IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, March 2026
Impact Factor: 8.2

Visionary BOT: An AI-Based Vision Assistant

Prof. S. B. Dhadake[1], Jaywardhan Thorat[2], Rohit Shimpi[3], Abhishek Paygude[4], Nitin Kolhe[5]
Assistant Professor, Smt. Kashibai Navale College of Engineering, Pune, India[1]
Undergraduate Students, Smt. Kashibai Navale College of Engineering, Pune, India[2-5]

Abstract: *With the rapid growth of digital work environments and prolonged screen usage, individuals increasingly face issues such as poor posture, mental fatigue, irregular routines, and reduced physical activity. Existing solutions often focus on isolated aspects of wellbeing and lack an integrated, interactive system capable of providing real-time assistance. This paper presents Visionary Bot, an AI-based smart robotic assistant designed to promote healthy work habits and improve daily productivity through continuous visual monitoring and intelligent feedback.*

The proposed system combines computer vision, deep learning, and embedded robotics to monitor user posture, detect emotional states, recognize food habits, and provide timely reminders for hydration, breaks, and routine management. Implemented using a Raspberry Pi 4B with a camera module, microphone, speaker, dual OLED displays acting as expressive robotic eyes, and MG90 servo motors for dynamic head movement, the system automatically aligns with the user. Lightweight deep learning models such as MediaPipe and CNN-based classifiers enable real-time processing on edge hardware.

Visionary Bot enhances human–computer interaction by physically tracking users, delivering voice-based feedback, and visually expressing responses, creating an engaging intelligent assistant. Experimental implementation shows the system effectively supports wellness monitoring while remaining affordable and compact, demonstrating the potential of AI-driven embedded robotic assistants in improving lifestyle management in educational, professional, and home environments..

Keywords*:* Computer Vision, Deep Learning, Embedded Systems, Human–Robot Interaction, Posture Detection, AI Assistant

I. INTRODUCTION

Rapid advancements in Artificial Intelligence (AI), embedded systems, and computer vision have enabled the development of intelligent systems capable of assisting humans in daily activities. Modern lifestyles, particularly among students and professionals, involve prolonged screen exposure and sedentary work habits, often leading to poor posture, mental fatigue, stress, and unhealthy routines. Conventional reminder applications and wearable devices provide limited contextual awareness and lack real-time interactive feedback.

To address these challenges, vision-based intelligent assistants have emerged as an effective solution for monitoring human behavior and promoting wellbeing. By integrating computer vision, deep learning, and human–computer interaction, such systems can observe user activities, analyze behavioral patterns, and provide personalized feedback automatically. Embedded AI platforms like Raspberry Pi allow these solutions to be implemented in compact, affordable, and energy-efficient hardware systems.

The proposed system, **Visionary Bot**, is an AI-powered smart wellness assistant designed to monitor user posture, detect emotional states, analyze food habits, and provide timely health reminders during long working sessions. The system uses real-time camera input, speech interaction, and lightweight deep learning models to understand user conditions and deliver corrective suggestions through voice and visual feedback. Servo-based head movement enables adaptive camera alignment with the user, improving monitoring accuracy and interaction.

Unlike traditional monitoring systems, Visionary Bot integrates physical health monitoring, emotional wellbeing analysis, and routine management into a single robotic platform. OLED displays functioning as expressive robotic eyes further enhance human–computer interaction, making the system more intuitive and engaging. The developed prototype

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, March 2026
Impact Factor: 8.2

demonstrates the feasibility of deploying AI-enabled embedded robotic assistants for improving productivity and encouraging healthier lifestyle habits in educational, professional, and home environments.

II. FEATURES

Real-Time Posture Monitoring

The system continuously captures user posture through a camera module and analyzes body alignment using computer vision techniques such as MediaPipe. Incorrect posture detection triggers corrective visual or audio feedback.

Emotion Recognition System

Facial expressions and speech inputs are analyzed using deep learning models to classify emotional states such as happy, neutral, or stressed, enabling personalized wellness suggestions.

Intelligent Food Recognition

The bot identifies food items from captured images using lightweight CNN-based models and provides basic nutritional awareness to promote healthy eating habits.

Smart Reminder and Planner Module

Automated reminders notify users about hydration, work breaks, meals, and activity resumption based on predefined schedules and system monitoring.

Servo-Based Adaptive Head Movement

Two MG90 servo motors enable dynamic head rotation, allowing automatic alignment between the camera and the user for improved tracking accuracy.

Interactive OLED Eye Display

Dual OLED displays function as robotic eyes, visually indicating system status, interaction mode, and attention feedback to enhance engagement.

Triangular Camera Mount Structure

A triangular structural frame positions the camera below the OLED eyes, ensuring stable placement and optimal viewing angle for human detection.

Voice Interaction and Audio Feedback

Integrated microphone and speaker modules provide voice-based alerts, reminders, and interactive communication with the user.

Edge AI Processing

AI inference operations are executed locally on the Raspberry Pi using optimized lightweight models, ensuring real-time response and user data privacy.

Modular and Expandable Design

The system architecture supports future integration of additional sensors, cloud connectivity, and advanced AI modules without significant redesign.

III. LITERATURE REVIEW

Recent advancements in Artificial Intelligence, Computer Vision, and Human–Computer Interaction have enabled the development of intelligent systems focused on human health monitoring and smart assistance.

Several researchers have explored **vision-based human posture estimation** using deep learning techniques. Vision-Based Human Pose Estimation via Deep Learning: A Comprehensive Survey highlights the effectiveness of CNN-based pose estimation models for detecting body alignment in real-time environments. Similarly, Development of Human Pose Recognition System Using Raspberry Pi and PoseNet demonstrated lightweight pose detection implementation on embedded platforms, proving feasibility for low-cost systems.

Research on posture monitoring systems such as Real-Time Posture Monitoring and Risk Assessment Using Computer Vision shows that continuous posture analysis can significantly reduce musculoskeletal risks during prolonged computer usage. However, most existing systems focus only on physical health monitoring.

Emotion-aware intelligent systems have also gained attention. The study Real-Time Speech Emotion Recognition using Deep Learning and Data Augmentation applied deep learning models for recognizing emotional states using speech signals, enabling adaptive human centered applications.

Food recognition and nutritional analysis using deep learning have been explored in works like Automatic Food Recognition Using Deep Convolutional Neural Networks with Self-Attention Mechanism and Nutritional Composition Analysis in Food Images: An Innovative Swin Transformer Approach, which demonstrate accurate classification of food items through CNN and transformer-based architectures.

Additionally, research in social and assistive robotics such as Systematic Review of Social Robots for Health and Wellbeing emphasizes the growing role of intelligent robots in improving lifestyle habits and user wellbeing.

Despite these advancements, existing solutions generally address **single-domain problems**, such as posture correction, emotion detection, or dietary monitoring independently. Very few systems integrate **physical health monitoring, emotional analysis, routine management, and interactive robotics** into a unified embedded platform.

The proposed **Visionary Bot** aims to bridge this gap by combining multimodal AI capabilities with an interactive robotic framework deployed on an affordable edge-computing device, enabling holistic wellbeing assistance in real-time environments.

Table 1: Software Frameworks and AI Modules Used in the Visionary Bot System

Sr. No.	**Software / Framework**	**Function / Purpose**	**Specification / Description**
1	Python Programming Language	Core programming language used for system development and AI integration	Python 3.x with libraries such as NumPy, OpenCV, and TensorFlow
2	OpenCV Library	Performs real-time image processing and computer vision operations	Used for video capture, image preprocessing, and face detection
3	MediaPipe Framework	Detects human body keypoints for posture monitoring	Lightweight pose estimation framework optimized for real-time performance
4	Convolutional Neural Network (CNN) Model	Classifies food items from captured images	Trained on food image datasets for basic food recognition
5	Speech Processing Module	Processes microphone input for voice interaction and emotion detection	Uses Python audio libraries such as PyAudio or SpeechRecognition
6	Text-to-Speech Engine	Generates audio alerts and reminders to communicate with the user	Google TTS / pyttsx3 for offline speech output
7	Raspberry Pi OS	Operating system controlling hardware and AI modules	Linux-based OS optimized for Raspberry Pi hardware
8	GPIO Control Library	Controls servo motors, OLED displays, and hardware peripherals	RPi.GPIO / GPIOZero libraries
9	Data Logging Module	Stores user activity and system logs	Local storage in CSV / JSON format

IV. HARDWARE REQUIREMENTS

The Visionary Bot system is implemented using compact and cost-effective embedded hardware components that enable real-time sensing, processing, and interaction.

1. Raspberry Pi 4 Model B

Acts as the central processing unit of the system. It performs AI model execution, image processing, sensor control, and system coordination.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, March 2026 Impact Factor: 8.2

2. Pi Camera Module
Used for capturing real-time video input for posture detection, emotion recognition, and food identification tasks.

3. MG90 Servo Motors (2 Units)
Two MG90 metal gear servo motors are used for robotic head movement, allowing automatic alignment of the camera toward the detected user for improved interaction and tracking.

4. OLED Displays (2 Units)
Dual OLED displays function as robotic eyes, enabling visual expression, status indication, and interactive feedback to enhance human–robot engagement.

5. Microphone Module
Captures voice input for speech interaction and emotion analysis based on vocal tone.

6. Speaker Module
Provides audio feedback including reminders, alerts, and system responses through synthesized speech.

7. Structural Frame with Camera Mount
A triangular front structure is designed below the OLED displays to securely position the camera, ensuring stable viewing alignment with the user.

8. Power Supply Unit
A regulated 5V power source is used to supply stable power to the Raspberry Pi, servos, and peripheral modules.

9. Connectivity Interfaces
GPIO pins, USB ports, Wi-Fi, and Bluetooth modules enable peripheral communication and optional cloud connectivity.

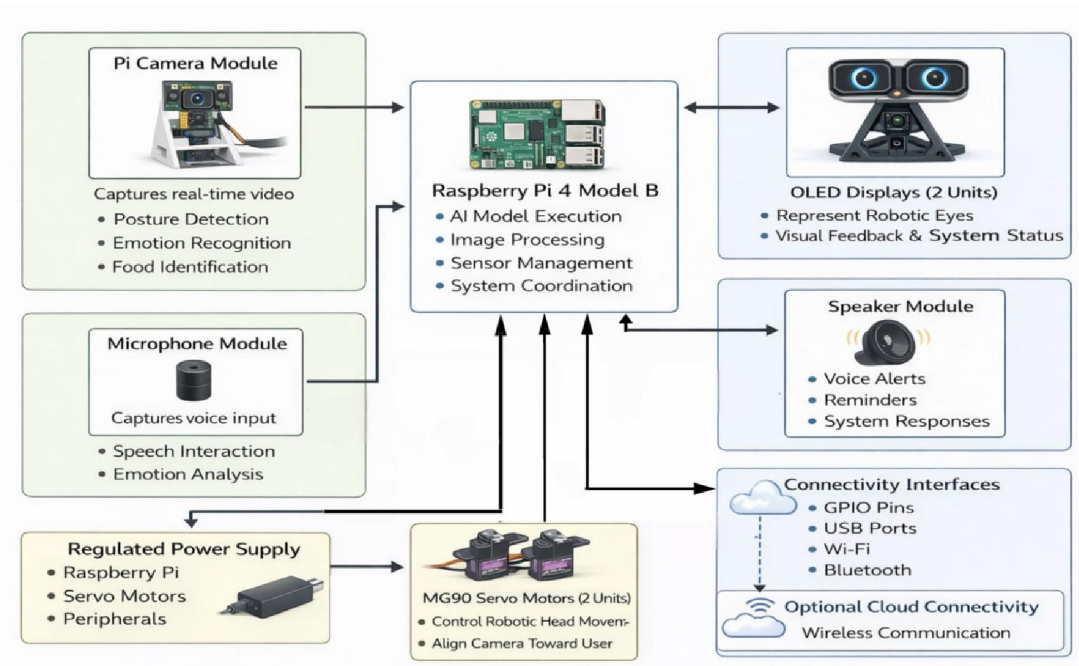

Fig. 1. System Architecture of Visionary Bot.

V. WORKING METHODOLOGY

The proposed Visionary Bot functions as an AI-enabled embedded wellness assistant that integrates computer vision, deep learning, and human–robot interaction to monitor user behavior in real time. The system is built on a Raspberry Pi

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, March 2026 Impact Factor: 8.2

platform equipped with a camera module, servo-based head movement mechanism, OLED eye displays, microphone, and speaker for interactive communication.

Initially, the system performs hardware initialization and loads the required artificial intelligence models for posture monitoring, emotion detection, and food recognition. The camera continuously captures live video input, while MG90 servo motors dynamically adjust the bot's head orientation to align the camera with the detected user, ensuring accurate observation and engagement.

The captured visual and audio data are processed using lightweight deep learning models such as MediaPipe and convolutional neural networks. These models extract body keypoints for posture analysis, facial features for emotion recognition, and object features for food identification. The processed outputs are evaluated to determine the user's physical posture condition and emotional state.

Based on the analyzed information, the decision module generates appropriate wellness responses. The system provides corrective posture alerts, emotional wellness suggestions, and nutritional guidance through voice notifications and visual expressions displayed via OLED screens acting as robotic eyes. A reminder module simultaneously tracks predefined schedules to notify users about hydration, breaks, and routine activities.

The monitoring process operates continuously in a feedback loop, enabling real-time interaction between the user and the robotic assistant. Relevant activity logs may be stored locally for performance analysis and future system enhancement. This methodology ensures intelligent automation, responsive interaction, and efficient real-time wellness monitoring within a compact embedded environment.

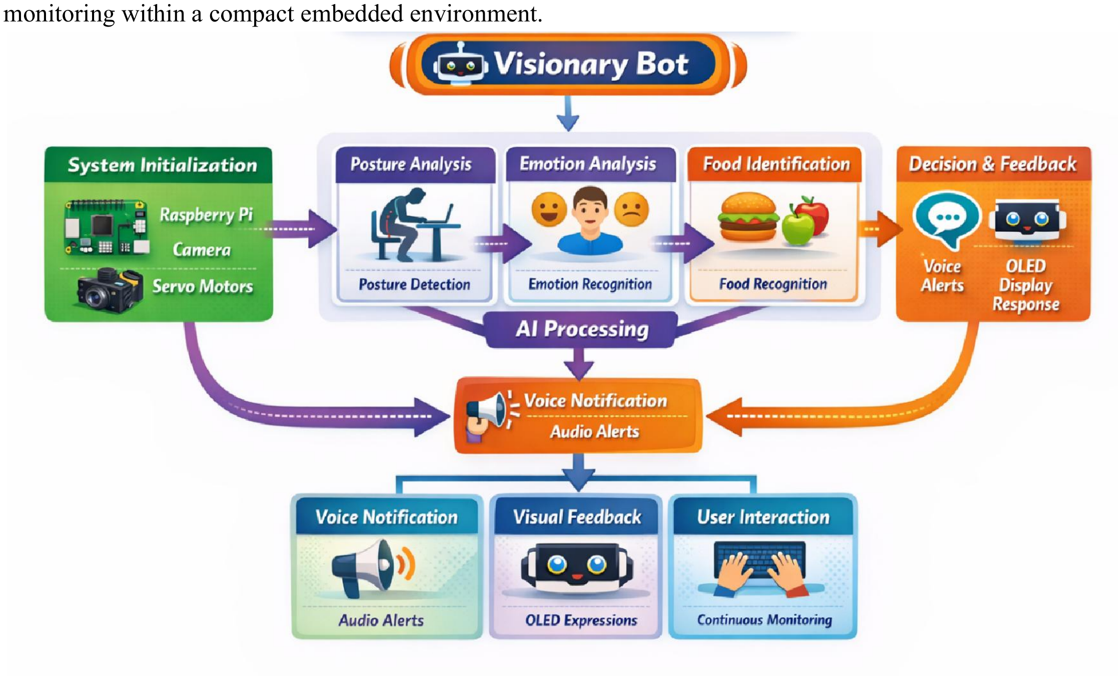

Fig. 2. Working Methodology of Visionary Bot

VI. CONCLUSION

The proposed *Visionary Bot* presents an innovative approach to integrating Artificial Intelligence with embedded robotic systems to enhance user wellbeing and productivity in modern digital environments. Prolonged screen usage and sedentary work habits often lead to poor posture, mental fatigue, and unhealthy routines. The developed system addresses these challenges by providing an intelligent and interactive assistant capable of monitoring user behavior and delivering real-time feedback.

The system combines computer vision, deep learning, and embedded hardware technologies to perform posture detection, emotion recognition, and food identification while also delivering timely reminders for hydration, breaks, and routine activities. The implementation using a Raspberry Pi 4, camera module, MG90 servo motors, OLED eye displays, microphone, and speaker demonstrates how advanced AI capabilities can be integrated into a compact and cost-effective robotic platform.

The adaptive head movement mechanism enables accurate user tracking, while the OLED displays and voice feedback enhance human–robot interaction, making the system more engaging and intuitive. The use of lightweight AI models such as MediaPipe and CNN-based classifiers ensures efficient real-time processing on edge hardware without heavy reliance on cloud computing.

Overall, the Visionary Bot prototype successfully demonstrates the feasibility of deploying AI-driven wellness assistants that promote healthier working habits, reduce physical strain, and improve daily productivity. The system highlights the potential of intelligent embedded robotic assistants for future smart environments, including homes, educational institutions, and workplaces.

Future improvements may include cloud-based analytics, personalized behavioral learning, integration of additional health sensors, and enhanced natural language interaction, enabling the system to evolve into a more advanced and fully autonomous personal wellbeing assistant.

REFERENCES

[1]. J. Lee and H. Tanaka, “Deep Learning in Food Image Recognition: Recent Advances and Applications,” *Applied Sciences*, vol. 15, no. 2, MDPI, 2025.

[2]. M. Li and P. Das, “Real-Time Speech Emotion Recognition Using Deep Learning and Data Augmentation,” *Artificial Intelligence Review*, Springer, 2025.

[3]. F. Rossi, M. Bianchi, and L. Conti, “Nutritional Composition Analysis in Food Images: An Innovative Swin Transformer Approach,” *Frontiers in Nutrition*, vol. 11, 2024.

[4]. G. Jocher, A. Chaurasia, and J. Qiu, “YOLOv8: In-Depth Analysis and Improvements,” *arXiv preprint arXiv:2404.12345*, 2024.

[5]. R. Li, Y. Zhao, and H. Sun, “Real-Time Posture Monitoring and Risk Assessment Using Computer Vision,” *IEEE Access*, 2024.

[6]. Chen, Y. Lin, and H. Zhang, “Automatic Food Recognition Using Deep Convolutional Neural Networks with Self-Attention Mechanism,” *Cognitive Computation and Systems*, Springer, 2024.

[7]. D. Y. Kim, S. Park, and J. Choi, “Systematic Review of Social Robots for Health and Wellbeing,” *ACM Transactions on Human-Robot Interaction*, vol. 13, no. 4, 2024.

[8]. L. Thomas, A. Brown, and R. Wilson, “Evaluating the Effectiveness of Apps Designed to Reduce Screen Time: A Systematic Review,” *Journal of Medical Internet Research (JMIR)*, vol. 25, 2023.

[9]. S. Zhang, X. Liu, Y. Wang, and J. Chen, “Vision-Based Human Pose Estimation via Deep Learning: A Comprehensive Survey,” *Applied Sciences*, vol. 13, no. 5, MDPI, 2023.

[10]. R. Kumar, S. Singh, and A. Verma, “Multifunctional Sitting Posture Detector Based on Face Tracking,” in *Proc. International Conference on Power, Control and Smart Energy Engineering (ICPCSEE)*, 2023.

[11]. A Patel, R. Shah, and K. Mehta, “Development of Human Pose Recognition System Using Raspberry Pi and PoseNet,” in *Proc. International Conference on Power, Control and Smart Energy Engineering (ICPCSEE)*, 2021