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.2025.70016

Visioneer: AI Guided Navigation for the Visually Impaired

Mohamed Areeb[1], Siddharth Kumar[2], Bhagyashree Patil[3]
Artificial Intelligence & Data Science Department, Shah & Anchor Kutchhi Engineering College, Mumbai, India

Abstract: Navigating independently in an unfamiliar environment poses significant challenges for individuals with visual impairments. Traditional aids like canes and guide dogs, while helpful, offer limited contextual awareness and situational understanding. This project, titled "Visioneer: AI-Guided Navigation for Visually Impaired," aims to bridge this gap through an intelligent, real-time assistive system that leverages advancements in Artificial Intelligence, Computer Vision, and Wearable Technology. The proposed solution is a pair of AI-enabled smart glasses integrated with a camera and a speaker, capable of detecting obstacles, identifying key landmarks, and delivering navigational cues via audio prompts. The system transforms visual data into meaningful auditory instructions, enabling users to move confidently and safely in dynamic environments.

At the core of the system lies the YOLO (You Only Look Once) object detection algorithm, which enables rapid and accurate identification of objects in the user's path. OpenCV is utilized for real-time image processing, enhancing the system's ability to interpret the environment efficiently. To provide seamless audio feedback, the pyttsx3 text-to-speech library is integrated, converting recognized objects and spatial cues into speech output. This combination of technologies ensures that the system remains both fast and functional, operating smoothly on portable hardware. By enhancing situational awareness and reducing dependency on others, the AI-Guided Navigation system empowers visually impaired individuals with a greater sense of autonomy, offering a practical, scalable solution for inclusive navigation.

Keywords: AI-guided navigation, visually impaired, YOLO, object detection, OpenCV, pyttsx3, assistive technology, smart glasses, deep learning, computer vision.

I. INTRODUCTION

The ability to perceive and interact with the surrounding environment through vision is something most people take for granted. For individuals with visual impairments, however, the absence or limitation of sight brings about significant challenges in performing even the most routine daily tasks—particularly those involving navigation and spatial awareness. Walking in public spaces, identifying obstacles, locating landmarks, and reacting to dynamic situations are all activities that become considerably more complex without visual input. Traditional mobility aids such as white canes and guide dogs offer a level of assistance, but their capabilities are inherently limited. These aids rely heavily on physical proximity and cannot provide comprehensive, real-time information about the surroundings, which is critical for safe and confident navigation.

In recent years, the advancement of Artificial Intelligence (AI), Computer Vision, and wearable technology has created new opportunities to enhance the quality of life for people with disabilities. Intelligent systems are now capable of interpreting complex visual data, recognizing patterns, and making decisions in real-time—all of which can be applied to create assistive technologies that are both powerful and practical. Among these developments, object detection algorithms, real-time image processing tools, and speech synthesis engines have shown remarkable potential in building applications tailored to the needs of the visually impaired.

This project, titled “Visioneer: AI-Guided Navigation for Visually Impaired,” focuses on designing and implementing a smart, AI-driven navigation system that can be worn in the form of glasses. The system integrates a front-facing camera to capture real-world visuals, the YOLO (You Only Look Once) algorithm for object detection, OpenCV for image processing, and the pyttsx3 library for converting textual information into speech. Through this combination, the system is able to perceive the user’s surroundings, identify obstacles or key objects, and communicate this information through clear audio cues. The ultimate goal is to allow visually impaired individuals to move more independently, safely, and efficiently through both familiar and unfamiliar environments.

This introductory chapter outlines the foundational elements of the project. It begins by providing the background and motivation behind the development of AI-based assistive technology. It then defines the specific problem being addressed, followed by a clear articulation of the objectives the project aims to achieve. Finally, it highlights the broader significance of the system, not only in terms of technical innovation but also in promoting inclusivity and accessibility in everyday life.

A. Background

Visual impairment affects millions of individuals across the globe, significantly limiting their ability to navigate independently, interact with their surroundings, and perform routine activities with confidence and ease. These limitations not only impact their mobility but also affect their social inclusion, employment opportunities, and overall quality of life. Traditionally, mobility aids such as white canes and guide dogs have been the primary solutions to assist visually impaired individuals in navigating their environment. While these aids are undoubtedly helpful, they offer only a basic level of assistance. White canes are limited to detecting immediate obstacles on the ground, and guide dogs, though intelligent and responsive, are expensive to train and maintain and cannot provide detailed contextual awareness.

