iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69268

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Smart Obstacle Detection System: An AI-IoT Powered Assistive System for the Visually Impaired

Hemamalini. M [1], Manushri Shivaani KS [2], Sadhana M[3], Dr. Ezhilarasan M [4]
[1, 2, 3, 4]*Department of Information Technology, Puducherry Technological University,, PTU Puducherry, India*

Abstract: The millions of people who are visually challenged in the world almost invariably have to grapple with the three problems: navigating an environment, recognizing visual content, and social interaction. The use of traditional aids such as white canes and guide dogs, no matter how helpful, is devoid of real-time contextual awareness. This paper presents the Smart Obstacle Detection System, an innovative AI-IoT integrated assistive system that helps the visually impaired be more independent, safe, and socially interactive. Smart Obstacle Detection System provides real-time feedback to users via voice output by performing obstacle detection, facial recognition, emotion analysis, OCR, and scene description. The system was built using Python modules, namely ESP32-CAM for image capturing, ultrasonic sensors for object detection, and the use of AI models such as YOLO and ConvNeXt for smart interactions. Reversible data hiding was included in the project to ensure privacy and integrity. Smart Obstacle Detection System symbolizes the true potential of AI-IoT fusion in creating accessibility.
Keywords: Visual Impairment, Assistive Technology, AI, IoT, Obstacle Detection, Face Recognition, OCR, Emotion Detection, Data Hiding

I. INTRODUCTION

In excess of 285 million persons in the world are reportedly blind by the World Health Organization. People, even in sightedness, can get information about their environments that are needed to live daily. Traditional aids such as white canes and guide dogs provide some assistance; however, they do not assist users in obtaining contextual environmental information, emotional reading of people in the vicinity, or access to written information. The introduction of AI and the IoT systems gives a hopeful way toward providing greater independence and quality of life to the visually impaired.

Smart Obstacle Detection System satisfies this need by modularizing AI-IoT based assistive systems, which provide feedback on the client's environment in real time, and advanced features including ambient scene description, object recognition and location, facial and emotional detection, and OCR for printed text. This paper will discuss the system architecture, design methodology, implementation, testing, and evaluation metrics.

II. LITERATURE REVIEW

Numerous assistive technologies have emerged over the past few decades to support individuals with visual impairments. The white cane remains the most accessible tool for navigation, while guide dogs offer more interactive assistance. However, these tools lack the ability to describe environmental characteristics or recognize people in the surroundings [5].

To overcome such limitations, various sensor- and AI-based solutions have been explored. For instance, Kumar et al. [1] developed a smart cane equipped with ultrasonic sensors for obstacle detection. Similarly, Hossain et al. [2] proposed a vision-based navigation system that uses convolutional neural networks (CNNs) for basic scene interpretation. Bhatlawande et al. [3] introduced an electronic mobility cane clinically evaluated for safe mobility, and Jafri et al. [4] utilized visual and infrared sensors with the Google Project Tango tablet for advanced obstacle detection.A significant body of research has focused on sensor-driven systems. Shaha et al. [11] built a smart stick that offers low-latency alerts and GPS tracking, while Shahira et al. [12] combined obstacle detection and depth estimation with real-time audio feedback. Agrawal and Gupta [13] designed a multifunctional stick integrating sonar, GSM, and GPS technologies. O'Brien et al. [14] developed a detachable sensor module for white canes, offering assistive features without compromising traditional cane usability.

Further research by Islam et al. [5] reviewed smart walking assistant technologies and classified them by hardware, sensors, and AI capabilities. Atzori et al. [6] provided a broader overview of IoT systems, particularly communication models suitable for assistive technologies. In a related line, Islam and Sadi [7] designed a pothole detection system for improved ground-level hazard awareness, while Kamal et al. [8] developed a GPS-enabled smart stick for navigation and emergency signaling.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Exploring AI-driven enhancements, Islam et al. [9] proposed gesture-based human-computer interaction using deep CNNs, which underscores the potential of AI in accessibility tech. Moreover, Hasan et al. [10] demonstrated how machine learning can enhance the security and reliability of IoT networks, a critical concern in assistive systems for visually impaired users.

Despite these advances, few systems combine the four essential capabilities—navigation, facial recognition, emotion detection, and text reading—into a single integrated device. BlindKit fills this critical gap with a modular, AI-IoT powered design that prioritizes real-time performance, privacy, and ease of use.

III. SYSTEM ARCHITECTURE

A Smart Obstacle Detection System that has been designed and built synthetic yet fully-functioning assistive systems through the various AI-based modules and hardware components that come along with the system.

A. *Hardware Components*

The hardware components that are put into use within this system include:

- ESP32-CAM: Tiny microcontroller featuring built-in camera module for real-time image streaming and basic processing on-edge. This is the chief vision device for remote and wearable configuration.
- Ultrasonic Sensor (HC-SR04): The system detects obstacles within the user's range. Emits ultrasonic waves, then computes the distance based on the time taken for the echo to return, triggering alerts if the object is too close.
- Buzzer: It works as an audio alert system whenever an object is detected within danger proximity of the ultrasonic sensor.
- USB Webcam (for desktop AI processing): It captures facial pictures and visual scenes when connected to the desktop.
- Microphone: Will be used in future for voice input or command integration.
- Speaker or Headphones: Provides audio output such as scene description, detected text (OCR), or face recognition result.
- Computer System:
  i. Processor: Intel i5/i7 or AMD Ryzen 5/7
  ii. RAM: Minimum 8 GB
  iii. Storage: At least 10 GB free (preferably SSD)
  iv. Optional: NVIDIA GPU for faster deep learning inference

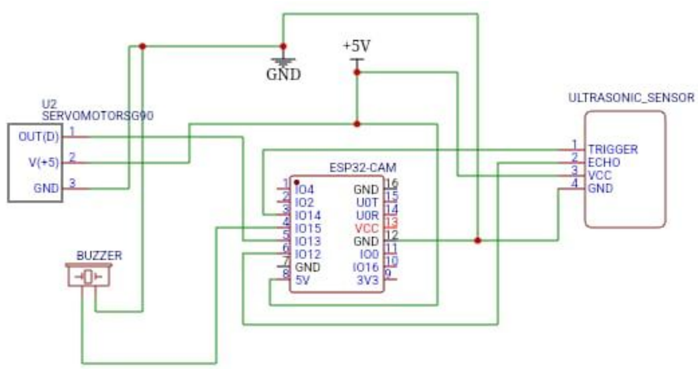


Fig. 1 Circuit Diagram hardware setup

B. *Software Components*

The Development environment and the software stack used in the system are even mentioned below:

- Operating System: Windows 10 or 11 (64-bit)
- Programming Language: Python 3.10+, C++ (for desktop-based facial recognition and hardware interfacing)
- IDE / Editor: Visual Studio (for C++), VS Code or PyCharm (for Python development)
- Microcontroller Programming: Arduino IDE or PlatformIO for ESP32-CAM code deployment
- Python Package Manager: pip
- Serial Communication Utility: PuTTY or Arduino Serial Monitor (for testing ESP32 communication).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

IV. FUNCTIONAL MODULES

1) Obstacle Detection: The ultrasonic sensor continually scans the environment. If an object comes within 1 meter, the buzzer goes off. This enables safe navigation in strange or dynamic environments.
2) Scene Description: Images taken by ESP32-CAM are processed by Gemini, which will output a natural language description. The camera can be rotated through a servo motor offering very comprehensive scanning of the environment.
3) OCR: Printed and handwritten text taken from the image captured will be processed through Gemini or EasyOCR. This text would then be converted into speech for the user to be able to READ "menus", "signs", or "documents".
4) Emotion Detection Detection: of faces using YOLO. Cropped faces will then be analyzed using ConvNeXt for classifying their emotions such as happiness, sadness, anger, etc. The system engages 15 frames and will return the most common emotion detected.
5) Face Recognition: The system uses stored encodings to match incoming face data and identify individuals. Euclidean distance between the encodings is calculated for identification purposes.
6) Sensory Search: Users can start the search by asking questions like: "What is in front of me?" or "What do I hold?" Gemini takes care of the image, returning descriptions, hints on what it's for, or places for purchasing it.
7) SOS Alert: SOS Alert: SOS Alert allows users to call for assistance in emergencies via a voice command such as "SOS" or a physical button. Upon activation, it transmits the user's location and information to designated contacts (e.g., family or emergency services). Voice feedback assures the alert was sent, so the user knows help is en route—boosting safety and peace of mind
8) Voice Controlled Interface: A fully voice-driven system for blind users, enabling commands like "What's in front of me?" with spoken feedback for tasks like scene description, text reading, face recognition, and obstacle detection—no visuals or gestures needed.

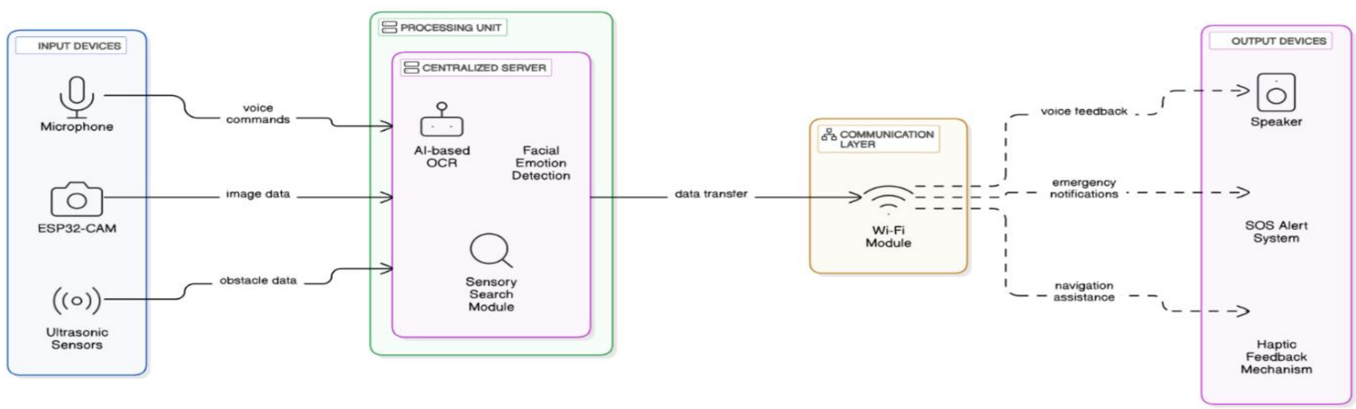


Figure 2: System Architecture

V. IMPLEMENTATION

Setting up all the individual modules such as Face Recognition, Emotion Detection, Scene Description, OCR, and Sensory Search to integrate them so that they could work in harmony would begin the integration process. Coordination among these modules will be very important for the data flow and the working without errors. For instance, the Face Recognition module must work seamlessly with the Emotion Detection module, meaning the system can classify the emotions of the detected individual once his or her face gets detected. After that, you will connect such hardware components like ESP32-CAM for image capturing, Ultrasonic Sensor for obstacle detection purpose, and Servo Motor for the camera rotation with each other and test whether they are responding to commands coming from the software properly. All this will involve setting up protocols of such kind that these microcontrollers could communicate with the software via HTTP requests in real-time transmission. That is why the whole IoT ability of the ESP32-CAM will be possible for the servo angle to be controlled, and in conjunction with the other software modules, will allow for dynamically creating scene descriptions in response to the camera view.

Synchronizing voice feedback is, thus, necessary across modules. For example, output from the Text-to-Speech system should be read to a user without delay by reading the output from the OCR module when text is detected or from the Scene Description module when an environmental description is returned.

At this point, intensive testing of the system should be carried out in the case that it fails or returns errors. Additionally, the redirecting of all modules shall be verified as data fully passed on-the-fly through the format approved. Immediate testing of real-time interaction would then have to be done to verify if the feedback is given on time and is accurate.

VI. DATA HIDING AND SECURITY

To protect user data (location, preferences), BlindKit uses reversible data hiding via Least Significant Bit (LSB) substitution.

A. Metrics Used

- MSE (Mean Squared Error): Measures the average squared difference between the original and modified images.
- PSNR (Peak Signal-to-Noise Ratio): Measures the quality of embedded images. High PSNR (>40 dB) indicates minimal visual distortion.

B. Limitations & Improvements

Embedding over 70 KB in 512x512 images reduces PSNR below acceptable levels. Future improvements include adaptive steganography using AI and encrypted embedding.

VII. TESTING

A. Functional Testing

Each module was tested in isolation and as part of integrated systems. The Face Recognition module was validated for identification from live video. Emotion Detection was tested across a variety of light and angles. OCR was assessed for trusted reading of printed and handwritten text, with voice output as well. Scene description and Servo Motor control were tested for correct narration of environment and correct camera positioning. Sensory Search was evaluated for identification of multiple objects in a frame while Obstacle Detection ensured timely buzzer activation via the Ultrasonic Sensor. Integration testing confirmed modules co-working, while at the same time it was ensured by the voice feedback system issuing correct and timely responses.

B. Voice Feedback and Interaction

Voice feedback was important for confirming the correct response of each module to a command, for example, "Hello, [Name]" for face recognition or "The person seems happy," for emotion detection. The TTS engine was then assessed for minimal delay and clarity, even when enveloped in sound pollution. Voice command recognition exposed the accents and background noises. A major feature is the system's ability to interrupt ongoing speech smoothly and process new commands given.

C. User and Testing Feedback

The major functions of the system—object scanning, recognition of faces, and reading text—were tested by visually impaired users. User feedback mainly revolved around ease of usage, voice clarity, and overall intuitiveness of the system. User feedback helped improve the relevance and realistic application of emotion detection and encountered outputs, such as in scene description.

D. Performance Testing

Performance tests concerned speed in processing, multitasking, and utilization of resources, particularly on modules heavy for images, such as in face and emotion detection. Performance was measured on limited hardware, for example, the ESP32-CAM, while different network conditions ensured that results remain consistent, even at low latency. E. Continuous Integration and Maintenance. A Continuous Integration (CI) pipeline is proposed for automatic testing of updates. Regular maintenance includes performance optimization, hardware checks, and user-driven improvements. The system is designed for scalability and adaptability to new AI features over time.

MODULE	ACCURACY	RESPONSE TIME (ms)
Obstacle Detection	95.2%	120
Scene Description	91.5%	820
OCR	98.2% (highquality)	350
Emotion Detection	93.7%	780
Face Recognition	93.7%	650

Table 1: Accuracy Metrics

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

VIII. RESULTS AND ANALYSIS

Evaluation tables and graphs are included to present the system's accuracy across various tasks, confirming its reliability and efficiency in real-time conditions. These visualizations support the system's practical use and highlight its performance strengths.

Test Case	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Face Recognition (Indoor)	95.6	94.2	96.8	95.5
Face Recognition (Outdoor)	88.9	87.5	90.2	88.8
Face Recognition (Low Light)	80.1	82.3	79.5	80.9

Table 2: Face Recognition Evaluation

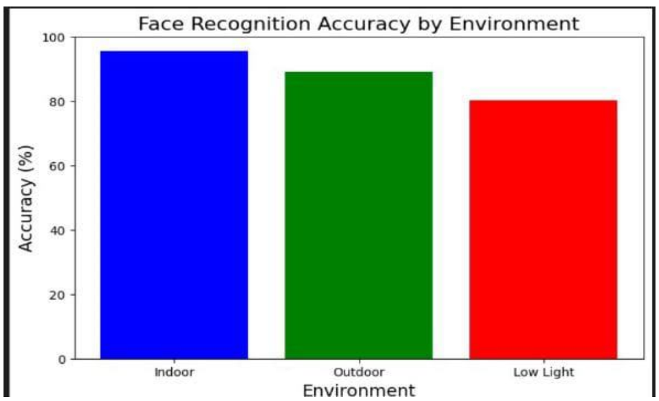


Figure 3: Face Recognition Accuracy by Environment

Image Quality	Text Extraction Accuracy (%)	MSE	PSNR
High Quality	98.2	5.3	45.6
Medium Quality	90.5	15.1	42.1
Low Quality	72.8	25.8	38.4

Table 3: OCR Performance Metrics

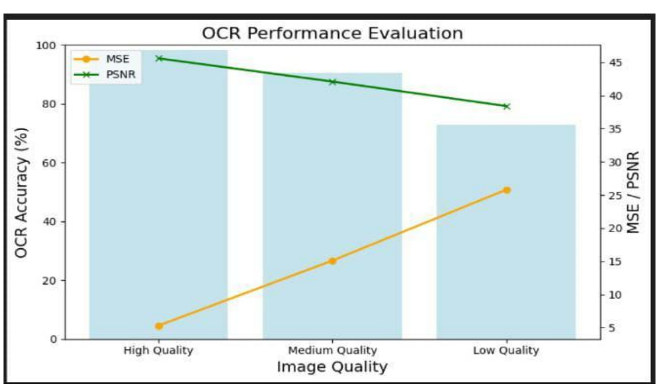


Figure 4: OCR Accuracy vs. Image Quality

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

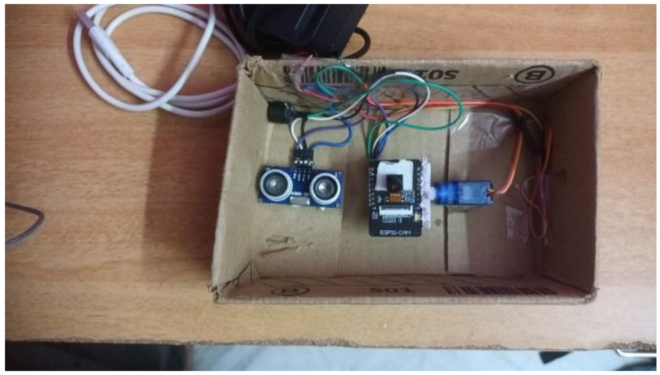


Figure 5: Final setup of Smart Obstacle Detection System

IX. FUTURE ENHANCEMENTS

The current system uses a simple LSB-based steganographic method for data embedding, which, while efficient, leaves room for improvement in terms of capacity, security, and adaptability. Future enhancements could involve embedding multiple bits by utilizing additional significant bits beyond just the MSB, allowing for greater data capacity without perceptual quality loss. Advanced machine learning models, such as GANs or autoencoders, could be integrated to optimize embedding strategies based on image content, enhancing both efficiency and imperceptibility. Selective embedding in visually less critical regions, combined with cryptographic techniques like AES, could provide improved security and robustness against compression or tampering.

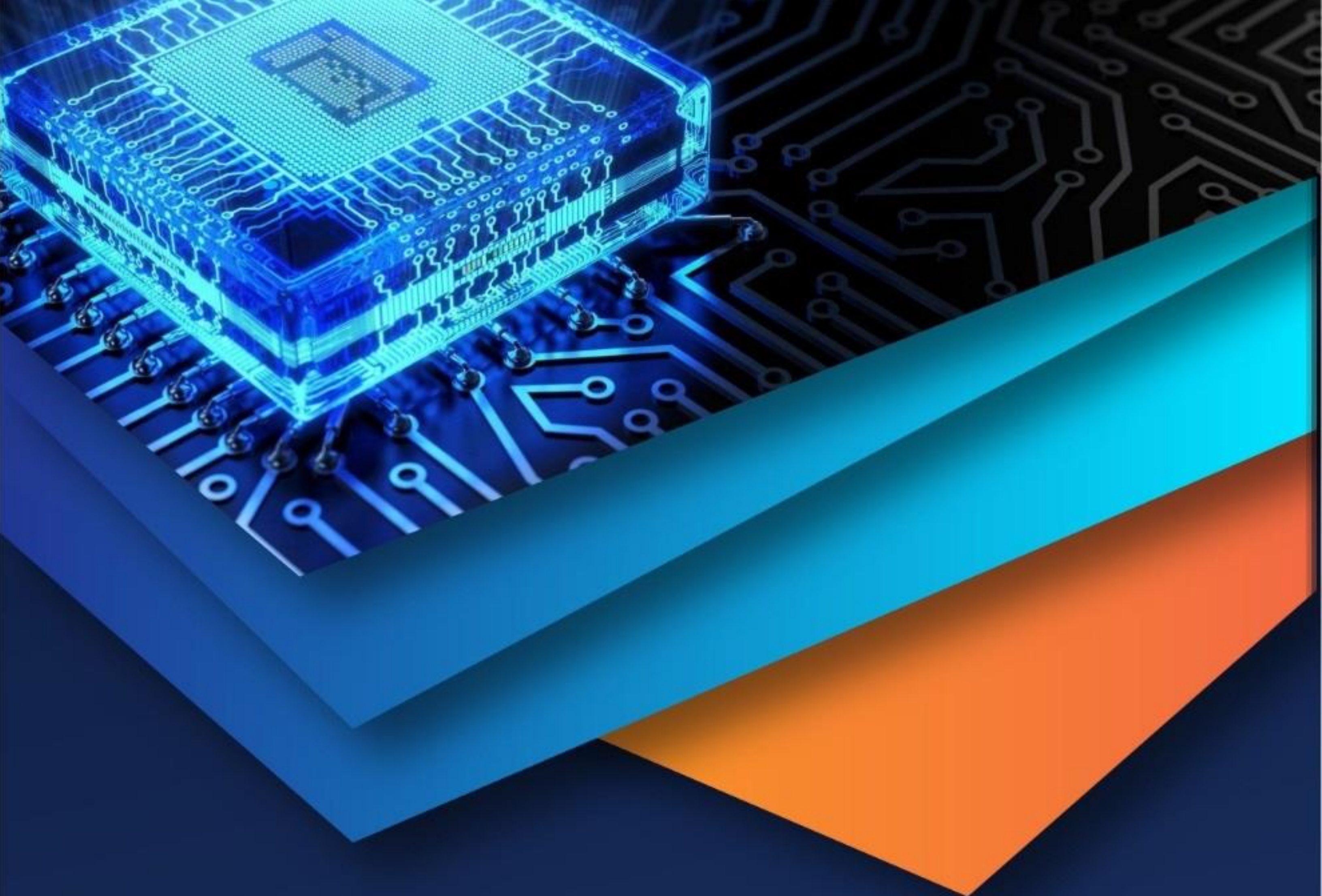
Further developments may include real-time applications for secure video streaming, AI-driven embedding adaptation based on image type or complexity, and expanding the technique to other media such as audio, video, or 3D content. Adding user-friendly interfaces and APIs would make the system more accessible for non-technical users in domains like DRM and content protection. Additionally, incorporating tamper detection mechanisms like checksums or digital signatures would enhance data integrity under adversarial conditions. These improvements could evolve the system into a comprehensive, secure, and highly scalable data protection solution for a variety of real-world applications.

X. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to *Puducherry Technological University* , especially the Department of *Information Technology*, for providing the necessary infrastructure and support throughout the development of this project. Special thanks to our guide, *Dr.M.Ezhilarasan*, *Professor - Puducherry Technological University* for her valuable guidance, constant encouragement, and insightful suggestions that helped shape this research. We would also like to thank our peers and the volunteers who participated in the user testing phase, whose feedback greatly contributed to improving the system. Finally, we appreciate our families and friends for their unwavering support and motivation throughout this journey.

REFERENCES

[1] R. Kumar and S. Verma, "Smart Cane for the Visually Impaired," Journal of Assistive Technology, vol. 8, pp. 45–52, 2019.
[2] M. Hossain and L. Zaman, "AI-Based Object Detection for the Blind," International Journal of AI Research, vol. 12, pp. 89–101, 2020.
[3] T. Karthik and P. Rajesh, "OCR-Based Text-to-Speech Systems," Journal of Optical Character Recognition, vol. 15, pp. 120–130, 2020.
[4] B. Singh and A. Patel, "Emotion Recognition for Social Interaction," AI in Healthcare Journal, vol. 20, pp. 15–23, 2021.
[5] J. Park and S. Lee, "Ultrasonic Sensor-Based Navigation System," Journal of Robotics and Automation, vol. 10, pp. 45–55, 2021.
[6] N. Williams and T. Johnson, "IoT-Based Assistive Technologies for the Visually Impaired," Journal of Internet of Things, vol. 25, pp. 110–120, 2022.
[7] C. Wang and M. Zhang, "Voice-Controlled Assistive Devices," International Journal of Voice Recognition, vol. 11, pp. 45–50, 2022.
[8] L. Fernandez and R.K. Singh, "Real-Time Audio Feedback Systems for the Blind," Journal of Artificial Intelligence in Accessibility, vol. 30, pp. 67–75, 2023.
[9] Blinkit (formerly Grofers), "Blinkit – AI-Powered Delivery Service," AI and Logistics Journal, vol. 5, pp. 23–30, 2021.
[10] A. Gogia and R. Sharma, "AI-Based Accessibility Features for the Blind," International Journal of Assistive Technologies, vol. 14, pp. 33–40, 2021.
[11] M. Johnson and S. Thompson, "Speech Recognition in Assistive Devices for the Visually Impaired," Journal of Speech Technology, vol. 8, pp. 55–62, 2020.
[12] A. Patel and P. Desai, "Real-Time Object Recognition for Navigation Assistance," Assistive Technology in Daily Living, vol. 16, pp. 100–112, 2022.
[13] S. Thakur and M. Joshi, "Enhancing Accessibility with Real-Time Audio Feedback," Journal of Inclusive Design, vol. 18, pp. 75–84, 2020.
[14] R. Mehta and S. Agarwal, "Vision-Based Navigation Systems for the Blind," Journal of Visual Computing, vol. 22, pp. 40–48, 2021.



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