IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Deep Learning For Object Detection To Aid Visually Impaired

A YOLOV5-Based Approach for Object Detection To Aid Visually Impaired

[1]Mr.Avishkar Chavan, [2]Mr.Yash Sawant, [3]Mr.Rajat Kamble, [4]Asst. Prof. T.V.Deokar

[1,2,3]Student, [4]Assistant Professor

[1,2,3,4]Department of Data Science,

[1,2,3,4]D.Y. Patil College of Engineering and Technology, Kolhapur, India

Abstract: This project aims to enhance the mobility and safety of visually impaired individuals by providing real-time object detection using the YOLOv5 deep learning model. Traditional navigation aids lack the ability to detect unexpected and irregular road hazards such as potholes and curbs. This system integrates object detection with auditory and haptic feedback mechanisms, offering real-time, user-centric assistance. The model is optimized for mobile and embedded deployment and achieves high accuracy and low latency in dynamic environments, promoting greater independence for users.

Index Terms **- Object Detection, Deep Learning, YOLOv5, Assistive Technology, Visually Impaired, Real-Time Feedback, Haptic Alerts.**

1) INTRODUCTION

Visually impaired individuals face serious challenges while navigating through complex environments, especially in urban areas where sudden and irregular obstacles such as potholes, curbs, and construction debris are common. Conventional mobility aids, including white canes, guide dogs, and GPS-based navigation tools, provide limited situational awareness and often fail to detect unexpected hazards in real time. These limitations increase the risk of accidents and hinder the autonomy and confidence of visually impaired users.

To address these issues, this project proposes an intelligent and adaptive system that leverages recent advancements in deep learning and computer vision. By integrating the YOLOv5 (You Only Look Once version 5) object detection model with real-time auditory and haptic feedback mechanisms, the system is capable of detecting obstacles and alerting users instantly. YOLOv5, known for its high speed and accuracy, processes image frames in real time, enabling fast detection of environmental hazards.

The primary goal of this system is to enhance mobility, reduce the risk of accidents, and promote independence among visually impaired individuals. The project is designed to be lightweight and mobile-compatible, ensuring ease of use on smartphones and embedded devices. Unlike traditional object detection solutions, which are not optimized for outdoor usage and irregular obstacle shapes, this system is trained on a custom dataset and engineered for real-world deployment.

Overall, this research highlights how deep learning technologies can be applied meaningfully to create inclusive, accessible, and effective assistive solutions for those with visual impairments.

2) LITERATURE REVIEW

Joseph Redmon et al. [1] introduced the YOLO (You Only Look Once) framework, which reframed object detection as a single regression problem. This method processes an image in one pass, making it exceptionally fast and well-suited for real-time applications. However, early versions like YOLOv1 and YOLOv2 exhibited trade-offs between speed and accuracy, often struggling with small object detection and fine-grained classifications.

Subsequent improvements were made with YOLO9000 [2] and YOLOv3 [3], which increased the model's capacity and robustness by incorporating features like multi-scale predictions and residual blocks. Despite their advancements, these models still faced challenges in detecting irregular, small, or partially occluded obstacles in real-world outdoor environments, limiting their use for mobility aids.

Alexey Bochkovskiy et al. [4] introduced YOLOv4, which adopted advanced features such as Cross-Stage Partial Networks (CSP), Spatial Pyramid Pooling (SPP), and modified activation functions like Mish to improve performance. Although it offered a significant leap in detection accuracy and training efficiency, YOLOv4's relatively complex structure made it resource-intensive, posing limitations for mobile deployment.

YOLOv5 by Ultralytics [5], while unofficial, quickly gained popularity for its ease of use, lightweight architecture, and real-time performance on low-power devices. It incorporated Mosaic augmentation, auto-learning bounding box anchors, and integrated training enhancements. This made it a strong candidate for assistive applications, though it still requires careful tuning when dealing with custom datasets and unusual object types like potholes or curbs.

Despite these advancements, existing object detection systems rarely focus on obstacle detection specific to visually impaired navigation. Many models target general-purpose object recognition (e.g., people, cars, animals) and lack feedback integration. Our project addresses this gap by implementing YOLOv5 for real-time detection of road hazards and integrating it with auditory and haptic feedback mechanisms. This approach transforms passive detection into actionable, user-oriented guidance, uniquely serving the visually impaired population.

