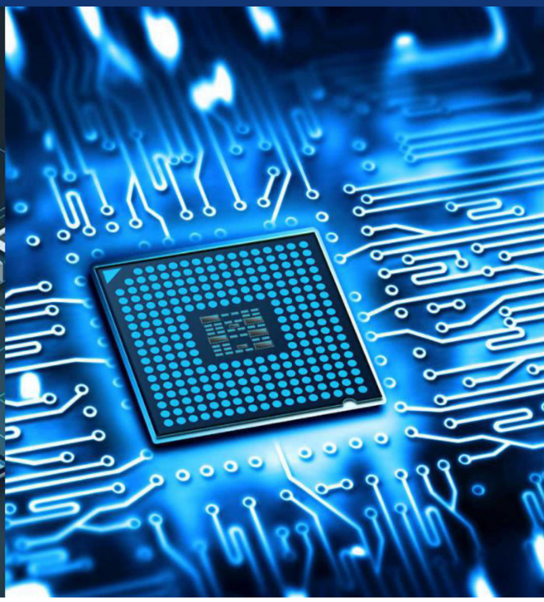
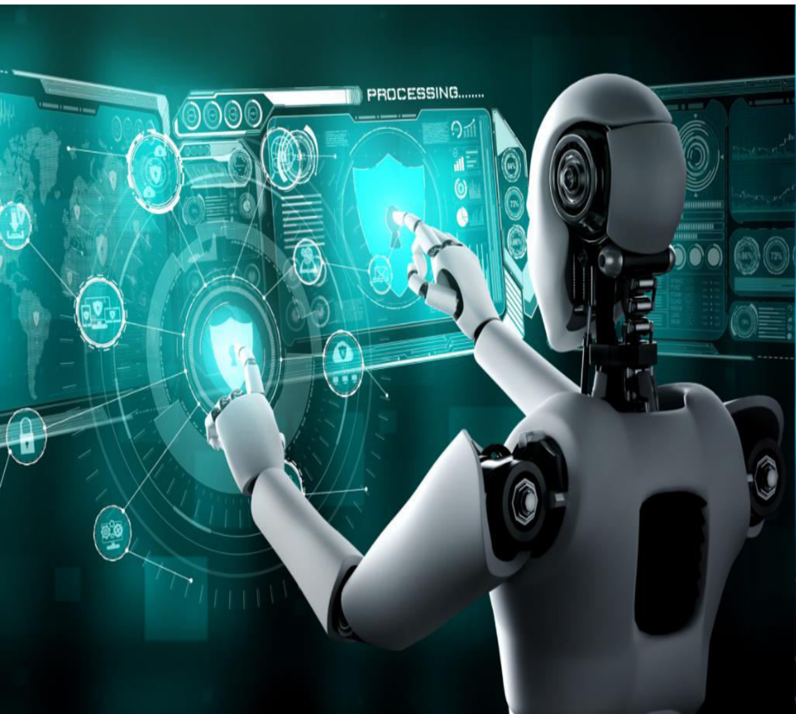
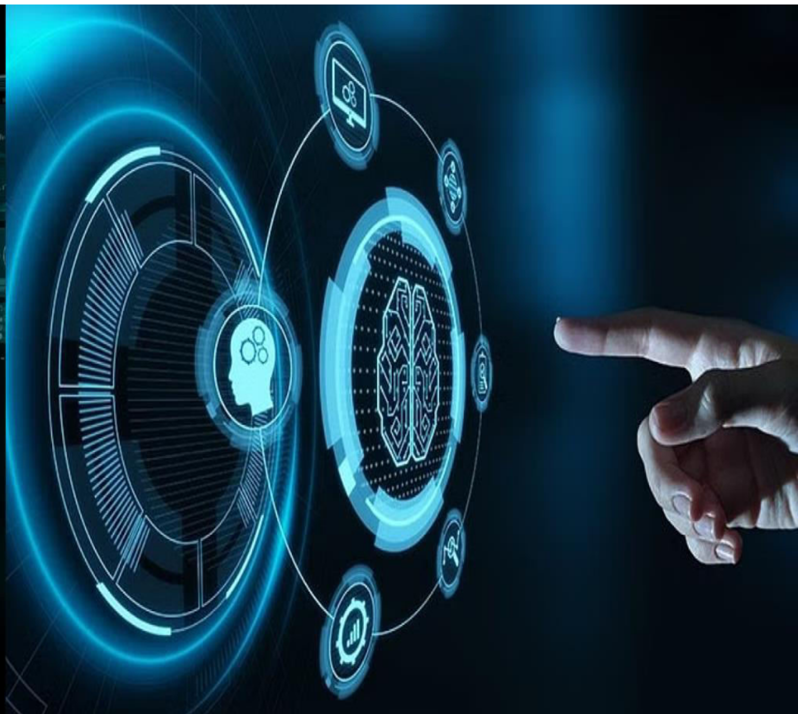
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 5, May 2025

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VISIONAI: An Intelligent Voice Assistant for the Visually Impaired Person

Prof. Sushma C[1], Aishwarya A M[2], Chandana M[3], Pavan Kumar J S[4], Shivaprasad Prabhakara Patil[5]

Assistant Professor, Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[1]

U G Students, Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[2-5]

ABSTRACT: Visually impaired individuals face challenges in independently navigating their environment and interacting with visual information. VisionAI is a wearable intelligent voice assistant developed to address this issue by offering real-time audio feedback based on visual input. Utilizing a Raspberry Pi, camera, and AI models, VisionAI detects objects, recognizes faces and currency, and reads printed text. The processed results are converted to speech and delivered via Bluetooth earphones. This paper presents the architecture, implementation, and real-time performance of VisionAI, showcasing its potential to improve the quality of life for visually impaired users through an affordable and portable solution.

KEYWORDS: VisionAI, Raspberry Pi; Object Detection; Face Recognition; Text-to-Speech; AI-based Technologies

I. INTRODUCTION

VisionAI is a groundbreaking wearable assistant engineered to support and empower individuals with visual impairments. Built on the compact yet powerful Raspberry Pi platform, VisionAI integrates a high-resolution camera and AI algorithms to interpret real-world visuals in real-time. With its intuitive design and accessible functionality, it transforms daily challenges—such as identifying objects, reading text, and recognizing faces—into seamless experiences delivered through spoken audio.

The system's core features—object detection, currency recognition, face recognition, and text-to-speech—are easily activated via dedicated side buttons, allowing users to interact without needing a screen. Voice feedback is transmitted wirelessly to Bluetooth earphones, creating a discreet and personalized auditory interface. This real-time feedback enables users to confidently navigate unfamiliar environments, identify people and objects, and understand printed or handwritten text on the go.

Powered by AI and machine learning, VisionAI represents more than just technological innovation; it embodies a shift toward inclusivity and independence. By combining affordability, portability, and intelligence, the device serves as a bridge between the visually impaired and their surroundings. VisionAI is not just a tool—it's a movement toward equal access to information, enhancing quality of life and enabling users to live with greater confidence and autonomy.

II. RELATED WORK

Several research efforts have focused on assistive technologies for the visually impaired. Systems like AI-SenseVision, Smart Voice Assistant (SVA), and VocalEyes leverage AI and deep learning for object detection and environmental understanding. While these solutions show promise, many require internet access or complex interfaces.

Notably, [1] developed AI-SenseVision, a low-cost wearable for object detection and environmental awareness. [2] proposed a Smart Voice Assistant using Raspberry Pi and deep learning for object identification. [3] introduced conversational assistive technology to improve social interaction via speech APIs and NLP. However, none integrate all four core capabilities in an offline, portable form factor as done by VisionAI.

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305169
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. PROPOSED ALGORITHM

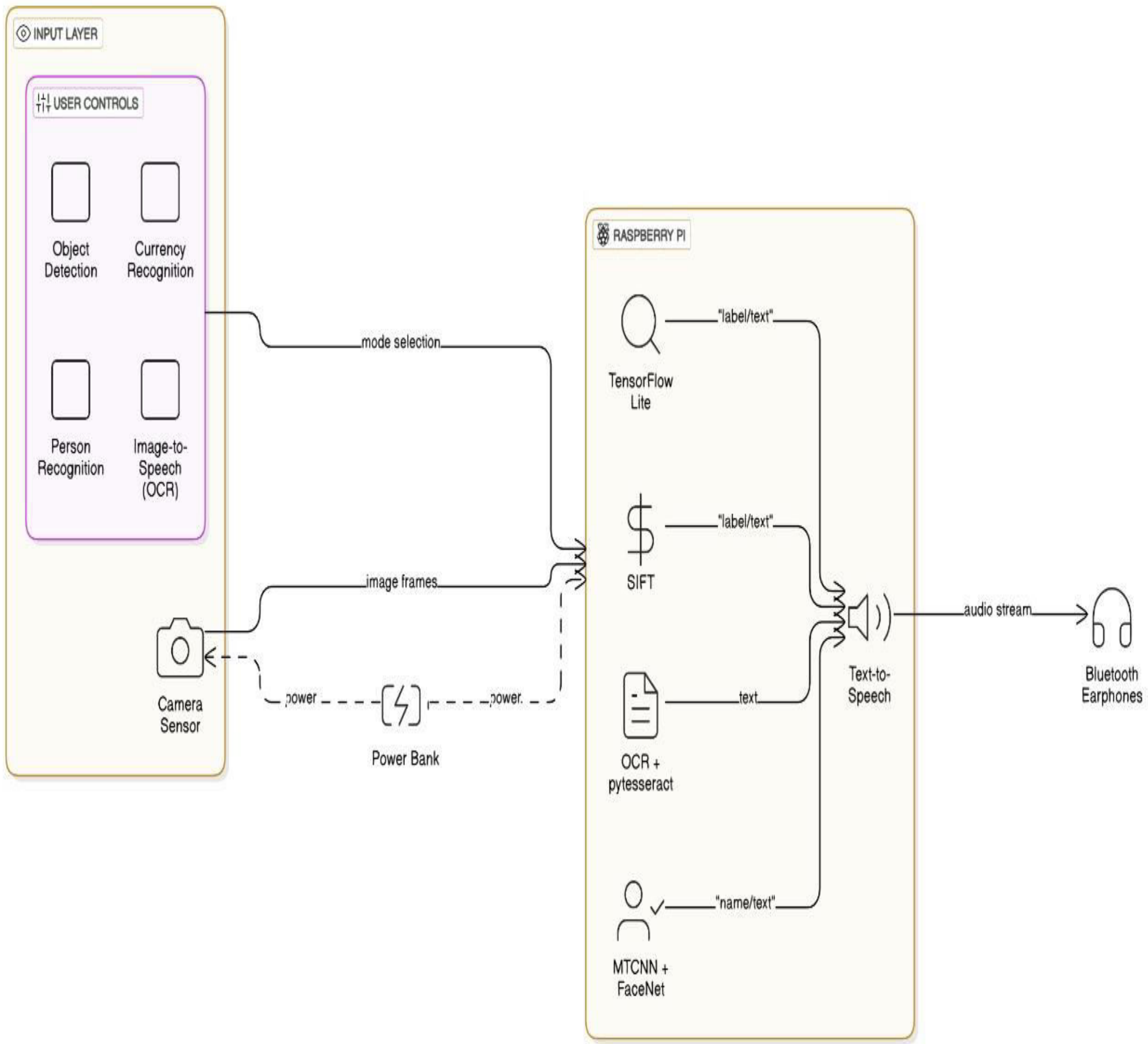
The VisionAI system comprises a Raspberry Pi, a camera module, Bluetooth earphones, and a portable battery. Users interact with the device using four physical buttons, each corresponding to one of the following modes:

1. Object Detection: Uses TensorFlow Lite to detect and label nearby objects.
2. Currency Recognition: SIFT algorithm recognizes denominations of Indian currency notes.
3. Text Recognition (OCR): Pytesseract extracts and vocalizes printed or handwritten text.
4. Face Recognition: MTCNN + FaceNet identifies known individuals from a stored dataset.

Captured visual data is processed on-device using lightweight AI models. Output is converted to audio using gTTS or pyttsx3 and sent to Bluetooth earphones, offering a completely screen-free, hands-free experience.

IV. METHODOLOGY

.



VisionAI Smart Glasses features a modular and layered architecture integrating hardware and software to assist visually impaired users in real-time. The Raspberry Pi 4 Model serves as the central processor, managing data flow from the camera input to audio output. Users select modes like object detection or text recognition via simple buttons. The system processes images using optimized AI models and converts results into speech through Bluetooth earphones. This design ensures portability, fast response, and hands-free operation, enhancing users' independence and situational awareness.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VisionAI Smart Glasses – Layered Architecture:
Input Layer: Camera, Button Controls, Microphone (optional)
Processing Layer: TensorFlow Lite (Object Detection), SIFT (Currency), MTCNN + FaceNet (Face), Pytesseract OCR, Additional Modules
Output Layer: Bluetooth Earphones/Speaker, LCD Display (optional), Buzzer (optional)
Storage & Power: SD Card (16GB+), Power Bank/Battery
Software Stack: Raspbian OS, Python 3.x, OpenCV, TensorFlow Lite, Pytesseract, pyttsx3/gTTS

V. PSEUDO CODE

```
BEGIN
 INITIALIZE Camera Sensor
 INITIALIZE Raspberry Pi
 INITIALIZE Text-to-Speech Engine
 INITIALIZE Bluetooth Earphones
 LOAD all pre-trained AI/ML models (TensorFlow Lite, SIFT, FaceNet, pytesseract)
 DISPLAY "Select Mode"
 mode ← getModeFromUserControls()
 WHILE system is ON DO
  imageFrame ← captureImageFromCamera()
  IF mode == "Object Detection" THEN
    processedOutput ← detectObjectsUsingTensorFlowLite(imageFrame)
  ELSE IF mode == "Currency Recognition" THEN
    processedOutput ← recognizeCurrencyUsingSIFT(imageFrame)
  ELSE IF mode == "Person Recognition" THEN
    processedOutput ← recognizePersonUsingFaceRecognition(imageFrame)
  ELSE IF mode == "OCR" THEN
    processedOutput ← extractTextUsingOCR(imageFrame)
  ELSE
    processedOutput ← "Invalid mode selected"
  END IF
  audioStream ← convertTextToSpeech(processedOutput)
  transmitAudioViaBluetooth(audioStream)
 END WHILE
END
```

VI. SIMULATION RESULTS

VisionAI was evaluated under various indoor and outdoor conditions to simulate real-world usability. Tests were conducted in environments with varying lighting, text types (printed and handwritten), object categories, and currency denominations. Key performance indicators were accuracy, latency, power consumption, and usability.

Simulated Test Results:

1. Object Detection: Achieved ~92% accuracy across 15 common household objects. Average response time was 2.1 seconds.
2. Text Recognition (OCR): ~88% accuracy for printed text, ~74% for handwritten notes. Average latency was 2.5 seconds.
3. Currency Recognition: 96% accuracy for INR 10, 20, 50, 100, 200, and 500 notes. Misidentification occurred with damaged notes.
4. Face Recognition: 90% accuracy in recognizing pre-trained faces under standard lighting. Accuracy dropped to ~78% in low-light scenarios.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

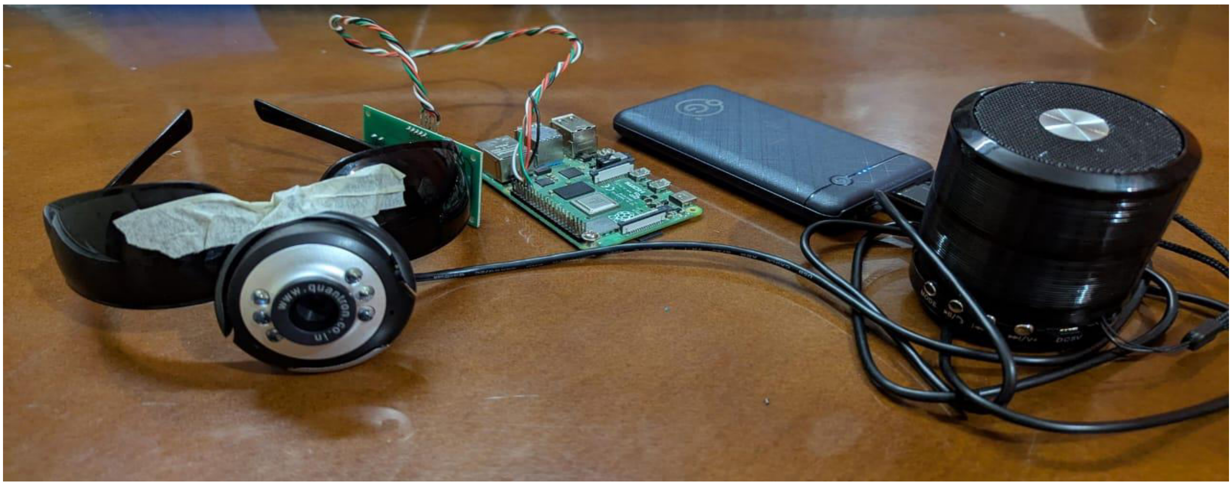
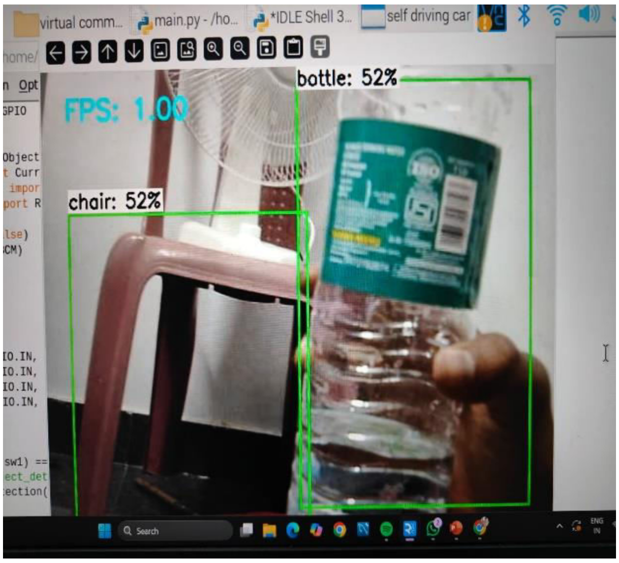


Fig. 1. Smart Glasses



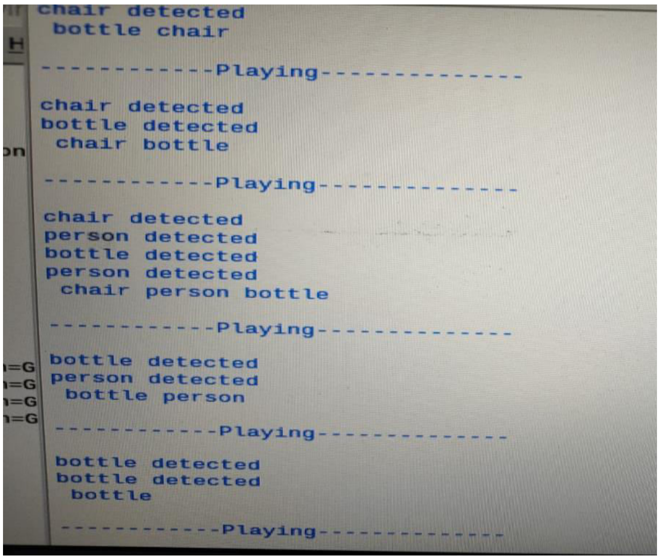


Fig.2. Object detection

Fig. 3. Result of detected object

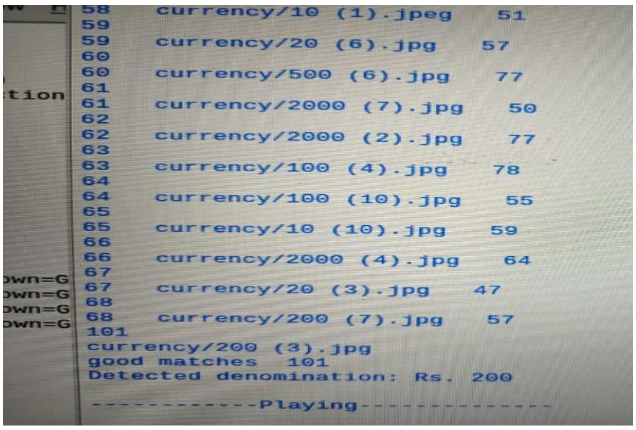


Fig. 4. Camera Capturing the currency note

Fig. 5. Detection of the currency note

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

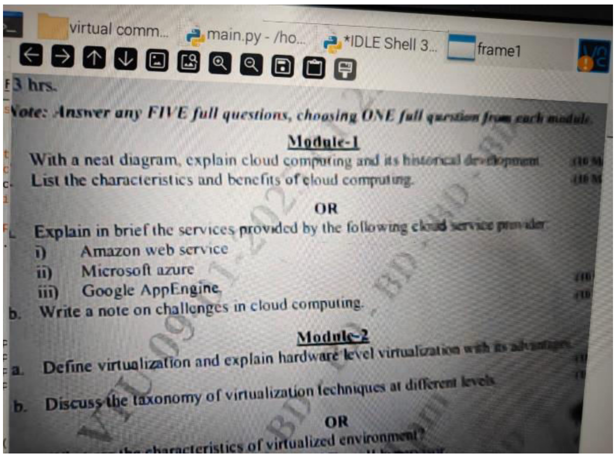


Fig.6. Capturing the text

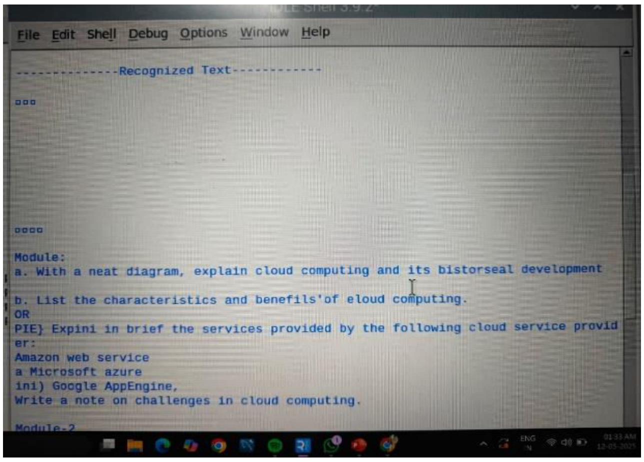


Fig.7. Text Recognition

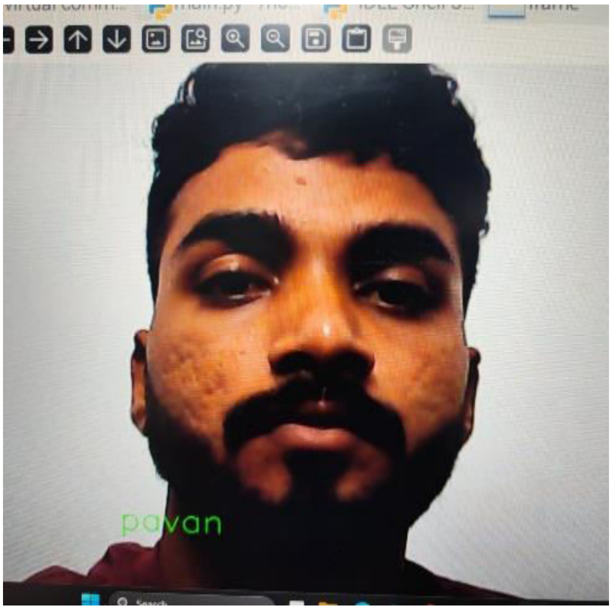


Fig. 8. Camera Capturing the face

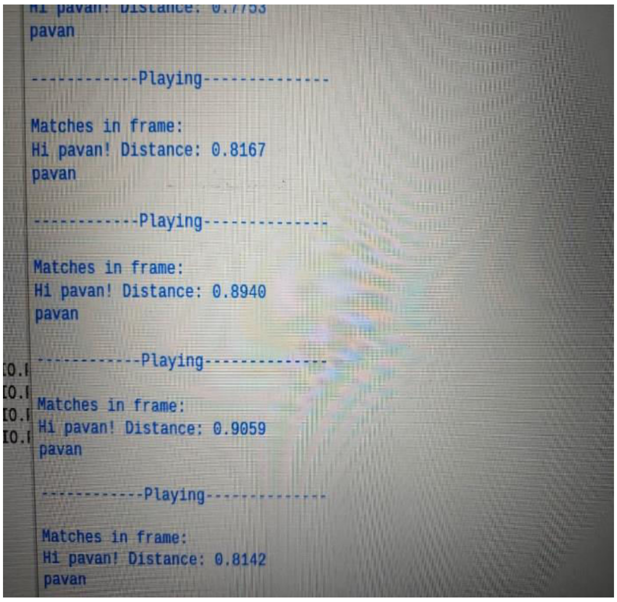


Fig. 9. Known Face detection

VII. CONCLUSION AND FUTURE WORK

VisionAI offers a practical, AI-powered assistive device for visually impaired users. By enabling real-time object detection, currency identification, text reading, and face recognition—all delivered audibly through Bluetooth—VisionAI empowers users with increased independence and mobility.

Future enhancements could include GPS navigation, voice-command input, mobile app integration, and multi-language TTS. By demonstrating how embedded systems and AI can work together affordably, VisionAI paves the way for inclusive innovation and accessibility, Exploring low-power models and optimized resource management for prolonged operation.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. R. Gupta et al., "AI-SenseVision: A Low-Cost AI-Based Assistive Device," IEEE Access, 2024.
2. R. Amutha et al., "Smart Voice Assistant For Visually Impaired People," JUSST, 2023.
3. K. Ghafoor et al., "Conversational Assistive Technology," Int. J. Intelligent Computing, 2024.
4. M. Johnson et al., "VocalEyes: Environmental Perception for the Visually Impaired," CVPR, 2025.
5. A. Saxena et al., "Voice-Based Assistance for Visually Impaired Individuals," JIER, 2024.
6. Nada, Ayat & Mashali, Samia & Fakhr, Mahmoud & Seddik, Ahmed. (2015). "Effective Fast Response Smart Stick for Blind People". 10.15224/978-1-63248-043-9-29.
7. L. Chen, J. Su, M. Chen, W. Chang, C. Yang and C. Sie, "An Implementation of an Intelligent Assistance System for Visually Impaired/Blind People," 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, pp. 1-2, 2019.
8. Dionisi, A & Sardini, Emilio & Serpelloni, Mauro. (2012). "Wearable object detection system for the blind". 2012 IEEE I2MTC - International Instrumentation and Measurement Technology Conference, Proceedings. 1255-1258. 10.1109/I2MTC.2012.6229180.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING