	600	17,43,042	1,46,10,229	8.38	41,620	2,04,621	5	1,25,436

III. METHODOLOGY

The field of AI-based object detection for visually impaired assistance has seen a variety of approaches and innovations over the years. Researchers have experimented with deep convolutional neural networks (DCNNs) and one stage detectors like RetinaNet to create systems capable of recognizing objects in indoor environments. These models are trained on datasets containing common household objects, allowing visually impaired users to navigate familiar spaces safely. For instance, systems have been developed that can identify furniture, appliances, and doorways, providing users with real-time audio cues that enhance spatial awareness. The RetinaNet model, for instance, has demonstrated strong performance in object detection tasks by utilizing a unique feature pyramid architecture that balances detection precision with computational efficiency. However, such systems are often limited in outdoor or variable environments due to the training data's specific focus on controlled indoor settings.

$$\begin{aligned}
 Precision &= \frac{True\ Positive}{True\ Positive + False\ Positive} \\
 &= \frac{True\ Positive}{All\ Observations}
 \end{aligned} \quad (1)$$

$$\begin{aligned}
 Recall &= \frac{True\ Positive}{True\ Positive + False\ Negative} \\
 &= \frac{True\ Positive}{All\ Ground\ Truth}
 \end{aligned} \quad (2)$$

In addition to DCNNs, other models such as MobileNet have been employed to support real time object detection on mobile and embedded devices. MobileNet's streamlined architecture makes it ideal for low-power applications, enabling systems that provide continuous assistance to visually impaired users without rapidly depleting battery life. MobileNet-based applications are typically lightweight, ensuring quick response times even on devices with limited processing power. Moreover, wearable devices that utilize models like YOLO and MobileNet can be embedded within everyday accessories, such as smart glasses or belts, to provide auditory cues with haptic feedback, to improve the overall user experience. These advancements highlight the range of approaches explored in recent literature and the ongoing pursuit of more refined and user-friendly assistive technologies for the visually impaired.

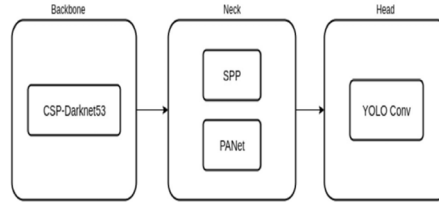


Fig 2. Major components of YOLOv5 architecture

IV. DATASETS AND EVALUATION METRICS

Datasets play an essential role in training and evaluating object detection models for visually impaired assistance systems. Publicly available datasets, such as the Common Objects in Context (COCO) and PASCAL VOC, are frequently used in model training due to their extensive variety of annotated objects and real-world scenarios. COCO, for example, includes images of everyday items that are commonly encountered in diverse settings, providing an ideal foundation for training models designed to operate in both indoor and outdoor environments. However, while these datasets offer a broad range of object classes, they may not be fully representative of the specific needs of visually impaired users. For this reason, custom datasets have been developed to simulate the conditions and obstacles commonly encountered by visually impaired individuals, such as images captured in low-light settings or containing objects from unique angles.

TABLE II: COMPARATIVE STUDY OF EXISTING LITERATURE FOR DETECTION FOR VISUALLY IMPAIRED USING DIFFERENT VERSIONS OF YOLO ALGORITHM

Ref.	Method	Dataset	Merits	Limitations
[2]	YOLO	PASCAL VOC, ImageNet 1000-class	High speed compared to RCNN and can be used for real time object detection.	
[3]	YOLO9000 (YOLOv2)	ImageNet and MS COCO dataset using Word Tree	Compared to YOLO, it can detect a broader range of object classes, and it achieves high accuracy and speed.	Not specific for visually impaired people and potholes.
[4]	YOLOv3	MS COCO dataset	Able to detect objects of different shape and size.	
[5]	YOLOv4	ImageNet and MS COCO dataset	High speed and accuracy compared to previous versions.	
[6]	YOLOv4	MS COCO dataset	It is specifically for object detection.	It is implemented on a laptop or pc which is not portable.
[8]	CNN	KITTI Road and Pothole detection dataset	It is specifically for the visually impaired to detect obstacles.	It uses CNN. It is not preferable for real time object detection compared to YOLO.
[9]	YOLO	Kaggle object detection dataset	It is portable and used in mobile phones	It requires internet. Not suitable for visually impaired