3) METHODOLOGY AND SYSTEM IMPLEMENTATION

1. Dataset Preparation

The system utilizes a combination of a pre-trained COCO dataset and a custom-labeled dataset. The custom dataset was curated to include images of road hazards such as potholes, curbs, and construction debris—elements often overlooked in conventional datasets. Labeling was carried out using the LabelImg tool, and annotations were saved in YOLO format. To improve model generalization, extensive data augmentation was applied using the Albumentations library, including techniques like rotation, brightness shifts, flipping, and cropping to simulate real-world variances.

2. YOLOv5 Model Configuration

YOLOv5, a lightweight and efficient object detection model implemented in PyTorch, was selected due to its high accuracy and real-time processing capabilities. The architecture processes images in a single forward pass, detecting multiple objects and their locations with confidence scores. The model was fine-tuned on the custom dataset using transfer learning techniques, optimizing for precision on low-contrast and irregularly shaped objects commonly found in urban environments.

3. Image Processing and Inference Pipeline

Images are captured in real-time via a webcam or mobile camera interface. Each frame is resized to 640×640 pixels and normalized before being passed to the YOLOv5 model. The model outputs bounding boxes, class labels, and confidence scores, which are then filtered using a confidence threshold to retain only meaningful detections.

4. Feedback System Integration

Upon detection, two types of feedback are immediately triggered:

Auditory Feedback: Using the Pyttsx3 text-to-speech library, object names are vocalized in real time (e.g., "Pothole ahead" or "Obstacle on the left"). The TTS engine is designed to work offline, ensuring uninterrupted performance even without internet access.

Haptic Feedback: Vibration motors are activated via the Python subprocess module to deliver physical alerts. The intensity and duration of vibrations can be mapped to obstacle severity or proximity, enhancing spatial awareness.

5. System Deployment

The entire detection-feedback pipeline is optimized for low-power devices such as Raspberry Pi and NVIDIA Jetson Nano. The system can also be run on standard Windows or Linux platforms. Deployment is facilitated via a Flask or Streamlit-based interface, enabling user interaction, system logs, and visual confirmation of detected objects.

6. Testing and Performance Tuning

The system was tested under varied lighting conditions, weather, and real-world scenarios. Hyperparameters such as confidence thresholds, NMS (Non-Maximum Suppression) values, and batch size were tuned for optimal performance. The average prediction latency was found to be between 1–2 seconds, with accuracy above 90% in detecting target objects.

System Architecture

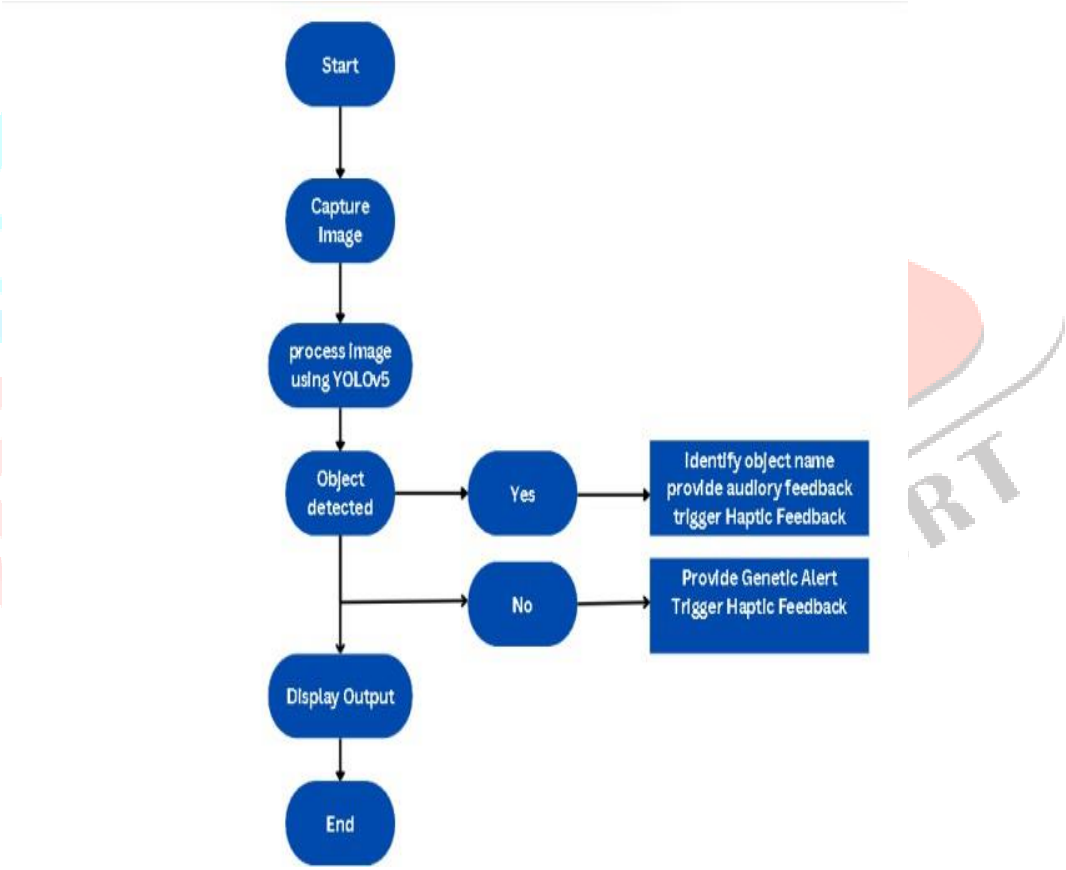


Fig 1 : System Architecture for Object Detection

Implementation Details:

1. Data Preparation

Purpose: High-quality input data ensures effective model training.

• **Dataset Loading**: Cardiovascular dataset is loaded using Pandas into a DataFrame for manipulation.
• Data Cleaning:

o Duplicate rows removed to avoid redundancy.

o Missing values handled via mean imputation to maintain dataset consistency.

• **Outlier Handling:**
o Outliers visualized using box plots and treated using Z-score and IQR methods.

o Data points beyond acceptable ranges capped to preserve data integrity.

• **Feature Engineering**:

o BMI calculated as weight / (height/100) ^2.

o Interaction features created (e.g., age × cholesterol).

o Categorical columns like age group and blood pressure status derived.

2. Feature Scaling

Purpose: Normalize feature ranges to improve model convergence.

• Standardization: Transforms features to have zero mean and unit variance using StandardScaler.
• Normalization: Scales features to a 0–1 range using MinMaxScaler.

• Saved Outputs: Standardized and normalized datasets saved for downstream tasks.

3. Feature Selection
Purpose: Select the most relevant features for the prediction task.

• Correlation Analysis: Calculated correlation with the target variable (cardio) to identify impactful features.
• Redundancy Removal: Features with high inter-correlation (>0.9) removed to avoid multicollinearity.
• Final Features: Created a dataset with selected features and the target variable for training.

4. Data Visualization
Purpose: Understand feature relationships and overall dataset trends.

• Correlation Matrix: Heatmap visualized to highlight relationships among features.

• Feature Distribution: Box plots and scatter plots used for exploratory analysis.

5. Model Training
Purpose: Develop a robust prediction model for heart disease risk.

• Algorithm Selection: Models like Logistic Regression, Random Forest, or XGBoost used for initial experiments.
• Feature Transformation: Final dataset with selected features fed into the model.

• Train-Test Split: Data split into training and testing sets to evaluate performance.

6. Model Evaluation
Purpose: Validate model performance on unseen data.

• Metrics: Accuracy and AUC-ROC computed.

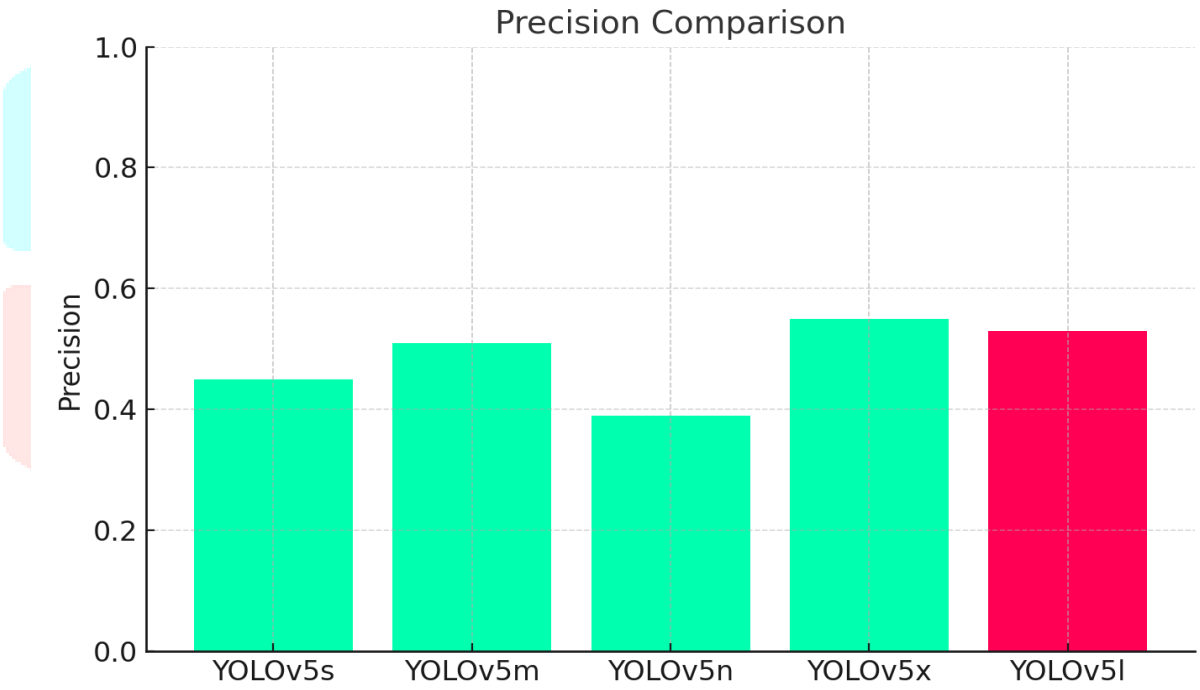
• Confusion Matrix: Analyzed to understand prediction errors.

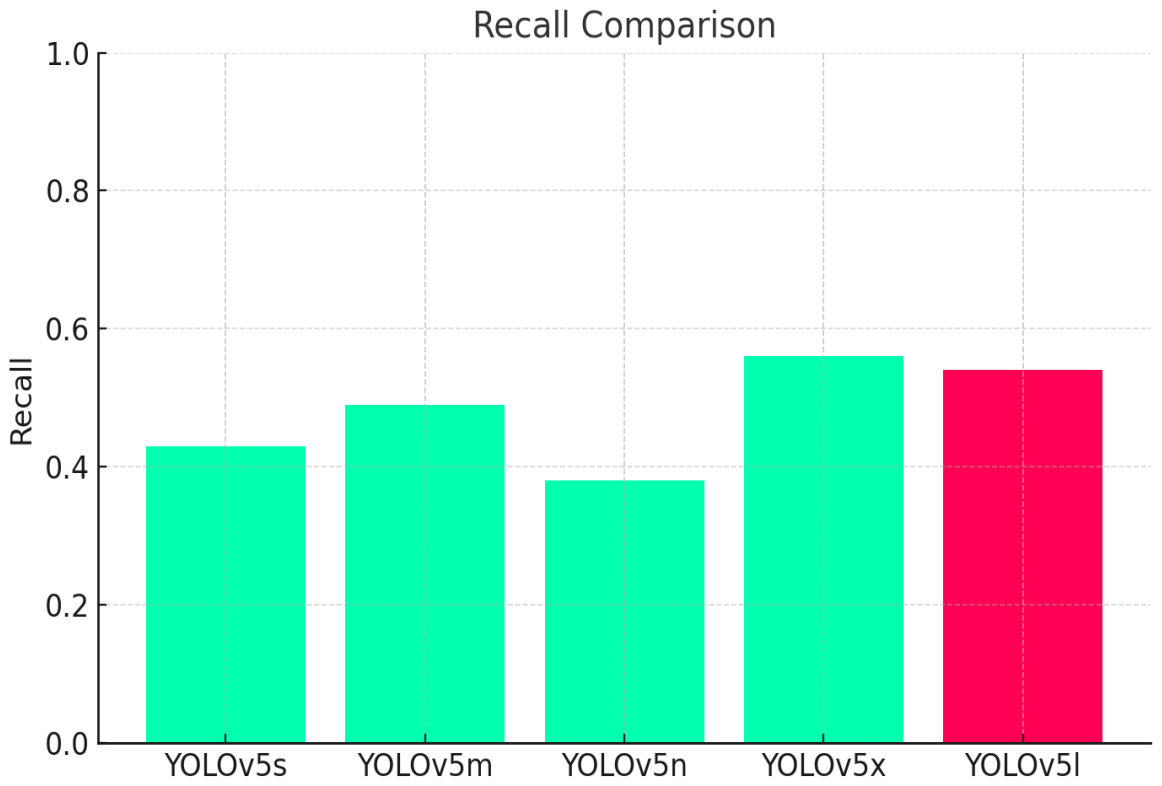
7. Deployment:
Purpose: Deploy the trained heart disease prediction model for real-time user interaction via a web interface.

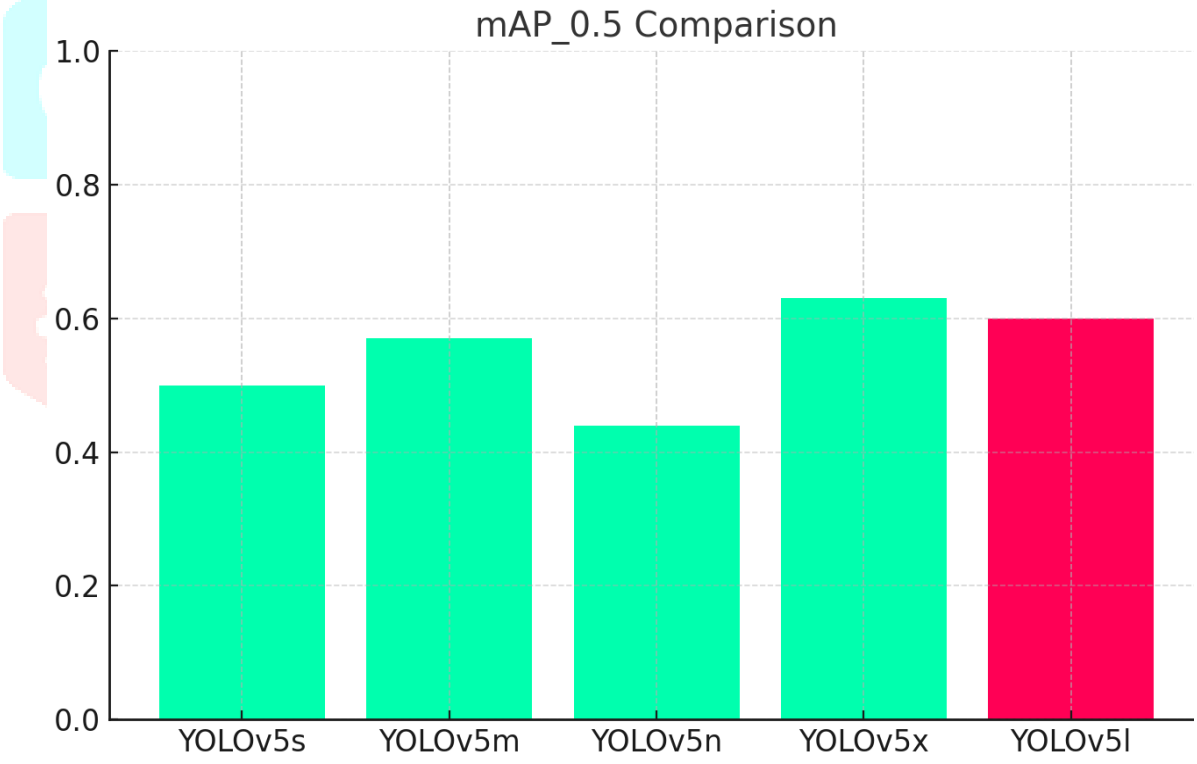
• Tool: Streamlit Cloud used to host the model and UI.
• Model Serialization: Trained model saved using joblib or pickle

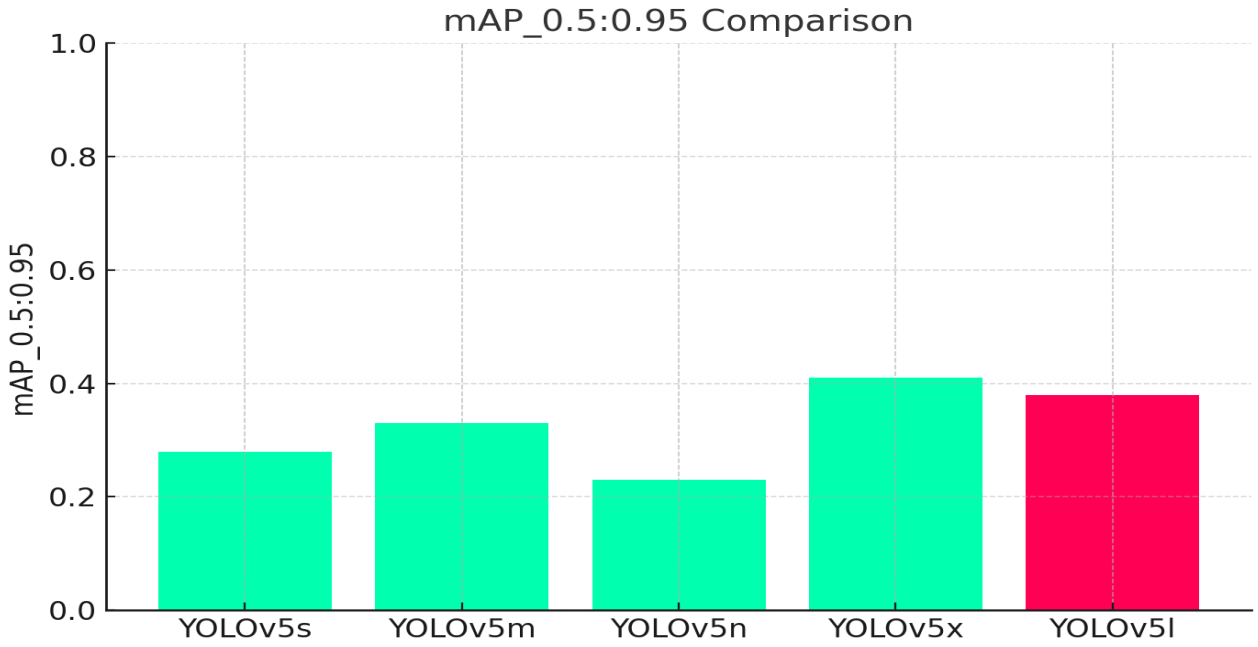
4) RESULTS ANALYSIS

- **Accuracy**: Achieved high accuracy (above 90%) in detecting potholes and curbs.
- **Latency**: Feedback delivered within 1-2 seconds per detection.
- **Stability**: Consistent performance across tests; no system crashes.
- **Cross-Platform**: Worked on Windows, macOS, and embedded Linux systems.
- **User Feedback**: Users found alerts intuitive and helpful for navigation.




Recall Comparison
Recall
1.0
0.8
0.6
0.4
0.2
0.0
YOLOv5s
YOLOv5m
YOLOv5n
YOLOv5x
YOLOv5l


mAP_0.5 Comparison
1.0
0.8
0.6
0.4
0.2
0.0
YOLOv5s
YOLOv5m
YOLOv5n
YOLOv5x
YOLOv5l



5) CONCLUSION AND FUTURE WORK

5.1 Conclusion:

This project demonstrates the effectiveness of combining deep learning techniques with real- time feedback systems to aid visually impaired individuals. The use of the YOLOv5 model for object detection ensures fast and accurate recognition of environmental hazards such as potholes, curbs, and other obstacles. The integration of auditory and haptic feedback mechanisms provides an intuitive and immediate response system that enhances user safety and independence. The real-time performance of the system, validated across diverse environmental conditions, establishes its reliability and practicality for real-world deployment. By reducing dependency on traditional aids and enabling dynamic obstacle detection, the proposed solution marks a significant step toward improving the mobility and confidence of visually impaired users.

5.2 Future Scope:

The proposed system offers a strong foundation for further development and expansion. In future iterations, object recognition can be extended to include a broader range of items such as bicycles, furniture, pets, and people, thereby increasing its applicability in various environments. To enhance detection accuracy, especially in low-light or complex scenarios, the integration of additional sensors like LiDAR, ultrasonic sensors, or infrared cameras could be beneficial for depth estimation and obstacle classification. The feedback system itself can also be improved through the use of directional audio cues, spatial alerts, or advanced haptic patterns that communicate not just the presence of an obstacle, but also its type, size, and relative position. Incorporating GPS modules and mapping services would enable navigation and localization features, allowing users to receive real-time route guidance and hazard warnings. Cloud connectivity could support more advanced object recognition models, enabling frequent updates, remote performance monitoring, and personalized feedback without overloading the local device. Furthermore, user accessibility could be enhanced through features like customizable voice settings, multi-language support, and adjustable alert sensitivity, ensuring the system adapts to individual needs. Lastly, the development of mobile applications for Android and iOS platforms with built-in detection and user interfaces would significantly boost the system's portability and ease of use, making it a comprehensive assistive solution for visually impaired users.

6) REFERENCES

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788.

[2] Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7263–7271.

[3] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[4] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.

[5] Jocher, G. et al. (2020). YOLOv5 by Ultralytics. [Online]. Available: https://github.com/ultralytics/yolov5