iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: II Month of publication: February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77551

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Smart Object Detection and Recognition System for Visually Impaired

Nishigandha S. Mankar[1], Yashashri R. Kadu[2], Yash P. Thakare[3], Abhay D. Bhakare[4], Shrikant D. Chaudhari[5], Prof. A.D.Chokhat[6]
[1, 2, 3, 4, 5]Student, [6]Assistant Professor at P R Pote Patil College of Engineering and Management, Amravati

Abstract: Visual impairment significantly restricts an individual's ability to perceive surroundings, recognize objects, and navigate safely in dynamic environments. This research proposes a Smart Object Detection and Recognition System designed to assist visually impaired individuals by leveraging advanced deep learning and computer vision techniques. The system captures real-time video input through a camera, processes the frames using optimized object detection algorithms such as Convolutional Neural Networks (CNN) and YOLO models, and identifies objects present in the environment with high accuracy. The detected object labels are converted into audio output through an integrated text-to-speech module, enabling users to receive immediate and meaningful auditory feedback. The proposed solution focuses on achieving real-time performance, high detection precision, and system portability while maintaining computational efficiency. Experimental evaluation under varying environmental conditions demonstrates reliable detection accuracy and minimal latency, making the system suitable for practical assistive applications. By integrating artificial intelligence with assistive technology, the proposed system aims to enhance independence, mobility, and overall quality of life for visually impaired individuals.

I. INTRODUCTION

The field of artificial intelligence and computer vision has witnessed remarkable progress over the past decade, enabling machines to interpret, analyse, and respond intelligently to visual information in real-world environments. Object detection and recognition, which involve identifying and classifying objects within images or video streams, have become fundamental components of numerous modern applications such as autonomous driving systems, robotic automation, security surveillance, medical imaging, and smart city infrastructure. These technologies rely heavily on deep learning architectures, particularly Convolutional Neural Networks (CNNs) and real-time detection frameworks like YOLO (You Only Look Once), which have demonstrated superior performance in terms of detection accuracy, speed, and adaptability under complex environmental conditions. Despite these technological advancements, visually impaired individuals continue to encounter significant barriers in performing everyday tasks such as identifying nearby objects, recognizing obstacles, and navigating unfamiliar surroundings safely. Conventional assistive tools, including white canes and guide dogs, provide limited environmental context and do not offer detailed object-level information. In response to this challenge, integrating intelligent vision systems into assistive technology presents a promising solution for enhancing independence and safety. The proposed Smart Object Detection and Recognition System aims to bridge the gap between advanced computer vision capabilities and practical accessibility needs by utilizing real-time image acquisition, deep learning-based object classification, and an efficient text-to-speech mechanism to deliver instant auditory feedback. The system is designed to operate with minimal latency while maintaining high detection precision across varying lighting conditions, object scales, and dynamic environments. By combining state-of-the-art deep learning algorithms with user-centric assistive design principles, this research seeks to develop a reliable, scalable, and portable solution that empowers visually impaired individuals with improved situational awareness, thereby contributing to inclusive technological innovation and socially impactful engineering advancements.

II. LITERATURE REVIEW

The domain of object detection and recognition has undergone substantial transformation over the past two decades, transitioning from traditional computer vision techniques to advanced deep learning-based methodologies. Early research in object detection primarily relied on handcrafted feature extraction approaches such as Haar-like features, Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and Speeded-Up Robust Features (SURF). These techniques utilized manually designed descriptors to capture edge, texture, and shape information from images, followed by classification using algorithms such as Support Vector Machines (SVM) or Adaboost classifiers.

Although these methods achieved moderate success in constrained environments, their performance significantly degraded under variations in illumination, object scale, rotation, background clutter, and occlusion. The lack of automatic feature learning limited their adaptability to complex real-world scenarios, especially in dynamic environments where real-time response is essential.