Both of these traditional solutions operate through physical contact or passive guidance, lacking the capability to dynamically interpret changing environments or offer real-time information about objects, people, or potential hazards in the user's path. This makes it particularly challenging for visually impaired individuals to move safely and independently in unfamiliar or crowded areas. With the rapid advancement of modern technologies—especially in the fields of Artificial Intelligence (AI), Computer Vision, and embedded systems—a significant opportunity has emerged to create intelligent assistive tools that can bridge this accessibility gap in a more meaningful way.

The integration of AI into wearable devices marks a transformative shift in the development of assistive technologies. These smart systems can now process visual data in real time, recognize objects and obstacles, track motion, and even understand patterns within the environment. By combining these capabilities with audio feedback systems, it becomes possible to build tools that translate visual information into verbal cues, empowering users with greater environmental awareness. *Visioneer: AI-Guided Navigation for Visually Impaired* is a project conceptualized and developed with this very purpose in mind. It aims to provide a real-time, intelligent navigation system embedded within a pair of wearable glasses. Equipped with a front-facing camera, object detection algorithms, image processing tools, and a text-to-speech engine, the system is designed to capture and interpret visual input from the environment and guide the user through clear, spoken instructions. By offering constant, hands-free, and context-aware support, Visioneer strives to make everyday navigation not only safer and more efficient but also truly empowering for individuals with visual impairment.

B. Problem Statement

Visually impaired individuals often face significant challenges in navigating unfamiliar environments due to a lack of accessible, real-time feedback about their surroundings. Existing solutions such as canes and guide dogs offer limited contextual information and are often insufficient in crowded or complex settings. While smartphone-based applications do exist, they typically require active user interaction and are not optimized for constant environmental scanning. The absence of a compact, intelligent, and real-time navigation solution leaves a crucial gap in accessibility tools for the visually impaired. This project aims to solve this problem by developing a wearable AI-based system that can detect objects, analyze the environment, and provide auditory guidance in real time.

C. Objectives

The primary goal of this project is to develop an intelligent, real-time navigation system tailored specifically for visually impaired individuals. The system aims to empower users by interpreting the visual environment and converting that data into helpful audio cues. To achieve this, the following specific objectives have been defined:

1) To design and develop a wearable AI-powered navigation system for visually impaired individuals:

The core objective is to create a compact, user-friendly wearable device—integrated into a pair of glasses—that can assist users with limited or no vision. The wearable form factor ensures convenience and hands-free operation. The device should be lightweight, minimally intrusive, and aesthetically acceptable while offering real-time feedback that aids navigation in both indoor and outdoor environments.

2) To implement real-time object detection using the YOLO algorithm for accurate environmental analysis:

For the system to be genuinely useful, it must be capable of identifying various objects and obstacles in the user's immediate surroundings. YOLO (You Only Look Once) is a state-of-the-art deep learning algorithm known for its high-speed and accurate object detection. Its real-time processing capabilities make it ideal for this application, allowing the system to recognize people, vehicles, walls, signboards, and other relevant objects in milliseconds, thereby enabling the user to react appropriately and safely.

3) To create a lightweight, low-latency system that functions effectively on portable hardware:
The entire system must be designed with portability and performance in mind. As it runs on wearable hardware like a Raspberry Pi or similar embedded device, resource efficiency and fast processing times are crucial. Optimization techniques are employed to ensure that object detection, audio output, and image processing occur with minimal delay, providing a smooth and uninterrupted user experience.

4) To improve the safety, independence, and confidence of visually impaired users during navigation:
Ultimately, the project aims to enhance the quality of life for visually impaired individuals by promoting independent navigation. By providing continuous, real-time updates about the environment, the system reduces reliance on external assistance and helps users navigate unfamiliar or crowded spaces more safely. This empowerment fosters a greater sense of autonomy, freedom, and self-confidence.

D. Significance

The significance of this project lies in its potential to create a meaningful impact on the lives of visually impaired individuals by leveraging the power of modern technologies like Artificial Intelligence, Computer Vision, and embedded systems. Vision impairment often restricts an individual's ability to move around freely, confidently, and independently, which in turn affects other aspects of life such as employment, education, and social interaction. Existing mobility aids, although useful, are not equipped to understand or communicate the complexity of dynamic environments. This project addresses this gap by introducing an AI-based assistive device that provides contextual, real-time navigation support.

The proposed system, *Visioneer*, stands out because it combines object detection, real-time image analysis, and speech output to simulate a visual guide for users. By doing so, it transforms passive navigation into an interactive, informative, and safe experience. Through voice prompts generated via pyttsx3 and visual input processed using YOLO and OpenCV, the system interprets the surrounding environment and narrates it to the user—thus effectively acting as a “digital companion” that can recognize, interpret, and communicate environmental cues. This is especially valuable in unfamiliar or crowded places where traditional aids fall short.

Another crucial aspect of this project’s significance is its emphasis on affordability, accessibility, and portability. While there are commercial AI-based solutions available in the assistive tech market, many of them are either too expensive, require internet connectivity, or lack real-time functionality. Visioneer is designed to run on low-cost, portable hardware, making it viable for widespread use—even in under-resourced or remote regions. Its offline processing capabilities further increase its usability in places with limited or no internet access.

From a societal perspective, Visioneer contributes to the broader movement toward inclusive technology. It aligns with global goals such as the United Nations’ Sustainable Development Goal (SDG) 10, which focuses on reducing inequality by ensuring equal access to technologies for people with disabilities. This project not only addresses a pressing real-world problem but also showcases how interdisciplinary innovation can be directed toward humanitarian causes.

In summary, Visioneer has the potential to enhance the autonomy, mobility, and quality of life for visually impaired individuals. By integrating AI into a practical, wearable solution, it makes a significant step toward inclusive, intelligent, and human-centered technology. It embodies the spirit of using engineering for social good—bridging the gap between vision and perception through smart assistance.

II. LITERATURE REVIEW

A. Smart Navigation System (Saranya et al., 2022)[1]

Takeaways: This study presents a smart navigation system that combines ultrasonic sensors and GPS technology to assist visually impaired individuals in movement. The system effectively detects nearby obstacles and alerts the user through vibration and voice commands. The integration of wearable hardware enhances the practicality of the solution.

Limitations: The reliance on ultrasonic sensors limits the system’s ability to accurately differentiate objects in complex environments. Additionally, the lack of AI-based image processing reduces the precision of real-time guidance, making it less effective in highly dynamic settings.

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

B. Smart Navigation System for Visually Impaired People (Benitha et al.)[2]

Takeaways: This research focuses on a sensor-based navigation aid that integrates RFID technology to assist users in structured indoor environments. The study highlights the importance of predefined pathways for guiding visually impaired users efficiently, ensuring minimal deviation from safe zones.

Limitations: The system is heavily dependent on RFID tags, which limits usability in open or unfamiliar spaces where RFID infrastructure is absent. The approach also lacks computer vision-based object recognition, restricting adaptability to unstructured environments.

C. Artificial Intelligence-Based Smart Navigation System (Mary Joans et al., 2022)[3]

Takeaways: This paper explores the application of AI-driven smart navigation, incorporating image processing and deep learning to detect objects. The study demonstrates how computer vision and convolutional neural networks (CNNs) can improve the precision of obstacle detection, enhancing navigation safety.

Limitations: The proposed system requires high computational power, making it less feasible for lightweight and portable hardware. The response time for real-time detection is also a challenge, which may affect usability in fast-paced environments such as crowded streets.

D. AI-Based Navigation System for Blind Persons (Shirol et al.)[4]

Takeaways: The study focuses on an AI-driven obstacle detection system using edge computing to reduce latency. By leveraging machine learning models, the system improves real-time detection accuracy, offering better guidance through voice-based alerts.

Limitations: The system lacks a speech synthesis module capable of context-aware navigation. Additionally, the need for continuous internet connectivity for certain AI features limits its offline usability, making it less practical in low-network areas.

E. Smart Assistive Navigation System for Visually Impaired People (Okolo et al., 2025)S[5]

Takeaways: This research presents an IoT-integrated navigation system, incorporating cloud-based AI processing to assist visually impaired individuals. The use of smart glasses with real-time feedback mechanisms enhances mobility and safety.

Limitations: The cloud dependency raises concerns about data privacy and security. Moreover, the real-time performance is limited by network latency, making the system less efficient in environments with poor internet connectivity.

III. SOFTWARE REQUIREMENT SPECIFICATION

A. Introduction

Software Requirement Specification (SRS):

The SRS defines the functional and non-functional requirements of the *Visioneer* project, guiding developers, designers, and testers. Visioneer is an AI-based navigation system for the visually impaired, using computer vision, deep learning, image processing, and text-to-speech to detect obstacles in real-time and deliver audio feedback. System accuracy, speed, and reliability are key to ensuring a smooth user experience.

B. Overall Description

1) Product Perspective

The *Visioneer* system is designed as a wearable assistive device, integrating hardware (camera, processing unit, speaker) and software (AI models, OpenCV, text-to-speech engine). The core AI model leverages YOLO (You Only Look Once) for real-time object detection, while OpenCV handles image preprocessing and frame extraction. The system runs on a portable computing device (e.g., Raspberry Pi, NVIDIA Jetson Nano) with minimal latency to ensure smooth operation.

2) Product Functions

The major functionalities of the software are:

- Object Detection: Identifies objects, obstacles, and dynamic elements in the user's environment using YOLO.
- Image Processing: Enhances video frames using OpenCV before feeding them into the object detection model.
- Text-to-Speech Conversion: Converts detected object information into audio instructions via pyttsx3.
- Real-time Feedback: Provides instantaneous spoken alerts to the user regarding surrounding objects and obstacles.

- Lightweight Operation: Ensures the system runs smoothly on embedded hardware with low computational power.

3) User Characteristics

The primary users of *Visioneer* are visually impaired individuals who rely on auditory cues for navigation. The system must be:

- Easy to use: Minimal user interaction required.
- Reliable: High accuracy in object detection and speech output.
- Non-intrusive: Lightweight and comfortable to wear.
- Responsive: Low latency for real-time assistance.

4) Constraints

- Limited processing power due to embedded hardware.
- Need for real-time performance optimization.
- Ensuring offline functionality without reliance on cloud services.
- Adapting to diverse environments with varying lighting and obstacles.

C. External Interface Requirements

1) Hardware Interfaces

- Camera Module: Captures real-time video feed (e.g., Raspberry Pi Camera, USB Camera).
- Processing Unit: Raspberry Pi, Jetson Nano, or equivalent to run AI models.
- Speaker/Earphones: Outputs text-to-speech instructions.
- Power Supply: Portable battery or direct power input for long usage.

2) Software Interfaces

- Operating System: Linux-based OS (Raspberry Pi OS, Ubuntu).
- Programming Language: Python for AI model, image processing, and TTS.
- Libraries & Frameworks:
- ➢ OpenCV: For image preprocessing and handling camera input.
- ➢ YOLO (You Only Look Once): For real-time object detection.
- ➢ pyttsx3: For offline text-to-speech synthesis.
- ➢ NumPy/Pandas: For data manipulation and handling.

3) Communication Interfaces

- The system does not require network connectivity for core functionality.
- Communication between components (camera, processing unit, speaker) occurs through onboard processing.
- The user interacts with the system solely via voice feedback; no touchscreen or keyboard input is needed.

D. System Requirements

1) System Feature 1: Real-Time Object Detection

- Description and Priority:

This feature is the core functionality of *Visioneer*, allowing the system to detect and recognize objects in real time using YOLO (You Only Look Once) and OpenCV. It ensures that visually impaired users can receive immediate and accurate information about their surroundings, helping them navigate safely. This feature has the highest priority as it directly impacts user safety and system usability.

- Stimulus/Response Sequences:
- Stimulus: The camera captures the environment and provides a real-time video feed.
- Response: The YOLO algorithm processes the frames, detects objects, and generates an audio description of the detected objects.
- Functional Requirements:
- ➢ The system must capture continuous real-time video from the camera.

- The AI model must detect objects with an accuracy of at least 85% in various environments.
- The detected objects must be classified and announced to the user within 1 second of detection.
- The system must be able to detect both static and dynamic objects, such as obstacles, people, and vehicles.
- The system must provide low-latency processing to ensure real-time feedback.

2) System Feature 2: Text-to-Speech (TTS) Audio Feedback

- Description and Priority:

This feature converts detected object data into spoken audio feedback using pyttsx3. It enables users to understand their surroundings through voice guidance, ensuring that they can make informed decisions while navigating. The priority level is high, as this feature is essential for the effective communication of detected objects to the user.

- Stimulus/Response Sequences:
- Stimulus: The system processes object detection results and determines relevant information to convey.
- Response: The pyttsx3 text-to-speech engine generates and plays an audio message, informing the user about nearby objects and potential obstacles.
- Functional Requirements:
- The system must convert detected object data into speech output within 1 second of detection.
- The TTS engine must support clear and natural-sounding voice synthesis.
- The system must allow for language customization to accommodate different users.
- The audio feedback should be non-intrusive and easily understandable in various environments.

3) System Feature 3

- Description and Priority
- Feature Name: Real-time Audio Feedback for Object Proximity
- This feature enables the system to provide real-time audio cues to the visually impaired user, indicating the type and distance of nearby objects. The feature uses the YOLOv8 object detection model to analyze the video feed and determine both object class and estimated distance from the user. The output is then converted to verbal cues through a wireless mini speaker.
- Priority: High This feature is critical for the core functionality of the Visioneer system, as it directly affects user navigation and situational awareness.
- Stimulus/Response Sequences
- Stimulus: The user begins walking, and the camera detects a new object in the frame.
  Response: The system identifies the object (e.g., "chair") and estimates the distance (e.g., "2 meters away"). The speaker plays the audio message: "Chair, 2 meters ahead."
- Stimulus: An object is rapidly approaching the user (e.g., a bicycle).
  Response: The system identifies the object and triggers an urgent response tone with the message: "Bicycle approaching quickly from the left!"
- Stimulus: Multiple objects are detected within proximity.
  Response: The system prioritizes the closest or most relevant object and issues a clear, concise message: "Wall, 1 meter ahead."
- Functional Requirements
- The system shall detect and identify objects in real-time using the YOLOv8 model.
- The system shall estimate and track object distance from the camera using bounding box size and frame reference.
- The system shall generate and play audio messages within 1 second of object detection.
- The system shall filter overlapping messages and prioritize critical alerts.
- The system shall maintain continuous feedback without user input while navigating.
- Non-Functional Requirements
- The system should achieve object detection and audio feedback latency under 1.5 seconds.
- The detection accuracy for commonly encountered objects (walls, doors, people, vehicles, etc.) should exceed 85%.
- The audio feedback should be loud and clear enough to be heard in moderately noisy environments (≥ 70 dB).

IV. PROJECT SCHEDULING AND PLANNING

The development of the Visioneer system has been structured into six distinct and systematic phases to ensure organized progress, timely execution, and efficient allocation of resources. Each phase focuses on a critical component of the system's life cycle, from conceptualization to validation.

1) *Phase 1: Literature Review and Initial Requirements*

Duration: 2 Weeks

Objective: To understand the existing systems and technologies aiding the visually impaired and identify gaps that Visioneer can fill. This phase involves surveying prior research, patent documents, and academic publications

2) *Phase 2: System Design and Model Selection*

Duration: 2 Weeks

Objective: To architect the overall system, including hardware and software components. This phase involves deciding the system flow, selecting YOLOv8 as the core object detection model, and finalizing hardware like the mini wireless camera and speaker.

3) *Phase 3: Data Collection, Preprocessing, and Model Testing*

Duration: 3 Weeks

Objective: To collect a relevant dataset for object detection in typical indoor and outdoor environments, perform preprocessing (resizing, labeling, noise removal), and test the YOLOv8 model for accuracy and speed. Modifications and fine-tuning are performed as needed based on early performance results.

4) *Phase 4: User Interface (UI) Development*

Duration: 2 Weeks

Objective: To design and implement a user-friendly interface that allows control, monitoring, and accessibility. The interface should be minimalistic, voice-based, and optimized for use by visually impaired users. This also includes options for manual override and audio feedback customization.

5) *Phase 5: Integration of Models with the UI*

Duration: 1 Week

Objective: To seamlessly integrate the YOLOv8 detection model with the UI and ensure real-time communication between the camera input, model processing, and speaker output. The integration focuses on reducing latency, maintaining performance stability, and ensuring synchronization of feedback.

6) *Phase 6: Large-Scale Testing and Validation*

Duration: 2 Weeks

Objective: To perform extensive testing in varied real-world scenarios including different lighting conditions, obstacle types, environments. The system is validated with feedback from actual users (especially visually impaired individuals) and optimized based on their input. This phase ensures that the system is reliable, accurate, and ready for deployment.

V. PROPOSED SYSTEM

The proposed system, **Visioneer**, is a real-time AI-based assistive solution designed to help visually impaired individuals navigate their environment safely and independently. The system leverages computer vision and audio output technologies to detect obstacles and objects in the user's surroundings and relay this information through voice cues.

A. *Architecture of Our Proposed System*

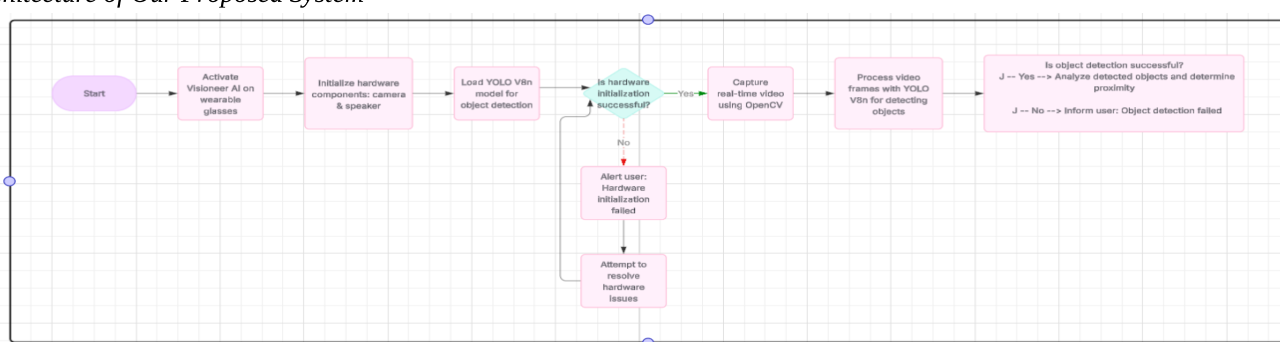


Fig.1 Architecture of System

The architecture of Visioneer is modular and consists of both hardware and software components working in tandem to provide seamless user support. The architecture can be divided into the following layers:

1) Input Layer – Camera Feed

- Hardware: A mini wireless camera mounted on glasses or a wearable frame.
- Function: Captures real-time video of the user's surroundings and transmits it wirelessly to a processing unit (laptop or embedded computer).

2) Processing Layer – Object Detection and Distance Estimation

- Core Algorithm: YOLOv8 (You Only Look Once, Version 8)
- Platform: Runs on a Python-based system, using GPU acceleration where available.

3) Output Layer – Audio Feedback

- Hardware: A wireless mini speaker or earphones.
- Function:
  - Converts object detection results into simple, understandable voice messages (e.g., "Wall ahead, 2 meters").
  - Uses Text-to-Speech (TTS) engine for dynamic audio output.
  - Prioritizes objects based on proximity and relevance (e.g., moving objects or obstacles).

4) User Interface Layer

- Platform: A lightweight UI on the processing unit (laptop or Raspberry Pi).
- Features:
  - Displays object detection feed for testing and debugging.
  - Allows toggling of voice modes (e.g., verbose, brief).
  - Emergency stop or manual override feature.
  - Voice activation or button interface for control by users.

5) Feedback and Control Loop

- Real-time feedback loop ensures continuous object detection and voice notification.
- System recalibrates on-the-fly based on movement, camera angle, and new objects entering the frame.

REFERENCES

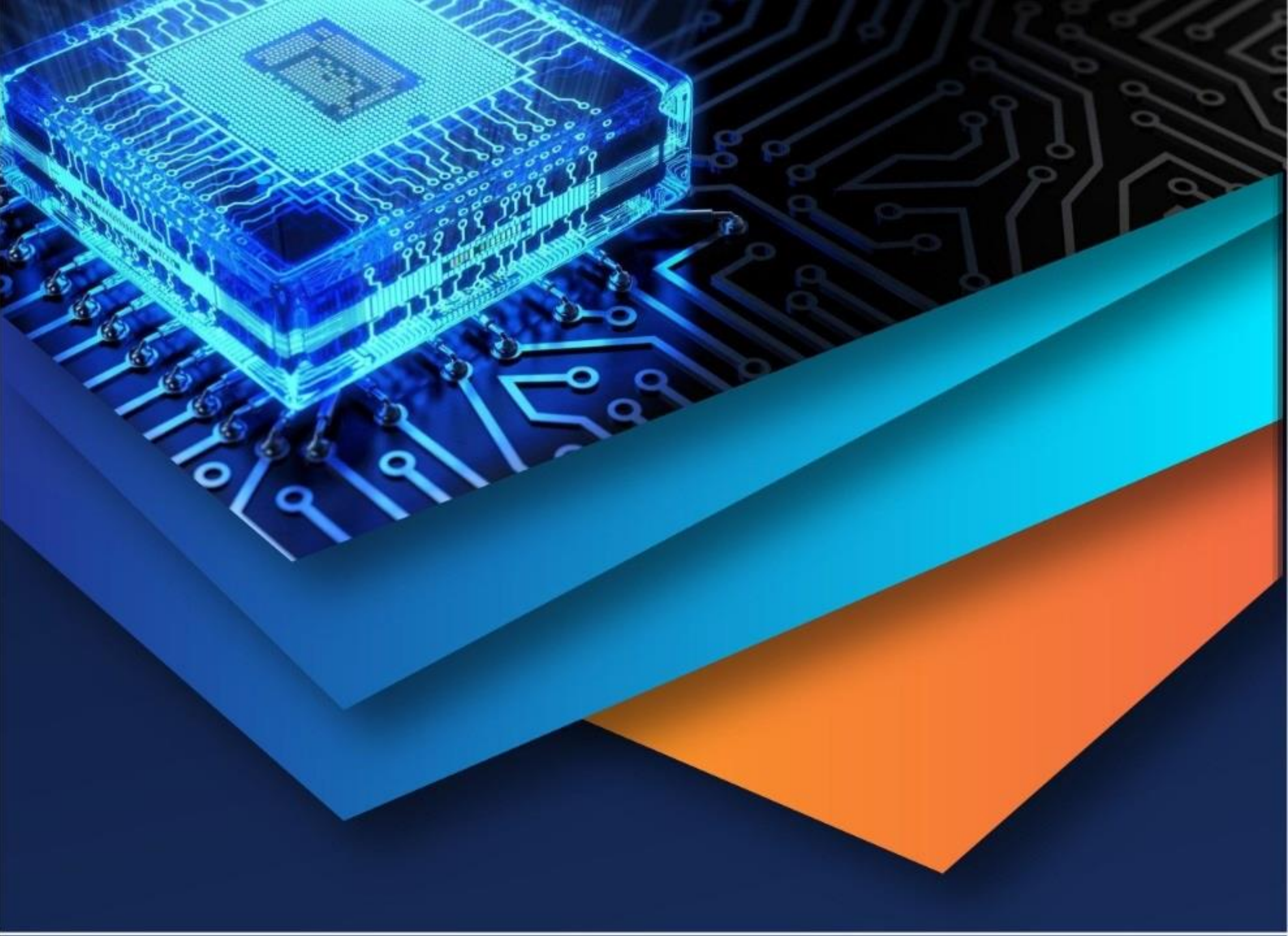
[1] Mrs. R. Saranya, Mrs. M. Devaki, K. Keerthika, S. Jaswanthra Samyukhta, T. Aarthy "Smart Navigation System" College of Engineering & Technology, Madurai, Tamil Nadu, India Vol: 04/Issue:08, August-2022.

[2] Mrs. V.S. Benitha.J.Sheryl, B.Mahalakshmi, "Smart Navigation System for Visually Impaired People" Mepco Schlenk Engineering College, Sivakasi, ISSN: 2455-1341

[3] Dr.S.Mary Joans, Aishwarya R, Meenu Sam, Sree Ranjane C, Subiksha S, "Artificial Intelligence based smart navigation system for blind people" Velammal Engineering College, Chennai, Tamil Nadu, India, Vol: 09/Issue:07, July-2022.

[4] Vikram Shirol, Mr, Aruna Kumar Joshi, Mr, Shreekant Jogar, Mrs Rajeshwari S G, "AI based - Navigation System for Blind Person" SKSVMACET, Laxmeshwar, Karnataka, India, ISSN: 2395-1990

[5] Gabriel Illuebe Okolo, Turke Althobaiti, Naeem Ramzan, "Smart Assistive Navigation System for Visually Impaired People" Northern Border University, Arar, Saudi Arabia. Published: 3 Jan, 2025.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY