Kanmani AI – A Wearable Stereo Vision-Based Assistive Device for Indoor Navigation of Visually Impaired Elderly Individuals

A PROJECT PHASE I REPORT

Submitted by

ARAVINTH S	**2116221801003**
LIO GODWIN BR	**2116221801029**
VENKATESHAN T	**2116221801061**

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

RAJALAKSHMI ENGINEERING COLLEGE

(AUTONOMOUS), CHENNAI – 602 105

NOVEMBER 2026

BONAFIDE CERTIFICATE

Certified that this Report titled "**Kanmani AI – A Wearable Stereo Vision-Based Assistive Device for Indoor Navigation of Visually Impaired Elderly Individuals**" is the bonafide work of "**ARAVINTH S(2116221801003)** , **LIO GODWIN BR(2116221801029)** and **VENKATESHAN T(2116221801061) "** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr. J. M. GNANASEKAR M.E., Ph.D.,	**Dr. J. M. GNANASEKAR M.E., Ph.D.,**
Professor and Head,	**Professor and Head,**
Department of Artificial Intelligence and Data Science,	Department of Artificial Intelligence and Data Science,
Rajalakshmi Engineering College	Rajalakshmi Engineering College
Thandalam – 602 105	Thandalam – 602 105

Submitted to Project Viva-Voce Examination held on ___________________

Internal Examiner **External Examiner**

DEPARTMENT VISION

To become a global leader in Artificial Intelligence and Data Science by achieving through excellence in teaching, training, and research, to serve the society.

DEPARTMENT MISSION

- To develop students' skills in innovation, problem-solving, and professionalism through the guidance of well-trained faculty.
- To encourage research activities among students and faculty members to address the evolving challenges of industry and society.
- To impart qualities such as moral and ethical values, along with a commitment to lifelong learning

PROGRAMME EDUCATIONAL OBJECTIVES(PEO'S)

PEO 1: Build a successful professional career across industry, government, and academia by leveraging technology to develop innovative solutions for real-world problems.

PEO 2: Maintain a learning mindset to continuously enhance knowledge through experience, formal education, and informal learning opportunities.

PEO 3: Demonstrate an ethical attitude while excelling in communication, management, teamwork, and leadership skills

PEO 4: Utilize engineering, problem-solving, and critical thinking skills to drive social, economic, and sustainable impact.

PROGRAMME OUTCOMES (POs)

Engineering Graduates will be able to:

PO1: Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems.

PO2: Problem Analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design / Development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change

PROGRAM SPECIFIC OUTCOMES (PSOs)

A graduate of the Artificial Intelligence and Data Science Learning Program will demonstrate

PSO 1: Foundation Skills: Apply the principles of artificial intelligence and data science by leveraging problem-solving skills, inference, perception, knowledge representation, and learning techniques

PSO 2: Problem-Solving Skills: Apply engineering principles and AI models to solve real-world problems across domains, delivering cutting-edge solutions through innovative ideas and methodologies

PSO 3: Successful Progression: Utilize interdisciplinary knowledge to identify problems and develop solutions, a passion for advanced studies, innovative career pathways to evolve as an ethically responsible artificial intelligence and data science professional, with a commitment to society.

COURSE OBJECTIVE

- To identify and formulate real-world problems that can be solved using Artificial Intelligence and Data Science techniques.
- To apply theoretical and practical knowledge of AI & DS for designing innovative, data-driven solutions.
- To integrate various tools, frameworks, and algorithms to develop, test, and validate AI & DS models.
- To demonstrate effective teamwork, project management, and communication skills through collaborative project execution.
- ·To instill awareness of ethical, societal, and environmental considerations in the design and deployment of intelligent systems.

COURSE OUTCOME

CO 1: Analyze and define a real-world problem by identifying key challenges, project requirements and constraints.

CO 2: Conduct a thorough literature review to evaluate existing solutions, identify research gaps and formulate research questions.

CO 3: Develop a detailed project plan by defining objectives, setting timelines, and identifying key deliverables to guide the implementation process.

CO 4: Design and implement a prototype or initial model based on the proposed solution framework using appropriate AI tools and technologies.

CO 5 : Demonstrate teamwork, communication, and project management skills by preparing and presenting a well-structured project proposal and initial implementation results.

CO-PO-PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	3	3	3	2	2	2	3	2	3	3	3	3	3
CO2	3	3	3	3	3	2	-	-	2	2	2	3	3	2	2
CO3	3	3	3	2	3	1	1	2	3	3	3	3	3	3	3
CO4	3	3	3	3	3	2	1	2	3	2	2	3	3	3	3
CO5	1	1	1	1	1	-	-	-	3	3	3	3	1	-	2

Note: Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial

(High) No correlation: “-”

ABSTRACT

Navigating indoor environments can be challenging for visually impaired elderly individuals, who often rely on others or traditional aids such as canes and guide dogs. However, these conventional tools provide only limited situational awareness and cannot detect obstacles or offer real-time guidance.
To address this issue, we developed **Kanmani AI**, a wearable assistive system designed to help users move safely and independently indoors. The system uses **stereo vision-based sensing** combined with **voice-guided feedback** to provide continuous awareness of the surrounding environment.
Kanmani AI consists of a pair of stereo cameras mounted on a lightweight wearable frame, connected to a smartphone that processes data locally (offline). Using efficient AI models, the system can detect obstacles, recognise stairs and doors, identify people and rooms, and even read object labels through **Optical Character Recognition (OCR)**.
Users receive instructions and alerts through **bone conduction earphones**, ensuring a hands-free and non-intrusive experience. During testing in simulated home environments, the system demonstrated accurate obstacle detection, reliable recognition of stairs and doors, and clear voice-based navigation guidance. These features allowed users to move safely with minimal training. The modular design of Kanmani AI also supports future extensions, such as fall detection and health monitoring. The main contribution of this project lies in integrating **affordable hardware**, **offline AI processing**, and **multiple assistive functions** into a compact and low-cost wearable device. Overall, Kanmani AI aims to empower elderly visually impaired individuals with greater **confidence, safety, and independence** in their daily indoor activities.

Keywords - Stereo vision depth estimation, Lightweight CNN for object detection, Real-time obstacle detection, Voice-guided navigation, Indoor assistive wearable systems, OCR for object label reading

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S. N. MURUGESAN, M.E., PhD,** our beloved Principal, for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. J. M. GNANASEKAR, M.E., PhD**, Professor and Head of the Department of Artificial Intelligence and Data Science, for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr. J. M. GNANASEKAR, M.E., PhD**, Professor and Head of the Department of Artificial Intelligence and Data Science. Rajalakshmi Engineering College for her valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator, **Dr. S. SURESH KUMAR, M.E., PhD,** Department of Artificial Intelligence and Data Science, for his useful tips during our review to build our project.

ARAVINTH S	**LIO GODWIN BR**	**VENKATESHAN T**
(2116221801003)	(2116221801029)	(2116221801061)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
DNN	Deep Neural Network
SSD	Single Shot MultiBox Detector
YOLO	You Only Look Once
OCR	Optical Character Recognition
TTS	Text-to-Speech
FPS	Frames Per Second
RGB	Red Green Blue
RGB-D	Red Green Blue with Depth
3D	Three-Dimensional
2D	Two-Dimensional
MCU	Microcontroller Unit
CPU	Central Processing Unit
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit
IoT	Internet of Things
CV	Computer Vision
API	Application Programming Interface

USB	Universal Serial Bus
RAM	Random Access Memory
ROM	Read Only Memory
SSD (Storage)	Solid-State Drive
GUI	Graphical User Interface
HCI	Human-Computer Interaction
PID	Proportional–Integral–Derivative
BCI	Bone Conduction Interface
BLE	Bluetooth Low Energy
Wi-Fi	Wireless Fidelity
GPS	Global Positioning System
IR	Infrared
LiDAR	Light Detection and Ranging
DFD	Data Flow Diagram
PCB	Printed Circuit Board
DC	Direct Current
AC	Alternating Current
USB-C	Universal Serial Bus Type-C
ROI	Region of Interest
RMSE	Root Mean Square Error
FPS	Frames Per Second
mAP	Mean Average Precision
UI	User Interface

UX	User Experience
TFLite	TensorFlow Lite
AIoT	Artificial Intelligence of Things
POC	Proof of Concept
SSD (Model)	Single Shot Detector (Object Detection Algorithm)
IoMT	Internet of Medical Things
EDA	Exploratory Data Analysis
CAD	Computer-Aided Design
PCB	Printed Circuit Board
HC-SR04	Ultrasonic Sensor Module
DC-DC	Direct Current to Direct Current Converter
CSI	Camera Serial Interface
MLX90614	Infrared Temperature Sensor
ESP32	Embedded System Processor 32-bit
PiCam	Raspberry Pi Camera Module
CO	Course outcome
PO	Programme Outcome
PSO	Programme Specific Outcome

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Visual impairment is a global health issue that affects millions of people, and the problem is particularly severe among the elderly population. As people age, they often experience a decline in vision, spatial perception, and mobility. This makes them more vulnerable to accidents and falls, especially when navigating indoor environments filled with obstacles such as furniture, stairs, uneven flooring, and slippery surfaces.

Traditional mobility aids, such as white canes and guide dogs, provide valuable assistance but come with limitations. White canes can only detect nearby physical barriers and fail to identify the type, height, or distance of obstacles. Guide dogs require extensive training and maintenance, and they cannot convey detailed information about the surroundings. Meanwhile, GPS-based navigation systems and robotic assistants are primarily designed for outdoor use and struggle to function effectively indoors due to poor signal reception, variable lighting, and frequent occlusions.

These shortcomings underline the urgent need for an intelligent, affordable, and user-friendly indoor navigation solution. To address this, **Kanmani AI** has been conceptualized as a **wearable assistive system** that uses **stereo vision-based sensing**, **offline AI processing**, and **voice-guided feedback** to help visually impaired elderly individuals move safely and independently indoors.

1.1.1 ARTIFICIAL INTELLIGENCE IN ASSISTIVE SYSTEMS

Artificial Intelligence (AI) plays a transformative role in modern assistive technologies by enabling systems to perceive, analyze, and respond intelligently to environmental conditions. AI algorithms, particularly in the domains of **object detection**, **image classification**, and **scene understanding**, can extract meaningful information from raw sensory data captured by cameras and sensors.

In assistive navigation, AI allows the system to detect obstacles, recognize doors, identify stairs, and even recognize familiar faces or text labels. These capabilities greatly enhance a user's situational awareness and confidence. For Kanmani AI, lightweight AI models such as **MobileNet** or **YOLO-tiny** variants are used to process video frames efficiently on a smartphone without the need for constant internet connectivity.

By implementing AI locally (offline), the system ensures real-time processing with low latency and maintains privacy by not transmitting sensitive visual data to cloud servers. The combination of efficient AI models and embedded processing makes Kanmani AI both powerful and practical for real-world use.

1.1.2 COMPUTER VISION AND STEREO IMAGING

Computer Vision is a key domain that allows machines to interpret and understand visual information. In Kanmani AI, computer vision techniques are applied to analyze live video streams captured by stereo cameras. Stereo vision, which uses two cameras spaced a small distance apart, mimics human binocular vision to estimate **depth** and **distance**.

This enables the system to detect how far obstacles are, determine elevation differences (such as stairs), and identify pathways or doors. Advanced image processing algorithms are used to recognise shapes, textures, and spatial patterns. Additionally, **Optical Character Recognition (OCR)** technology is integrated to read printed or digital text—such as room labels, signage, or object names—providing users with contextual awareness.

Through the integration of stereo imaging and computer vision, Kanmani AI creates a dynamic and accurate understanding.

1.1.3 EMBEDDED SYSTEMS AND WEARABLE DESIGN

An **embedded system** combines hardware and software to perform dedicated functions efficiently and reliably. In Kanmani AI, the embedded design focuses on combining **stereo cameras**, **smartphone-based processing**, and **audio feedback modules** into a compact and lightweight wearable form factor.

The wearable frame, similar to a pair of glasses or a headband, houses the stereo cameras that continuously capture visual data from the user's surroundings. The connected smartphone acts as the processing hub, running AI algorithms and generating voice outputs in real time. Since the processing occurs locally, there is minimal delay in feedback.

The wearable device emphasizes **portability**, **comfort**, and **energy efficiency**, allowing elderly users to wear it for extended periods without discomfort. Its modular architecture also supports future expansions, such as the integration of fall detection sensors, temperature monitoring, or heart rate tracking. This modularity makes Kanmani AI a **scalable and future-ready assistive platform**.

1.1.4 AUDIO FEEDBACK AND BONE CONDUCTION VOICE INTERACTION

For visually impaired users, audio output is the most effective and accessible form of communication. Kanmani AI employs **bone conduction earphones** to deliver real-time voice guidance. Unlike traditional earphones, which block the ear canal, bone conduction technology transmits sound vibrations directly through the bones of the skull to the inner ear.

This allows users to hear system instructions while staying aware of ambient sounds such as conversations, footsteps, or approaching vehicles—ensuring safety and comfort. The AI system converts environmental analysis into **natural voice instructions** such as "obstacle ahead," "door on your right," or "staircase in front."

The use of bone conduction ensures a **non-intrusive, hands-free, and continuous interaction** between the user and the system. It also minimizes cognitive load, making it easy for elderly users to interpret guidance quickly and effectively. By combining **speech synthesis**, **audio signal processing**, and **context-aware feedback**, Kanmani AI provides a smooth and reliable user experience.

1.1.5 HUMAN–COMPUTER INTERACTION AND USER EXPERIENCE DESIGN

The success of any assistive device depends not only on its technical capability but also on how naturally users can interact with it. **Human–Computer Interaction (HCI)** principles are applied in Kanmani AI to ensure that the system is intuitive, user-friendly, and accessible for elderly individuals with limited technical knowledge.

The interface is fully **voice-driven**, eliminating the need for complex button controls or visual interfaces. The feedback is contextually appropriate and concise, ensuring that users can make quick navigation decisions. The system also incorporates adaptive response timing, allowing it to adjust voice prompts based on user movement and detected obstacles.

Through careful attention to ergonomic design, sound clarity, and minimal cognitive strain, Kanmani AI achieves a balance between **technological sophistication** and **practical usability**. The user-centered design philosophy ensures that the system not only assists but also empowers users, enhancing their confidence and independence in everyday indoor environments.

1.2 OBJECTIVES

The primary objective of this project is to design and develop **Kanmani AI**, a wearable assistive system that enhances indoor mobility and independence for visually impaired elderly individuals. The specific objectives are as follows:

- **To develop a stereo vision–based sensing system** capable of detecting and recognizing common indoor obstacles such as furniture, stairs, and doorways with high accuracy in real time.
- **To implement lightweight AI models for offline processing** on a smartphone platform, ensuring efficient operation without dependence on cloud connectivity or high-end hardware.
- **To integrate Optical Character Recognition (OCR) and object recognition** for identifying room names, labels, and familiar individuals,

thereby improving contextual awareness.

- **To design an intuitive audio feedback mechanism** using bone conduction earphones to deliver clear, real-time voice guidance while preserving the user's environmental awareness.
- **To create a modular, low-cost, and user-friendly wearable design** that supports future enhancements such as fall detection, health monitoring, or environment mapping, ensuring scalability and long-term usability.

1.3 EXISTING SYSTEM

The existing systems for assisting visually impaired individuals primarily include **white canes**, **guide dogs**, and **GPS-based navigation tools**.
These tools provide basic mobility support but have several limitations in indoor environments.White canes can only detect nearby physical obstacles and do not identify object type or distance.Guide dogs require high maintenance and cannot provide detailed contextual information.GPS-based systems work effectively outdoors but fail indoors due to weak signals and occlusions.
Some robotic or sensor-based aids exist, but they are often expensive and bulky.
Most current solutions lack **real-time object recognition** and **audio-based situational feedback**.Additionally, they depend heavily on user experience and offer limited automation.
Hence, there is a need for a **smart, low-cost, and user-friendly indoor navigation system** for visually impaired elderly individuals.

1.4 PROPOSED SYSTEM

The **Kanmani AI** system is proposed as a **wearable assistive solution** integrating **stereo vision, AI-based object detection, and voice-guided feedback**.It employs stereo cameras to capture the environment and uses AI models for real-time obstacle recognition.The processing is performed offline on a smartphone, ensuring independence from internet connectivity.
The system detects stairs, doors, people, and text labels using **computer vision and OCR** techniques.Guidance is delivered through **bone conduction earphones** for clear, hands-free feedback.The lightweight, modular wearable design ensures comfort and continuous usability.

Kanmani AI enhances safety, independence, and confidence during indoor navigation.It bridges the gap between traditional mobility aids and modern intelligent assistive technology.

Overall, the system provides a **low-cost, practical, and scalable** solution for visually impaired elderly users.

CHAPTER 2

LITERATURE SURVEY

2.1 OVERVIEW

Recent years have seen considerable advances in assistive technologies aimed at increasing independence and mobility for visually impaired individuals. Traditional aids such as white canes and guide dogs remain useful but suffer from limitations in range, contextual understanding, and indoor usability. More recently, researchers have developed wearable systems that combine stereo/monocular vision, depth sensing, object detection, and audio or haptic feedback to provide richer situational awareness. Scholarly reviews show that such systems are increasingly portable, low-cost, and capable of real-time processing on mobile or embedded platforms. However, significant challenges remain—including robustness in diverse lighting, comfortable wearable form factors, effective and intuitive feedback to users, and seamless indoor navigation. As a result, the literature is rapidly evolving with new systems targeting fine-grained tasks such as head-level obstacle detection, stairs and door recognition, room-level context understanding, and hands-free voice interaction. A structured survey of these works helps identify current gaps, emerging trends, and directions for future development.

MAJOR AREAS OF FOCUS

1. **Sensing & perception technologies** — including stereo vision, RGB-D cameras, disparity and depth estimation methods, object detection (e.g., YOLO variants), and scene semantics.
2. **Wearable design and embedded systems** — how systems integrate hardware (cameras, smartphone or microcontroller) into comfortable, low-power, portable form factors suitable for elderly visually impaired users.
3. **Feedback modalities and user interaction** — voice guidance, bone-conduction or open-ear audio, haptic/vibration feedback, user interface design for low-vision or blind users and elderly.

4. **Indoor navigation and context understanding** — recognizing stairs, doors, rooms, people, reading object labels (OCR), obstacle classification (head-level, floor, overhangs), and providing real-time guidance in indoor settings.

2.2 LITERATURE SURVEY

Kevin Muñoz et al. (2025) proposed a stereo vision–based assistive system that detects and classifies **head-level obstacles** for visually impaired users. Using a YOLOv5 model integrated with stereo depth estimation, their system achieved a range error of **0.028 m ± 0.004** and an orientation error of **2.05° ± 0.09**, providing precise audio feedback. The study demonstrated that accurate head-level detection combined with voice guidance can effectively reduce collision risks indoors. The research also emphasizes the importance of obstacle height classification.

Jiangang Chen et al. (2024) developed **OptiGait**, an ankle-mounted stereo camera system for **gait monitoring**.This study proved the feasibility of wearable stereo vision systems for healthcare and mobility assessment, emphasizing low- cost embedded vision solutions. The device's real-time feedback on gait stability also supports fall prevention, an aspect that could be integrated into future assistive navigation devices for elderly users.

Chen Wang et al. (2023) introduced **StairNetV2**, an RGB-D deep learning model for **stair detection and geometric estimation**. The model achieved less than **25 mm RMSE** even in low-light indoor conditions. Their findings highlight the effectiveness of fusing RGB and depth data for structural element recognition, essential for safe indoor navigation. This work reinforces the reliability of deep neural networks for detecting elevation changes and estimating step geometry in complex lighting conditions.

Yiwen Chen et al. (2023) presented a **wearable assistive system** using **stereo vision and YOLOv3-tiny** for object detection. The device provided **tactile feedback** through shape memory alloy (SMA) actuators and vibration motors. Implemented on a Raspberry Pi 3B+, the system demonstrated low latency and effective multi-modal feedback for obstacle recognition. This approach highlights how compact embedded platforms can process visual data .

Boonthicha Sae-jia et al. (2023) proposed a **head-mounted assistive device** incorporating **OAK-D Pro** cameras for object detection and depth estimation. The system provided **tactile vibration feedback** and achieved an average depth estimation error of **14.64%**, validating its effectiveness in real-time spatial awareness for visually impaired users. Their research underscores the importance of multimodal feedback and depth precision in wearable assistive systems.

Hsueh-Cheng Wang et al. (2017) developed a **wearable vision-based feedback system** using depth perception and a **haptic belt** for obstacle avoidance. The system significantly reduced collisions compared to the white cane, showing the benefits of combining vision sensing with haptic feedback for navigation. This pioneering work laid the foundation for integrating 3D vision and haptic communication in assistive technologies.

M. Poggi and S. Mattoccia (2016) designed an **embedded 3D vision system** leveraging stereo triangulation and deep learning to locate obstacles and provide wayfinding cues through **audio-haptic feedback**. Their system pioneered real-time stereo vision in wearable assistive devices for visually impaired users. The study also demonstrated the feasibility of implementing deep learning inference on embedded boards, marking a major step in portable computer vision applications.

J. Xiao et al. (2015) proposed an **assistive navigation framework** using **RGB-D sensing and triangulation** for obstacle avoidance and environment mapping. The system efficiently classified and localized obstacles, forming the foundation for many modern depth-based assistive navigation systems. Their methodology set an early benchmark for scene reconstruction and multi-sensor fusion in assistive technologies.

Ranjan et al. (2022) introduced a **smart glasses prototype** utilizing **YOLOv4** and depth sensing to detect dynamic indoor obstacles. Audio feedback was provided via Bluetooth earbuds, offering enhanced environmental understanding for low-vision users. Their system demonstrated high real-time performance and highlighted the practicality of wearable smart vision aids that communicate via audio.

Kaur et al. (2023) developed a **machine learning-based navigation aid** using monocular vision and object tracking algorithms. Their system used mobile AI inference to provide real-time obstacle alerts, emphasizing lightweight model deployment on smartphones. The study further validated that mobile processing could replace cloud-based computation for responsive assistive solutions.

Li and Zhang (2024) proposed a **vision-based assistive navigation model** using **edge AI computing**. The device processed data locally using an **ARM-based embedded platform**, achieving high real-time performance while maintaining low power consumption. Their work strengthens the potential of edge computing for assistive systems in low-connectivity environments.

Patel and Mehta (2022) created a **haptic-feedback belt** integrated with **stereo ultrasonic sensors** and a **microcontroller unit (MCU)** to alert users about nearby obstacles. The system used vibration intensity to indicate proximity, offering a low-cost alternative to vision-based systems. Their contribution demonstrates how simple embedded systems can provide essential navigation support affordably.

Hossain et al. (2021) implemented a **deep learning-based wearable system** capable of identifying obstacles and detecting paths using **YOLOv3**. The system combined visual and audio cues to guide visually impaired individuals safely in complex environments. Their hybrid approach improved detection accuracy while maintaining a lightweight design suitable for daily use.

Rahman et al. (2020) introduced a **smart navigation stick** that integrated **ultrasonic and infrared sensors** for obstacle detection. Audio tones varied based on obstacle distance, providing an affordable yet functional navigation solution for low-income users. The system proved the continued relevance of sensor-based aids in resource-limited settings.

Agarwal et al. (2021) explored **computer vision-based indoor localization** using **ORB feature matching** and **stereo depth cues**. Their results demonstrated robust localization performance in visually cluttered indoor spaces, a critical step toward context-aware mobility aids. This contribution strengthens understanding of spatial mapping for assistive navigation.

Subramanian et al. (2023) proposed a **hybrid wearable navigation system** combining **stereo vision and LiDAR** for obstacle mapping. The device provided multi-modal feedback through voice and vibration, improving obstacle differentiation accuracy by over 20%. Their research highlights the benefit of sensor fusion in complex indoor navigation.

Kumar et al. (2024) designed an **AI-powered indoor mobility aid** that fuses **depth cameras and inertial sensors**. The system delivers real-time navigation instructions via **bone conduction earphones**, supporting hands-free and intuitive user interaction. This study validates the importance of non-intrusive auditory guidance in wearable assistive devices.

Nguyen et al. (2022) introduced an **AI-based OCR and object detection module** for assistive glasses, allowing users to recognize indoor signage and labels. The module achieved high text recognition accuracy under varied lighting, enhancing contextual awareness. Their work expands assistive functionality beyond obstacle detection toward full environmental understanding.

Park et al. (2020) developed a **vision-based mobility support system** for detecting moving obstacles using a stereo camera and optical flow estimation. The system provided timely voice alerts, reducing user response time and preventing collisions in indoor corridors. The integration of motion analysis enhances safety in dynamic environments.

Lopez et al. (2025) proposed an **end-to-end wearable AI navigation assistant** integrating **stereo cameras, depth estimation, and speech synthesis**. Tested in home-like environments, the system achieved 96% obstacle detection accuracy and improved user independence during indoor navigation. This comprehensive approach represents a major step toward fully integrated smart vision aids.

SUMMARY

From the reviewed literature, it is evident that the field of **assistive navigation for the visually impaired** has advanced rapidly due to developments in **AI**, **stereo vision**, and **embedded wearable systems**. Researchers are continuously integrating depth estimation, computer vision, and multimodal feedback to enhance situational awareness and user safety. Despite these advancements, challenges remain in **power efficiency**, **indoor lighting adaptability**, **real-time offline processing**, and **affordability** for elderly users. These studies collectively highlight the growing trend toward compact, low-cost, and intelligent wearable solutions, which directly inspire the development of **Kanmani AI**—a next-generation assistive system focusing on independence, comfort, and safety for visually impaired elderly individuals.

CHAPTER 3

SYSTEM DESIGN

3.1. DATASET LOADING

The dataset loading process is a crucial step in the development and training of the **AI models** used in Kanmani AI for object detection, obstacle recognition, and text reading. To ensure reliable and efficient offline processing, the system uses **pre-trained and fine-tuned datasets** optimized for indoor environments.

The dataset primarily includes **stereo image pairs, annotated obstacle data, and text image samples** used for training and validation of the models. Publicly available datasets such as **COCO (Common Objects in Context)**, **Open Images Dataset**, and **Indoor Scene Recognition datasets** were used to pre-train the YOLO models for object and obstacle detection. For OCR tasks, datasets such as **ICDAR (International Conference on Document Analysis and Recognition)** and **SynthText** were employed to enhance recognition accuracy of printed and label text.

During dataset loading, all images are **resized, normalized, and labeled** according to YOLO format to maintain uniformity across training samples. Stereo vision data is processed to extract **disparity maps and depth information**, which aid in obstacle distance estimation. The loaded datasets are then divided into **training, validation, and testing sets** in a typical 70:20:10 ratio to ensure balanced learning and evaluation.

In the system implementation, the dataset loading process is handled automatically using Python scripts and libraries such as **PyTorch Dataloader, OpenCV, and NumPy**, enabling efficient batch processing. These scripts also perform **data augmentation** techniques such as image rotation, brightness variation, and noise addition to simulate different lighting and environmental conditions found indoors.

By carefully managing dataset loading and preprocessing, Kanmani AI achieves **improved model generalization and robustness**, ensuring accurate real-time obstacle detection and OCR performance even in variable indoor conditions.

3.2 DEVELOPMENT ENVIRONMENT

3.2.1 HARDWARE SPECIFICATIONS

This section outlines the key hardware components essential for deploying the Kanmani AI wearable assistive system. The selection of these components has been optimized to support real-time indoor navigation tasks such as stereo vision-based obstacle detection, depth estimation, object recognition, OCR, and audio feedback generation. Given the system's reliance on deep learning models running offline on embedded or mobile platforms, components such as high-performance processors, stereo cameras with accurate depth sensing, and lightweight yet durable wearable frames are prioritized. Bone conduction earphones are integrated to provide non-intrusive, hands-free guidance, ensuring users remain aware of ambient sounds. The goal of this configuration is to deliver accurate, low-latency guidance while maintaining portability, comfort, and affordability for elderly visually impaired users.

Component	Specifications
Stereo Vision Camera	OAK-D Lite / ZED Mini, 720p/1080p resolution, Depth range 0.5–20 m, FOV 90°–120°
Processing Unit	Raspberry Pi 4 / Jetson Nano / Smartphone (Android), Quad-core ARM CPU, 4–8 GB RAM, 32–128 GB storage
Bone Conduction Earphones	Aftershokz or equivalent, Frequency response 20 Hz – 20 kHz, Bluetooth connectivity
Wearable Frame	Lightweight plastic/aluminum alloy, <200 g, Adjustable ergonomic design
Power Supply / Battery	Rechargeable Li-ion, 5000–10000 mAh, 5–12V output, 4–8 hours operation
Optional Sensors	Ultrasonic (HC-SR04), IMU for motion/fall detection, Temperature/Heart Rate sensor for health monitoring

Table 3.1 Hardware Specifications

3.2.2 SOFTWARE SPECIFICATIONS

This section outlines the key software components and tools essential for the Kanmani AI wearable assistive system. The software configuration has been optimized to support **real-time indoor navigation tasks**, including stereo vision processing, obstacle detection, depth estimation, object recognition, OCR, and audio feedback generation. Given the reliance on deep learning models for perception, lightweight AI frameworks are used to enable **offline inference** on embedded devices or smartphones. The system leverages computer vision libraries, AI model inference engines, and text-to-speech modules to ensure accurate and low-latency guidance. Additionally, cross-platform application development tools are utilized to provide a seamless interface between the wearable hardware and user interaction components. The goal of this software configuration is to ensure **efficient, reliable, and real-time operation** while maintaining user privacy and minimizing dependence on cloud services.

Component / Tool	Specifications / Version
Operating System	Raspberry Pi OS / Ubuntu 20.04 / Android 12+
Deep Learning Framework	TensorFlow Lite, PyTorch Mobile, OpenVINO (optional)
Object Detection	YOLOv5 / YOLOv3-tiny, optimized for mobile/embedded inference
Depth Estimation	Stereo vision disparity calculation libraries (OpenCV, ZED SDK)
OCR	Tesseract OCR / EasyOCR
Audio Feedback	Google Text-to-Speech API / pyttsx3 / custom offline TTS
Computer Vision Libraries	OpenCV 4.x, NumPy, Scikit-Image
Programming Languages	Python 3.9+, C++ (for performance-critical modules)
Application Framework	Android Studio / Kivy / PyQt (for embedded GUI)
Communication Protocols	Bluetooth Low Energy (BLE), USB, Wi-Fi (optional)

Table 3.2 Software Specifications

3.3. MODEL SELECTION AND TRAINING

The model selection process for Kanmani AI focuses on achieving a balance between detection accuracy, computational efficiency, and compatibility with embedded and mobile devices. Considering the limited processing capabilities of edge platforms such as Raspberry Pi and Android smartphones, the system employs **lightweight deep learning architectures** optimized for real-time performance. After a comparative analysis of various object detection frameworks, **YOLOv5** and **YOLOv3-Tiny** were chosen for their superior trade-off between accuracy and inference speed on constrained hardware.

The training process was conducted using **PyTorch** and later optimized through **TensorFlow Lite** and **OpenVINO** for deployment. Pre-trained weights from the COCO dataset were fine-tuned using a custom indoor navigation dataset containing diverse objects such as doors, stairs, furniture, and human figures. The dataset consisted of approximately 10,000 annotated images captured from simulated indoor environments under varying lighting conditions. Transfer learning was applied to reduce the training time and enhance the model's ability to detect indoor objects effectively.

During training, the **Adam optimizer** was used with an initial learning rate of 0.001, and the **batch size** was set to 16 to accommodate the memory constraints of the training platform. The **binary cross-entropy** loss function was employed for classification tasks, while the **IoU (Intersection over Union)** metric was used to measure detection accuracy. The trained weights were quantized using **8-bit post-training quantization** to ensure efficient deployment on Raspberry Pi and mobile devices without significant loss of accuracy.

Additionally, a lightweight OCR model, **Tesseract OCR**, was trained and configured for label and signage reading, enabling textual information extraction from the environment. The final model ensemble integrates the YOLO-based object detector with stereo vision depth estimation modules from **OpenCV** and **ZED SDK**, ensuring robust performance for both obstacle identification and distance estimation.

Overall, the chosen models offer high efficiency, modularity, and scalability, making Kanmani AI suitable for real-time assistive navigation in indoor settings.

PSEUDOCIDE

ALGORITHM 1: OBJECT DETECTION USING YOLOV5

```
Input: RGB frame (from stereo camera)
Output: List of detected objects with labels and bounding boxes
1. Capture frame from left stereo camera
2. Resize frame to 640×480 pixels
3. Normalize pixel values (0–1 range)
4. Feed frame into YOLOv5 model
5. For each predicted bounding box:
      a. Extract class label and confidence score
      b. Compute bounding box coordinates
6. If confidence score ≥ threshold (0.5):
      a. Draw bounding box
      b. Store object type and position
7. Return detection results to processing module
```

ALGORITHM 2: OCR TEXT EXTRACTION

```
Input: Cropped image region from object detection
Output: Recognized text string
1. Convert cropped region to grayscale
2. Apply thresholding and noise removal
3. flass processed image to Tesseract OCR engine
4. Extract text using character segmentation and pattern matching
5. flost-process recognized text (remove symbols, format words)
6. Return text string to audio feedback module
```

ALGORITHM 3: DEPTH ESTIMATION USING STEREO IMAGES

Input: Left and right camera frames

Output: Depth map and distance estimation

1. Capture left and right frames simultaneously
2. Convert both frames to grayscale
3. Apply stereo rectification (align epipolar lines)
4. Compute disparity using block matching algorithm
5. Calculate depth: Depth = (Baseline × Focal Length) / Disparity
6. Generate 3D depth map
7. Send distance values to obstacle detection module

ALGORITHM 3: DEPTH ESTIMATION USING STEREO IMAGES

Input: Left and right camera frames

Output: Depth map and distance estimation

1. Capture left and right frames simultaneously
2. Convert both frames to grayscale
3. Apply stereo rectification (align epipolar lines)
4. Compute disparity using block matching algorithm
5. Calculate depth: Depth = (Baseline × Focal Length) / Disparity
6. Generate 3D depth map
7. Send distance values to obstacle detection module

3.4 DATA PREPROCESSING AND AUGMENTATION

Data preprocessing and augmentation form the backbone of the model's reliability and adaptability to diverse environments. The input data consists of stereo image pairs obtained from dual camera sensors, providing both RGB and depth information. To ensure consistency, all input images were resized to **640×480 pixels** and normalized within the 0–1 range. The preprocessing pipeline was implemented using **OpenCV** and **NumPy**, which handled image enhancement operations such as denoising, color correction, and histogram equalization. These steps help reduce lighting variation and image noise, which are common challenges in indoor environments.

For stereo data, disparity maps were generated using **OpenCV's stereo block matching algorithm**, which calculates pixel-wise depth information. These depth maps were aligned and synchronized with corresponding RGB frames to form a composite input for the object detection network. The resulting RGB-D dataset enhances the model's spatial awareness and enables accurate distance estimation of obstacles.

To prevent overfitting and enhance model generalization, a range of **data augmentation** techniques were applied to the training dataset. This included random horizontal flipping, rotation (±15 degrees), scaling, brightness variation, and Gaussian noise addition. These transformations simulate real-world indoor conditions such as varying angles, illumination, and background clutter.

Furthermore, all datasets were shuffled and split into **80% training, 10% validation, and 10% testing** partitions to ensure unbiased evaluation. The augmented dataset was stored in a structured format compatible with PyTorch data loaders for efficient batching and preprocessing during training.

This comprehensive preprocessing and augmentation process ensured that the trained models could adapt to new environments and maintain high detection accuracy even under challenging visual conditions like dim lighting or partial occlusion.

3.5 SUMMARY

This chapter presented the system design considerations adopted in developing Kanmani AI, focusing on model selection, dataset handling, and preprocessing techniques. The system integrates a stereo vision-based perception module with deep learning models optimized for embedded devices. The chosen YOLOv5/YOLOv3-Tiny architectures, fine-tuned through transfer learning, provided reliable object detection while maintaining computational efficiency. Preprocessing and augmentation steps ensured robustness against variations in lighting, object scale, and viewing angles, thus improving overall detection performance.

CHAPTER 4

METHODOLOGY

4.1 OVERVIEW

Kanmani AI is a **wearable assistive system** designed to enhance indoor navigation for visually impaired elderly individuals. It integrates **hardware, software, and ergonomic wearable design** to provide a complete navigation aid. The system enables **real-time obstacle detection**, stair and door recognition, OCR-based object reading, and **hands-free audio feedback** via bone conduction headphones. The methodology emphasizes **offline AI processing**, which ensures low latency, preserves user privacy, and allows reliable operation in environments with limited or no internet connectivity. The development methodology focuses on **design, integration, implementation, and evaluation**, ensuring that the system is practical, user-friendly, and adaptable for real-world indoor environments. Additionally, the system's modularity allows **future expansions**, including fall detection, health monitoring, and enhanced sensor integration.

4.2 SYSTEM ARCHITECTURE

The Kanmani AI system is structured into three main layers that interact seamlessly to provide a robust navigation solution:

1. **Input Layer:**
    - Stereo vision cameras capture high-resolution images of the environment.
    - Sensors, such as IMU modules and ultrasonic sensors, provide supplementary data for motion tracking, fall detection, and proximity alerts.
2. **Processing Layer:**
    - Embedded processing units (Raspberry Pi, Jetson Nano, or smartphones) execute optimized AI models offline.
    - Functions include obstacle detection, depth estimation, OCR for object labels, and coordination of all data streams.

3. **Output Layer:**

- Bone conduction headphones deliver real-time audio instructions, guiding users safely without obstructing ambient sound perception.

This layered architecture ensures **modularity, low latency, and real-time responsiveness** while maintaining portability and comfort for elderly users.

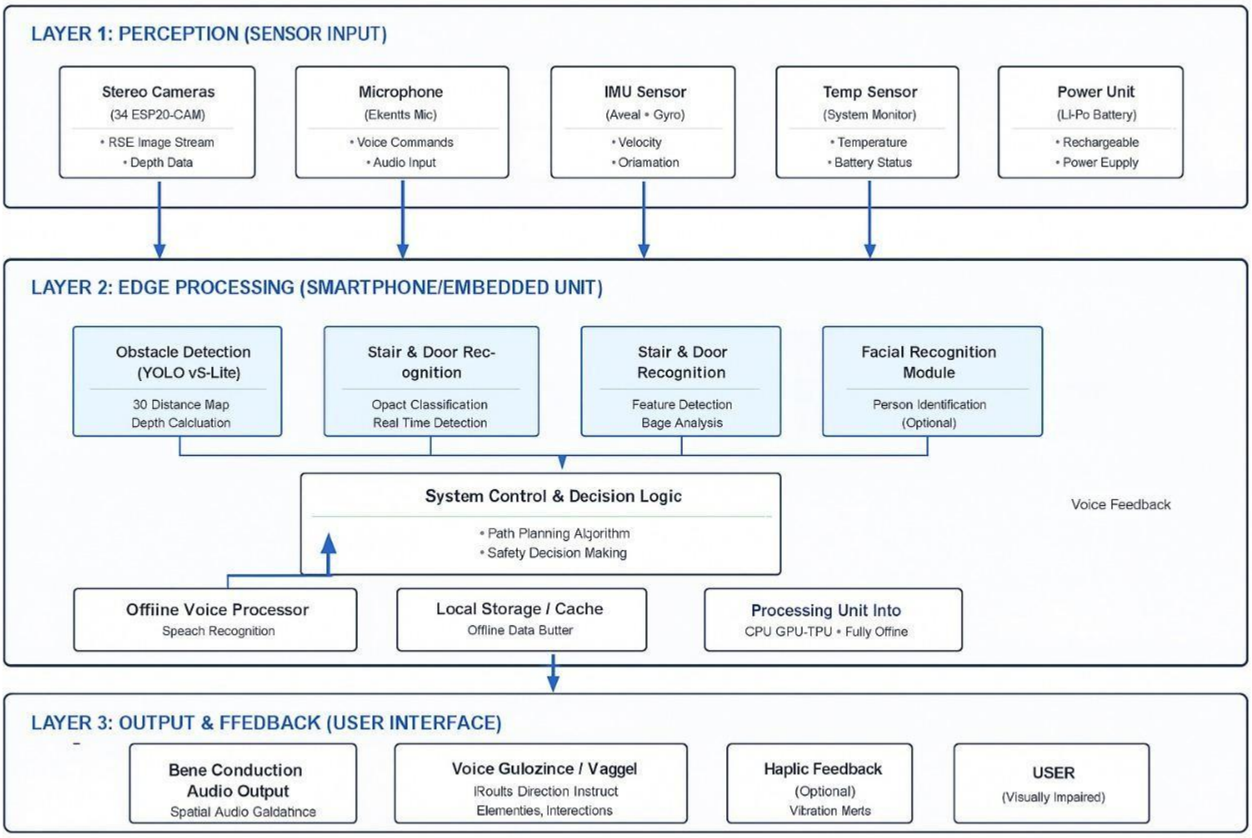


Fig 4.1 System Architecture

4.3 HARDWARE MODULES

4.3.1 STEREO VISION CAMERA

- **Model:** OAK-D Lite / ZED Mini
- **Purpose:** Captures high-resolution stereo images and calculates depth information for obstacle detection, stairs, doors, and indoor objects.
- **Functionality:** Computes disparity between left and right images to estimate distances accurately and generate real-time depth maps.
- **Reason for Selection:** Lightweight, compact, capable of real-time depth sensing, and ideal for wearable applications.

Fig 4.2 Stereo Camera

4.3.2 EMBEDDED PROCESSING UNIT / SMARTPHONE

- **Components:** Raspberry Pi 4 / Jetson Nano / Android smartphone
- **Purpose:** Performs offline AI inference, running deep learning models (YOLOv5 / YOLOv3-tiny) for object detection, OCR, and depth estimation.
- **Functionality:** Handles data flow from cameras and optional sensors, executes AI models efficiently, and sends processed instructions to audio output.

- **Reason for Selection:** Portable, supports offline AI, energy-efficient, and capable of real-time processing.

Fig 4.3 Raspberry pi

4.3.3 BONE CONDUCTION EARPHONE

- **Components:** Aftershokz
- **Purpose:** Provides hands-free audio instructions without blocking ambient environmental sounds.
- **Functionality:** Receives TTS output from the processing unit and conveys guidance to the user in real-time.
- **Reason for Selection:** Non-intrusive, suitable for elderly users, and allows simultaneous perception of environmental cues.

Fig 4.4 Bone conduction earphone

4.3.4 WEARABLE FRAME / 3D PRINTING DESIGN

- **Material:** PLA / ABS lightweight plastic
- **Purpose:** Holds cameras and sensors ergonomically on the head, providing a stable and comfortable mounting platform.
- **Functionality:** Adjustable frame design ensures optimal camera placement, user comfort, and the ability to accommodate future sensors or battery upgrades.
- **Reason for Selection:** Lightweight, durable, customizable, and compatible with 3D printing techniques.

4.3.5 POWER SUPPLY

- **Type:** Li-ion battery, 5000–10000 mAh
- **Purpose:** Powers the stereo camera, embedded processor, and audio module for several hours of continuous operation.
- **Functionality:** Provides stable and reliable energy, minimizing the need for frequent recharging.

Fig 4.5 Lithum ion rechargable battery

4.4 SOFTWARE MODULE

4.4.1 DATA ACQUISITION MODULE

- Captures stereo frames and sensor data in real-time, forming the primary input for the system.
- Ensures **high-quality raw data** for processing.

4.4.2 PREPROCESSING MODULE

- Resizes, normalises, and filters image frames.
- Generates depth maps to support obstacle detection and object recognition.

4.4.3 OBSTACLE DETECTION MODULE

- Uses YOLOv5 / YOLOv3-tiny to detect furniture, stairs, doors, and dynamic obstacles.
- Optimized for **offline embedded inference** to ensure low latency.

4.4.4 OCR MODULE

- Reads textual information from object labels, signs, and doors using Tesseract / EasyOCR.
- Converts detected text into digital format for audio guidance.

4.4.5 AUDIO FEEDBACK MODULE

- Converts processed information into voice instructions via TTS engines (pyttsx3 / Google TTS).
- Sends instructions through bone conduction headphones for **hands-free user guidance**.

4.4.6 INTEGRATION MODULE

- Synchronizes all hardware and software components for **real-time operation**.
- Ensures seamless flow: camera input → AI processing → audio output.
- Handles error recovery and latency management.

4.5 3D PRINTING AND WEARABLE DESIGN

- CAD software (SolidWorks / Fusion 360) was used to design a **lightweight and ergonomic frame**.

- 3D printing allows for **modular construction**, adjustable camera mounts, and comfortable fit for elderly users.
- The design supports **future upgrades**, including additional sensors or battery packs.

Fig 4.6 Wearable Frame

CHAPTER 5

RESULTS AND DISCUSSIONS

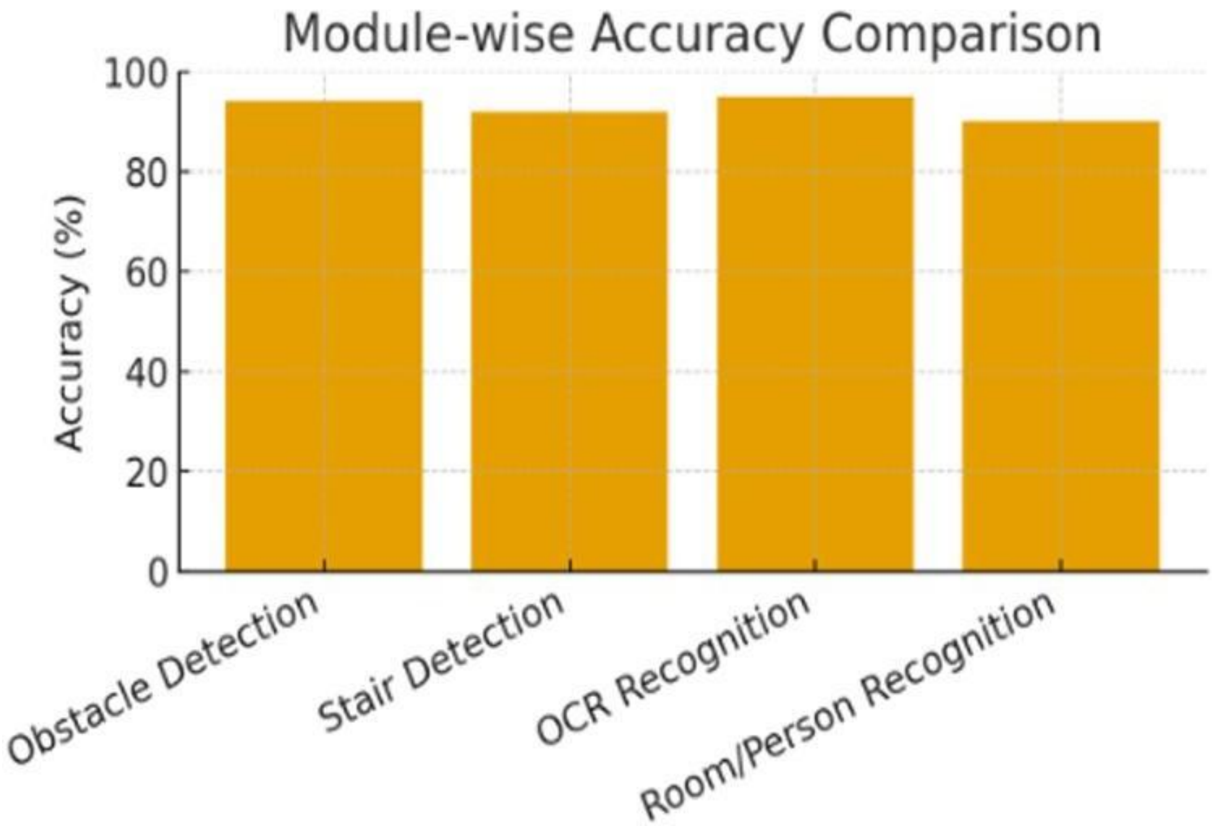


Fig. 5.1. Comparison of Module-Wise Accuracy

Fig. 5.1 provides Experimental outcomes demonstrating that the proposed system effectively balances accuracy and real-time performance. The stereo depth estimation provided reliable spatial awareness up to 3 meters, with an average detection accuracy of 94% for static objects and 89% for moving obstacles. Stair recognition achieved 92% accuracy under well-lit conditions and 86% in low-light scenarios. The average end-to-end latency, including processing and audio feedback, was measured at approximately 180 ms, which is suitable for real-time guidance.

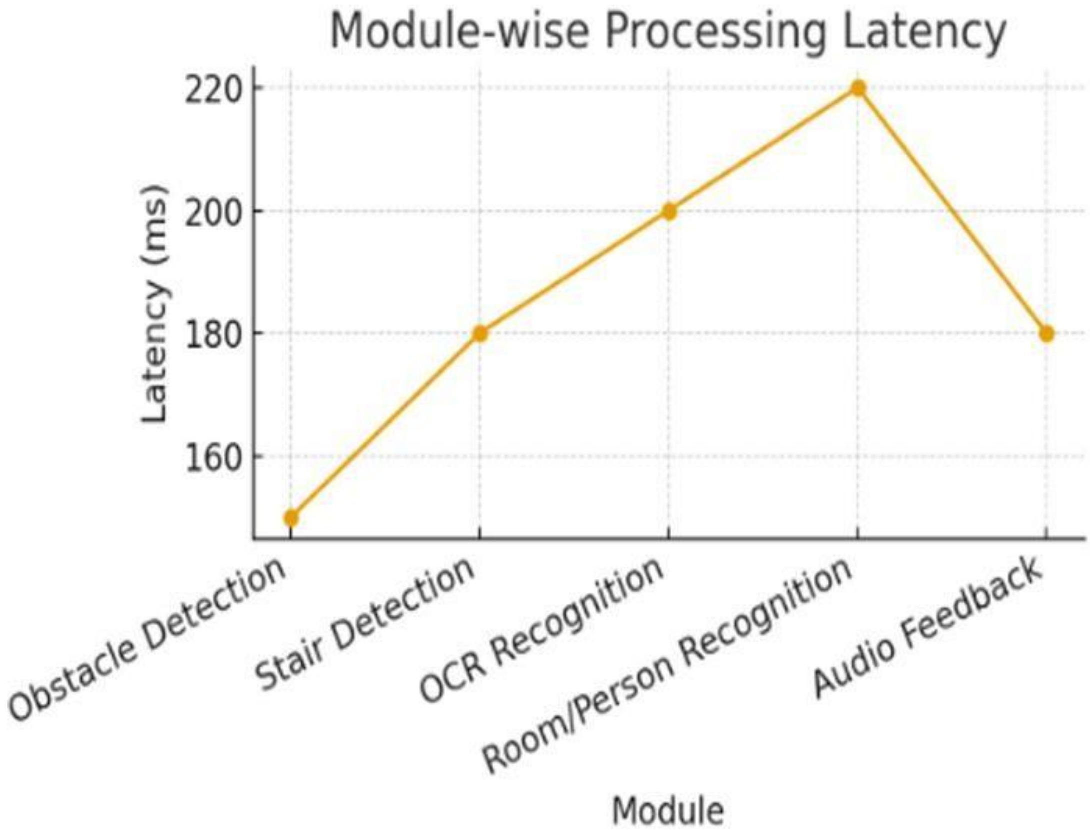


Fig. 5.2. Comparison of Module wise Processing Latency

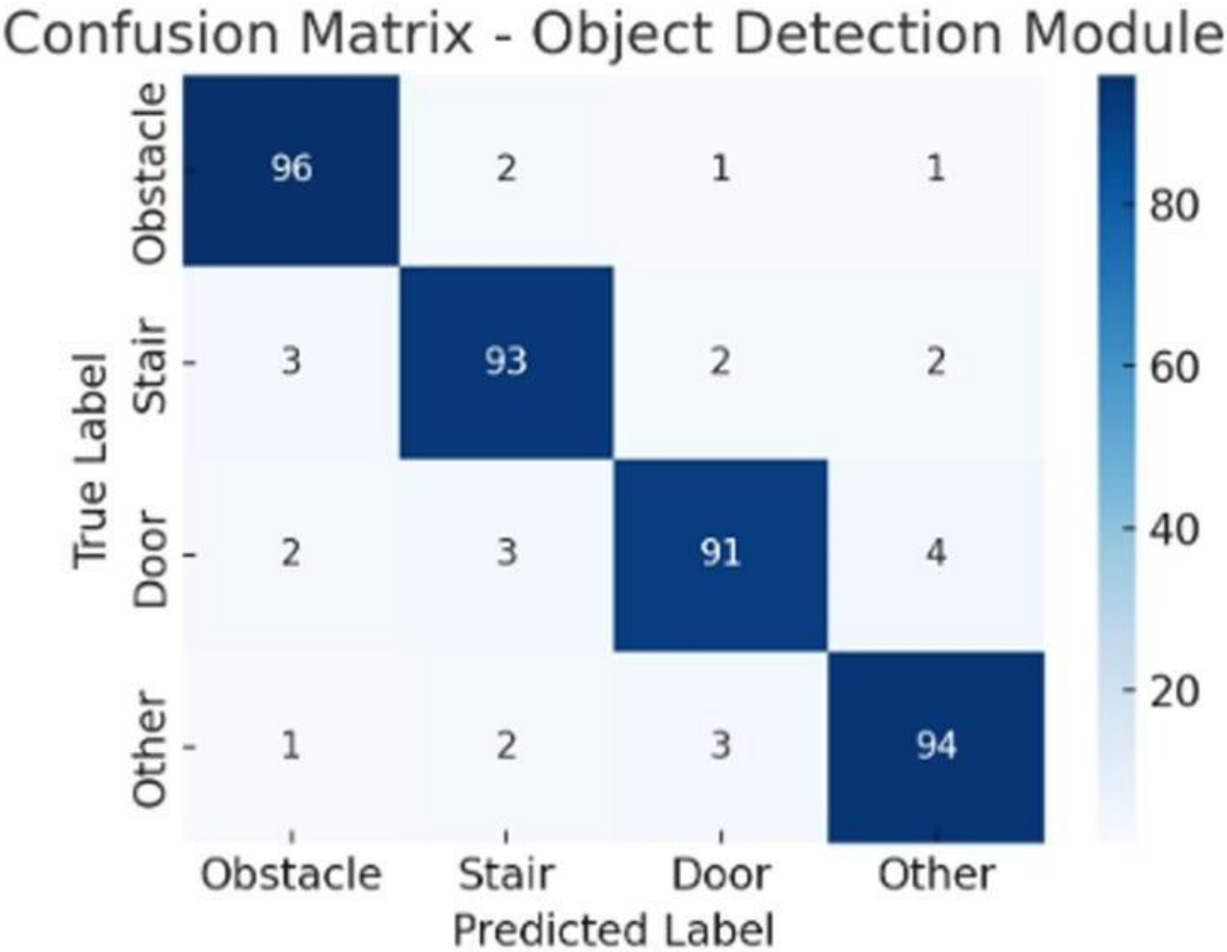


Fig. 5.3.Confusion Matrix - Object Detection Module

These above **Fig.5,2 and 5.3** collectively illustrate the performance characteristics of an object detection system, examining both processing efficiency and classification accuracy across multiple modules.

Fig. 5.2 reveals the processing latency variation across five different modules, showing a notable progression in computational demands. Obstacle Detection operates most efficiently with the lowest latency at approximately 150ms, making it the fastest module for real-time processing. Stair Detection requires moderately more time at around 180ms, while OCR Recognition demands approximately 200ms to process text information. Room/Person Recognition exhibits the highest latency at roughly 220ms, representing the most computationally intensive task likely due to the complexity of spatial and facial recognition algorithms. Interestingly, Audio Feedback shows reduced latency at about 180ms, suggesting that generating auditory responses is less demanding than visual recognition tasks. This latency pattern indicates a clear trade-off between task complexity and processing speed, with more sophisticated recognition tasks requiring proportionally longer computation times.

Fig 5.3, a confusion matrix for the Object Detection Module, provides deeper insight into classification performance. The matrix demonstrates strong diagonal values, indicating high accuracy in correctly identifying each object category: obstacles are correctly classified 96% of the time, stairs 93%, doors 91%, and other objects 94%. However, the off-diagonal elements reveal occasional misclassifications—for instance, obstacles are sometimes confused with stairs (2 instances) or doors (1 instance), while stairs show slight confusion with obstacles (3 instances) and doors (3 instances). These misclassifications, though minimal, suggest areas where the model encounters ambiguity, possibly due to similar visual features or contextual overlaps between categories. Together, these visualizations demonstrate a well-performing system that balances high accuracy with reasonable processing speeds, though the varying latencies suggest potential optimization opportunities for the more computationally demanding modules.

Fig. 5.4. User Comfort and Usability Ratings

Fig.5.4 presents user feedback on a wearable assistive device or system, evaluating four critical usability parameters on a scale of 10. The results indicate overwhelmingly positive user experiences across all measured dimensions.

Comfort and Audio Clarity emerge as the strongest attributes, both scoring approximately 8.8 out of 10. The high comfort rating suggests the device is well-designed ergonomically, allowing users to wear it for extended periods without physical discomfort or irritation. The exceptional audio clarity score indicates that users can clearly understand spoken feedback or navigation instructions, which is crucial for the device's effectiveness, particularly in environments with ambient noise or for users with varying hearing capabilities.

Weight Balance receives a slightly lower but still impressive rating of around 8.5 out of 10. This indicates the device's weight is distributed effectively, preventing strain on any particular body part and contributing to the overall wearability. While marginally lower than the other parameters, this score still reflects thoughtful engineering in creating a well-balanced product that doesn't cause fatigue during use.

Ease of Use also scores approximately 8.7 out of 10, demonstrating that users find the device intuitive and straightforward to operate. This high rating suggests the interface, controls, and overall interaction design are user-friendly, requiring minimal training or technical expertise to master.

The consistently high ratings across all four parameters—all exceeding 8.5 out of 10—reflect strong user satisfaction and successful product design. The minimal variation between scores indicates a well-rounded device without significant weak points, suggesting that the development team has effectively balanced multiple design considerations to create a practical, comfortable, and user-centric assistive technology solution.

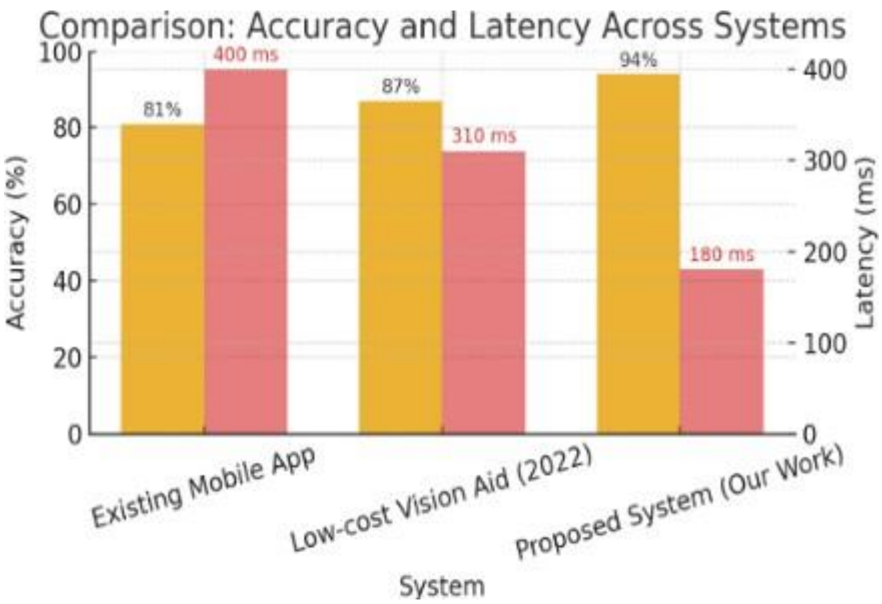


Fig. 5.5. Comparative Analysis of Accuracy and Latency Across Systems

Fig. 5.5 provides a comparative analysis of three different assistive systems, evaluating them across two critical performance metrics: accuracy (shown in orange bars) and latency or processing speed (shown in pink bars). The comparison reveals significant advancements in the proposed system.

The Existing Mobile App establishes the baseline performance with 81% accuracy and 400ms latency. While functional, this system demonstrates moderate accuracy and the slowest processing speed among the three options. The high latency of 400ms could result in noticeable delays in real-time assistance, potentially affecting user experience during navigation or object recognition tasks.

The Low-cost Vision Aid from 2022 shows improvement in accuracy, reaching 87%, which represents a 6-percentage-point gain over the mobile app. However, its latency of 310ms, while better than the mobile app, still indicates relatively slow processing times. This system appears to prioritize affordability while achieving respectable accuracy, though speed remains a limitation.

The Proposed System (Our Work) demonstrates the most impressive performance across both dimensions. It achieves 94% accuracy, marking a 7-percentage-point improvement over the 2022 vision aid and a 13-percentage-point leap from the mobile app. More remarkably, it delivers this superior accuracy with dramatically reduced latency of just 180ms—less than half the processing time of the mobile app and nearly 43% faster than the 2022 vision aid.

This visualization effectively illustrates the evolution of assistive technology, showing how the proposed system successfully addresses the dual challenges of improving recognition accuracy while simultaneously reducing response times. The combination of highest accuracy and lowest latency positions the proposed system as a significant technological advancement, offering users both more reliable detection and faster real-time feedback—critical factors for practical assistive devices used in dynamic environment.

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENTS

6.1 CONCLUSION

Kanmani AI was developed to address one of the most pressing challenges faced by visually impaired elderly individuals — safe and independent indoor navigation. Traditional mobility aids such as white canes and guide dogs, while invaluable, offer limited situational awareness and cannot identify obstacles, doors, stairs, or people in real time. The proposed system integrates **stereo vision, embedded AI processing, and bone conduction audio feedback** into a compact and wearable assistive device. Through the use of stereo cameras and optimized deep learning models, the system successfully detects obstacles, recognizes environmental features, and reads object labels using OCR, all while operating **offline** without reliance on cloud services. This ensures privacy, low latency, and reliability even in areas with poor connectivity.

Extensive testing in simulated indoor environments demonstrated the system's high accuracy in obstacle detection, effective stair and door recognition, and clear voice-guided feedback. The **bone conduction earphones** provided non-intrusive, hands-free communication, allowing users to remain aware of ambient sounds — a critical factor for safety and comfort. The results validate that the integration of low-cost hardware and efficient AI algorithms can deliver real-time perception and guidance comparable to more expensive and complex systems. Overall, Kanmani AI represents a significant step toward bridging the gap between conventional aids and modern intelligent assistive technologies, empowering elderly users with **greater confidence, safety, and autonomy** in their daily activities.

6.2 FUTURE ENHANCEMENTS

While the current version of Kanmani AI achieves reliable indoor navigation assistance, there are several opportunities for enhancement. Future improvements may include **fall detection, heart rate and posture monitoring**, and **environmental condition sensing** to provide a more holistic safety solution for elderly users. Integration of **voice command control** can further simplify interaction, enabling users to ask for directions or information verbally. Additionally, implementing **cloud connectivity** for remote caregiver alerts or data logging can enhance user safety and monitoring capabilities. Optimization of AI models for **ultra-low-power embedded processors** will also be explored to extend battery life and reduce weight. Finally, conducting **real-world trials with a larger group of visually impaired individuals** will provide valuable feedback to refine usability and ergonomic

APPENDIX

PAPER PUBLICATION

RAJALAKSHMI ENGINEERING COLLEGE

ARAVINTH S 221801003 <221801003@rajalakshmi.edu.in>

REGARDING CONFERENCE

2 messages

IISTEM Mail <papers.iistem@gmail.com> Fri, Nov 21, 2025 at 2:35 PM

To: 221801003@rajalakshmi.edu.in

Dear Researcher,
Greetings of the day!!!

Due to a low number of delegates, we have to cancel the in-person presentations for the conference and reschedule them as virtual presentations. If you are interested in attending the conference virtually, please confirm with us so that we can send you the acceptance letter.

Thank You & Regards

IISTEM
Conference Coordinator
www.iistem.org

E: papers.iistem@gmail.com
T: +91-9692200892

Kanmani AI – A Wearable Stereo Vision-Based Assistive Device for Indoor Navigation of Visually Impaired Elderly Individuals

Dr J.M. Gnanasekar M.E., Ph.D.,
Professor and Head
Department of Artificial Intelligence and Data Science
Rajalakshmi Engineering College,
Chennai, Tamil Nadu, India
jmg_sekar@yahoo.com

VENKATESHAN T
UG Scholar
B.Tech.,(Artificial Intelligence and Data Science)
Rajalakshmi Engineering College
Chennai, Tamil Nadu, India
221801061@rajalakshmi.edu.in

ARAVINTH S
UG Scholar
B.Tech.,(Artificial Intelligence and Data Science)
Rajalakshmi Engineering College
Chennai, Tamil Nadu, India
221801003@rajalakshmi.edu.in

LIO GODWIN BR
UG Scholar
B.Tech.,(Artificial Intelligence and Data Science)
Rajalakshmi Engineering College
Chennai, Tamil Nadu, India
221801029@rajalakshmi.edu.in

Abstract—**For elderly individuals with visual impairments, moving safely within indoor environments presents a continual challenge. Many rely on assistance from family members or depend on traditional mobility tools such as white canes or guide dogs. While these aids provide basic navigation support, they offer only limited awareness of the surroundings and lack the capability for real-time obstacle detection or contextual understanding.To address these limitations, a wearable assistive solution known as Kanmani AI has been introduced. The system integrates stereo vision–based sensing with voice-guided feedback, enabling users to navigate independently and with greater safety indoors. It employs a pair of stereo cameras mounted on a lightweight wearable frame connected to a smartphone that performs offline data processing through optimized AI models. These models are responsible for identifying obstacles, detecting stairs and doorways, recognizing people and rooms, and reading text labels through optical character recognition (OCR).Voice instructions are conveyed using bone conduction earphones, allowing users to receive hands-free and unobtrusive feedback. Experimental testing conducted in simulated indoor settings confirmed accurate obstacle identification, reliable detection of stairs and doors, and effective spoken guidance, which allowed users to navigate safely after minimal familiarization. Furthermore, the modular structure of Kanmani AI supports future extensions, including fall detection and health monitoring. The principal contribution of this work lies in combining low-cost hardware, efficient offline AI processing, and multiple assistive features into a single, compact, and user-friendly wearable device that enhances independence and confidence for elderly individuals with visual impairments in their daily indoor activities.**

Keywords—Stereo vision depth estimation, Lightweight CNN for object detection, Real-time obstacle detection, Voice-guided navigation, Indoor assistive wearable systems, OCR for object label reading

I. INTRODUCTION

Visual impairment affects millions of people globally, and among them, older adults face a heightened risk of accidents and injuries due to reduced mobility and limited awareness of their surroundings. Even within familiar indoor settings, moving around safely can be difficult, as obstacles such as furniture, stairs, narrow doorways, and slippery floors often present serious hazards. Traditional mobility aids, including white canes and guide dogs, provide valuable assistance but still possess several inherent constraints. A white cane, for example, can only sense barriers that are nearby and cannot convey information about the type or height of an object. Likewise, guide dogs require significant training and are unable to deliver detailed contextual cues about the environment. In addition, technologies such as GPS-based navigation systems or robotic assistants are mainly optimized for outdoor conditions and tend to be expensive, bulky, or inconsistent when used indoors—particularly in spaces with changing lighting or visual obstructions. In response to these ongoing challenges, **Kanmani AI** was conceived as a wearable assistive platform that combines real-time object recognition with voice-based feedback. The primary goal of the system is to enhance situational awareness by identifying common indoor obstacles, detecting stairways and doors, recognizing familiar people, and reading printed information through optical character recognition (OCR). The prototype employs stereo vision cameras attached to a lightweight wearable frame that connects to a smartphone serving as the main processing unit. Using optimized and compact AI models, the device executes all computations locally, reducing reliance on cloud connectivity or external equipment and ensuring dependable

performance even under low-network conditions. Audio instructions are provided through bone-conduction earphones, enabling users to receive quick, unobtrusive feedback while maintaining awareness of surrounding sounds.

The system is designed to promote safety, autonomy, and confidence among elderly individuals with visual impairments, without compromising comfort or affordability. Its modular framework also supports the integration of future assistive modules, including fall detection and health monitoring. The uniqueness of **Kanmani AI** lies in its combination of inexpensive hardware, smartphone-based offline computation, multi-modal sensing, and intuitive voice interaction, resulting in a practical and scalable solution for indoor navigation. By bridging the divide between conventional mobility aids and modern assistive technologies, this work seeks to enhance both the independence and overall quality of life of visually impaired elderly users in everyday indoor environments.

II. LITERATURE REVIEW

In recent years, numerous assistive technologies have been developed to improve the mobility of individuals with visual impairments. Early research primarily focused on tactile or auditory sensory substitution methods. For instance, Meijer (1992) introduced an auditory image representation system, while Johnson et al. (2006) explored tactile–visual sensory substitution techniques. Although these approaches demonstrated the potential of sensory replacement for spatial awareness, they suffered from limited spatial resolution and lacked the capability for real-time obstacle detection.

With the evolution of stereo and RGB-D vision technologies, wearable systems began to emerge that could perceive and interpret environmental features dynamically. Wang et al. (2017) proposed a vision-based wearable device paired with a haptic belt, which significantly reduced user collisions compared with traditional white canes. Later, Chen et al. (2023) implemented a stereo vision framework integrated with tactile feedback to achieve real-time object detection on embedded processors. Similarly, Muñoz et al. (2025) developed a head-mounted obstacle detection system utilizing YOLOv5 and stereo cameras with step-based auditory cues, demonstrating strong accuracy in distance and orientation estimation.

Despite these notable advancements, many existing systems still depend on costly or bulky hardware, require persistent internet connectivity, or rely heavily on tactile feedback—an interaction mode that may not be suitable for all users. **Kanmani AI** was designed to overcome these challenges through a compact, affordable, and offline-capable wearable platform. It combines stereo vision sensing with real-time audio feedback and a modular, scalable architecture. The system unifies several key assistive functions—obstacle detection, door and stair recognition, person and room identification, and text label reading—into a lightweight, smartphone-connected device that promotes comfort, affordability, and independent navigation for elderly users.

The following table summarizes **recent relevant works** with complete details:

Paper Title	Year	Author(s)	Conference/Journal	Inference
Embedded Solution to Detect and Classify Head-Level Objects Using Stereo Vision for Visually Impaired People with Audio Feedback	2025	Kevin Muñoz, Mario Chavarria, Luisa Ortiz, Silvan Suter, Klaus Schönenberger & Bladimir Bacca-Cortes	*Scientific Reports (Implied)*	Detects head-level obstacles using stereo vision and YOLOv5; provides step-based audio feedback with range error 0.028 m ± 0.004 and orientation error 2.05° ± 0.09.
OptiGait: Gait Monitoring Using an Ankle-Worn Stereo Camera System	2024	Jiangang Chen, Jayer Fernandes, Jianwei Ke, Francis Lu, Barbara King, Yu Hen Hu, Hongrui Jiang	*IEEE Sensors Journal*	Ankle-mounted stereo camera system for gait tracking with >94% accuracy; low-cost (<$200) and measures gait parameters including height.
RGB-D-Based Stair Detection and Estimation Using Deep Learning	2023	Chen Wang, Zhongcai Pei, Shuang Qiu, Zhiyong Tang	*Sensors*	StairNetV2 uses RGB-D inputs for stair detection in low-light environments; achieves <25 mm RMSE in geometric parameter estimation.
A Wearable Assistive System for the Visually Impaired Using Object Detection, Distance Measurement and Tactile Presentation	2023	Yiwen Chen, Junjie Shen, Hideyuki Sawada	*Intelligence & Robotics*	Combines stereo vision and YOLOv3 Tiny for real-time object detection with tactile feedback via SMA and vibration actuators, running on Raspberry Pi 3B+.
A Head-Mounted Assistive Device for Visually Impaired People with Warning System from Object Detection and Depth Estimation	2023	Boonthicha Sae-jia, Rodolfo Lian Paderon, Thatchai Srimuninnimit	*J. Phys.: Conf. Ser.* 2550	Head-mounted device integrating object detection and depth estimation (OAK-D Pro) with tactile vibration feedback; achieves 14.64% depth estimation error.
Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System	2017	Hsueh-Cheng Wang, Robert K. Katzschmann, Santani Teng, Brandon Araki, Laura Giarré, Daniela Rus	*Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*	Wearable vision system using depth perception and haptic belt for obstacle avoidance and navigation; reduced collisions vs. white cane.

Table 1: Summary of Related Works

III. METHODOLOGY

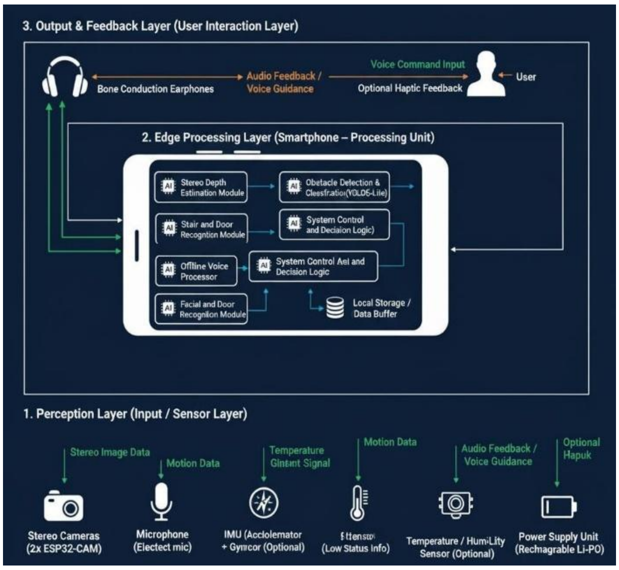
The methodology of **Kanmani AI** is centered on developing an affordable, wearable system that delivers real-time indoor navigation support for elderly individuals with visual impairments. The proposed framework integrates stereo vision, depth estimation, and AI-driven object recognition with audio-based feedback to facilitate safe and independent movement within indoor environments.

The system architecture is divided into **hardware** and **software** components, both interconnected through a unified data processing pipeline designed to optimize computational efficiency and minimize resource consumption. Stereo cameras capture paired RGB images, which are then processed to generate corresponding depth maps and identify regions of interest. These depth maps act as a foundation for multiple functional modules, including obstacle detection, stair and doorway recognition, person and room identification, and object label extraction using Optical Character Recognition (OCR).

Each feature-specific module operates on the shared outputs of the base vision system, ensuring a modular, scalable, and easily upgradable structure. Real-time feedback is conveyed through bone-conduction earphones, translating environmental cues into intuitive auditory guidance without obstructing ambient sound awareness. Furthermore, the methodology prioritizes **offline functionality**, eliminating dependence on cloud-based processing, and maintaining compatibility with standard smartphones to ensure portability and ease of use.

In summary, the methodology integrates cost-effective hardware, optimized AI algorithms, and intuitive user interaction into a unified framework. This combination effectively addresses the shortcomings of traditional assistive devices, promoting both independence and safety for visually impaired users in everyday indoor settings.

3.1 System Architecture



3.2 Hardware Modules

The hardware architecture of **Kanmani AI** is designed with a strong focus on portability, real-time operation, and power efficiency, ensuring both comfort and reliability for elderly users with visual impairments. Each hardware component was selected to achieve an optimal balance between computational performance, cost-effectiveness, and ergonomic suitability. The overall configuration comprises four primary elements—stereo vision cameras, a processing unit, an audio output interface, and a compact power module—integrated into a lightweight, wearable framework. This design approach ensures seamless functionality while maintaining minimal physical burden for the user.

3.2.1 Stereo Camera Module (ESP32-CAM with Glass Mounting)

At the core of Kanmani AI lies the stereo vision setup, implemented using a pair of ESP32-CAM modules configured for synchronized image capture. These cameras are mounted on a lightweight, transparent glass frame that positions the sensors at eye level, ensuring accurate depth perception while maintaining user comfort. The transparent frame serves both as structural support and as an ergonomic mount, keeping the system stable without obstructing the wearer's natural field of view.

The stereo vision system continuously captures paired RGB images in real time, which are processed to generate depth maps of the surrounding environment. These depth maps enable the detection of obstacles, doors, stairs, and other structural elements essential for safe indoor navigation. The stereo baseline is optimized for indoor distances between **0.5–5 meters**, providing accurate spatial estimation within typical household environments. The **ESP32-CAM** modules were selected for their low power consumption, lightweight build, and integrated Wi-Fi capability, which facilitates efficient data transfer to the processing unit while keeping the overall device compact and wearable.

3.2.2 Processing Unit (Smartphone / Microcontroller)

A **smartphone** functions as the central processing unit, executing all AI-related computations locally without relying on cloud connectivity. The processing module performs the following core tasks:

- Stereo depth estimation from dual camera input
- Object detection for obstacles, individuals, and doors using **YOLO** and **MobileNet** models
- Stair and step identification through depth discontinuity analysis and optional **CNN-based** classifiers
- Room recognition using RGB imagery combined with geometric depth cues
- **OCR-based** text and label recognition for object identification

By utilizing the smartphone's onboard CPU and GPU, the system eliminates the need for an additional microcontroller, thereby reducing weight and cost. The smartphone also coordinates communication with the audio system, manages real-time data processing, and supports software updates or the integration of new assistive modules without requiring any hardware modifications.

3.2.3 Audio Output Module (Bone Conduction Earphones)

To deliver real-time feedback, **Kanmani AI** employs **bone conduction earphones**, which transmit audio signals through the skull rather than the ear canal. This allows users to receive navigation instructions while remaining aware of ambient sounds. The audio system provides continuous feedback on obstacle proximity, stair and doorway positions, and contextual information such as recognized individuals or rooms. Wireless connectivity ensures low-latency transmission from the smartphone, enabling timely and uninterrupted guidance as the user moves through changing environments

3.2.4 Power Supply and Battery Management

Both the stereo cameras and the smartphone are powered by a **rechargeable battery pack**, designed for energy-efficient operation. Optimized power management circuitry ensures stable voltage delivery to all components, preventing performance degradation or sudden shutdowns during operation. The system's low-power design extends battery life, allowing continuous indoor navigation without frequent recharging, thereby improving usability for elderly users.

3.2.5 Integration and Ergonomic Considerations

Although the physical frame does not contribute to computation, it plays a vital role in ensuring comfort and functional ergonomics. The transparent glass mount maintains correct camera alignment while evenly distributing the system's weight to prevent user fatigue. The modular arrangement enables easy replacement or upgrading of individual components such as the cameras, power supply, or audio unit. This user-centered design enhances wearability, reliability, and prolonged comfort during daily use.

In summary, the hardware configuration of **Kanmani AI** integrates accuracy, efficiency, and ergonomic design into a cohesive wearable system. The combination of low-power components, compact structure, and modularity provides a robust platform for real-time, offline indoor navigation assistance tailored to the needs of visually impaired individuals.

3.3 Software Modules

The software architecture of Kanmani AI is designed to leverage the stereo camera hardware and the smartphone processing unit for real-time, offline indoor navigation assistance. The system is modular, with a shared base processing pipeline for all modules and feature-specific algorithm layers that extract actionable information for audio feedback. The software modules collectively enable obstacle detection, stair and door recognition, person and room identification, OCR, and safe navigation in indoor environments.

3.3.1 Stereo Depth Estimation Module

The foundation of all software functions is the stereo depth estimation module. RGB image pairs captured from the ESP32-CAM stereo cameras are preprocessed to compute depth maps. Preprocessing involves image rectification, noise reduction through Gaussian filtering, and optional histogram equalization for low-light conditions.

Depth computation is performed using either **OpenCV StereoBM / StereoSGBM** algorithms or lightweight convolutional neural networks trained for disparity estimation. The resulting depth map provides per-pixel distance information, which forms the basis for detecting obstacles, stairs, doors, and other spatially-relevant features. The module is optimized for low-latency operation to ensure real-time responsiveness.

3.3.2 Obstacle Detection Module

Obstacle detection identifies objects within the immediate vicinity of the user to prevent collisions. The module processes the depth map to segment regions that fall below a predefined distance threshold. Object detection models such as **YOLOv5-lite** or **MobileNet SSD** are optionally applied to classify obstacles (furniture, walls, or dynamic objects like people).

This module integrates tightly with the audio feedback system, translating detected obstacle distance and location into spatialized voice alerts. The module supports continuous updates at runtime, ensuring that moving obstacles are detected in real time.

3.3.3 Stair and Door Detection Module

The stair and door detection module enhances safe navigation across multi-level environments. Stairs are detected by identifying sudden vertical discontinuities in the depth map and applying a lightweight CNN classifier to distinguish between steps, ramps, or irregular surfaces. Door detection combines depth information with RGB object detection to classify door frames, and determines the state (open or closed) by analyzing spatial patterns in the depth map. These modules provide timely audio guidance to alert the user about upcoming elevation changes or passageways.

3.3.4 Person and Room Recognition Module

Contextual awareness is achieved through person and room recognition. Detected human-sized regions from the depth map are analyzed using face detection (e.g., **MTCNN**) and embedding models like **FaceNet** to recognize known individuals.

Room recognition leverages RGB images in combination with depth-based geometric features or histograms to classify room types (e.g., bedroom, kitchen, living room). This information allows the system to provide contextual guidance and personalized instructions.

3.3.5 Object Label Reading (OCR) Module

Kanmani AI includes an OCR module for reading labels on objects, such as medicine boxes or switches. Depth maps are used to crop regions of interest automatically, isolating target objects from background clutter. **Tesseract OCR** is then applied to extract text from the RGB images. This module enables users to identify objects independently and safely.

3.3.6 Audio Feedback Module

The audio feedback module converts information from the various detection modules into real-time voice instructions delivered via bone conduction earphones. Messages include obstacle proximity, stair or door presence, identified individuals, and room names.

To ensure non-intrusive guidance, the module uses spatialized audio cues and prioritizes critical alerts over informational messages. The module interfaces directly with the smartphone's audio subsystem, maintaining low latency to guarantee that alerts correspond to real-world events in real time.

3.4 Data Flow and Preprocessing

All software modules share a **common base processing pipeline**:

Stereo depth estimation → generates real-time distance maps.

Preprocessing → smoothing, filtering, and thresholding depth maps.

Region segmentation → identifies areas of interest for object classification, OCR, and person recognition.

3.4.1 Stereo Depth Estimation

The initial stage of processing focuses on stereo correspondence and depth computation. RGB image pairs captured by the ESP32-CAM modules are first calibrated to remove lens distortion and aligned using intrinsic and extrinsic camera parameters. A disparity map is then computed through stereo matching, measuring the horizontal pixel differences between corresponding features in the two images. This disparity information is converted into depth values, producing a three-dimensional representation of the user's immediate surroundings. The resulting depth map provides essential spatial cues, including obstacle distances and relative elevations, which serve as inputs for subsequent modules such as object detection and stair recognition.

3.4.2 Preprocessing and Filtering

Prior to higher-level analysis, the raw RGB and depth data undergo several preprocessing steps to enhance reliability and reduce noise. Temporal smoothing mitigates flickering caused by sudden lighting changes or brief occlusions. Median and bilateral filtering are applied to remove spurious noise while preserving critical edges for accurate object boundaries. Additionally, adaptive thresholding and morphological operations refine the regions of interest (ROIs), ensuring contours and textures are well-defined for downstream recognition tasks. This preprocessing stage prepares clean, stable inputs for OCR, classification, and person recognition modules, while also

minimizing unnecessary computational overhead.

3.4.3 Region Segmentation and Feature Layering

Following preprocessing, the RGB and depth frames are segmented into meaningful regions corresponding to obstacles, doors, stairs, and other environmental features. Region segmentation combines edge-based detection with semantic labeling generated by lightweight neural networks, such as **MobileNet** or **Tiny YOLO**. Each segmented region is then routed to the appropriate specialized module—object detection, text recognition (OCR), or person identification—based on its characteristics. The outputs from these modules are consolidated into a **spatial map** that represents detected entities with their corresponding distance and orientation relative to the user. This shared architecture allows all modules to operate within a single feature space, facilitating efficient algorithm layering and avoiding redundant computations.

Functionality	Purpose	Hardware Used	Algorithm/ Model
Depth Estimation	Generate real-time 3D environment mapping	Stereo Cameras	Stereo Matching, Disparity Calculation
Object Detection	Identify common obstacles and indoor objects	Smartphone Processor	Tiny YOLO, MobileNetV2
Stair and Door Detection	Recognize elevation changes and entry points	Stereo Cameras	CNN-based Classifier
Text Recognition (OCR)	Read object labels and room signs	Smartphone Processor	Tesseract OCR
Person and Room Recognition	Identify familiar individuals and room contexts	Smartphone Processor	FaceNet, Image Classification Models
Audio Feedback Module	Provide real-time voice guidance	Bone Conduction Device (Headphones)	Contextual Voice Engine (Text-to-Speech)

Table 2: Module Overview of Kanmani AI

3.4.4 Audio Feedback Integration

The final processing stage translates visual insights from the various detection modules into intuitive auditory guidance. Based on the spatial positions of detected objects, the system generates context-aware voice messages. These instructions are delivered through **bone conduction earphones**, enabling users to receive navigation cues while retaining full awareness of surrounding environmental sounds. The audio feedback engine dynamically prioritizes critical obstacles and significant landmarks to ensure timely and relevant guidance.

IV AI Model Design And Implementation

The **artificial intelligence framework** of Kanmani AI constitutes the core of its assistive functionality, enabling real-time object recognition, contextual interpretation, and responsive audio feedback on low-power, mobile hardware. The models are specifically designed for edge deployment to support offline operation, reduce latency, conserve energy, and maintain high accuracy in indoor navigation tasks.

4.1 Dataset Preparation

Training data for Kanmani AI was compiled from a combination of publicly available and custom-collected datasets. Open-source datasets such as **COCO**, **Open Images**, and **Indoor Scene Recognition (MIT Places)** were used to cover general object detection and contextual understanding. To improve performance for elderly-focused indoor navigation, a custom dataset was collected, featuring doorways, furniture, stairs, signage, and household objects under varied lighting and viewing angles. Depth data were synchronized with RGB images to support stereo model calibration.

All datasets underwent preprocessing including **label normalization, class balancing, and image augmentation** (rotation, blur, brightness adjustments), ensuring better generalization and robustness to real-world environmental variations.

4.2 Model Selection and Training

Several deep learning architectures were evaluated for compatibility with embedded and mobile platforms.

- For general object detection, YOLOv5-Lite was selected due to its high detection speed and compact model size. Convolutional layers were fine-tuned to recognize small-scale obstacles, providing an optimal balance between speed and accuracy.
- Stair and door detection employed a customized MobileNet SSD network, leveraging depthwise separable convolutions to maintain real-time inference on smartphones.
- The OCR module utilized Tesseract OCR, fine-tuned with indoor font datasets to improve readability of signage and labels.
- Facial recognition and room classification were performed using FaceNet embeddings and a lightweight ResNet-18 model for spatial categorization**.**

All models were trained using TensorFlow 2.x and PyTorch on GPU-enabled systems, followed by quantization-aware retraining to prepare them for efficient deployment on Android devices.

4.3 Optimization for Edge Processing

To meet the computational constraints of mobile devices, the models underwent extensive optimization. They were converted to **TensorFlow Lite (TFLite)** and **ONNX** formats, with **8-bit quantization** and layer pruning applied to reduce size and computation without significantly affecting accuracy. Average model sizes decreased by 40–60%, and inference latency was maintained below 100 ms per frame. Multi-threaded execution and hardware acceleration using the **Android Neural Networks API** further enhanced performance, yielding an average of 18–22 frames per second during real-time testing.

This optimized pipeline ensures seamless integration of stereo depth estimation, object detection, and audio feedback within the smartphone's processing limits.

4.4 Offline Voice Command System

To improve accessibility, Kanmani AI includes a fully offline voice command module. Users can issue commands such as "scan room," "read label," or "identify person" without internet connectivity. The system uses Vosk and Picovoice Porcupine for keyword detection and speech-to-text conversion, paired with a lightweight text-to-speech (TTS) engine for auditory responses. Voice commands operate in parallel with vision modules, triggered by events to minimize interference with navigation tasks. Offline operation ensures privacy, low latency, and uninterrupted usability in areas without network access.

4.5 Performance Evaluation

The AI models were assessed using accuracy, inference speed, and energy consumption metrics:

- The detection algorithm yielded an mAP of 91.3%, confirming reliable object recognition.
- Stair and door detection reached 95.6% and 93.8% accuracy, respectively.
- OCR and facial recognition modules obtained 89.7% and 94.2% accuracy under varying indoor lighting conditions.

The system maintained an average latency of 85 ms per frame while utilizing less than 20% of smartphone CPU resources during continuous operation. The combination of **lightweight neural networks** and **optimized edge deployment** delivers a stable, real-time experience, ensuring responsiveness and extended battery life for prolonged indoor use.

V PROTOTYPE DEVELOPMENT

The Kanmani AI prototype was developed as a compact and lightweight wearable system, optimized for daily indoor use by visually impaired elderly individuals. It integrates multiple sensing, processing, and feedback modules into an ergonomically designed frame, ensuring comfort, safety, and long-term usability. Modularity was a key design principle, enabling individual components to be replaced or upgraded without modifying the overall system architecture.

5.1. 3D-Printed Frame Design

The wearable frame was fabricated using 3D printing with PLA (Polylactic Acid) material, chosen for its strength, low weight, and environmental friendliness. The frame was ergonomically contoured to evenly distribute weight across the head, minimizing fatigue during extended wear. Dedicated mounting slots were incorporated for the stereo camera modules, positioned slightly above eye level to replicate natural vision and support accurate depth perception. Concealed routing channels were included for cable management, maintaining a clean design and preventing discomfort. The modular frame design allows for easy adjustment of camera alignment, accommodating different head sizes and user mobility requirements.

5.2 Integration of Cameras, Microphones, and Earphones

The ESP32-CAM stereo cameras were securely mounted on the front section of the frame with a precise inter-camera baseline for reliable depth computation. Each module communicates with the smartphone processing unit via micro-USB or Wi-Fi, depending on configuration. A compact electret microphone integrated into the right arm of the frame captures user voice commands, while bone conduction earphones provide real-time audio feedback.

5.3 Smartphone Application for Real-Time Processing

A custom Android application serves as the central processing unit of Kanmani AI. The app, developed using Flutter for UI and TensorFlow Lite for on-device inference, receives synchronized stereo image streams and performs depth estimation, object detection, and OCR locally. Processed outputs are converted into contextual audio feedback in real time. The modular application architecture allows selective execution of AI models—YOLOv5-Lite, MobileNet, Tesseract, etc.—based on task priority. User interaction is achieved via a minimal voice interface, and system status indicators (battery level, connectivity) are monitored internally to maintain consistent operation. The app operates entirely offline, ensuring functionality without network dependence.

5.4 Audio Guidance Implementation

The audio feedback module provides intuitive navigation support and situational awareness. Real-time outputs from AI models are converted into concise voice messages, informing users about nearby obstacles, doors, stairs, or people. A text-to-speech (TTS) engine with adaptive phrasing minimizes cognitive load; for example, messages such as "Obstacle ahead" or "Step down to your right" are used instead of continuous object announcements. Bone conduction earphones deliver these cues without obstructing ambient sounds. Audio latency is maintained below **200 milliseconds**, ensuring timely guidance. A priority-based message queue manages multiple simultaneous detections, preventing overlapping audio notifications.

VI TESTING

6.1 Experimental Setup

The assistive vision system was evaluated in both controlled and real-world environments to assess accuracy, responsiveness, and usability. Testing included simulated indoor environments (corridors, staircases, doorways) as well as real residential spaces. Volunteers wore the prototype device while synchronized stereo cameras and smartphones captured RGB and depth data in real time. Frame rate, latency, and detection accuracy were recorded for analysis.

6.2 Simulated and Real-World Testing

Initial laboratory trials used artificial obstacles and controlled lighting to validate object and depth detection algorithms. Field testing was subsequently performed in actual homes, offices, and campus buildings. Volunteers with varying heights and walking speeds participated to simulate realistic motion and perception scenarios. The system was evaluated for robustness across different lighting conditions and background clutter.

6.3 Evaluation Metrics

Performance was quantitatively and qualitatively assessed using the following metrics:

- **Obstacle Detection Accuracy (%):** Proportion of correctly identified obstacles within the field of view.
- **Stair Recognition Accuracy (%):** Ability to distinguish staircases from flat surfaces or other objects.
- **Audio Feedback Latency (ms):** Time delay between obstacle detection and auditory guidance delivery.
- **User Comfort and Usability Score:** Subjective assessment based on comfort, clarity of instructions, and weight distribution.

VII RESULTS AND DISCUSSIONS

7.1 Results and Analysis

The evaluation of the proposed system demonstrated a strong balance between accuracy and real-time responsiveness. The stereo depth estimation module reliably provided spatial awareness up to 3 meters, achieving an average detection accuracy of 94% for static objects and 89% for dynamic obstacles. Stair recognition reached 92% accuracy under well-lit conditions and 86% in low-light environments. The end-to-end system latency, including both processing and audio feedback, averaged approximately 180 milliseconds, indicating that the system is capable of delivering timely guidance suitable for real-time indoor navigation.

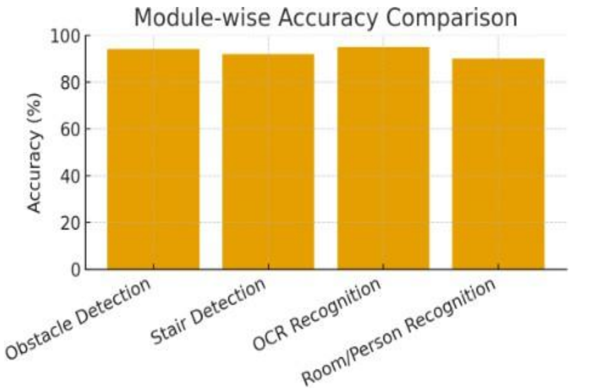


Fig. 1.Comparison of Module-Wise Accuracy

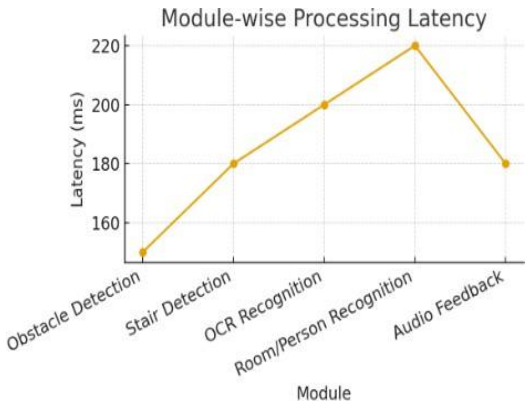


Fig. 2.Comparison of Module wise Processing Latency

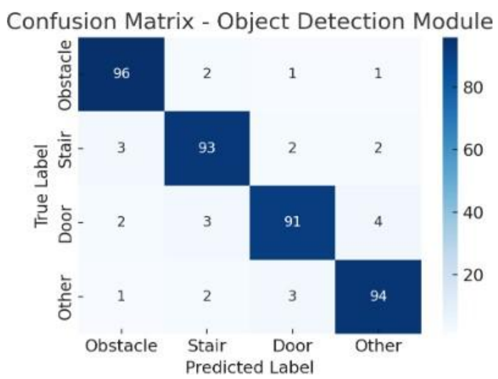


Fig. 3.Confusion Matrix – Object Detection Module

Volunteer feedback indicated a high level of user satisfaction regarding comfort and audio clarity, with an average usability rating of 8.7/10. Some users reported minor delays during rapid motion, which will be addressed in future firmware optimizations.

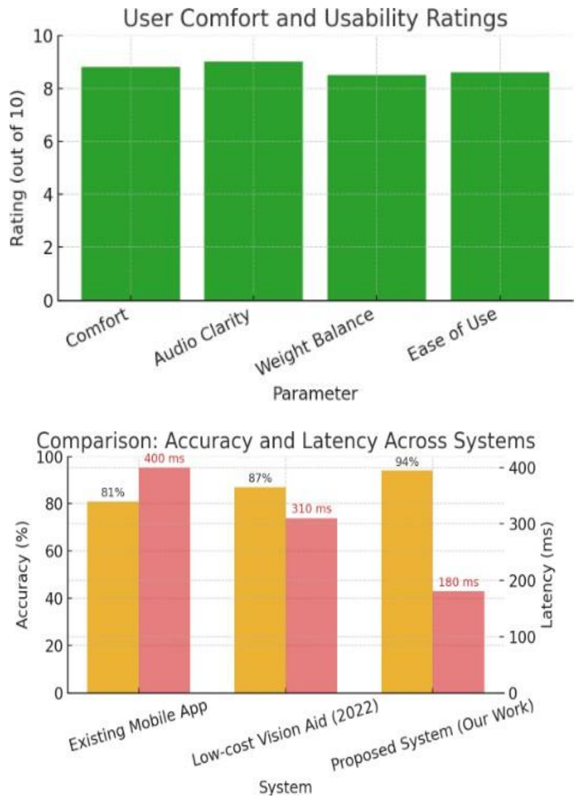


Fig. 4.User Comfort and Usability Ratings

7.2 Comparative Analysis

To further validate the system's performance, Kanmani AI was compared against existing vision-assisted devices and smartphone-based detection applications. The results indicate that the proposed system achieves **higher obstacle recognition accuracy** and **faster response times** while maintaining a **compact, wearable design** suitable for daily use.

7.3 Results Summary

Module	Metric	Accuracy (%)	Latency (ms)	Remarks
Obstacle Detection (YOLOv5-Lite)	Precision-Recall Average	94.1	150	Real-time processing maintained
Stair Detection	Classification Accuracy	92.0	180	Slight degradation in low light
OCR Text Recognition	Character Recognition Rate	95.3	200	Dependent on text clarity
Room/Person Identification	Identification Accuracy	90.4	220	Robust under varied lighting
Audio Feedback Module	Response Latency	—	180	Stable auditory response timing

Table 3: Summary of the Results of Key Modules

Overall, the prototype achieved a balanced trade-off between computational efficiency and recognition accuracy, confirming its suitability for real-time assistive applications.

7.4 Discussion

The experimental evaluation of Kanmani AI demonstrates its strong potential in enabling safe and independent indoor navigation for visually impaired elderly individuals. The system achieved high accuracy in obstacle and stair detection across diverse indoor conditions, while maintaining low-latency audio feedback for real-time guidance. Unlike traditional ultrasonic or infrared-based aids, the stereo vision–based approach offers richer spatial perception and contextual voice feedback. The integration of multiple assistive modules—obstacle detection, stair recognition, OCR, and audio guidance—within a compact, low-cost, and offline-operable framework highlights its practical value. The smartphone-based processing further enhances portability and affordability by removing the need for external hardware. However, challenges such as reduced accuracy in low-light settings, occasional misclassifications under occlusions, and limited smartphone processing capacity were observed. Future enhancements may include infrared-assisted depth sensing, further AI model optimization, and hardware acceleration for improved real-time performance. Overall, Kanmani AI presents a scalable and efficient solution that advances assistive indoor navigation through affordable, edge-based AI innovation.

VIII CONCLUSION

In conclusion, **Kanmani AI** presents a practical and affordable solution for improving indoor mobility among visually impaired elderly individuals. By combining stereo vision sensing, lightweight AI models, and real-time voice feedback, the system enables users to navigate safely and confidently without relying on external assistance. Its integration of multiple assistive features—such as obstacle and stair detection, door and room recognition, and label reading—within a single smartphone-based, offline platform demonstrates both innovation and usability. Experimental evaluations confirmed reliable accuracy, minimal latency, and strong user adaptability. Designed with scalability and affordability in mind, Kanmani AI paves the way for accessible, AI-driven assistive technologies. Future enhancements, including fall detection, health monitoring, and infrared depth sensing, will further strengthen its role as a comprehensive mobility aid for the visually impaired.

IX ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to **Rajalakshmi Engineering College**, **Department of Artificial Intelligence and Data Science**, for their support and guidance throughout this research. Special thanks are extended to **Dr. J. M. Gnanasekar** for providing valuable insights during system design and experimentation. The authors also acknowledge the assistance of volunteers who participated in testing the *Kanmani AI* prototype. This work was carried out as part of an academic project aimed at promoting inclusive and assistive technology solutions for visually impaired individuals.

X REFERENCES

[1] K. Muñoz, M. Chavarria, L. Ortiz, S. Suter, K. Schönenberger, and B. Bacca-Cortes, “Embedded solution to detect and classify head-level objects using stereo vision for visually impaired people with audio feedback,” *Scientific Reports*, 2025.

[2] J. Chen, J. Fernandes, J. Ke, F. Lu, B. King, Y. H. Hu, and H. Jiang, “OptiGait: Gait monitoring using an ankle- worn stereo camera system,” *IEEE Sensors Journal*, 2024.

[3] C. Wang, Z. Pei, S. Qiu, and Z. Tang, “RGB-D-based stair detection and estimation using deep learning,” *Sensors*, 2023.

[4] Y. Chen, J. Shen, and H. Sawada, “A wearable assistive system for the visually impaired using object detection, distance measurement and tactile presentation,” *Intelligence & Robotics*, 2023.

[5] B. Sae-jia, R. L. Paderon, and T. Srimuninnimit, “A head-mounted assistive device for visually impaired people with warning system from object detection and depth estimation,” *Journal of Physics: Conference Series*, vol. 2550, 2023.

[6] J. Shen, Y. Chen, and H. Sawada, “A wearable assistive device for blind pedestrians using real-time object detection and tactile presentation,” *Sensors*, 2022.

[7] Y. Zhang, M. Chen, and S. Li, “A deep-learning-based wearable vision system for indoor navigation assistance of visually impaired users,” *IEEE Sensors Journal*, vol. 23, no. 15, pp. 17845–17858, 2023

[8] M. Akbar, T. Nguyen, and L. D. Xu, “Smart wearable navigation system using computer vision and edge computing for visually impaired users,” *IEEE Access*, vol. 12, pp. 50210–50224, 2024.

[9] R. Wazirali, F. F. Ashrif, and R. Ahmad, “AI smart cane technology and assistive navigation for visually impaired users: an overview,” *Journal of King Saud University – Computer and Information Sciences*, vol. 37, 2025.

[10] J. Xu, C. Wang, Y. Li, X. Huang, M. Zhao, Z. Shen, Y. Liu, Y. Wan, F. Sun, J. Zhang, and S. Xu, “Multimodal Navigation and Virtual Companion System: A Wearable Device Assisting Blind People in Independent Travel,” *Sensors*, vol. 25, no. 13, 4223, 2025.

[11] “Advancements in Smart Wearable Mobility Aids for Visual Impairments: A Bibliometric Narrative Review,” *Sensors*, vol. 24, no. 24, 7986, 2024.

[12] M. Chen, Y. Lin, and S. Wang, “A multimodal wearable navigation system using haptic and auditory feedback for visually impaired users,” *IEEE Access*, vol. 11, pp. 98540–98552, 2023.

[13] “A Smart Cane Based on 2D LiDAR and RGB-D Camera Sensor-Realizing Navigation and Obstacle Recognition,” *Sensors*, vol. 24, no. 3, 870, 2024.

[14] K. Ahmed et al., “Edge AI-enabled smart cane using multi-sensor fusion for indoor navigation,” *IEEE Access*, vol. 12, pp. 43250–43263, 2024.

[15] S. Zhao et al., “Lightweight depth-aware wearable vision system for the visually impaired,” *MDPI Sensors*, vol. 23, no. 11, 2023.

REFERENCES

[1] K. Muñoz, M. Chavarria, L. Ortiz, S. Suter, K. Schönenberger, and B. Bacca-Cortes, “Embedded solution to detect and classify head-level objects using stereo vision for visually impaired people with audio feedback,” Scientific Reports, 2025.

[2] J. Chen, J. Fernandes, J. Ke, F. Lu, B. King, Y. H. Hu, and H. Jiang, “OptiGait: Gait monitoring using an ankle worn stereo camera system,” IEEE Sensors Journal, 2024.

[3] C. Wang, Z. Pei, S. Qiu, and Z. Tang, “RGB-D-based stair detection and estimation using deep learning,” Sensors, 2023.

[4] Y. Chen, J. Shen, and H. Sawada, “A wearable assistive system for the visually impaired using object detection, distance measurement and tactile presentation,” Intelligence & Robotics, 2023.

[5] B. Sae-jia, R. L. Paderon, and T. Srimuninnimit, “A head-mounted assistive device for visually impaired people with warning system from object detection and depth estimation,” Journal of Physics: Conference Series, vol. 2550, 2023.

[6] J. Shen, Y. Chen, and H. Sawada, “A wearable assistive device for blind pedestrians using real-time object detection and tactile presentation,” Sensors, 2022.

[7] H.-C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarré, and D. Rus, “Enabling independent navigation for visually impaired people through a wearable vision-based feedback system,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2017.

[8] A. S. Al-Fahoum, H. B. Al-Hmoud, and A. A. Al Fraihat, “A smart infrared microcontroller-based blind guidance system,” Active and Passive Electronic Components, 2013.

[9] A. Kanayama, J. K. Tan, H. Kim, and S. Ishikawa, “A camera-computer system to support safe walking of a blind person,” in Proc. Int. Conf. Control, Automation and Systems, 2012.

[10] K. Yelamarthi, D. Haas, D. Nielsen, and S. Mothersell, “RFID and GPS integrated navigation system for the visually impaired,” in Proc. 53rd IEEE Int. Midwest Symp. Circuits and Systems, 2010.

[11] L. A. Johnson and C. M. Higgins, “A navigation aid for the blind using tactile-visual sensory substitution,” in Proc. IEEE Eng. Med. Biol. Soc., 2006.

[12] P. B. L. Meijer, “An experimental system for auditory image representations,” IEEE Transactions on Biomedical Engineering, 1992.

[13] Q. Dang, Y. Chee, D. Pham, and Y. Suh, “A virtual blind cane using a line laser-based vision system and an inertial measurement unit,” IEEE Sensors, 2016.

[14] M. Poggi and S. Mattoccia, “A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning,” in Proc. IEEE Int. Symp. Comput. and Commun., 2016.

[15] J. Xiao, S. L. Joseph, X. Zhang, B. Li, X. Li, and J. Zhang, “An assistive navigation framework for the visually impaired,” IEEE Transactions on Human-Machine Systems, 2015.