iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76233

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 XII Dec 2025- Available at www.ijraset.com

AI-Powered Smart Wearable Glass for the Visually Impaired with Voice-based Navigation and Scene Awareness

Dr. Gauri Ghule[1], Manasvi N. Pawar[2], Yash G. Patil[3], Pushpraj R. Patil[4], Dhanashri G. Pawar[5]
Department of Electronics and Telecommunication, Vishwakarma Institute of Technology, Pune, India

Abstract: Visually impaired people often face difficulties in navigation and understanding their surroundings independently. Traditional tools like white canes and guide dogs lack in providing information. With recent advancements in artificial intelligence and wearable technology, more effective solutions are becoming possible. In this project, we introduce a pair of AI-powered smart glasses intended to assist visually impaired individuals that deliver real-time audio descriptions of the environment. The system integrates an ESP32-S3 microcontroller with an OV2640 camera to capture images. These images are sent to a server where AI models (like gpt-4o for image captioning and gtts for speech) analyze the scene and turn it into spoken descriptions. The audio is then delivered wirelessly through a Bluetooth earpiece. Inspired by previous work such as Vis Buddy and other smart glasses, our design focuses on being low-cost, easy to use, and practical in real life. The glasses respond to button presses, making them interactive and user-friendly. User testing shows that this system helps users become more independent and aware of their surroundings. Overall, this work aims to make daily life easier and more accessible for visually impaired people using modern, affordable technology.

Keywords: Smart Glasses, ESP32-S3, Image Captioning, GPT-4o, Text-to-Speech, Bluetooth Audio, Wearable Devices, smart human integration, computer vision, enabling technology

I. INTRODUCTION

Individuals unable to see, as the blind or visually impaired, tend to encounter enormous difficulties in executing everyday tasks like mobility independently, identifying objects, and distinguishing between individuals. A considerable portion of the world's population experiences moderate to severe visual impairment, and most of them use assistive aids to accomplish basic mobility functions [1], [2]. Visual impairment can be caused by some of the underlying reasons, like corneal opacities, cataracts, glaucoma, and uncorrected refractive errors, and impacts millions of people across the globe [3]. Over time, traditional aids like walking sticks and guide dogs have been the conventional solutions for mobility support. Though these aids are crucial in navigation, they do also have major limitations — e.g., white canes only respond to immediate obstacles in physical reach and cannot perceive elevated or far-away objects [6], and guide dogs necessitate high maintenance costs and cannot provide in-depth information about the environment or specific objects [4]. Such limitations tend to make visually impaired people dependent on caretakers for directions, something that is not always feasible, leading to limited mobility, low confidence, and isolation [4].

Advances in wearable technology, artificial intelligence, and computer vision have made it possible to develop novel assistive devices [2], [3]. Cameras with tiny camera modules like the Pi Camera Modules and ESP32-CAM, combined with AI software, can take photographs, analyze them, and provide real-time audio descriptions to facilitate independent mobility and environmental perception [2], [8]. Building on this idea, our research seeks to develop a compact, low-cost, and transportable AI-driven smart glasses system.

The envisioned smart glasses are also specifically made to be friendly for use, and they are triggered by a simple push- button, which is a characteristic that is in conformity with other low-interaction aid devices. Upon triggering, the camera captures the scene in front, the embedded processor identifies the principal object or scene, and this is converted into speech for immediate replay to the user [2], [6]. This efficient method gives the user immediate perception of the environment without the constant need for a guide, enabling the user to carry out object recognition and scene identification tasks more independently [3], [6].

In contrast to intricate multi-module setups incorporating haptic feedback, diverse sensors, or sophisticated large vision–language models [5], [16], [17], our approach emphasises minimal hardware, affordability, and core functionalities [13], [14]. This focus ensures wider adoption, especially in low-resource settings, and offers a practical pathway for real-world deployment [15].

Our paper proposes an "AI-Powered Smart Glasses for Visually Impaired People" solution that overcomes the limitations of conventional aids without incurring the sophistication of expensive research prototypes. The single- button access facilitates a low learning curve for the user [11], [15], and the real-time audio description fills the blind spot between auditory and visual perception for the user. This study makes an addition to the emerging body of wearable assistive technology through the provision of a well-targeted, cost-effective, and practicable solution that has high potential for widespread adoption. In the remainder of this paper, we present the literature review, hardware and software architecture, discuss the implementation process, and analyze the system's performance in realistic usage scenarios [14], [15].

II. LITERATURE SURVEY

The history of assistive technology for the visually impaired has evolved considerably in recent decades, driven by rapid advancements in artificial intelligence (AI), machine learning (ML), and embedded systems. Traditional aids such as white canes and guide dogs remain important, but they cannot provide fine-grained scene interpretation or contextual awareness. This has spurred the development of wearable AI-based solutions designed to enhance mobility, accessibility, and independence.

Early innovations demonstrated the feasibility of integrating smart glasses with onboard processors and cameras to capture real-time images, process them locally or via cloud resources, and provide spoken feedback. Geoggarous et al. [1] developed AIris, a wearable assistive device capable of face recognition, text reading, environmental identification, and scene analysis, balancing local and cloud processing for performance efficiency. Similarly, Patel et al. [2] proposed an affordable design built with a Raspberry Pi, a camera module, and pre-trained deep learning models for offline object recognition and OCR, enabling deployment in areas with poor connectivity.

Mukhiddinov and Cho [3] enhanced low-light adaptability in smart glasses by integrating image enhancement modules before object detection, while VisBuddy [4] extended functionality with voice-controlled commands, OCR with column detection, and online content retrieval. Cabrera et al.

[5] combined YOLO-based detection with LLM-driven reasoning to provide navigation instructions via haptic feedback, avoiding audio dependence. Gillani et al. [6] emphasized personalization by enabling recognition database updates and integrating proximity sensors for safety alerts. Long et al. [7] provided a bibliometric review showing the increasing shift toward multimodal sensing and hybrid edge– cloud architectures.

In mobile-based approaches, Sandnes et al. [8] developed SceneRecog, a smartphone application leveraging EfficientNet-Lite for scene classification, demonstrating the feasibility of AI navigation aids without specialized hardware. For indoor contexts, Baskar et al. [9] built a Raspberry Pi– and Arduino-based obstacle detection system combining ultrasonic and PIR sensors with vibration motors and audio output, achieving reliable detection of static and moving objects within a 35 cm range.

Some works extend beyond navigation to educational and cognitive domains. Liu et al. [10] examined the effect of data- driven feedback on programming learning for visually impaired students, finding reduced cognitive load and improved performance with real-time analytics. Porquis et al.

[11] designed the ABBI wearable for visually impaired children, using auditory beacons to enhance spatial cognition. Thapa et al. [12] developed a Raspberry Pi–based Electronic Travel Assistant integrating ultrasonic sensors, GPS, and camera modules for outdoor navigation.

Esra Hassan [13] developed low-cost smart glasses for visually impaired students, focusing on text recognition using MATLAB/Simulink and Tesseract OCR. The glasses also have a modular design, allowing for expansion to other assistive tasks. Patil et al. [14] created an IoT-based navigation system that combines GPS, ultrasonic sensors, and voice guidance for real-time obstacle detection. Makanyadevi et al. [15] introduced an AI-powered dual-assistive solution, pairing YOLOv8-enabled smart glasses with ultrasonic- equipped shoes for 360° hazard awareness. Islam and Sadi [16] built Generative AI-assisted Standalone Wearable Reading Glasses, integrating PaddleOCR, an offline fine- tuned LLM, Whisper ASR, and Festival TTS into a $50 self-contained device that operates entirely offline.

Advanced object recognition techniques are also emerging. Reddy et al. [17] created a deep-learning–based Smart Blind Assistant for obstacle detection, while Sudarshan et al. [18] explored Vision Transformer (ViT) architectures for enhanced detection accuracy. Sudhakar et al.

[19] implemented a ResNet50–LSTM–based image captioning system, achieving 73% accuracy, outperforming VGG16-based baselines. Sindhwani and Srikanthan [20] addressed system-level efficiency with an automated, application-specific OS optimisation framework, reducing CPU overhead by up to 25%.

Beyond core vision tasks, Yadav and Singh [21] examined PEGASIS-based protocols in wireless sensor networks for energy-efficient communication in distributed sensing, relevant to wearable networks. Rasheed et al. [22] reviewed AI-driven AR glasses for multilingual communication, proposing a three-tier edge–cloud architecture for translation. Lu et al. [23] applied bootstrap-based multiple hypothesis testing for robust evaluation of speech enhancement algorithms, beneficial for improving wearable audio feedback.

Researchers surveyed low-light object detection methods that use deep learning and IoT, emphasising affordability and sensor fusion [24]. Wang et al. [25] developed an OpenCV-based framework for optimising deep learning with big data, enhancing image processing speed through hardware-aware parallelism. Agrawal et al. [26] also created a pipeline for converting speech to text and summarizing text, allowing for concise audio output from long-form content.

A. Synthesis and Gaps

Across these studies, notable trends include:

1) Integration of multimodal feedback (audio, haptics, tactile) for adaptable accessibility.
2) Increasing use of contextual AI (VLMs, LLMs) for richer environmental understanding.
3) Persistent focus on low-light performance and robust perception in dynamic environments.
4) Adoption of hybrid edge–cloud architectures balancing latency, power, and accuracy.

B. However, Several Gaps Remain

1) Privacy-preserving on-device inference for large models is underdeveloped.
2) Lack of standardised testing with visually impaired participants in real-world scenarios.
3) Need for low-cognitive-load UX designs that reduce mental effort during navigation.
4) Fail-safe navigation protocols to handle unexpected hardware or software failures.

The proposed project addresses these gaps by combining low-light enhancement, efficient vision–language reasoning, privacy-conscious offline inference, and multimodal feedback into a cost-effective wearable platform that ensures real-time responsiveness and scalability.

III. DESIGN METHODOLOGY

A. Methodology

The projected device aims to develop AI and cloud-based smart glasses that assist visually impaired individuals by providing descriptions of surroundings, objects, and situations using real-time image processing and auditory information feedback [1]–[3], [6], [15]. The methodology involves several steps. The first step begins with flowcharting and system design [2], [6]. It is distributed into several parts as setting up the microcontroller for capturing and sending the image to server with sufficient amount of detailing [2], [6], setting up a server which receives and sends images [16], implementing the ML model for image processing considering all the factors as accuracy, time constraint and adequate functionality with it [4], [8], [18], receiving and playing the audio using the A2DP technology [23]. Suitable hardware components, including the custom ESP32-S3 microcontroller, OV2640 camera, battery power supply, push button for image capture, and Bluetooth earpiece for audio output, are selected along with compatible software such as gTTS, openAI, Arduino IDE, C++ and Python [5], [7], [9], [20]. In the hardware setup, the ESP32-S3 is connected to the OV2640 camera to capture images and transmit them over Wi-Fi, while the Bluetooth earpiece is used for delivering audio output [13], [15].The software development includes programming of the ESP32- S3 to capture and transmit images upon button press to a cloud-based server, where AI models process the images to generate scene descriptions [17], [18], [19], [24]. These descriptions are then converted to speech using gTTS and sent back to the ESP32-S3 for playback through the earpiece using Bluetooth [26]. Testing and iteration involve verifying the proper functioning of image capture, server communication, object recognition, and audio playback, along with user trials involving visually impaired participants to gather sufficient sample-spaced feedback on accurate usability and functioning [1], [4], [7], [14]. In the last phase of optimization and refining, the task is to improve the latency, code optimization on the microcontroller, minimize crashing out and approximated errors, improve comfort, and mount all the pieces of hardware on sturdy, comfortable goggles or glasses [6], [11], [12], [15]. The device acts as an improvement in senses and intuition by leveraging the combination of technology and hearing ability [1], [3], [6], [15], [22].

B. *Algorithm*

1) Step 1: System Initialization
- Initialize the ESP32-S3 microcontroller.
- Connect and configure the OV2640 camera module.
- Set up Bluetooth or Wi-Fi module.
- Initialize push button and audio output system (e.g., speaker or Bluetooth earpiece).
- Check available PSRAM for memory allocation.

2) Step 2: Image Capture
- Wait for user input (e.g., push button press).
- On trigger, capture an image using the OV2640 camera.
- Compress the image to reduce data size (if needed).

3) Step 3: Send Image to AI Processing Unit
- Establish a connection with the AI processing unit (local server or cloud).
- Transmit the captured image via Wi-Fi or Bluetooth to the server.

4) Step 4: AI Image Processing

(On external processor using Python + Flask/FastAPI)
- Receive image on the server.
- Load and run the BLIP model or other image captioning/object detection model.
- Extract a meaningful description of the scene or object (e.g., “A red car parked on the left side of the road”).

5) Step 5: Convert Text to Speech
- Use gTTS or any TTS engine to convert the description into speech.
- Generate an audio file or real-time stream.

6) Step 6: Send Audio Feedback
- Transmit the audio back to the ESP32.

C. *Flowchart*

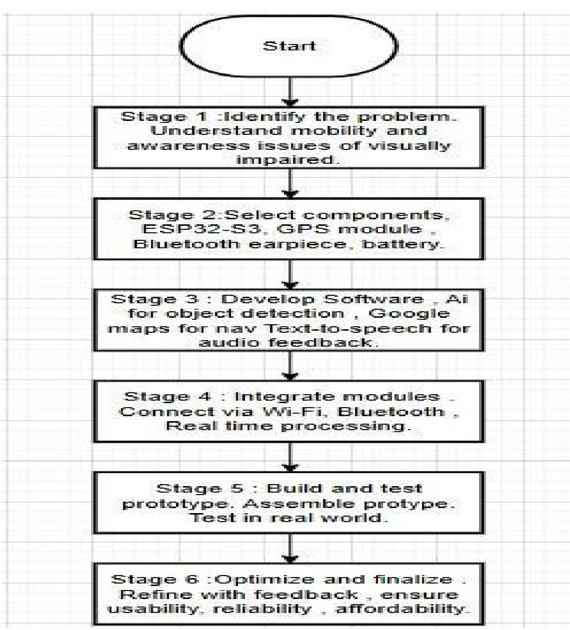


Fig. 1. Flowchart

D. Block Diagram

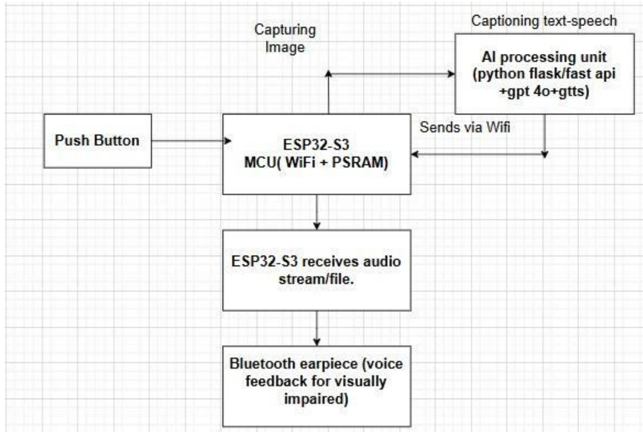


Fig. 2. Block Diagram

E. Circuit Diagram

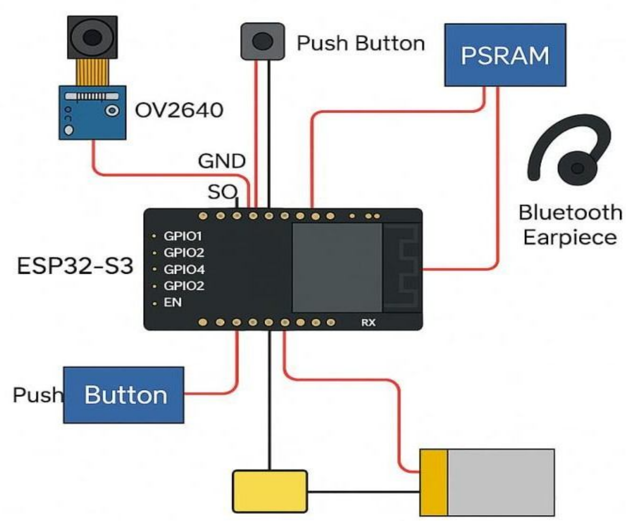


Fig. 3. Circuit Diagram

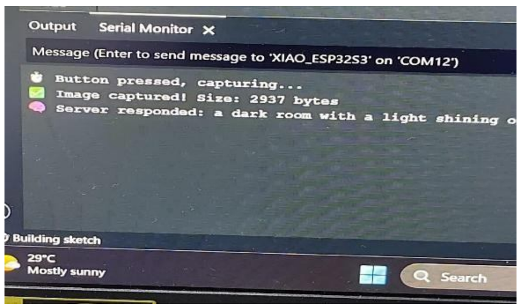
F. Experimental Arrangement

IV. RESULT & DISCUSSION

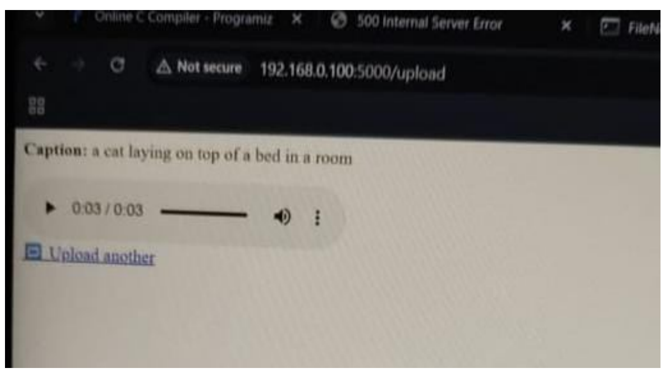
After testing the Smart Wearable Glass, we observed the following:

1) Obstacle Detection: The sensors properly detected objects such as walls, poles, or furniture in the user's path. This prevented accidents.
2) Voice Feedback: The voice feedback was clear and assisted the user in real-time, thus boosting their confidence while walking.
3) Scene Recognition: The AI system was able to recognize environments such as rooms, streets, or parks, and notify the user accordingly.
4) User Testing: When tested by the visually impaired, they could move more independently and safely with reduced fear. In general, the project functioned well in real-life situations and was an assistive tool. Although still a prototype, it holds much promise for use in the real world after additional refinement.

A. *Serial Monitor Output Showing System Response*



B. *Web Interface Displaying the Generated Caption*



The functionality of the system was tested using both hardware-level serial output and a web-based interface. As shown in Fig. 1, the ESP32-S3 is able to capture an image upon user request and get a significant description from the server. In Fig. 2, the web interface based on Flask displays the generated caption together with audio synthesized through gTTS, rounding off the real-time image-to-speech loop.

V. CONCLUSION

The project aims to bridge the disconnect between visual impairment and independent mobility by the use of modern technologies like Artificial Intelligence, sensors, and voice interaction. Through the integration of obstacle detection, scene recognition, and voice guidance, the Smart Wearable Glass allows visually impaired users to perceive their environment in real-time. Our prototype was able to demonstrate successfully that visually impaired individuals can receive environmental feedback in the form of sound, which tells them about obstacles, objects, or locations around them. The system is an artificial visual guide, giving users verbal information and notifications.

A. Major technologies Employed

1) Ultrasonic sensors to measure distance and sense obstacles.

2) Camera module and AI vision processing algorithms for detecting objects and recognizing scenes.

3) Google Text-to-Speech (gTTS) to produce voice output. 4] Microcontroller ESP32-S3 for control and processing.

Overall, the project shows that wearable assistive technology has the potential to significantly increase the safety, confidence, and mobility of visually impaired users, paving the way for increased independence and integration into society.

REFERENCES

[1] Evangelos Geoggarous, Diongsysia Danai Brilli, Konstantina Nikita, Nikos Melanitis, Stefania Tsilivaki, “AIris: An AI-powered Wearable Assistive Device for the Visually Impaired.” August 2024

[2] Smit Patel, Nirav Satani, Sandip Patel, “AI-Powered Glasses for Visually Impaired Persons.” International Journal of Recent Technology and Engineering (IJRTE), July 2020.

[3] Mukhriddin Mukhiddinov, Jinsoo Cho, “Smart Glass System Using Deep Learning for the Blind and Visually Impaired”, November 2021.

[4] Shiloah Elizabeth D, Ishwarya Sivakumar, Nishaali Meenakshisundaram, Ishwarya Ramesh, Sunil Retmin Raj, “VisBuddy- A smart wearable assistant for the visually challenged.”, Jan 2022.

[5] Miguel Altamirano Cabrera, Yara Mahmoud, Luis Moreno, Muhammad Haris Khan, Dzmitry Tsetserukou, Issatay Tokmurziyev, “LLM-Glasses: GenAI-driven Glasses with Haptic Feedback for Navigation of Visually Impaired People,” August 2025.

[6] Syeda Anshrah Gillani, Mahmoud Aljawarneh, Abdul Akbar Khand, Mirza Samad Ahmed Baig, Muhammad Hamzah Siddiqui, Shahid Munir Shaha, “AI-based Wearable Vision Assistance System for the Visually Impaired: Integrating Real-Time Object Recognition and Contextual Understanding Using Large Vision-Language Models”, 2023.

[7] Liumei Long, Xiaochen Zhang, Xiaoyu Huang, Wujing Li, XingXu, Yiran Ding, “Advancements in Smart Wearable Mobility Aids for Visual Impairments: A Bibliometric Narrative Review,” 2024.

[8] Frode Eika Sandnes, Bineeth Kuriakose, Raju Shrestha, “SceneRecog: A Deep Learning Scene Recognition Model for Assisting Blind and Visually Impaired Navigate using Smartphones” IEEE International Conference on Systems, Man, and Cybernetics, 2021.

[9] V. Vijaya Baskar, Ikben Ghosh, S.Karthikeyan, R. J.Hemalatha, T.R.Thamizhvani, “An Indoor Obstacle Detector to Assist the Visually Impaired Person in Real-Time with a Navigator”, 2021 International Conference on Computational Performance Evaluation (ComPE), North- Eastern Hill University, Shillong, Meghalaya, India. Dec 1- 3, 2021.

[10] Meishan Liu, Liye Zhu, Jinming Wang, Yizhou Qian, ”The Impact of Data-Driven Feedback on Programming Learning: A Cognitive Load Perspective”, 2025, the 7th International Conference on Computer Science and Technologies in Education.

[11] Lope Ben Porquis, Sara Finocchietti, Giorgio Zini, Giulia Cappagli, Monica Gori, Gabriel Baud-Bovy, “ABBI: A Wearable Device for Improving Spatial Cognition in Visually-Impaired Children”, IEEE BIOCAS 2017.

[12] Amrita Thapa, Sutapa Debbarma, Bijoy Kumar Upadhyaya, “Design of a Raspberry Pi-Based Electronic Travel Assistant for Visually Impaired Persons”, 2022 IEEE Calcutta Conference.

[13] Esra Ali Hassan, “Smart Glasses for the Visually Impaired People”, January 2016.

[14] Anurag Patil, Abhishek Borse, Anish Borse, Ajay Phad, Dr. Simran Khiani, “A Low-Cost IoT-based Navigation Assistance for Visually Impaired Persons”, Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023).

[15] Makanyadevi K, Avinash M V, Chandhrakiran S V, Kavitha S, Bhalaram Krishna S A, Dhiyanesh S. “AI- Powered Smart Glasses and Shoes for the Visually Impaired”. Proceedings of the 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI- 2025).

[16] Md. Repon Islam, Muhammad Sheikh Sadi, “Generative AI-assisted Standalone Wearable Reading Glasses for the Visually Impaired”, 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON).

[17] Vijaya Vardan Reddy S, B.Sathyasri, J.Joselin Jeya Sheela, Malarvizhi. C Vanaja, Shyam S, “Smart Blind Assistant using Deep Learning for the Visually Impaired Users”, Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS 2023).

[18] Sudarshan S, Mohana, Ambika G, Kavitha A, Nataraj K, Sudhangowda B S, “Object Detection using Vision Transformer and Deep Learning for Computer Vision Applications”, Proceedings of the 7th International Conference on Intelligent Sustainable Systems (ICISS-2025).

[19] Sudhakar J, Viswesh Iyer V, Sree Sharmila T, “Image Caption Generation using Deep Neural Networks”, 2022 International Conference for Advancement in Technology (ICONAT).

[20] Sindhwani, Srikanthan, “Framework for Automated Application-Specific Optimization of Embedded Real-Time Operating Systems”, International Conference(ICICS 2005).

[21] R. K. Yadav, Anurag Singh, “Comparative Study of PEGASIS-based Protocols in Wireless Sensor Networks”, International Conference,2016.

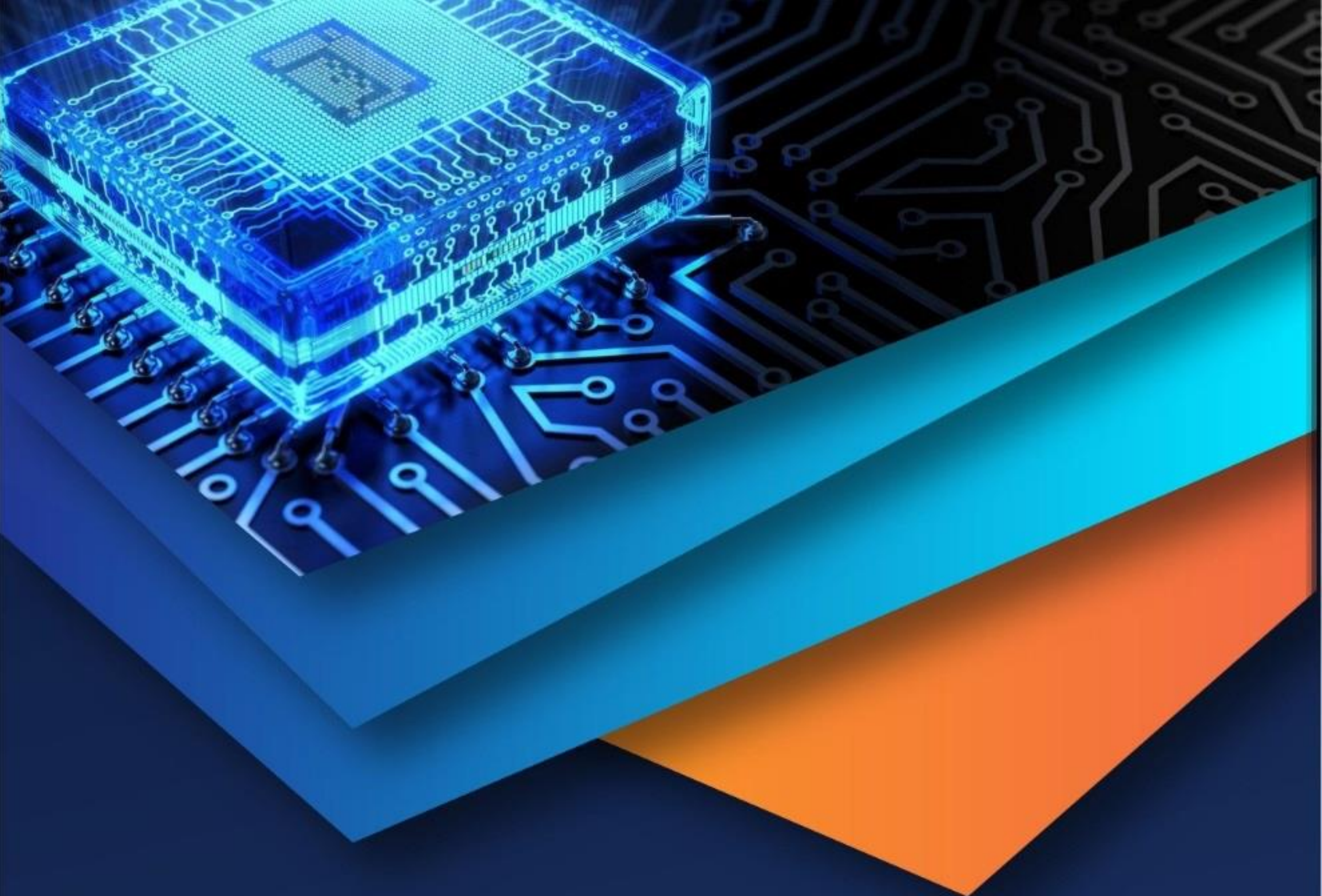
[22] Zainab Rasheed, Sameh Ghwanmeh, Aziz Almahadin, “A CRITICAL REVIEW OF AI-DRIVEN AR GLASSES FOR REAL TIME MULTILINGUAL COMMUNICATION”, 2024 Advances in Science and Engineering Technology International Conferences (ASET).

[23] Zhihua Lu, Philipp Heidenreich, Abdelhak M. Zoubir, “OBJECTIVE QUALITY ASSESMENT OF SPEECH ENHANCEMENT ALGORITHMS USING BOOTSTRAP- BASED MULTIPLE HYPOTHESIS TESTS.”, International Conference (ICASSP 2010).

[24] Sai Dikkatwar, Abhishek Mhaske, Harsh Ninawe, Harshit Khadia, Shraddha S.Gugulothu, “Recent Trends in Low Light Object Detection in Indoor Spaces for Assisting the Visually Impaired”, 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT).

[25] Weijun Wang1, Junquan Wang, Zhiqiang Zhang, Da Shi, “OpenCV Implementation of Image Processing Optimization Architecture of Deep Learning Algorithm based on Big Data Processing Technology”, Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022).

[26] Poorva Agrawal, Kashish Sharma, Keyur Dhage, Isha Sharma, Nitin Rakesh, Gagandeep Kaur, “Speech-to-Text Conversion and Text Summarization”, 2024 First International Conference on Technological Innovations and Advanced Computing (TIACOMP).



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