Additionally, these custom datasets often incorporate various augmentations to enhance model robustness, including variations in brightness, noise, and blurring, all of which improve the model's ability to perform reliably under different environmental conditions. The effectiveness of these object detection models is assessed through evaluation metrics like Mean Average Precision (mAP), Intersection over Union (IoU), and Frames Per Second (FPS).

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (3)$$

TABLE III. PARAMETER/METHODS USED IN TRAINING THE YOLOV5 MODEL

Parameter/Method name	Value
Mode	Single GPU
Number of GPU	0
Data set size	425
Batch size per GPU	16
Batch accumulate	1
Total batch size	16
Effective batch size	16
Iterations per epoch	50
Gradient updates per epoch	50
Image size	640 × 640
Learning rate	0.001
Optimizer	Adam optimizer
Loss function	A combination of binary cross-entropy, focal, and IOU loss function

The mAP metric is critical as it provides a comprehensive measure of a model's accuracy across multiple object classes. IoU, on the other hand, assesses the overlap between predicted and actual bounding boxes, indicating the model's localization precision. In practical terms, these metrics are crucial for determining a model's suitability for real-time application in assistive devices, as higher mAP and IoU scores correlate with more reliable object recognition. FPS, a performance metric unique to real-time applications, is used to gauge the model's ability to process images quickly enough to provide seamless feedback. Together, these metrics help researchers balance accuracy with speed and computational efficiency, enabling the development of practical, real-world solutions for visually impaired assistance.

V. ARCHITECTURE OF OBJECT DETECTION MODELS

Several deep learning models have been applied in assistive technology for the visually impaired, with each offering distinct advantages in terms of performance and suitability for various tasks. RFCN, for example, incorporates residual connections within a fully convolutional network structure, making it highly effective in retaining spatial details and delivering accurate object boundaries. By reducing the number of parameters and computational requirements, RFCN achieves a balance between accuracy and efficiency, making it suitable for portable devices with limited resources. Mask RCNN, an extension of Faster RCNN, enhances object detection by introducing a segmentation branch that enables pixel-level object identification, allowing for greater detail and accuracy in complex scenes. Although this model is computationally intensive, it excels in environments where precise localization and segmentation are necessary, such as densely populated areas or indoor environments filled with obstacles.

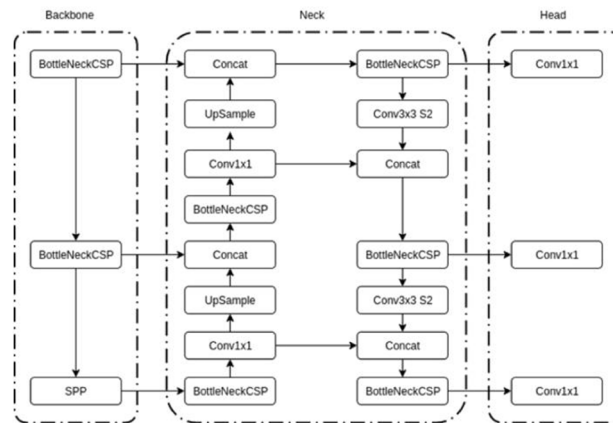


Fig 4. YOLOv5 network architecture

1. YOLO: It offers a more streamlined approach to object detection by performing localization and classification in a single pass, significantly increasing processing speed. YOLO divides an image into a grid and predicts bounding boxes and class probabilities for each cell, which allows it to process images at high speeds without sacrificing too much accuracy. This characteristic makes YOLO highly suitable for real-time applications on edge devices, where speed is crucial for user safety and timely feedback. Moreover, YOLO's flexibility in handling varied object sizes and orientations makes it a versatile solution for visually impaired assistance, particularly in dynamic outdoor settings. Each of these architectures brings unique strengths to the field, and selecting the appropriate model depends on the specific requirements of the assistive application, such as the need for precision, speed, or resource efficiency.

2. R-CNN: The Region-based Convolutional Neural Network (R-CNN) was the first paper in the R-CNN family, and demonstrated how CNNs can be used to immensely improve the detection performance. R-CNN uses a class agnostic region proposal module with CNNs to convert detection into classification and localization problem. A mean-subtracted input image is first passed through the region proposal module, which produces 2000 object candidates. This module finds parts of the image which has a higher probability of finding an object using Selective Search. 3. SPP-Net: He proposed the use of Spatial Pyramid Pooling (SPP) layer to process image of arbitrary size or aspect ratio. They realized that only the fully connected part of the CNN required a fixed input. SPP-net merely shifted the convolution layers of CNN before the region proposal module and added a pooling layer, thereby making the network independent of size/aspect ratio and reducing the computations. The selective search algorithm is used to generate candidate windows. Feature maps are obtained by passing the input

image through the convolution layers of a ZF-5 network. The candidate windows are then mapped on to the feature maps, which are subsequently converted into fixed length representations by spatial bins of a pyramidal pooling layer.

3. Fast R-CNN: One of the major issues with R-CNN/SPP-Net was the need to train multiple systems separately. Fast R-CNN solved this by creating a single end-to-end trainable system. The network takes as input an image and its object proposals. The image is passed through a set of convolution layers and the object proposals are mapped to the obtained feature maps. This layer is connected to 2 fully connected layer and then branches out into a N+1-class SoftMax layer and a bounding box regressor layer, which has a fully connected layer as well. The model also changed the loss function of bounding box regressor from L2 to smooth L1 to better performance, while introducing a multi-task loss to train the network.

4. Faster R-CNN: Even though Fast R-CNN inched closer to real time object detection, its region proposal generation was still an order of magnitude slower (2 sec per image compared to 0.2 sec per image) suggested a fully convoluted network as a region proposal network (RPN) in that takes an arbitrary input image and outputs a set of candidate windows. Each such window has an associated objectless score which determines likelihood of an object. Fast R-CNN is trained from the region proposals of RPN from first step. Till this point, the networks do not have shared convolution layer. Now, we fix the convolution layers of the detector and fine-tune the unique layers in RPN.

5. FPN: Use of image pyramid to obtain feature pyramid (or featurized image pyramids) at multiple levels is a common method to increase detection of small objects. Even though it increases Average Precision of the detector, the increase in the inference time is substantial. Lin et al. proposed the Feature Pyramid Network (FPN), which has a top-down architecture with lateral connections to build high-level semantic features at different scales. The FPN has two pathways, a bottom-up pathway which is a Convnet computing feature hierarchy at several scales and a top-down pathway which up samples coarse feature maps from higher level into high-resolution features.

6. R-FCN: Dai et al. proposed Region-based Fully Convolutional Network (R-FCN) that shared almost all computations within the network, unlike previous two stage detectors which applied resource intensive techniques on each proposal. They argued against the use of layer, instead of RoIPool layer, to avoid pixel level misalignment due to spatial quantization.

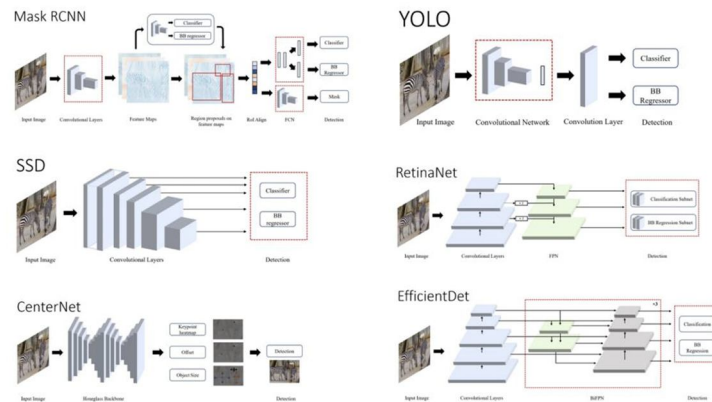


Fig 5. Illustration of the internal architecture of different object detectors

VI. IMPLEMENTATION IN ASSISTIVE SYSTEMS

The implementation of object detection models within assistive systems designed for visually impaired individuals involves a complex interplay of software and hardware integration to fully connected layers and instead used convolutional layers. However, deeper layers in the convolutional network are translation invariant, making localization tasks, them ineffective for 7. Mask R-CNN: Mask R-CNN extends on the Faster R-CNN by adding another branch in parallel for pixel-level object instance segmentation. The branch is a fully connected network applied on RoIs to classify each pixel into segments with little overall computation cost. It uses similar basic Faster R-CNN architecture for object proposal, but adds a mask head parallel to classification and bounding box regressor head. One major difference was the use of RoIAlign ensure that these systems are both functional and user-friendly. A prominent hardware choice for these applications is the Raspberry Pi, a compact and affordable microcontroller that provides sufficient processing power to host AI models in a portable form.

Equipped with a Pi camera and connected to various sensors, the Raspberry Pi serves as the central processing unit in many assistive systems, handling real-time image capture, object detection, and feedback delivery. In a typical setup, the object detection model, such as RFCN or YOLO, is loaded onto the Raspberry Pi, which then processes images captured by the Pi camera. The Raspberry Pi's versatility allows developers to integrate a range of sensory components, such as ultrasonic sensors, which are used to measure the distance of objects from the user. This spatial data, combined with visual information, is translated into auditory feedback delivered through headphones, giving the user real-time updates on their immediate surroundings. This auditory feedback is crucial for providing the visually impaired with spatial awareness, allowing them to respond effectively to detected obstacles or objects. In addition to standalone hardware like the Raspberry Pi, wearable devices using the YOLO model have become increasingly popular due to YOLO's high-speed processing capability and efficiency. The YOLO model's architecture allows it to perform object detection in a single pass, which significantly reduces latency and makes it ideal for real-time applications on mobile or embedded devices with limited processing power. This setup is particularly beneficial in urban settings where users need immediate feedback to navigate through crowded and dynamic environments. By integrating the YOLO model, these devices can operate continuously, providing real-time updates without significant lag. This streamlined approach enables visually impaired users to move independently, receiving timely notifications about obstacles and path clearances as they walk. In each of these implementations, the primary objective is to ensure the system is not only accurate and responsive but also comfortable and accessible, empowering users to confidently interact with their surroundings without the need for a guide.

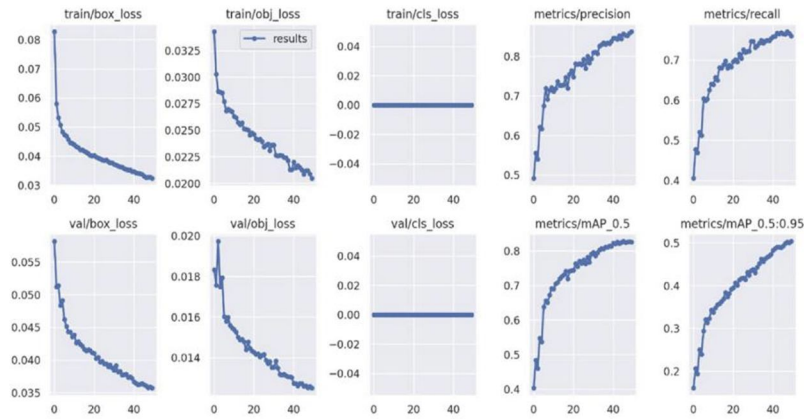


Fig. 6. Plotting the results of three losses and precision/recall metrics

VII. COMPARATIVE ANALYSIS OF MODELS

Evaluating object detection models for visually impaired assistance highlights key trade-offs in speed, accuracy, and resource efficiency. Mask R-CNN offers detailed, pixel-level segmentation and performs well in complex or cluttered environments, making it suitable for scenarios where high accuracy is critical. However, its heavy computational demands limit its use in mobile or wearable devices. In contrast, YOLO (You Only Look Once) prioritizes real-time performance through a single-pass detection process. Its ability to quickly detect and classify objects makes it ideal for portable assistive tools, though it may underperform in highly detailed or crowded scenes. For practical applications, especially in mobile settings, YOLO or lightweight alternatives like MobileNet are preferred due to their efficiency. In some cases, hybrid models combining features from both one-stage and two-stage detectors offer a balanced solution. Ultimately, the choice of model depends on the specific needs of the assistive system—whether it demands precision, speed, or a compromise between the two—to ensure safe and effective navigation support for visually impaired users.

The practical deployment of object detection models in assistive devices also depends on hardware constraints and user experience. Visually impaired users often rely on compact, battery-powered devices, which limits the feasibility of using high-complexity models. Therefore, the model must not only deliver accurate results but also operate efficiently on devices with limited processing power. Models like MobileNet-SSD are gaining attention for their lightweight architecture, enabling acceptable accuracy with minimal resource consumption. Furthermore, real-time feedback is essential, as even slight delays in object recognition can compromise user safety during navigation. Developers must also consider environmental variability, such as changes in lighting,

weather, or object orientation, which can affect model performance. As a result, the selection and fine-tuning of object detection models for assistive technology require a careful balance of technical capability, responsiveness, and adaptability to real-world conditions.

TABLE V: COMPARISON OF LIGHTWEIGHT MODELS

Model	Year	Top-1 Acc%	Latency (ms)	Parameters (Million)	FLOPs (Million)
SqueezeNet	2016	60.5	-	3.2	833
MobileNet	2017	70.6	113	4.2	569
ShuffleNet	2017	73.3	108	5.4	524
MobileNetv2	2018	74.7	143	6.9	300
PeeleNet	2018	72.6	-	2.8	508
ShuffleNetv2	2018	75.4	178	7.4	597
MnasNet	2018	76.7	103	5.2	403
MobileNetv3	2019	75.2	58	5.4	219
OFA	2020	80.0	58	7.7	595

VIII. FUTURE TRENDS

Future advancements in AI-based assistive technology for the visually impaired are expected to focus on increased efficiency, portability, and multisensory feedback systems. Lightweight models designed specifically for low-power devices, such as MobileNet's newer iterations, hold promise for developing more power efficient and faster systems that can run continuously. Researchers are also exploring multisensory feedback that combines audio, haptic, and even olfactory cues to create a more immersive and informative experience for users. Such multimodal approaches would allow visually impaired individuals to interpret their surroundings in a richer, more intuitive manner, improving their situational awareness and autonomy. The integration of Augmented Reality (AR) with object detection is another promising trend. AR enhanced assistive devices could provide visual or tactile orientation, feedback on object proximity, or movement, potentially transforming how users interact with their environment. Additionally, edge AI and federated learning could allow these systems to adapt and improve over time without requiring constant cloud connectivity. Such advancements signify an exciting frontier in assistive technology, one where AI-driven systems for the visually impaired become more powerful, accessible, and integrated into everyday life, helping users navigate their surroundings with a newfound level of independence and confidence.

IX. CONCLUSION

This survey has examined various AI-based object detection models and their applications in assistive technology for the visually impaired, highlighting their capabilities, limitations, and practical implementations. While models such as Mask RCNN provide precise object detection, their high computational demands make them less suitable for mobile applications. YOLO, on the other hand, balances speed and accuracy, making it ideal for real-time assistive devices on portable hardware. By assessing the unique requirements of visually impaired users and the capabilities of current AI models, this paper sheds light on the potential of these technologies to transform accessibility and mobility. As research continues, the development of more refined, lightweight, and user-centered solutions will likely bring us closer to a future where visually impaired individuals can navigate independently and with greater confidence in an increasingly accessible world. The potential of AI-driven object detection in transforming accessibility for the visually impaired is immense. As research in this area advances, we can expect assistive devices to become increasingly compact, efficient, and user friendly, integrating seamlessly into everyday life. The future of assistive technology lies not only in improving the technical performance of these models but also in refining the user experience through multisensory feedback, augmented reality integration, and personalized learning.

These advancements will empower visually impaired individuals to navigate their environments with newfound independence, safety, and confidence. With continuous improvements in AI, hardware miniaturization, and multimodal interaction, we are moving toward a world where assistive technology is both an indispensable tool and an accessible means of enhancing quality of life for millions of visually impaired individuals around the globe.

REFERENCES

- [1] Kumar, D. & Thakkar, Hiren Kumar & Merugu, Suresh & Gunjan, Vinit & Gupta, Suneet. (2022). Object Detection System for Visually Impaired Persons Using Smartphone. 10.1007/978-981-16-3690 5_154.
- [2] K. Jivrajani et al., "AIoT-Based Smart Stick for Visually Impaired Person," in *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-11, 2023, Art no. 2501311, doi: 10.1109/TIM.2022.3227988.
- [3] K. Naqvi, B. Hazela, S. Mishra, and P. Asthana, "Employing real-time object detection for visually impaired people," in *Data Analytics and Management (Lecture Notes on Data Engineering and Communications Technologies)*, vol. 54, A. Khanna, D. Gupta, Z. Pólkowski, S. Bhattacharyya, and O. Castillo, Eds. Singapore: Springer, 2021, pp. 285–299, doi: 10.1007/978-981-15-8335-3_23.
- [4] N. E. Albayrak, "Object recognition using TensorFlow," in *Proc. IEEE Integr. STEM Educ. Conf. (ISEC)*, Aug. 2020, p. 1–5, doi: 10.1109/ISEC49744.2020.9397835.
- [5] Alagarsamy, Saravanan & Rajkumar, Dhiliphan & Syamala, Ketepalli & Niharika, Ch & Rani, D. & Balaji, K.. (2023). An Real Time Object Detection Method for Visually Impaired Using Machine Learning. 1-6. 10.1109/ICCCI56745.2023.10128388.
- [6] S. Sameer, P. Madan, S. Kannan, V. J. Upadhye, H. Patil and S. Rajkumar, "AI based Object Detection for Assisting the Visually Impaired People," 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Lalitpur, Nepal, 2024, pp. 512 518, doi: 10.1109/ICMCSI61536.2024.00080.
- [7] Aung, Myo Min & Maneetham, Dechrit & Crisnapati, Padma & Thwe, Yamin. (2024). Enhancing Object Recognition for Visually Impaired Individuals using Computer Vision. *International Journal of Engineering Trends and Technology*. 72. 297-305. 10.14445/22315381/IJETT V72I4P130. on *Computer Vision and Pattern Recognition*, pp. 248–255, ISSN: 1063-6919.
- [8] C. Ananth, P. Thenmozhi, D. J. Vadakkan, S. Kavitha and T. A. Kumar, "Artificial Intelligence based Visual Aid with Live Tracking of Visually Impaired People," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 573-577, doi:10.1109/ICTACS56270.2022.9988614.
- [9] M. H. Zaidi, A. Yasir, A. S. Singh and A. K, "Implementation Of Object Detection and Identification For Visually Impaired," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 2533-2538, doi: 10.1109/ICACITE53722.2022.9823670.
- [10] A. Widayani, H. Kusuma and D. Purwanto, "Visually Impaired Person Detection Using Deep Learning for Dangerous Area Warning System," 2022 International Seminar on Intelligent Technology and Its Applications (ISITIA), Indonesia, 2022, pp. Surabaya, 204-208, doi: 10.1109/ISITIA56226.2022.9855268.
- [11] H. Shah, R. Shah, S. Shah and P. Sharma, "Dangerous Object Detection for Visually Impaired People using Computer Vision," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670992.
- [12] J. doi: Hemavathy, A. Sabarika Shree, S. Priyanka and K. Subhashree, "AI Based Voice Assisted Object Recognition for Visually Impaired Society," 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 2023, pp. 1-7, doi:10.1109/ICDSAAI59313.2023.104524 56.