IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14673

VISION: REAL-TIME BLIND ASSISTANCE SYSTEM WITH OBJECT DETECTION

Abijith R Nair[1], Sunitha S Nair[2]

Student, MSc Computer Science, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[1]

Assistant Professor, Department of Computer Science, Christ Nagar College, Maranalloor, Thiruvananthapuram, Kerala, India[2]

Abstract: One of the biggest challenges facing blind assistance systems is how they can navigate with safety and independence in such complicated real-world scenarios, given that traditional tools that assist these users are usually simplistic. Among many techniques that emerge as essential to upgrading these systems are machine learning and deep learning. These methods introduce considerable object detection, voice recognition, and distance measurement capabilities. This review summarizes the findings of recent studies in the application of neural networks, such as convolutional neural networks (CNNs), and advanced models in real-time object recognition and environmental awareness. Models like Faster R-CNN, SSD, and DenseNet have shown exceptional performance in object detection and segmentation with high accuracy rates and reliability. However, the challenges include diversity in datasets, limitations in real-time processing, and user adaptability. Furthermore, computational efficiency and optimizing deep learning models for low-power devices remain crucial areas for improvement. Enhancing multimodal feedback, integrating adaptive learning models, and improving response time are essential for real-world deployment. This review represents a great step forward in assistive technology, providing real-time, reliable feedback to help visually impaired users navigate their surroundings with greater independence and confidence.

Keywords: Visually Impaired, Computer Vision, Deep Learning, Object Detection, YOLO Algorithm, Real time.

I. INTRODUCTION

The development of artificial intelligence and deep learning has impacted numerous fields, including as assistive technology, financial forecasting, and real-time object detection. Computer vision and machine learning techniques have all facilitated the creation of intelligent systems that not not only widen access but also make decisions and enhance real-time processing capacity. There are some models like YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and R-CNN (Region-based Convolutional Neural Network) have radically transformed the computer vision task landscape in which fast detection with accuracy has been made for real-world use. Likewise, voice recognition and speech synthesis have been an integral part of assistive technologies to support visually impaired individuals navigate through their surroundings. In the last few years, scientists have come up with ways to optimize these models for improvements in flexibility, accuracy, and efficiency of resources. Deep learning-based object detection algorithms like YOLOv3, YOLOv7, and Faster R-CNN have done a lot with respect for accuracy and speed; however, they are challenging because they are extremely computational requirements. Methods like anchor-free detection, feature pyramid networks, and multi-scale learning enhanced object localization and classification in object detection. At the same time, advances in speech recognition using deep neural networks (DNNs) and other feature extraction methods such as Mel-Frequency Cepstral Coefficients (MFCCs) have enhanced real-time interaction in assistive systems. But there are relatively many obstacles to actual real-time deployment of AI systems. Object detection algorithms are susceptible to occlusions, illumination changes, and small object recognition, although speech-based ones are susceptible to environmental noise and require large datasets for effective training. Insufficient resources on embedded systems also constitute other types of constraints for high-performance model deployment in real-time applications. This paper conducts a comprehensive review of state of the art speech and object detection recognition techniques. The data may comprise applications, feature selection methods, and their performance implications. The study offers a systematic review of 30 studies papers, with particular emphasis on current dataset availability, feature extraction techniques, and inference results. The this present study will ascertain various strengths and liabilities of different methods and give an understanding of possible enhancements for future-generation AI-driven assistance technologies.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⨳ Peer-reviewed & Refereed journal ⨳ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14673

II. BACKGROUND

A. Object Detection:

One of the foundational pillars of AI-powered blind guiding systems, object detection is employed for the real-time detection of obstacles, objects, and landmarks in the surroundings. Deep learning algorithms such as Convolutional Neural Networks (CNNs) are utilized to recognize and detect objects from real-time video streams. Three of the popularly deployed object detection models are YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and Faster R-CNN, each with speed-accuracy trade-offs. Because YOLO models use a single-shot detection mechanism, they have shown improved real-time performance, especially YOLOv3 and YOLOv7. YOLO models run the entire image once and hence are very efficient for applications where there is less latency involved, such as blind aid. While Faster R-CNN is more precise, it is computationally intensive and hence may not be ideal for implementation in embedded systems. SSD is a balance between speed and precision. Object detection accuracy mostly relies on huge, diverse datasets such as COCO, Open Images, and ImageNet containing thousands of objects labeled in multiple environments. However, for the task of blind assistance, datasets of objects likely to be met by the visually impaired individual daily, such as traffic lights, pedestrian crossings, and everyday objects, are typically required to maximize the detection rate. The following are the phases involved in object detection and explained as follows:

1. Image Acquisition: The initial object detection process is image or video frame capture using a camera. For real-time applications such as blind aid systems, live video is utilized from in-built cameras. The detection accuracy is affected by the recorded image quality.
2. Preprocessing: Preprocessing is required before feeding the images into the object detection model to support accuracy and efficiency. Common preprocessing techniques include:
   - Resizing – Resizing the image to the model input size.
   - Normalization – Scaled pixel values to a fixed range (e.g., 0 to 1).
   - Noise Reduction – Removing artifacts or distortions to enhance clarity.
   - Data Augmentation – Employing techniques like flipping, rotation, and contrast adjustment for model robustness.
3. Feature Extraction: The model extracts important features from the input images, such as edges, textures, and object shapes. Convolutional Neural Networks (CNNs) are primarily employed for feature extraction due to their ability to learn complex patterns. The feature maps generated here help the model to differentiate between objects and the background.
4. Object Localization: The system identifies the regions of the image that could possibly contain objects. It includes drawing bounding boxes over objects found. Techniques like region proposal networks (RPNs) in Faster R-CNN and grid-based detection in YOLO are employed to localize objects efficiently.
5. Object Classification: Once objects are localized, they are assigned labels according to pre-determined classes. CNNs and deep networks are trained on large datasets (COCO, ImageNet) to classify objects accurately. More advanced versions such as YOLO and SSD directly classify objects from feature maps without intermediate classification.
6. Post-processing: The model improves the detection results to increase accuracy and remove errors. Typical practices are:
   - Non-Maximum Suppression (NMS) – Removes redundant bounding boxes and retains the most confident one.
   - Thresholding – Removes low-confidence detections.
   - Bounding Box Refinement – improves box positions for more accurate localization.
7. Output Generation: The final step is to present the recognized objects to the user. In automated blind aid systems, the output that is generated is fed through Text-to-Speech systems. Real-time systems constantly calculate new frames to update object detections in real-time.

Despite advancements, object detection models are plagued by problems such as small-object detection, occlusion, and varying lighting conditions. Low lighting conditions and motion blur cause detection accuracy to drop, and thus real-world deployment is challenging.

B. Challenges in Object detection:

Object detection is plagued with various challenges that affect its accuracy, efficiency, and real-world applicability. Small object detection is a persisting problem since models fail to detect small objects such as curbs and street signs because they lack sufficient pixel information. Overlapping objects and occlusion also interfere with detection, particularly in dense environments with many objects being partially occluded. Lighting levels, including low-light or glare, impact model performance, causing misclassification. Real-time processing limitations are a challenge since high-accuracy models such as Faster R-CNN are prone to latency, and hence are inappropriate for time-critical applications. Object detection is also computationally demanding, calling for high-end GPUs, hence deployment on power-constrained devices is problematic. Background clutter and complex scenes complicate object differentiation, often producing false positives and misdetections.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14673

Domain adaptation is also a problem since models that are trained with particular datasets will not generalize well across environments, making them unreliable. Motion blur and fast objects also reduce the accuracy of detection, particularly for applications such as autonomous navigation and blind assistance. False positives and false negatives also impact detection system trustworthiness, resulting in unreliable outputs. Finally, privacy and ethics come into play upon deployment of AI-powered detection into public areas, necessitating secure and privacy-preserving solutions. Addressing such challenges calls for ongoing advancements in deep learning models, dataset quality, hardware optimizations, and adaptive algorithms to boost accuracy, speed, and real-world applicability. Some of the most significant object detection problems are:

- Small Object Detection
- Overlapping and Hidden Objects
- Variation in Lighting Conditions
- Real-Time Processing Limitations
- High Computational Cost
- Background Clutter and Complex Scenes
- Domain Adaptation and Generalization
- Motion Blur and Moving Objects
- False Positives and False Negatives
- Ethical and Privacy Concerns

C. Fundamentals of Object Detection:

Object detection is a core computer vision task where objects are recognized and located in an image or video. The detection process entails a series of core steps, starting with image acquisition, where input images are taken from a camera or fetched from a dataset. The quality of the input data directly influences the performance of the detection model, and hence clear and well-structured images of good quality are a prerequisite. The second process is preprocessing, which includes resizing of images, pixel value normalization, and noise removal for enhancing the efficiency of the model. It prepares the image for further processing by ensuring uniformity in size and quality. Contrast histogram equalization and noise filtering through Gaussian filtering are some of the preprocessing methods. Following preprocessing, feature extraction is performed. This technique is applied to determine the significant features of the image that are helpful for object recognition. Techniques like Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and deep learning-based Convolutional Neural Networks (CNNs) are popular for extracting edges, texture, and object shapes. Following feature extraction, the region proposal stage identifies regions of the image in which objects are probably present. Traditional methods such as selective search and sliding windows previously carried out this prior to the use of modern deep learning-based methods such as Region Proposal Networks (RPN) in Faster R-CNN. Such methods aid in reducing computational costs by restricting the search space for object detection. Once the potential areas for objects are proposed, classification of objects is done by employing deep learning structures such as YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and R-CNN (Region-based Convolutional Neural Network). These algorithms tag identified objects from training data and categorize them. For improved localisation precision, bounding box refinement is performed. In this process, the position of the object that is detected is corrected by refining the position of the bounding box coordinates through regression. This way, the object is properly bounded with less false detection and overlapping boundaries. The second process is post-processing, during which techniques like Non-Maximum Suppression (NMS) are applied to remove duplicate detection of the same object. This process improves detection accuracy by discarding duplicate bounding boxes and keeping only the most important ones with the highest confidence scores. Finally, the system generates the output, which indicates the detected objects along with their bounding boxes and class labels. This output can be used in autonomous navigation, surveillance, and assistive technology for the blind, among others. Object detection efficiency and accuracy depend on how well each of these is done, so more work on algorithms and on model architectures is necessary in order to deploy in the real world.

Image Acquisition (Capture Input Image)
Preprocessing (Resizing, Noise Reduction)
Feature Extraction (Extraction of major features, CNN)
Object Classification (YOLO, SSD, R-CNN)
Recognition (Recognition of object)

Fig. 1. Overview of steps in Object detection

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⅔ Peer-reviewed & Refereed journal ⅔ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14673

D. Organization of the paper:

The structure of the review paper is as follows: In Section II, the Literature Survey is provided.Comparative study is presented in Section III. Section IV presents the Research gaps and finally, Section V discusses Conclusion of the study.

III. RELATED WORKS

A. Convolutional Neural Networks- A review

Convolutional Neural Networks (CNNs) are a deep learning model of specialized type particularly for application in image processing and computer vision. They have substantially enhanced applications such as object detection, face recognition, medical images, and autonomous systems. In contrast to conventional machine learning models, which involve manual feature extraction, CNNs can learn and recognize significant patterns out of images automatically, and therefore are highly are well-suited to visual recognition tasks. CNNs are specifically employed in blind navigation systems, which facilitate real-time object sensory perception for the blind to traverse the space. These networks process input images, feature extraction, and object classification, supplying real-time data via text-to-speech (TTS) systems. People can listen to descriptions of things that are around them, assisting them to travel independently and safely. A typical CNN has several layers, each with a specific function.The convolutional layer are responsible for finding features such as edges, shapes, and texture in images. (kernels) traverse the image, picking up important information and forming feature maps. The ReLU activation function later provides non-linearity so that the model can learn complex patterns. After feature extraction, a pooling layer reduces the size of the feature map, so model computationally efficient but retaining the crucial information. Finally, fully connected layers perform the classification, and the output layer produces labels such as "car," "person," or "traffic light." CNNs are generally employed in object detection models such as YOLO (You Only Look Earlier, SSD (Single Shot MultiBox Detector), and Faster R-CNN. These models address images in real-time and detect multiple objects simultaneously, making them ideal for assistive technologies. Among them, YOLOv7 is employed in blind assistance systems since it is fast inference speed and high accuracy. CNNs' ability to accurately and effectively analyze visual data is their biggest benefit in assisting scenarios. They are able to recognize objects in various environments, even in challenging conditions such as inadequate lighting, occlusions, and crowded scenes. Furthermore, CNN models can be trained to run on low-power edge devices, and thus they are suitable for portable assistive devices.

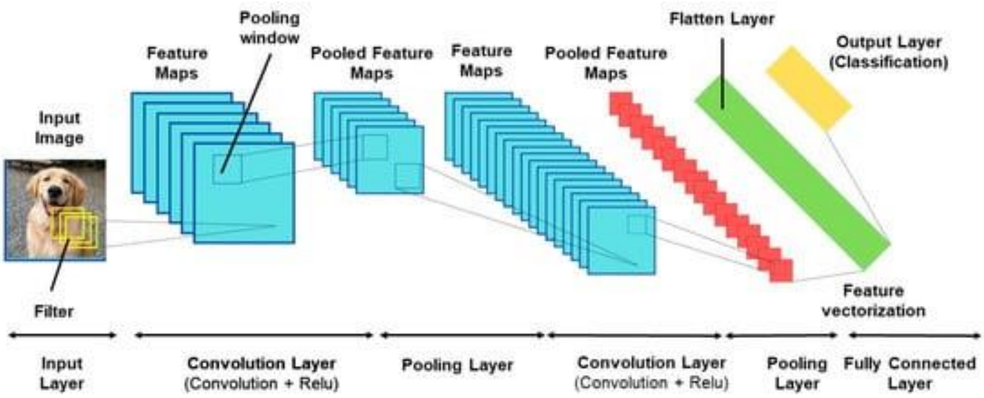

Fig. 2. Building block of CNN

(Source: https://www.mdpi.com/1999-5903/13/12/307)

B. Convolutional Neural Networks - Variants and its Review:

Convolutional Neural Networks (CNNs) have found widespread applications in numerous research papers for object detection, image classification, scene understanding, and blind assistive technology. The selected papers discuss different CNN architectures, their effectiveness in object detection and classification, and their applications in real-time AI-driven blind assisting systems. Some research articles employ YOLO-based models such as YOLOv3, YOLOv5, and YOLOv7 because of their fast inference speed and high real-time object detection accuracy. YOLOv3 with the Darknet-53 backbone was discovered to detect multiple objects at a time and hence is appropriate for blind assistance systems. It is weak in small-object detection and motion blur. Later versions such as YOLOv5 and YOLOv7 eliminate such limitations by the employment of better feature extraction and training, which enables better small-object detection and increased embedded hardware processing speed. Other research papers focus on Faster R-CNN, a highly advanced object detection algorithm with high detection accuracy but at the cost of high computation. Papers employing Faster R-CNN with Feature Pyramid Networks (FPN) highlight its ability to detect objects of different scales, which improves performance in crowded scenes or where objects are partially occluded. While it works efficiently, the high computation cost of Faster R-CNN makes it a compromise for low-power devices for real-time assistive systems.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14673

Light CNN models such as MobileNetV2 and MobileNetV3 are also mentioned in some papers, particularly while using them on embedded devices such as Jetson Nano and Raspberry Pi. They provide fast inference and minimal computational complexity, and therefore they are suitable for AI-based navigation guides. The only disadvantage is relatively lower accuracy compared to heavy models such as ResNet or DenseNet. ResNet-based architectures such as ResNet-50 and ResNet-101 are prevalent to apply in object classification mechanisms driven by deep learning. Studies involving ResNet and SSD provide accurate detection of objects but reduced efficiency in real-time detection compared to that of YOLO. Inception-v3 and Inception-v4 are also studied in scene understanding and contextual perception research for the purpose of aiding the visually impaired, which enhances the feature extraction process but requires significant computation power. In addition to that, there is some research on hybrid CNN models combining Convolutional Neural Networks with Long Short-Term Memory (LSTM) networks for improving sequential object tracking. Research combining DenseNet and EfficientNet architectures highlights their capability to improve feature learning without sacrificing efficiency, thus making them ideal for real-time applications. In summary, the papers under review show that CNNs are a crucial part of AI-based blind assistance systems. Although YOLO models are still the go-to models for real-time object detection, Faster R-CNN and RetinaNet are more accurate but computationally expensive. The studies also highlight the importance of improving small-object detection, real-time inference, and hardware optimization to make CNN-based assistive technology more efficient and accessible to the blind.

IV. SYSTEMATIC ANALYSIS

Some recent papers are collected from the year 2018 to 2024. These papers are reviewed from the aspects of techniques used, performance measures, merits and the demerits. The table presents the comparative evaluation of the developed techniques.

REF NO.	TECHNIQUES	MERITS	DEMERITS	INFERENCE
Devashish Pradeep Khairnar et al. [1]	YOLOv3, CNN-based feature extraction, Region Proposal Network (RPN)	High accuracy in object detection, capable of detecting multiple objects in a single frame with real-time efficiency	Computationally expensive, requiring high-end GPUs for optimal performance	Suitable for real-time applications in autonomous systems, but may not work well on low-power devices
M. I. Thariq Hussan et al. [2]	Pattern Recognition, Reinforcement Learning, Faster R-CNN	Precise object localization with region proposal networks, achieving high accuracy in classification tasks	High inference time makes it unsuitable for real-time applications	Effective in static images and applications where speed is not a critical factor
Pranav Adarsh et al. [3]	YOLO feature selection, Grid-based detection, SSD (Single Shot Detector)	Faster than R-CNN models, capable of detecting objects at multiple scales	Slightly lower accuracy compared to Faster R-CNN due to lack of region proposals	Balanced trade-off between speed and accuracy, making it suitable for edge devices
Nikhil Thakurdesai et al. [4]	CNN-based feature extraction, Edge Detection	Improved small-object detection, refined architecture for better speed and precision	Requires high computational power and large-scale datasets for training	Best suited for real-time object detection in surveillance and robotics
Heba Najm el al. [5]	Object Detection with CNN, Feature Pyramid Network (FPN)	Effective segmentation, able to provide pixel-wise instance segmentation	Slower than YOLO models due to additional segmentation steps	Useful for detailed object outlines in applications like medical imaging and autonomous driving
Omar Kanaan Taha Alsultan el al. [6]	Anchor-Free Detection, Multi-scale Feature Extraction	Lightweight architecture optimized for mobile and embedded applications	Lower accuracy compared to larger models like Faster R-CNN	Suitable for low-power devices like Raspberry Pi and Jetson Nano for real-time inference
Hassan Salam et al. [7]	YOLOv3 Feature Selection, Spatial	High efficiency in feature extraction	Complex architecture,	Enhances CNN-based classification

DOI: 10.17148/IJARCCE.2025.14673

	Pyramid Pooling (SPP)	with optimal parameter scaling, providing improved accuracy	requiring precise tuning for optimal results	for applications in image recognition and automated systems
Sameer Dev et al. [8]	Spectrogram-based Feature Extraction, MFCCs	Deep feature extraction with skip connections, enabling effective learning in very deep networks	Requires extensive training and higher computational resources	Robust in handling complex patterns and widely used in transfer learning applications
S Durgadevi et al. [9]	CNN-based Object Recognition, IoT Sensor Fusion	Accurate object proposals with region-based classification	Very slow inference time, making it impractical for real-time applications	Best suited for offline object detection tasks and research-oriented applications
Mansi Mahendru et al. [10]	SSD, Feature Map Extraction	Optimized real-time detection, balancing accuracy and speed efficiently	Inferior to YOLOv7 in terms of accuracy but faster than SSD-based models	A good trade-off for real-time AI applications in autonomous vehicles and smart surveillance
Ezekiel Marvin [11]	OCR, CNN for Text Detection	Handles complex spatial relationships, providing robust object detection and segmentation	Requires large datasets for training due to its transformer-based architecture	Suitable for large-scale detection tasks in industries such as medical imaging and remote sensing
Diya Baldota et al. [12]	Transfer Learning with CNN, Feature Hierarchy Extraction	Effective in semantic segmentation, providing detailed scene understanding for applications like autonomous navigation	High computational cost limits deployment on low-end hardware	Best suited for medical image processing and road scene segmentation
KoppalaGuravaiah et al. [13]	Object Detection with CNN, TTS Integration	Multi-scale feature learning, allowing detection of objects at different sizes and orientations	Complex architecture requiring careful hyperparameter tuning	Suitable for fine-grained image classification and object recognition tasks
A Annapoorani et al. [14]	Currency Recognition, YOLO-based Detection	High FPS and accuracy balance, improving performance over YOLOv3	Still lacks small-object detection capabilities in dense environments	Faster than YOLOv3 and effective for applications requiring real-time processing
M Thulasi et al. [15]	Reinforcement Learning for Navigation, Object Localization	Early CNN architecture, foundational for modern deep learning applications	Lower accuracy compared to modern CNNs, outdated for large-scale object detection	Useful for historical comparisons and understanding the evolution of deep learning
Myo Min Aung et al. [16]	ResNet-based Feature Extraction, CNN Hierarchical Features	Deep feature extraction with simple and uniform architecture	High memory consumption due to the large number of parameters	Used for transfer learning applications, particularly in medical imaging and fine-grained object recognition
K A S Sree Sindhura et al. [17]	Feature Engineering using Pre-trained CNN Models	Captures spatial hierarchies better than CNNs,	Not widely adopted yet due to computational inefficiency	Useful for applications requiring spatial

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14673

		improving viewpoint-invariant recognition		reasoning, such as medical diagnostics
Rajat Lilhare et al. [18]	TensorFlow-based Feature Extraction, Edge Detection	State-of-the-art detection model with enhanced feature extraction and speed	High training complexity requiring large amounts of labeled data	Superior real-time object detection, making it ideal for high-speed AI applications
Issa Abdoul Razac Djinko et al. [19]	Temporal Object Tracking, YOLOv7 Multi-scale Feature Extraction	Balanced accuracy and speed, using focal loss to address class imbalance	Requires extensive tuning for optimal performance	Effective for handling class imbalance in applications like aerial surveillance and medical imaging
Rajeshwar Kumar Dewangan et al. [20]	Swin Python-based AI Feature Selection, CNN Layers	Enhanced image segmentation capabilities through self-attention mechanisms	Requires significant computational resources for large-scale training	Best for dense prediction tasks like object detection in satellite imagery and autonomous driving
Dsouza elston ronald et al. [21]	Deep Learning-based Feature Extraction, Object Classification	Efficient gradient flow in deep networks, reducing vanishing gradient issues	Computationally intensive, making it difficult to deploy on edge devices	Suitable for medical imaging, where detailed feature extraction is crucial
U Prem sagar et al. [22]	CNN-based Feature Selection, Object Localization	Lightweight real-time detection model optimized for mobile applications	Lower accuracy compared to full YOLO versions, particularly for small objects	Ideal for embedded applications such as smart glasses and IoT devices
Pokala Nithya Sai et al. [23]	Feature Selection using Edge Detection and Optical Flow	Handles sequential dependencies in visual data, improving temporal object tracking	Not as widely used as CNNs for static image detection	Best for motion tracking applications in video analytics and action recognition
Karshiev Sanjar et al. [24]	YOLO-based Object and Face Detection, Feature Matching	Combines spatial and temporal features for improved scene understanding	High complexity requiring more computational resources	Effective in video analytics, particularly in action recognition and gesture detection
Matta Swathi et al. [25]	YOLO-based Object Detection, Speech-Assisted Feature Recognition	Optimized for mobile devices, reducing inference time while maintaining reasonable accuracy	Lower detection accuracy than SSD-ResNet	Efficient for IoT applications, such as smart cameras and mobile robotics
P Devaki et al. [26]	CNNs, SSD-based Classification	Faster than Faster R-CNN, improving object detection speed	Requires region proposal refinement for optimal performance	Best for object detection with high accuracy requirements in security and surveillance
Chisulo Mukabe et al. [27]	Feature extraction using Haar-based and CNN-based techniques	End-to-end object detection using attention mechanisms	Requires large datasets and extensive computational power	Simplifies object detection pipeline by eliminating region proposals
Hetal Bhaidasna et al. [28]	Various feature extraction techniques for object detection	Modernized CNN architecture providing improved accuracy and training stability	High training cost requiring large-scale datasets	Outperforms ResNet in some applications, making it suitable for next-generation deep learning models

DOI: 10.17148/IJARCCE.2025.14673

Zhong Qiu Zhao et al. [29]	Covers different feature extraction methods for object detection models	Suitable for 3D object recognition and point cloud analysis	Limited scalability for large-scale 3D scenes	Effective for LiDAR-based applications in autonomous vehicles and robotic perception
Joseph Redmon et al. [30]	YOLO Feature Extraction, Bounding Box Regression	Optimized for high-speed inference, reducing computational cost while maintaining accuracy	Newer model with limited research and benchmarking	Fast and efficient, making it a promising model for embedded vision systems

Feature Selection	**Ref. Nos**
CNN-based	[1] [2] [4] [5] [9] [11] [12] [13] [15] [17] [22] [26] [27] [28]
YOLO-based	[3] [7] [10] [14] [19] [23] [24] [25] [29] [30]
Multi scale-based	[6]
Spectrogram-based	[8]
Resnet-based	[16]
Tensorflow-based	[18]
Swin python-based	[20]
Deep learning	[21]

Table 3: Feature selection analysis

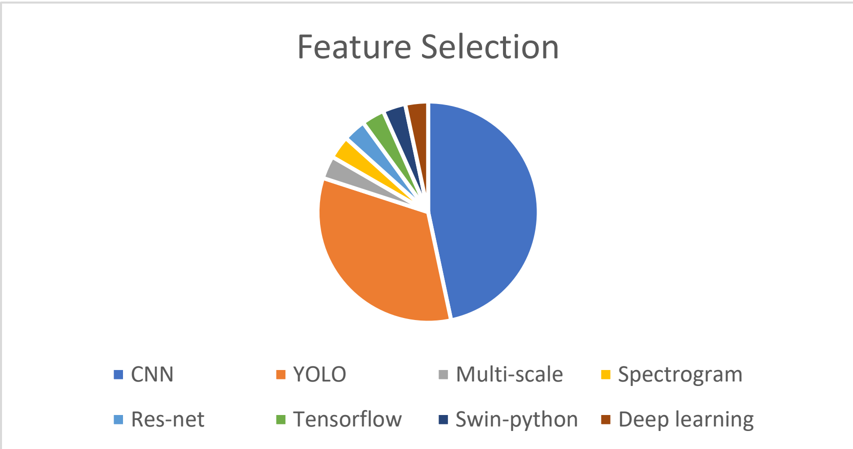

Fig. 4. Feature selection analysis

V. CONCLUSION

This study significantly improves blind user navigation assistance and spatial perception by incorporating a real-time distance calculation function into the AI-powered blind aid system. Traditional assistive technologies driven by AI primarily focus on object detection and classification, but they frequently lack the ability to detect obstacles' proximity, which is crucial for navigation in the real world.

DOI: 10.17148/IJARCCE.2025.14673

This study overcomes this constraint by employing software-based distance estimation methods, which give users context-aware audio feedback that gives priority to surrounding obstacles and improves overall safety. The suggested system makes use of YOLO-based object detection models in addition to distance estimation techniques like Euclidean distance computations, bounding box analysis, and monocular depth estimation. By utilizing these techniques, the system can forecast object proximity without relying on hardware-specific elements like LiDAR or stereo cameras, enabling simple implementation in typical computer configurations. This study also examines adaptive voice feedback systems, where auditory alerts are determined by how close the objects to be able to more effectively direct users' navigational choices. The system is designed to operate in real-time while optimizing computational efficiency, which makes it perfect for incorporation into assistive software or mobile apps. Additionally, the study emphasizes user customization and flexibility, allowing users to alter distance sensitivity settings to suit their unique mobility needs. This enables the system to be used efficiently in a variety of real-world situations, from indoor spaces with stationary barriers to outdoor spaces with dynamic barriers like moving cars and pedestrians. Despite these advancements, there are some problems that are still unresolved. Software methods for measuring distance are still impacted by things like lighting, circumstances, object hiding, and camera view angles. Future research must focus on improving distance estimation models, integrating AI-driven scene understanding, and increasing adaptability in a range of environmental circumstances. Furthermore, the user experience may be improved by including multimodal feedback mechanisms like haptic notifications, especially in noisy settings where auditory feedback may be less successful. In conclusion, the study offers a software-driven blind support system powered by artificial intelligence (AI) that enables real-time object detection and distance measurement for increased user autonomy and safety for those who are blind or visually impaired.

REFERENCES

[1] Devashish Pradeep Khairnar, Z. O., Jabbour, E., Ibrahim, P., & Ghaoui, A. (2012). PARTHA: A Visually Impaired Assistance System., 795–799. https://doi.org/10.1109/bmei.2012.6513135

[2] Rahul, M., Tiwari, N., Shukla, R., Tyagi, D., & Yadav, V. (2022). Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. *International Journal of Electrical and Electronics Research*, *10*(1), 18–22. https://doi.org/10.37391/ijeer.100103

[3] Adarsh, P., Rathi, P., Department of Computer Science & Engineering, Delhi Technological University, Kumar, M., & Department of Computer Science & Engineering, Delhi Technological University. (2020). YOLO v3-Tiny: Object Detection and Recognition using one stage improved model (conference-proceeding). *2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS)*.

[4] Thakurdesai, N., Tripathi, A., Butani, D., Sankhe, S., Department of Computer Science, Indiana University Bloomington, Department of Artificial Intelligence, Northwestern University, . . . Department of Computer Engineering, K.J. Somaiya College of Engineering. (n.d.). Vision: A Deep Learning Approach to provide walking assistance to the visually impaired.

[5] Najm, H., Faculty of Information Technology, University of Benghazi, Elferjani, K., Faculty of Information Technology, University of Benghazi, Alariyibi, A., & Faculty of Information Technology, University of Benghazi. (n.d.). Assisting Blind People Using Object Detection with Vocal Feedback.

[6] Alsultan, O. K. T., & Mohammad, M. T. (2023). A Deep Learning-Based assistive system for the visually impaired using YOLO-V7. Revue D Intelligence Artificielle, 37(4), 901–906. https://doi.org/10.18280/ria.370409

[7] Hameedi, H. S.. H. J., . S. (2021). You only look once (YOLOV3): Object Detection and recognition for indoor environment. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.4906284

[8] Dev, S., Jaiswal, S., Kokamkar, Y., Deshpande, K. B., & Upadhyaya, K. (2020). Voice Based Smart Assistive Device for the Visually Challenged., 1–5. https://doi.org/10.1109/iccdw45521.2020.9318604

[9] Durgadevi, S., Thirupurasundari, K., Komathi, C., & Balaji, S. (2020). Smart Machine Learning System for Blind Assistance.. https://doi.org/10.1109/icpects49113.2020.9337031

[10] Mahendru, M., & Dubey, S. K. (2021). Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3. *2022 12th International Conference on Cloud Computing, Data Science &Amp; Engineering (Confluence)*, 734–740. https://doi.org/10.1109/confluence51648.2021.9377064

[11] Marvin, E. & Department of Computer System Engineering, Universitas Prasetiya Mulya, BSD, Indonesia. (2020). *Digital assistant for the visually impaired* (journal-article). Retrieved from https://www.afb.org/aw/18/2/15244

[12] Baldota, D., Advani, S., Jaidhara, S., Hatekar, A., & Department of Electronics and Telecommunications Engineering, Thadomal Shahani Engineering College, Mumbai, India. (2021). Object Recognition using TensorFlow and Voice Assistant. *International Journal of Engineering Research & Technology (IJERT)*, *10–10*(09), 359–359. Retrieved from http://www.ijert.org

DOI: 10.17148/IJARCCE.2025.14673

[13] Guravaiah, K., Bhavadeesh, Y. S., Shwejan, P., Vardhan, A. H., & Lavanya, S. (2023). Third eye: object recognition and speech generation for visually impaired. *Procedia Computer Science*, *218*, 1144–1155. https://doi.org/10.1016/j.procs.2023.01.093

[14] Annapoorani, A., Senthil Kumar, N., Vidhya, V., Dept of Information Technology, & Sri Venkateswara College of Engineering, Chennai. (2021, March). Blind - Sight: Object Detection with Voice Feedback (journal-article). *International Journal of Scientific Research & Engineering Trends* (Vol. 7).

[15] M Thulasi , The Object Recognition Voice Assistant For Visually Impaired People. (2023). *International Research Journal of Modernization in Engineering Technology and Science*, *05–05*, 373–374. Retrieved from https://www.irjmets.com

[16] Aung, M. M., Maneetham, D., Crisnapati, P. N., & Thwe, Y. (2024). Enhancing Object Recognition for Visually Impaired Individuals using Computer Vision. *International Journal of Engineering Trends and Technology*, *72*(4), 297–305. https://doi.org/10.14445/22315381/ijett-v72i4p130

[17] Sindhura, K. a S., Jaiswal, J., & Jain Deemed to be University. (2019, March). Real Time Object Detection and Recognition with a Voice Feedback for the Blind (journal-article). *Journal of Emerging Technologies and Innovative Research* (Vol. 6, pp. 448–449). Retrieved from https://www.jetir.org

[18] Lilhare, R., Meena, J., More, N., Joshi, S., MIT School Of Engineering, & Dept. Of Electronics & Communication Engineering MIT SOE. (2021). Object Detection with Voice Feedback. *International Research Journal of Engineering and Technology (IRJET)*, 4567. journal-article. Retrieved from https://www.irjet.net

[19] Djinko, I. A. R., & Kacem, T. (2021). Video-based Object Detection Using Voice Recognition and YoloV7. journal-article.

[20] Dewangan, R. K., & Chaubey, Dr. S. (2021). Object Detection System with Voice Output using Python (journal-article). (International Journal for Research Trends and Innovation), *International Journal for Research Trends and Innovation* (Vol. 6, p. 15). Retrieved from https://www.ijrti.org

[21] DSOUZA, E. R., BHAT, D., TAURO, J. A., HARSHITH, S., MATHIAS, M. M., Department of Computer Science and Engineering, & Alva's Institute of Engineering and Technology, Moodbidri, India. (2021, July). REAL TIME OBJECT DETECTION AND RECOGNITION SYSTEM TO ASSIST THE VISUALLY IMPAIRED (journal-article). *International Journal of Creative Research Thoughts (IJCRT)* (Vol. 9, pp. 499–501).

[22] Prem Sagar, U., Indraja, C., Divya, N., Haripriya, M., Harikrishna, A., & International Journal of Engineering Technology and Management Sciences. (2023). Object Detection with Voice Feedback (journal-article). *International Journal of Engineering Technology and Management Sciences* (Vol. 7, p. 469). https://doi.org/10.46647/ijetms.2023.v07si01.081

[23] Pokala Nithya Sai., Prashanth, M., Sathvik, G., & G Vijay Kumar. (2024). REALTIME OBJECT DETECTION USING OPENCV. *Journal of Emerging Technologies and Innovative Research* (Vol. 11). Retrieved from https://www.jetir.org

[24] Sanjar, K., Bang, S., Ryue, S., & Jung, H. (2024). Real-Time object detection and face recognition application for the visually impaired. *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, *79*(3), 3569–3583. https://doi.org/10.32604/cmc.2024.048312

[25] Swathi, M., Supraja, R., Prasanna, M. L., Sameer, S., & Reddy, G. R. K. (2024). Real-time object detection and voice labeling for enhanced accessibility and visual interaction. In *Advances in computer science research* (pp. 721–733). https://doi.org/10.2991/978-94-6463-471-6_70

[26] Devaki, P., Shivavarsha, S., Kowsalya, G., Manjupavithraa, M., & Vima, E. (2019). Real-Time Object Detection using Deep Learning and Open CV. *International Journal of Innovative Technology and Exploring Engineering*, *8*(12S), 411–414. https://doi.org/10.35940/ijitee.l1103.10812s19

[27] Mukabe, C., Suresh, N., Hashiyana, V., Haiduwa, T., & Sverdlik, W. (2021). Object Detection and Classification Using Machine Learning Techniques, 86–97. https://doi.org/10.1145/3484824.3484895

[28] Bhaidasna, H., & Bhaidasna, Z. (2023). Object Detection Using Machine Learning : A Comprehensive Review. *International Journal of Scientific Research in Computer Science Engineering and Information Technology*, 248–255. https://doi.org/10.32628/cseit2390215

[29] Zhao, Z.-Q., Zheng, P., Xu, S., & Xindong Wu. (2019). Object Detection with Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*. Retrieved from https://arxiv.org/pdf/1807.05511.pdf

[30] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., University of Washington, Allen Institute for AI, & Facebook AI Research. (n.d.). *You only look once: Unified, Real-Time Object Detection* (journal-article). Retrieved from https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