The emergence of deep learning marked a major breakthrough in the field of computer vision. Convolutional Neural Networks (CNNs) introduced automated hierarchical feature extraction, enabling systems to learn spatial patterns directly from raw image data without manual intervention. Region-based approaches such as R-CNN, Fast R-CNN, and Faster R-CNN improved localization accuracy by incorporating region proposal mechanisms, significantly enhancing detection precision compared to traditional sliding window techniques. However, these multi-stage detectors required substantial computational resources and were relatively slower, making them less suitable for real-time assistive applications. To address this limitation, single-stage detectors such as YOLO (You Only Look Once) and Single Shot Detector (SSD) were introduced, offering a unified detection framework that simultaneously predicts bounding boxes and class probabilities. YOLO, in particular, demonstrated exceptional real-time performance by framing object detection as a regression problem, thereby achieving high speed with competitive accuracy. Subsequent versions further improved detection performance in small object recognition and complex scenes, making them highly relevant for practical deployment in embedded systems.

In the context of assistive technologies for visually impaired individuals, several research efforts have attempted to integrate object detection models into wearable devices, smartphone applications, and smart navigation systems. Some systems focus primarily on obstacle detection using ultrasonic sensors or depth cameras, while others combine GPS-based navigation with voice guidance. However, many existing solutions face challenges such as high computational demand, limited object categories, dependency on cloud processing, latency issues, and reduced efficiency under low-light conditions. Recent studies emphasize lightweight neural network architectures, model optimization techniques such as pruning and quantization, and edge computing deployment to enable real-time processing on low-power devices like Raspberry Pi and mobile processors. Despite these advancements, there remains a need for a balanced system that ensures high detection accuracy, fast inference speed, portability, and user-friendly audio feedback integration. Building upon prior research in deep learning-based detection and assistive system design, the proposed work aims to develop a robust, scalable, and real-time object detection and recognition framework tailored specifically to the practical requirements of visually impaired users, thereby contributing to the advancement of intelligent and inclusive assistive technologies.

III. METHODOLOGY

A. Dataset Selection and Model Training

The development of the proposed Smart Object Detection and Recognition System begins with careful dataset selection and structured model training. High-quality datasets play a crucial role in ensuring the robustness and generalization capability of deep learning models. Standard benchmark datasets such as MS COCO (Common Objects in Context) are utilized because they contain a large variety of real-world objects captured under diverse lighting conditions, backgrounds, and object scales. In addition to publicly available datasets, custom datasets may also be created to include specific objects relevant to visually impaired users, such as doors, stairs, chairs, vehicles, and everyday household items. Data preprocessing techniques including image resizing, normalization, and augmentation (rotation, flipping, brightness adjustment) are applied to improve model adaptability and prevent overfitting. The selected dataset is then used to train deep learning architectures such as Convolutional Neural Networks (CNN) integrated with object detection frameworks like YOLO. The training process involves multiple epochs where model weights are iteratively updated using backpropagation and optimization algorithms such as Adam or SGD (Stochastic Gradient Descent). Loss functions such as classification loss and bounding box regression loss are minimized to improve object localization and recognition accuracy. Transfer learning techniques are employed by utilizing pre-trained weights from large-scale datasets, which significantly reduce computational cost and training time while maintaining high precision. This approach enhances convergence speed and allows the model to perform effectively even with limited hardware resources.

B. Model Validation, Testing, and Optimization

After successful training, the model undergoes a rigorous validation and testing phase to evaluate its performance under diverse environmental conditions. The validation process ensures that the model generalizes well beyond the training dataset and does not suffer from overfitting. Testing is conducted in real-world scenarios including indoor and outdoor environments with variations in lighting intensity, object occlusion, motion, and background complexity. This stage is essential to determine whether the system can reliably function in dynamic and unpredictable surroundings encountered by visually impaired users.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Performance evaluation metrics such as precision, recall, F1-score, and mean Average Precision (mAP) are calculated to measure detection accuracy. Additionally, detection speed is assessed using Frames Per Second (FPS) to confirm real-time processing capability. Since assistive systems require immediate response to ensure user safety, maintaining a balance between accuracy and speed is critical. To enhance inference efficiency, optimization techniques such as model pruning, weight quantization, and hardware acceleration using GPU or edge computing devices are implemented. These optimizations reduce computational complexity and memory usage while preserving detection performance, enabling deployment on resource-constrained devices like Raspberry Pi or embedded systems.

C. Real-Time Integration and Audio Feedback Mechanism

The final stage of the methodology focuses on integrating the trained and optimized model into a real-time object detection framework. A camera module continuously captures live video frames, which are preprocessed and passed to the detection model for instant analysis. The model identifies objects within each frame, generates bounding boxes, and assigns class labels with confidence scores. To avoid repetitive or unnecessary announcements, filtering mechanisms may be implemented to prioritize high-confidence detections and suppress redundant outputs. This ensures clarity and usability for visually impaired users. Once an object is detected and classified, the corresponding label is transmitted to a text-to-speech (TTS) engine, which converts textual output into clear and understandable audio signals. The audio feedback is delivered through speakers or headphones, providing immediate environmental awareness. The system is designed to operate with minimal latency to prevent delays that could compromise user safety. The overall integration emphasizes reliability, scalability, and user-centric design principles, ensuring that the assistive solution remains intuitive, portable, and adaptable to different deployment platforms. By combining real-time detection with efficient auditory communication, the methodology ensures maximum usability and practical effectiveness for visually impaired individuals.

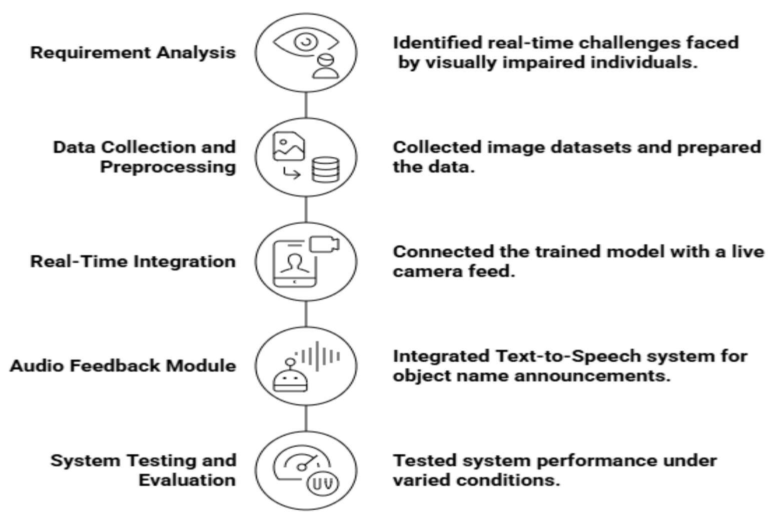


Fig.1. System Architecture

IV. APPLICATIONS

The proposed Smart Object Detection and Recognition System has significant applications in the field of assistive technology, particularly for visually impaired individuals. The primary application of this system is to enhance independent mobility and environmental awareness by providing real-time audio descriptions of surrounding objects. By detecting obstacles such as vehicles, stairs, doors, chairs, and other everyday objects, the system enables users to navigate safely in both indoor and outdoor environments. This reduces dependency on human assistance and improves confidence while performing daily activities.

In addition to personal mobility assistance, the system can be integrated into wearable devices such as smart glasses, head-mounted cameras, or portable embedded systems, making it suitable for continuous real-time usage. It can also be adapted into smartphone-based applications, where users can utilize their device cameras for object detection and receive instant voice feedback. Furthermore, the system can be enhanced to support obstacle distance estimation and contextual scene understanding, making it more intelligent and user-friendly.

Beyond assistive applications, the core technology of this system can also be extended to other domains such as smart surveillance, industrial automation, robotics, and healthcare monitoring. The integration of deep learning-based object detection with real-time processing makes the solution scalable and adaptable across multiple platforms. Thus, while the primary focus remains on supporting visually impaired individuals, the system demonstrates broad applicability in various real-world intelligent vision-based systems.

V. LIMITATIONS

Although the proposed Smart Object Detection and Recognition System demonstrates promising performance in assisting visually impaired individuals, certain limitations remain. One of the primary limitations is dependency on lighting conditions and camera quality. In extremely low-light, overly bright, or shadow-heavy environments, detection accuracy may decrease due to reduced image clarity.

Similarly, blurred frames caused by rapid movement can affect object recognition performance. While deep learning models are robust, they are not entirely immune to environmental variations.

Another limitation involves computational constraints and hardware dependency. Real-time object detection requires significant processing power, especially when running advanced models such as YOLO or CNN-based detectors. When deployed on low-power embedded systems like Raspberry Pi, performance may slightly decrease in terms of detection speed (FPS). Additionally, the system's efficiency depends on proper optimization techniques, and without hardware acceleration, latency may increase, which could affect user experience. Furthermore, the system is limited by the categories of objects included in the training dataset. If an object is not part of the trained classes, the model may fail to recognize it or misclassify it. Small, partially occluded, or overlapping objects can also reduce detection accuracy. The current system primarily focuses on object identification and does not fully incorporate advanced scene understanding, depth estimation, or contextual reasoning, which are important for more intelligent decision-making. These limitations highlight potential areas for further improvement and research enhancement.

VI. FUTURE SCOPE

The proposed Smart Object Detection and Recognition System lays a strong foundation for intelligent assistive technology; however, several advanced enhancements can further expand its functionality and real-world applicability. One of the most significant future improvements involves incorporating depth perception and spatial awareness mechanisms into the system. By integrating stereo cameras, ultrasonic sensors, LiDAR modules, or deep learning-based monocular depth estimation models, the system can estimate the distance between the user and surrounding objects. This enhancement would allow the system not only to identify objects but also to provide directional and proximity-based warnings such as "Obstacle two meters ahead" or "Step down approaching." Such distance-aware guidance would substantially increase safety and navigation accuracy in both indoor and outdoor environments, especially in crowded urban settings.

Another major area of development lies in optimizing the system for wearable and embedded platforms. Future work can focus on deploying lightweight neural network architectures such as MobileNet, EfficientNet, or YOLO-Nano variants to reduce computational load while maintaining high detection accuracy. Implementing advanced model compression techniques such as quantization, pruning, and knowledge distillation can significantly enhance performance on low-power devices like Raspberry Pi, NVIDIA Jetson Nano, or custom AI edge processors. Additionally, integration into smart glasses, body-mounted cameras, or compact assistive modules would improve portability and continuous usability. The incorporation of cloud-based synchronization and over-the-air model updates could allow periodic performance improvements without requiring hardware changes, ensuring long-term adaptability and scalability.

Further advancements may include the addition of contextual scene understanding and intelligent decision-making capabilities. Instead of simply detecting isolated objects, future systems can analyze object relationships and environmental context using advanced neural networks such as Transformer-based vision models. For example, the system could identify complex scenarios such as "pedestrian crossing the road," "vehicle approaching from left," or "open doorway on right side." Integration with GPS-based navigation systems could enable real-time route guidance combined with environmental awareness, offering a comprehensive mobility assistant.

Moreover, multilingual voice support, customizable audio preferences, emotion-sensitive feedback tones, and user-adaptive learning mechanisms could enhance personalization. Research can also explore integrating obstacle prediction models that anticipate potential hazards based on motion patterns.

VII. CONCLUSIONS

This research presents a Smart Object Detection and Recognition System specifically designed to assist visually impaired individuals by leveraging advancements in deep learning and computer vision. The proposed system integrates real-time image acquisition, efficient object detection models such as CNN and YOLO, and a text-to-speech module to provide immediate auditory feedback. By converting visual information into meaningful audio guidance, the system enhances situational awareness and promotes independent mobility for users. The modular architecture ensures scalability and adaptability, allowing deployment across different hardware platforms including embedded systems and portable devices.

Extensive training, validation, and optimization processes were implemented to achieve a balance between detection accuracy and real-time performance. Performance metrics such as precision, recall, and inference speed demonstrate that the system is capable of operating efficiently under diverse environmental conditions. Although certain limitations exist, such as dependency on lighting conditions and hardware capabilities, the proposed framework establishes a reliable foundation for intelligent assistive technology. The experimental observations indicate that the integration of optimized deep learning models significantly improves detection consistency while maintaining low latency, which is crucial for real-world usability.

Overall, the system contributes toward inclusive technological innovation by bridging the gap between advanced artificial intelligence and practical accessibility solutions. By integrating deep learning-based object detection with user-centric design principles, the proposed approach not only addresses current challenges faced by visually impaired individuals but also opens avenues for future research in intelligent navigation, contextual scene understanding, and wearable assistive systems. The developed solution represents a meaningful step toward enhancing autonomy, safety, and quality of life for visually impaired users through smart vision-based assistance. Furthermore, the adaptability of the proposed framework ensures its potential extension into broader assistive ecosystems, reinforcing the role of artificial intelligence in building socially impactful and human-centered technologies.

VIII. ACKNOWLEDGMENT

The authors sincerely express their gratitude to the Department of Computer Science and Engineering, P. R. Pote Patil College of Engineering and Management, Amravati, for providing the necessary infrastructure, technical resources, and academic environment required for the successful completion of this research work. We are especially thankful to our project guide for their continuous guidance, valuable insights, constructive feedback, and constant encouragement throughout the development and evaluation of the Smart Object Detection and Recognition System for Visually Impaired. Their mentorship significantly contributed to strengthening the technical depth and practical implementation of the project. We also extend our appreciation to all faculty members, colleagues, and peers who supported us through discussions, suggestions, and motivation during various stages of research and experimentation. Furthermore, we acknowledge the contributions of the open-source research community, whose publicly available datasets, frameworks, and deep learning tools greatly facilitated the design, training, and optimization of the proposed system.

REFERENCES

[1] M. Suresh, Sandhyarani, et al., “Smart Vision and Intelligent Object Recognition System for the Visually Impaired,” in Proceedings of SoCPaR 2023, Lecture Notes in Networks and Systems (LNNS), vol. 1245, Springer, 2023.

[2] U. Masud, T. Saeed, H. M. Malaikah, et al., “Smart Assistive System for Visually Impaired People: Obstruction Avoidance Through Object Detection and Classification,” IEEE Access, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3146320.

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

[4] F. Lan, G. Zhai, and W. Lin, “Lightweight Smart Glass System with Audio Aid for Visually Impaired People,” in Proceedings of IEEE TENCON, 2015, pp. 1–4.

[5] P. Sudhakar, K. S. D. Ushasri, et al., “Object Detection and Alert System for Visually Impaired People,” International Journal of Engineering Science and Advanced Technology (IJESAT), July 2023.

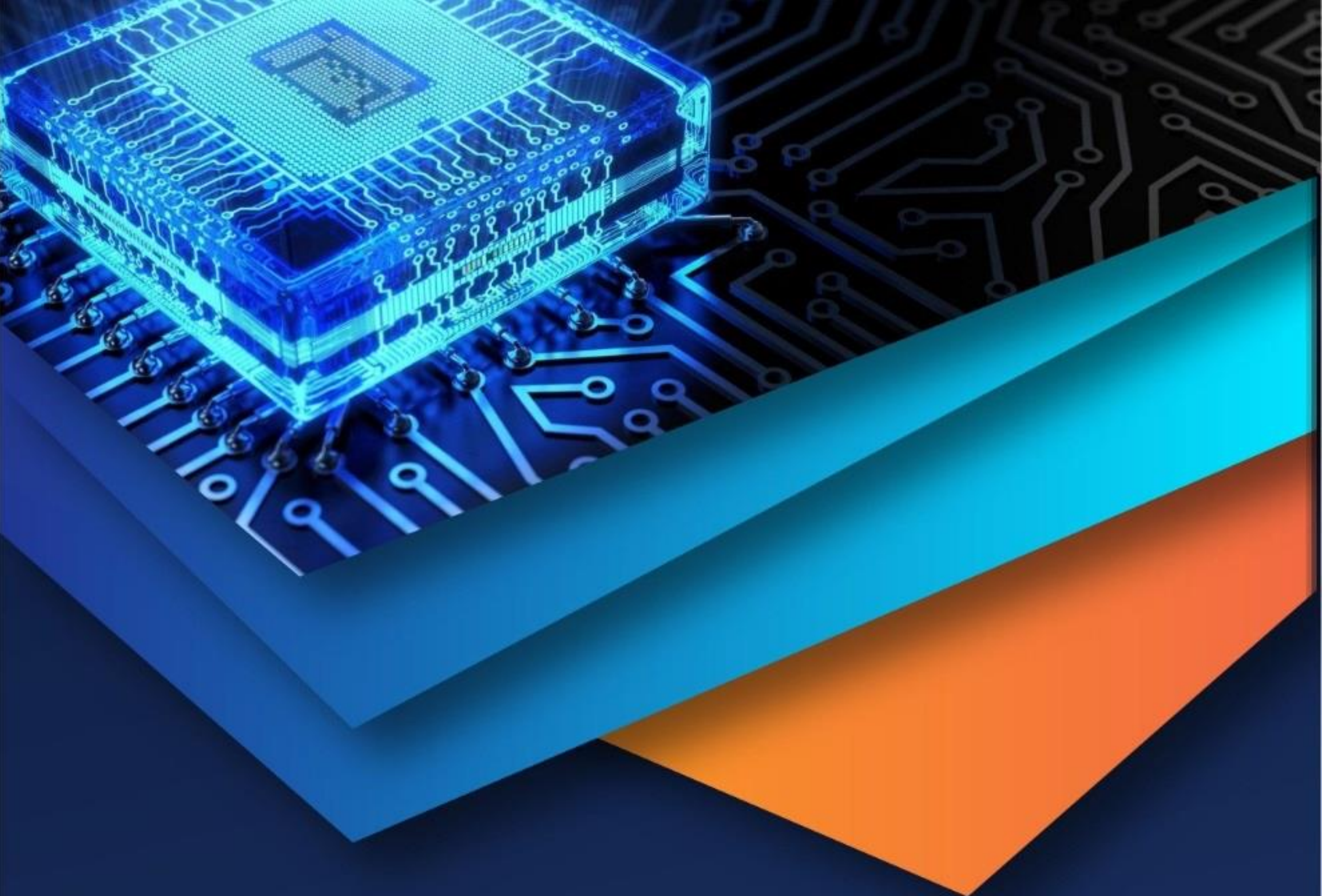
[6] L. Kavyashree, et al., “Review of Smart Cane for Visually Impaired Using Object Detection,” International Journal of Research Publication and Reviews (IJRPR), vol. 6, no. 1, Jan. 2025.

[7] A. Al-Khalifa and H. Al-Razgan, “Assistive Technologies for the Blind: A Survey,” International Journal of Computer Applications, vol. 85, no. 3, pp. 1–7, Jan. 2014.

[8] D. Dakopoulos and N. G. Bourbakis, “Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 1, pp. 25–35, Jan. 2010.

[9] S. Liu, D. Huang, and Y. Wang, “Revisiting Single Shot Multibox Detector,” in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019, pp. 5933–5942.

[10] M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in Proceedings of the International Conference on Machine Learning (ICML), 2019, pp. 6105–6114.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY